NATIONAL SCIENCE BOARD

SCIENCE & ENGINEERING INDICATORS 2024

R&D

Publications Output: U.S. Trends and International Comparisons

NSB-2023-33

December 11, 2023

This publication is part of the 2024 edition of the *Science and Engineering Indicators (Indicators)* suite of reports. *Indicators* is prepared under the guidance of the National Science Board by the National Science Foundation's National Center for Science and Engineering Statistics. The *Indicators* suite consists of a summary report called *The State of U.S. Science and Engineering*, more-detailed thematic reports with supporting data, and a data tool that provides state-level indicators.

Table of Contents

List of Sidebars

List of Tables

List of Figures

Executive Summary

Key takeaways:

- The United States remains a highly influential nation in science and engineering (S&E) research, as measured by the volume of peer-reviewed scholarly publications and the rate of citations to those publications.
- In 2022, China remained the largest producer of publications, followed by the United States, then by India. China's top scientific field in terms of number of articles produced was engineering (25% of all publications), whereas the top field in the United States was health sciences (37%), and India's top field was computer and information sciences (21%).
- Analysis of funding acknowledgments shows that from 2018 to 2022, the scientific fields most frequently acknowledging federal funding are chemistry, biological and biomedical sciences, astronomy and astrophysics, and physics.
- The United States, the European Union (EU-27), and China currently produce a high number of highly cited articles, relative to their overall production.
- Open access (OA) has become an increasingly important feature of the publication landscape, in terms of output and impact, as shown by the growth of fully OA articles.
- International collaborations with U.S. authors of S&E publications have increased over the last 15 years, and China is the most frequent U.S. partner.
- International collaboration in the fast-growing field of artificial intelligence helps show an important research network and the most important collaborations in terms of absolute and relative size.

The primary method of disseminating research findings is through publication of peer-reviewed journal articles and conference proceedings (i.e., *publication output*). Data on publication output indicate a continued increase in global research activity, a growth in the proliferation and impact of some categories of OA research, and an internationally connected research ecosystem.

Global publication output reached 3.3 million articles in 2022, based on data from the Scopus database of S&E publications. The regions, countries, or economies with the largest volume of S&E publications in 2022 were China, with 27% of global output, and the United States, with 14%. From 2012 to 2022, the global yearly publication total grew by 59%. In terms of growth for these two largest producers, China and the United States had noticeably different expansion in their levels of overall production (growing by 173% and 6%, respectively).

Beyond differences at the level of region, country, or economy, the number of OA publications has increased dramatically in the last 10 years. In 2022, nearly 1.6 million articles were OA (classified in one of four OA categories), compared with about 1.5 million traditional closed-access journal articles. Just 10 years prior, OA articles accounted for around a third of all articles with a known access status. This growth is also clear with respect to impact, where OA research as a whole has a higher proportion of highly cited articles relative to the size of OA scholarship.

When an article is cited by a high number of subsequent articles by other authors, it is deemed to have exceptional scientific impact. Analyzing the distribution of highly cited articles based on the authors' locations, the United States has a long-standing record of producing a disproportionate share of such articles, although its share has decreased in recent years. China's share of those articles grew consistently over the past 20 years, and its scientific impact is on par with that of the EU-27. That impact varies by scientific discipline. In 2020, publications by authors in the United States in materials science, geosciences, and physics had relatively higher scientific impact than those in other fields. For publications by authors in China for the same year, those in the social sciences tended to have higher scientific impact than those in any other field.

International collaborations continue to grow in their share of global scientific publications. From 2012 to 2022, the share of articles from authors affiliated with institutions in multiple regions, countries, or economies increased by 19%. In 2022, the United States was involved in a high number of international collaborations (40% of U.S. articles produced included an international coauthor). Other top producers like China (19%), India (24%), and the United Kingdom (67%) varied in the concentration of international collaborations among their respective total outputs.

Introduction

Research publications and presentations at conferences represent the main mechanisms for disseminating research findings. Presentations appear in the published research literature in conference proceedings. Published literature is an indicator of scientific activity and global research partnerships. Additionally, analysis of how published literature is cited provides insight into the impact of research output. Scientific publications serve as a key linkage enabling public uses of scientific output (Yin et al. 2022).

This report presents data on research publication output by region, country, or economy and scientific field; impact measures; and international collaboration. The first section examines comparative region, country, or economy data on publication output across science and engineering (S&E) fields and includes a sidebar on federal funding acknowledgments. The second section provides an analysis of scientific impact as measured by bibliographic citations in research publications. The third section focuses on collaboration among researchers in the United States and those in other regions, countries, or economies through examining coauthoring and citation patterns. This section also includes a sidebar on the artificial intelligence (AI) publication output and collaboration network.

Bibliometric Data Preliminaries

This report analyzes S&E publications and citations using bibliometric data in Scopus, a database of scientific literature with English-language titles and abstracts (Science-Metrix 2021a). Because research activities are complex and multifaceted processes, the knowledge and social benefits that they produce are difficult to measure directly. Bibliometric data, including publications and citations, provide invaluable indicators of research output due to their ubiquity across regions, countries, or economies and time. Nonetheless, bibliometric analyses of publication and citation data remain proxy measures for the knowledge and social benefits produced by research activities, so they carry certain limitations. Publications themselves may represent differing “amounts” of research output because differences in field conventions or incentive structures may result in more or fewer publications covering novel research findings. In addition, publications do not represent all types of research products, such as data sets (Franzoni, Scellato, and Stephan 2011; Sugimoto and Larivière 2018).

This report analyzes nearly 44 million English-language articles published from 2003 to 2022. The analysis included papers published in conference proceedings and research articles published in peer-reviewed scientific and technical journals (collectively referred to as *articles*). The analysis excluded editorials, errata, letters, and other materials that do not typically present new scientific data, theories, methods, apparatuses, or experiments. The analysis also excluded working papers and preprints, which typically have not yet been peer reviewed, and articles published in journals that lack substantive peer review, sometimes referred to as *predatory journals* (Grudniewicz et al. 2019). Even with robust coverage and filtering, bibliometric data may retain biases or gaps in coverage, including a bias toward English-speaking regions, countries, or economies. In terms of interpretability, longer-term trends are the best way to view publications-related data. Year-to-year differences may be due to the process by which the information is indexed in Scopus. Additional details regarding document selection, limitations, and sources of bias are available in the **Technical Appendix**.

Information about how research was produced—such as the field; region, country, or economy of origin; and collaboration—may also be inferred from bibliometric data. For example, author affiliation data were used for determining publication output by region, country, or economy through fractional counting and international collaborations through whole counting. The supplemental tables include calculations using both whole and fractional counting for the various indicators to illustrate the difference in results. Articles were categorized by S&E fields corresponding to the 14 fields of science in the National Center for Science and Engineering Statistics (NCSES) Taxonomy of Disciplines (TOD) (Science-Metrix 2019). Additional details regarding fractional and whole counting, field categorization, and limitations are available in the **Technical Appendix**.

Publication Output by Region, Country, or Economy and by Scientific Field

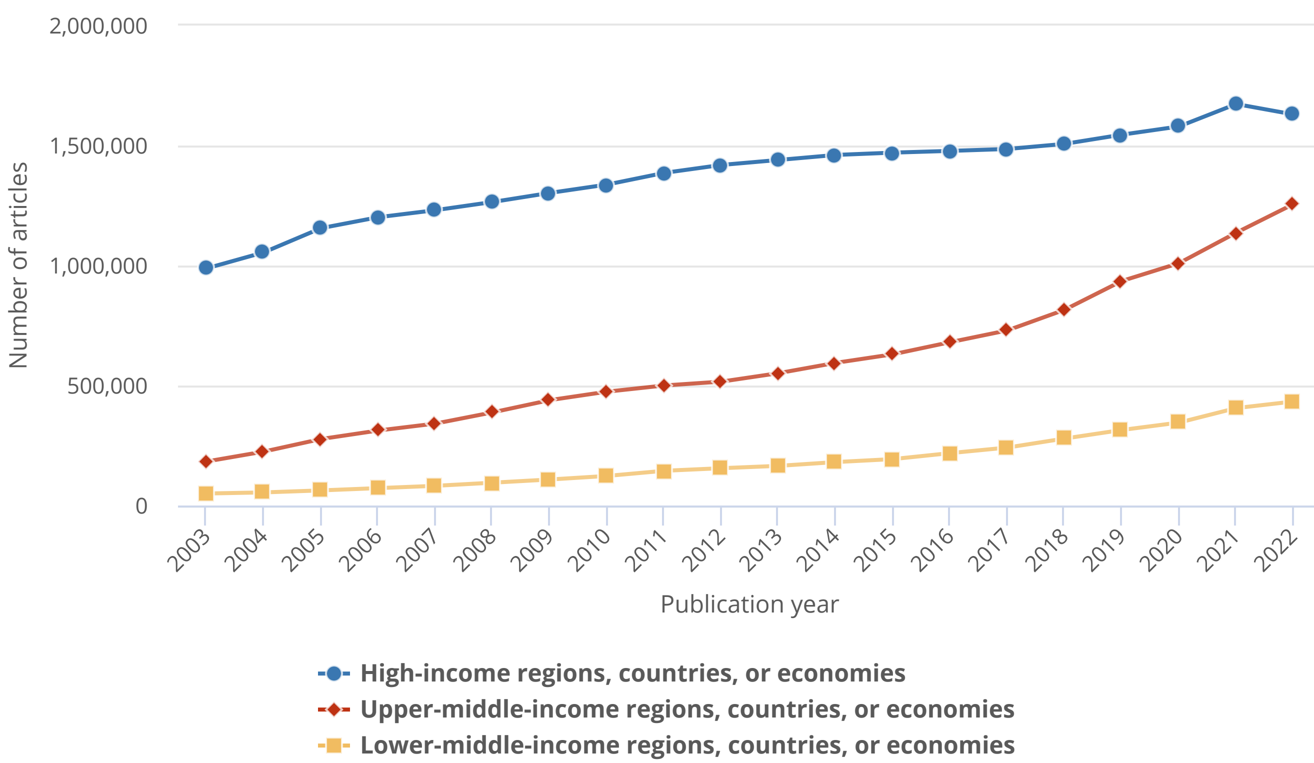
This section of the report outlines trends over time in publication output across regions, countries, or economies and by fields of science. This section also provides insights into the research contributions of different regions, countries, or economies and how the focus of their scientific publications has changed over time. In addition, the section highlights variations in the distribution of publications across scientific fields for different regions, countries, or economies and examines trends over time in closed-access and open-access (OA) publications. This section also summarizes federal funding acknowledgments as a source of data to shed light on published research that received federal funding. (See sidebar **Using Funding Acknowledgments to Track Federally Funded Research Over Time**.)

Output by Region, Country, or Economy

Total worldwide S&E publication output reached 3.3 million articles in 2022, based on entries in the Scopus database.[1] Approximately 86% of publications in 2022 came from regions, countries, or economies with high-income and upper-middle-income economies (**Figure PBS-1**). The number of publications from all income-level groups grew between 2003 and 2022 (Table SPBS-2). Also, the number of publications from upper-middle-income economies grew more quickly than the number from high-income economies during the more recent period between 2010 and 2022.

Figure PBS-1

S&E publications, by income group: 2003–22



Note(s):

Article counts refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles produced by authors from different regions, countries, or economies, each region, country, or economy receives fractional credit on the basis of the proportion of its participating authors). Data are not directly comparable with data from *Science and Engineering Indicators 2022*; see the Technical Appendix for information on data filters. Low-income regions, countries, or economies are not included in this figure because of their low publication output. Data by region, country, or economy and income group are available in Table SPBS-2. Regions, countries, or economies are allocated to income groups based on World Bank data, using their current designation. For example, all of China's publications from 2003 to 2022 are counted as part of the upper-middle-income category because that is China's current designation.

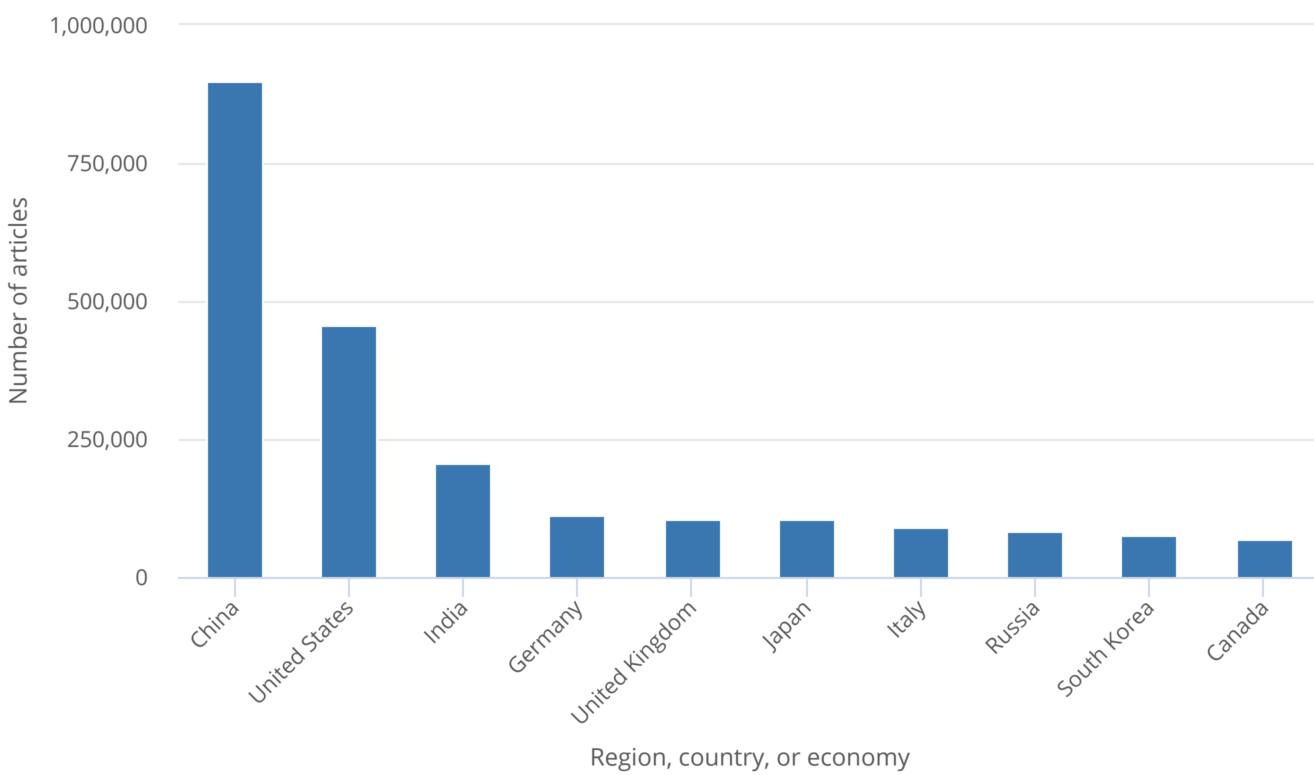
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023; World Bank Country and Lending Groups, accessed May 2023.

Science and Engineering Indicators

In 2022, six regions, countries, or economies each produced more than 100,000 articles: China, the United States, India, Germany, the United Kingdom, and Japan. Together, these leading regions, countries, or economies accounted for over 50% of the worldwide total in 2022 (**Figure PBS-2**; **Table PBS-1**).[2] In absolute numbers, the growth in worldwide annual publication output (from 2.0 million in 2010 to 3.3 million in 2022) was driven in particular by two countries: China (42% of additional publications during that period) and India (11%) together accounted for more than half of that increase in publications (**Figure PBS-3**). Russia, South Korea, Iran, and Brazil made notable contributions to the growth in the number of publications from the rest of the world from 2010 to 2022 (**Figure PBS-3**; Table SPBS-2). Generally, the set of the top 15 producers of S&E articles was the same each year between 2010 and 2022, with the exception of Iran replacing Taiwan in the top 15 beginning in 2014 (**Table PBS-1**; Table SPBS-2).

Figure PBS-2

S&E publications for 10 leading regions, countries, or economies: 2022



Note(s):
Article counts refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles produced by authors from different regions, countries, or economies, each region, country, or economy receives fractional credit on the basis of the proportion of its participating authors). Data by all countries, regions, and economies are available in Table SPBS-2.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Table PBS-1

S&E publications in all fields for 15 largest producing regions, countries, or economies: 2012 and 2022

(Number and percent)

Rank	Region, country, or economy	2012	2022	2022 world total (%)
na	World	2,105,157	3,344,037	na
1	China	329,067	898,949	26.88
2	United States	430,164	457,335	13.68
3	India	78,135	207,390	6.20
4	Germany	105,639	113,976	3.41
5	United Kingdom	98,685	105,584	3.16
6	Japan	109,040	103,723	3.10
7	Italy	64,131	90,586	2.71
8	Russia	36,532	84,252	2.52
9	South Korea	56,101	76,936	2.30
10	Canada	59,762	69,052	2.06
11	Spain	54,680	67,100	2.01
12	Brazil	48,166	67,031	2.00
13	France	72,431	65,888	1.97
14	Australia	46,728	62,305	1.86
15	Iran	31,462	60,940	1.82

na = not applicable.

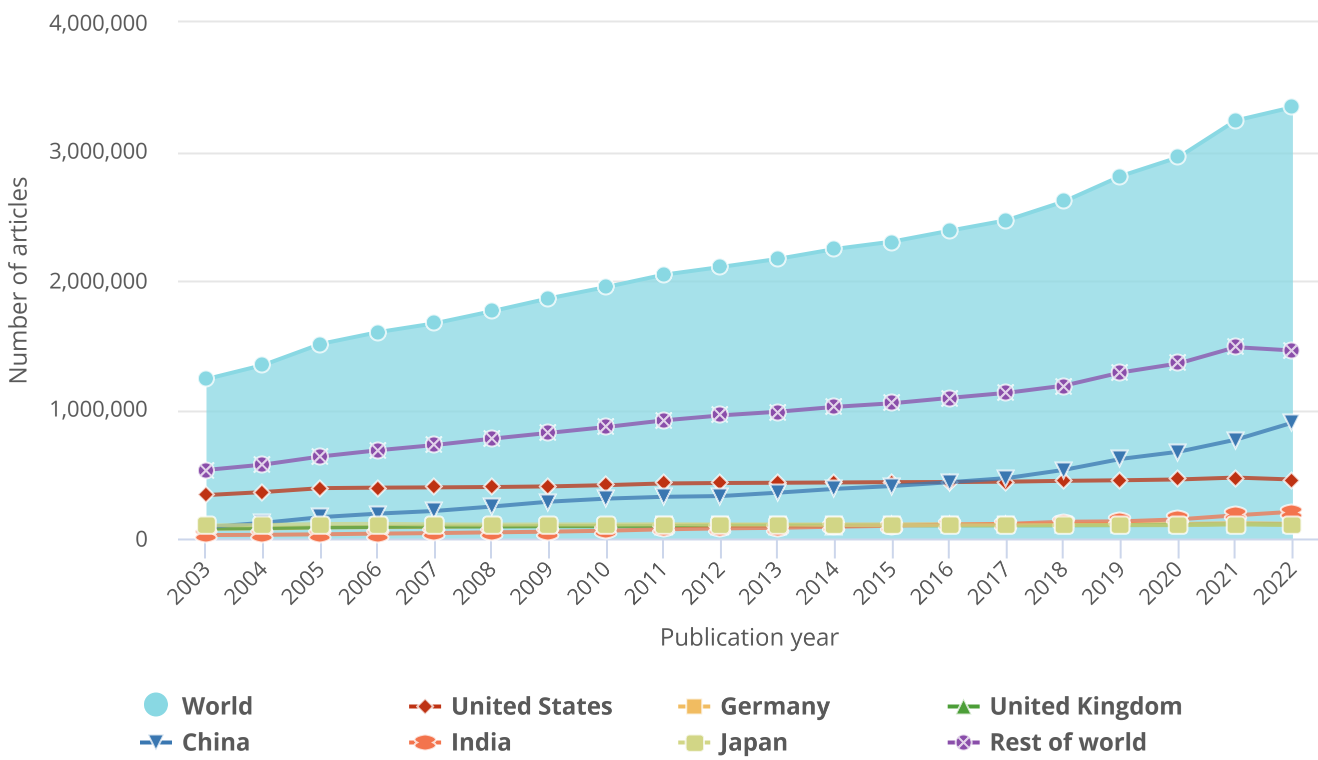
Note(s):
The regions, countries, or economies are ranked based on the 2022 total. Article counts refer to publications from conference proceedings and peer-reviewed journal articles in S&E and indexed in Scopus (see Technical Appendix for more details). Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles from multiple regions, countries, or economies, each region, country, or economy receives fractional credit on the basis of the proportion of its participating authors). Detail may not add to total because of regions, countries, or economies that are not shown. Proportions are based on the world total excluding unclassified addresses (data not presented). Details and other regions, countries, or economies are available in Table SPBS-2.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Figure PBS-3

S&E publications, by selected region, country, or economy and rest of world: 2003–22



Note(s):
Article counts refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles produced by authors from different countries, each country receives fractional credit on the basis of the proportion of its participating authors). Data for all regions, countries, and economies are available in Table SPBS-2.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

The U.S. trend of moderate but increasing publication output varied by state. The National Science Board's (NSB's) State Data Tool (**https://ncses.nsf.gov/indicators/states/**) provides state-level data based on each state's doctorate population and research and development (R&D) funding. Indicators include academic S&E article output per 1,000 science, engineering, and health doctorate holders in academia (NSB 2021a) and academic S&E article output per $1 million in academic S&E R&D funding (NSB 2021b).

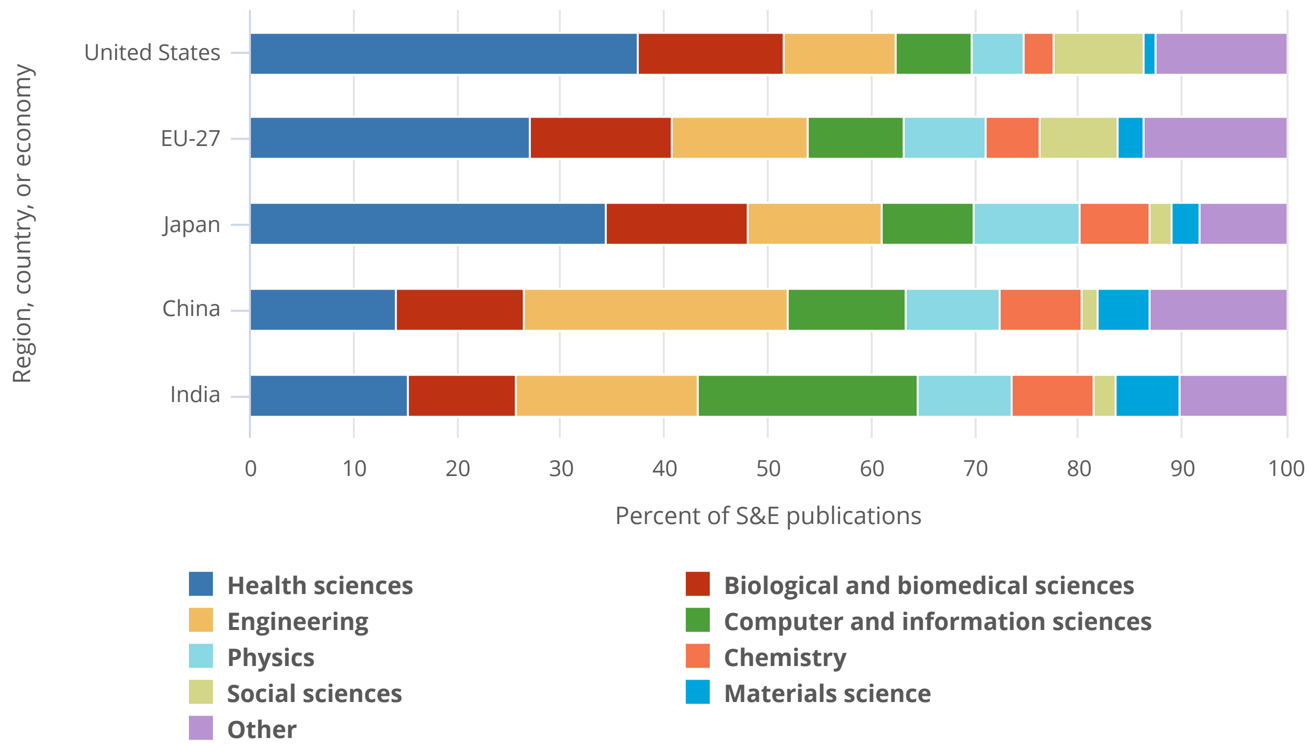
Output by Scientific Field

The distribution of publications by field of science across region, country, or economy may indicate research priorities and capabilities. Health sciences was the field of science in which most articles were published in 2022, representing almost a quarter of all publications (Table SPBS-2 and Table SPBS-10). Other fields with large numbers of publications included engineering (17% of publications in 2022), biological and biomedical sciences (13%), and social sciences (5%) (Table

SPBS-2, Table SPBS-5, Table SPBS-8, and Table SPBS-16). In the United States, the European Union (EU-27), and Japan, health sciences publication output in 2022 far exceeded that of any other field.[3] Meanwhile, of the other top producers, publications from China were most highly concentrated in engineering (25%), and publications from India were published predominantly in computer and information sciences (21%) (**Figure PBS-4**).

Figure PBS-4

Distribution of national S&E research portfolios across scientific fields, by selected region, country, or economy: 2022



EU-27 = European Union.

Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles from multiple countries, each country receives fractional credit on the basis of the proportion of its participating authors). See Table SPBS-1 for countries included in the EU; beginning in 2020, the United Kingdom was no longer a member of the EU. See Table SPBS-3 through Table SPBS-16 for data on all regions, countries, and economies by each S&E field.

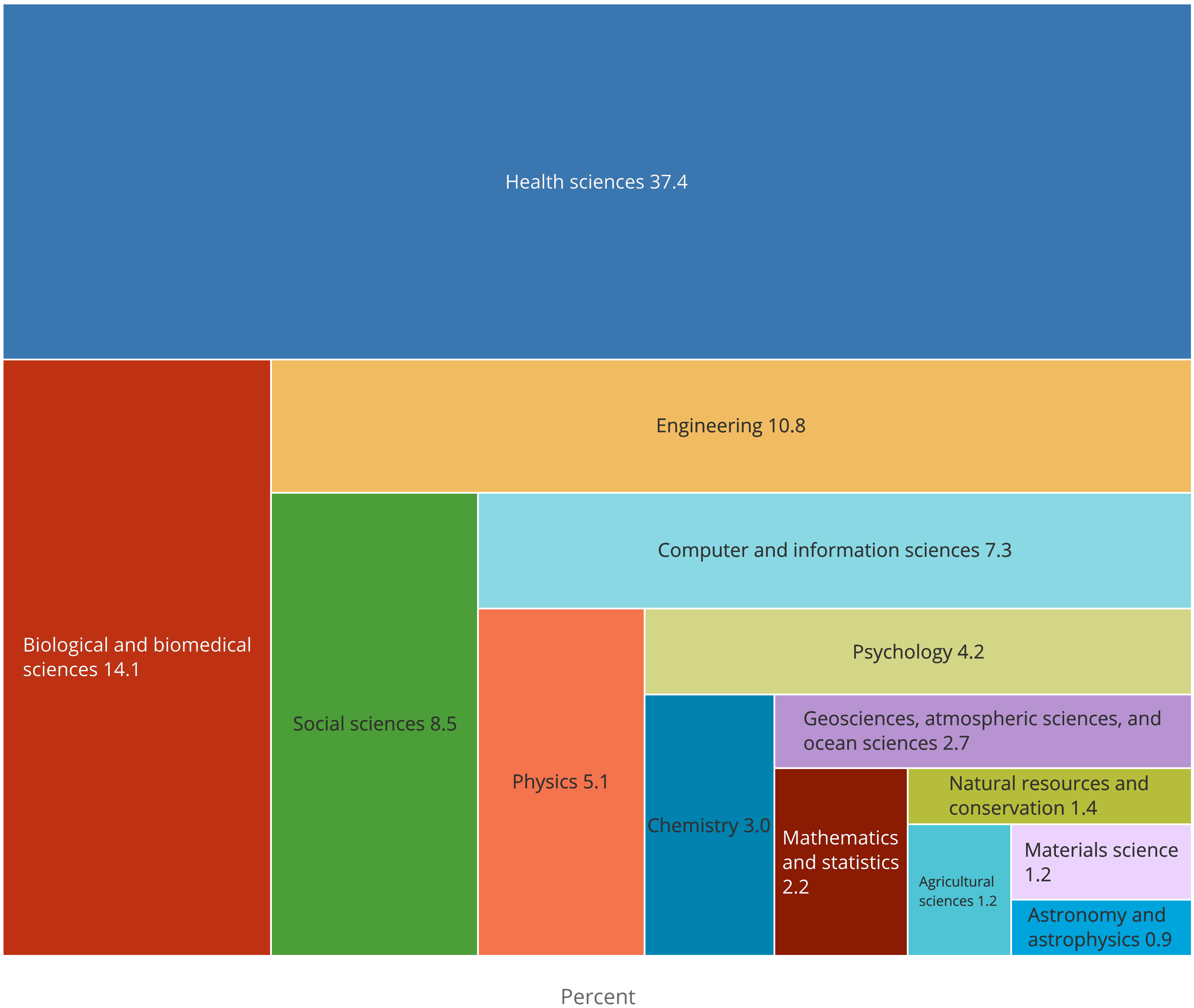
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Fields within life sciences were dominant in the United States in 2022, with more than half of all U.S. publications in health sciences (37%) or biological and biomedical sciences (14%) (**Figure PBS-5**). There were fewer U.S. publications in engineering (11%), computer and information sciences (7%), and physics (5%). In comparison with the United States, China had a stronger focus on publications in engineering and in the physical sciences and information sciences. In 2022, 25% of China's publications were in engineering, 11% were in computer and information sciences, and 9% were in physics (**Figure PBS-6**). Compared with the United States, China had a lower percentage of its publications in health sciences (14%) and biological and biomedical sciences (12%). In 2022, China also had a much lower percentage of its publications in social sciences (1%) when compared with the United States (8%).

Figure PBS-5

U.S. S&E publication portfolio, by field of science: 2022



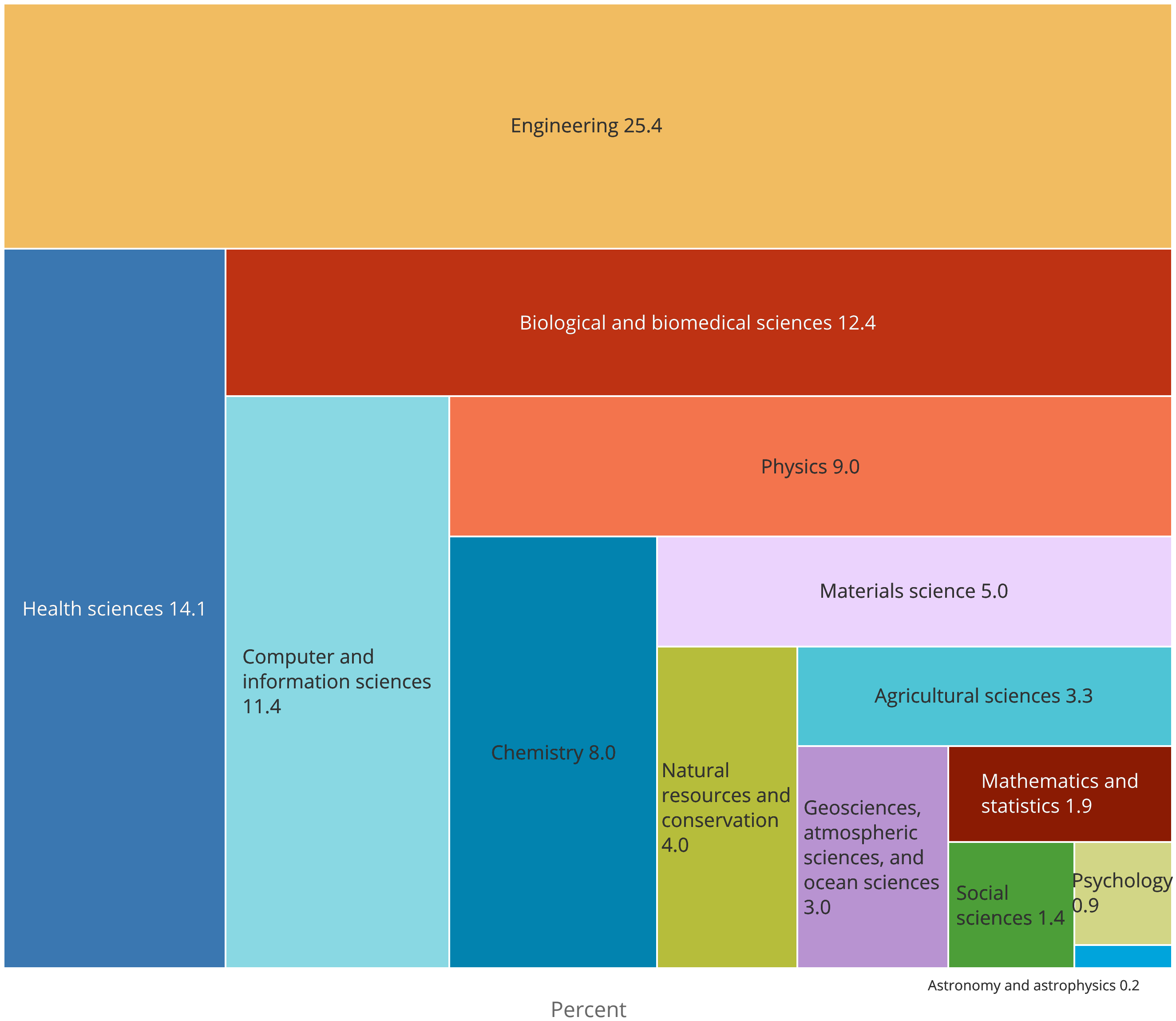
Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles from multiple countries, each country receives fractional credit on the basis of the proportion of its participating authors). See Table SPBS-3 through Table SPBS-16 for data on all regions, countries, and economies by each S&E field.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Figure PBS-6

S&E publication portfolio from China, by field of science: 2022



Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles from multiple countries, each country receives fractional credit on the basis of the proportion of its participating authors). See Table SPBS-3 through Table SPBS-16 for data on all regions, countries, and economies and by each S&E field.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

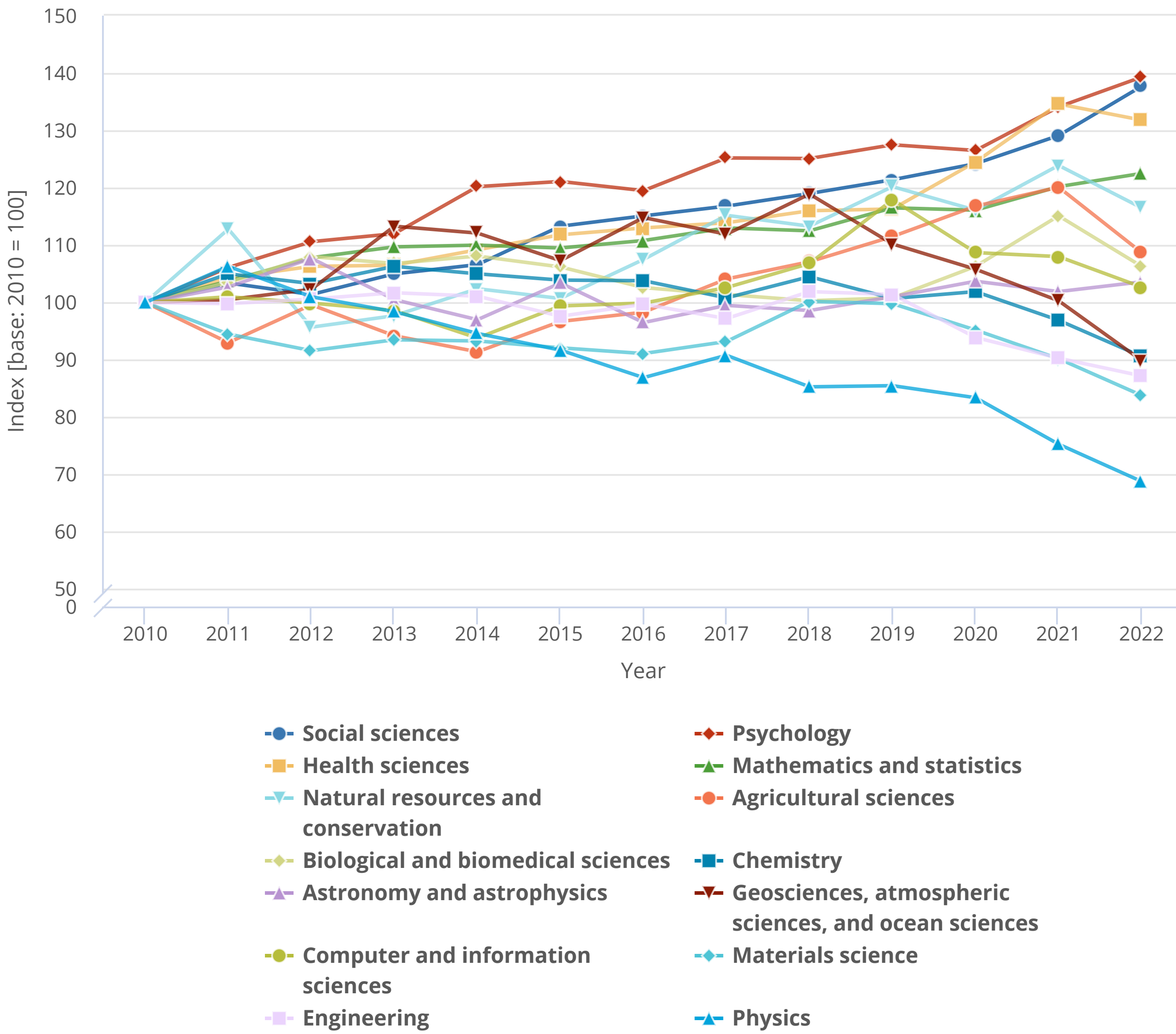
All the leading regions, countries, or economies saw an increase in their output of health sciences publications between 2010 and 2022. This increase is to be expected, given the context of increasing publication rates in general over that period, with overall number of publications increasing by 71% (Table SPBS-2), while publications in health sciences increased by 66% (Table SPBS-10). Russia had the highest relative growth rate among the 20 leading regions, countries, or economies in health sciences, increasing its publication output by almost 450% between 2010 and 2022 (Table SPBS-10). China and Iran each increased their output of health sciences publications by more than 250% over this period, while India's health sciences publication output increased by more than 180%. The United States increased its output of health sciences publications by 32% over this period, while Germany, France, the United Kingdom, and Japan had the smallest increases, each with less than 20%.

Leading regions, countries, or economies also saw increases in engineering publications. The fastest growing between 2010 and 2022 were India (up 378%) and Russia (up 230%) (Table SPBS-8). China increased its output of engineering publications by 176% from 2010 to 2022, while France, the United States, and Japan all saw declines in newly published engineering articles per year over this period (3%, 13%, and 26%, respectively).

In the United States, publication output varied from that of other regions, countries, or economies with respect to scientific fields. Of the fields not already mentioned, the fastest growing from 2010 to 2022 were psychology (up 39% from 2010 to 2022) and the social sciences (up 38%) (**Figure PBS-7**). Meanwhile, fields with the largest decreases in U.S. publications included physics (down 31% from 2010 to 2022) and materials science (down 16%).

Figure PBS-7

Index of U.S. publications, by field: 2010–22



Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a fractional count basis (i.e., for articles from multiple countries, each country receives fractional credit on the basis of the proportion of its participating authors). See Table SPBS-3 through Table SPBS-16 for data on all regions, countries, and economies by each S&E field.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

SIDEBAR

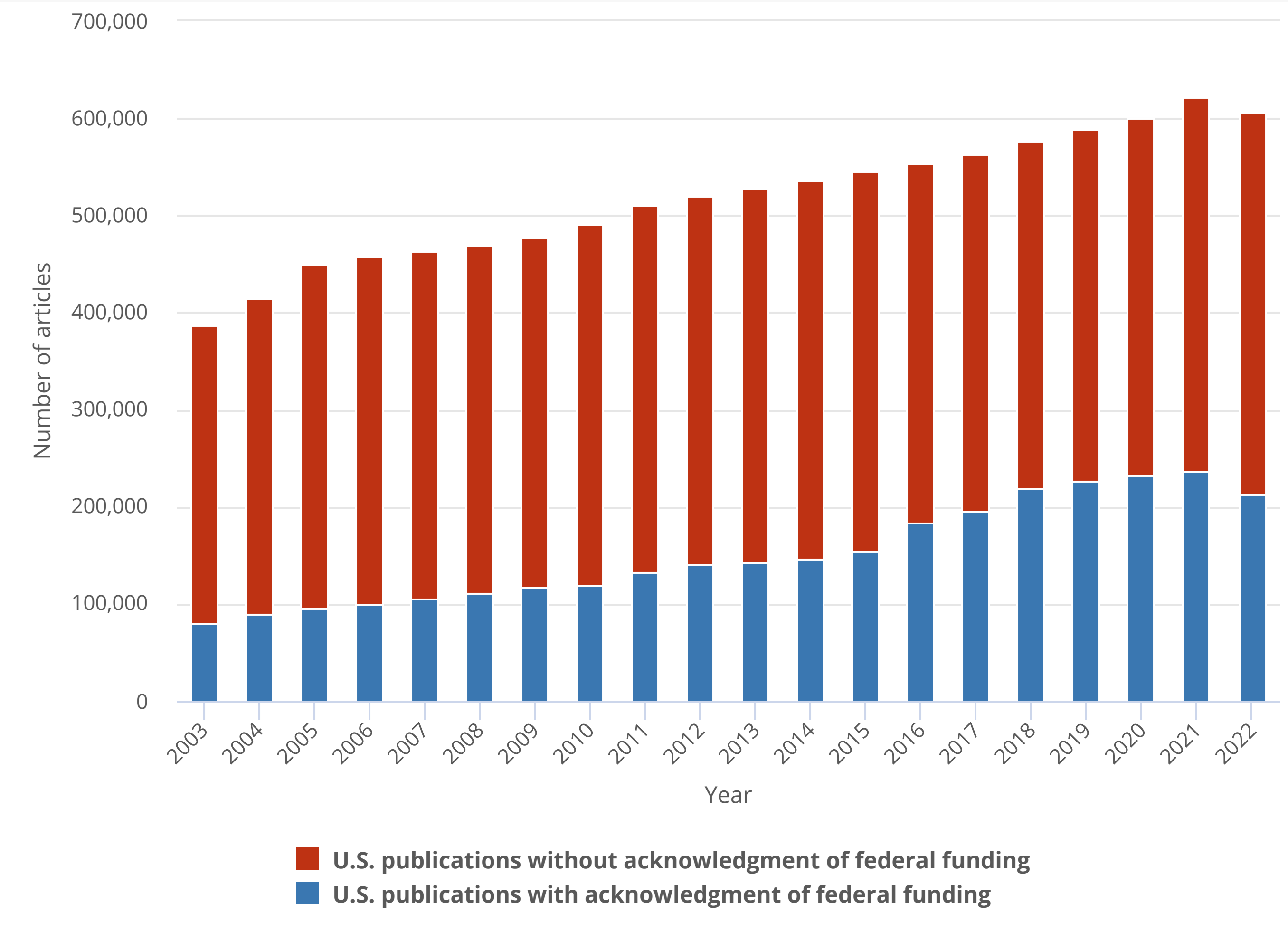
Using Funding Acknowledgments to Track Federally Funded Research Over Time

Federally funded research is an important component of the research ecosystem and is often envisioned as a means of supporting science performed for public benefit that may not otherwise be motivated by commercial interest (Bornmann 2013; Stephan 2012; Yin et al. 2022). Federal research funding supports applied and basic research (see *Indicators 2022* report "[2022] **Research and Development: U.S. Trends and International Comparisons**") and has long been linked to successful expansions in scientific production—through the increased productivity and impact of individual researchers and laboratories (Ebadi and Schiffauerova 2016) and the national scale (Leydesdorff and Wagner 2009). This sidebar explores funding acknowledgments, as recorded in Scopus, as an emerging source to help illustrate the extent to which published research is supported by federal agencies and the trends in federally funded research. Specifically, the share of published research acknowledging support by federal funding was highest in chemistry and smaller in other fields, such as the social sciences (**Table PBS-A**). These differences may be driven by factors such as the resource costs to conduct research and by field differences, such as the overall frequency of publication, team size, and cultural differences among the disciplines. The time period analyzed in this sidebar is 2018–22, unless otherwise indicated.

Funding acknowledgments can shed light on the ability and priorities of federal funding to support discovery as measured by peer-reviewed journal articles and conference proceedings. However, some benefits and limitations of this emerging data source are important to highlight so as to accurately interpret these trends. Each peer-reviewed journal article and conference proceeding in the Scopus database includes a field for funding acknowledgments that are extracted by algorithmic (software) means. In some cases where the acknowledgments field is incomplete, funding information from agencies is also used to identify funded publications in Scopus. Using this field, it is possible to observe the conversion of federal funds to published research outputs, but a direct linkage between funding inputs and published discoveries remains challenging. First, extraction of this information into a structured field is a relatively new effort and is most complete for the most recent 4 years. Figure SPBS-1 shows how funding acknowledgment sections have grown in coverage since 2003 and that funding information was indexed for 68% of all publications in 2022.[*] Many factors may have contributed to this growth in addition to improved extraction, including increasing pressure and requirements from funders to include funding acknowledgments, standardization of acknowledgment language, and incentives to demonstrate high publication output—because future funding is tied to past conversion of funds into publications—while receiving funding.[†] Last, this inquiry helps explore research that acknowledges any federal funding but does not only account for publications that source all their funding from a single source. In practice, a publication may be generated using funding from multiple sources within the federal government, or from additional sources in state government, local government, or the private sector.

Figure PBS-A

U.S. S&E publications with and without acknowledgments of U.S. federal funding: 2003–22



Note(s):
Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Whole counting is used. An article is considered to be federally funded if the funding information tied with the publication record in Scopus links it with one of the U.S. federal agencies. Not all Scopus publications have funding information available, and coverage has evolved with time. For more information, see Figure SPBS-1. For a breakdown of federally funded papers by funding agency, see Table SPBS-90.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Figure PBS-A tracks the growth of federally funded publications relative to the total research production in the United States. Other than a small downturn from 2021 to 2022, every year has seen an increase from the previous year in the number of publications that acknowledge funding support from federal agencies. The most comprehensive data from the past 4 years show variation among subject areas in the percentage of publications that acknowledge federal support. **Table PBS-A** shows number and share of publications appearing between 2018 and 2022 that acknowledged funding from federal sources and those acknowledging funding from other sources. During this time, more than 50% of publications in the following subject fields acknowledged federal funding support: chemistry (55% of publications), biological and biomedical sciences (53%), astronomy and astrophysics (53%), and physics (52%). Only two subject areas have less than 30% of publications with federal funding acknowledged: agricultural sciences (28%), and social sciences (15%). Otherwise, all other fields had between 30% and 50% of their publications acknowledging federal funding.

Table PBS-A

U.S. S&E publications, by U.S. federal funding status and field: 2018–22

(Number and percent)

Field	U.S. publications (total)	U.S. publications (federally funded)	Percentage of federally funded publications	U.S. publications (other funding)	Percentage of publications acknowledging funding from another source
Health sciences	1,004,671	318,838	31.7	239,606	23.8
Biological and biomedical sciences	447,843	237,472	53.0	151,511	33.8
Engineering	356,520	128,106	35.9	67,692	19.0
Computer and information sciences	235,765	79,218	33.6	29,528	12.5
Social sciences	202,900	29,694	14.6	30,282	14.9
Physics	191,700	98,716	51.5	39,480	20.6
Geosciences, atmospheric sciences, and ocean sciences	109,183	49,647	45.5	31,728	29.1
Psychology	107,480	34,853	32.4	23,678	22.0
Chemistry	103,217	56,361	54.6	30,797	29.8
Mathematics and statistics	63,733	27,634	43.4	12,644	19.8
Natural resources and conservation	52,010	17,979	34.6	14,311	27.5
Materials science	42,610	19,859	46.6	10,046	23.6
Agricultural sciences	42,419	11,701	27.6	10,384	24.5
Astronomy and astrophysics	34,358	18,050	52.5	8,382	24.4

Note(s):
Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Whole counting is used. An article is considered to be federally funded if the funding information tied with the publication record in Scopus links it with one of the U.S. federal agencies. Not all Scopus publications have funding information available, and coverage has evolved with time. For more information, see Figure SPBS-1. For a breakdown of federally funded papers by funding agency, see Table SPBS-90.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

In conclusion, federal funding plays an important role in the current research environment in the United States. Of the 606,144 articles published in journals and conference proceedings in 2022, 35% acknowledged support from federal agencies (**Figure PBS-A**). Ultimately, acknowledgment of federal funding can help show trends in the conversion of grants into published research over time and show variation at the subject or field level.

* Missing data in funding fields in a Scopus entry may mean that the research did not receive funding, the authors did not cite any funding despite receiving it, or the algorithm was unable to extract the acknowledgment. Of the articles from 2003 that had an entry for funding acknowledgment (27% had text in the funding field in Scopus), around 76% acknowledged a federal funding source. Comparatively, of the publications in 2022 with indexed funding information (68%), 52% acknowledged a federal source. The growth of coverage of funding not being tied to federal funding acknowledgments provides evidence that the data source has become more dependable over time. Data for the percentage of publications with indexed funding sources by year and field can be found in Figure SPBS-1.

† Table SPBS-90 displays the number of articles and conference proceedings acknowledging federal funding at the agency level and sub-agency level. These counts represent the number of supported articles as acknowledged and attributed in Scopus from 2003 to 2022.

Output and Open Access

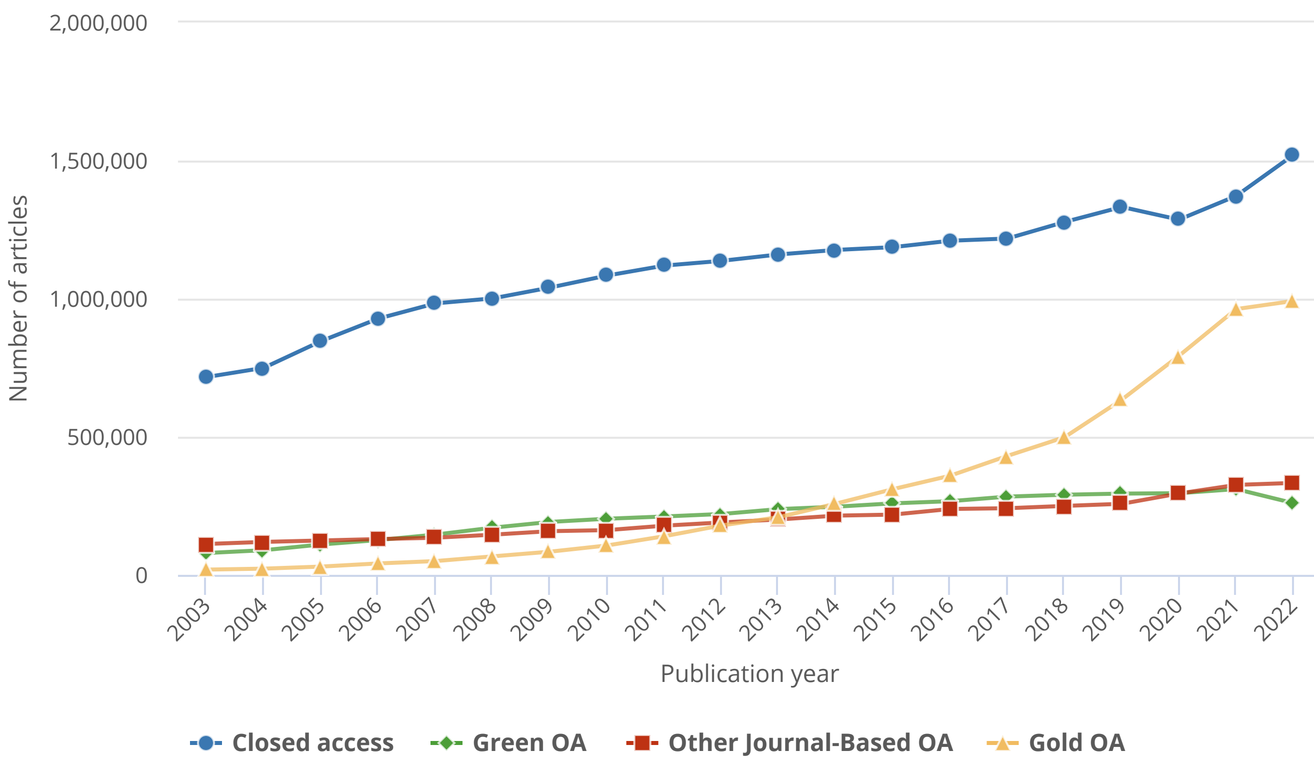
There is growing support for the availability of S&E publications through OA sources among government and private funders, institutions, and scientists themselves. Some of these funders have imposed requirements on their grantees to publish their research results in OA journals. In the United States, the Office of Science and Technology Policy announced that all federal agencies should update their public access policies as soon as possible to ensure that results of their funded research are publicly available, with full implementation of these policies by the end of 2025 (Brainard and Kaiser 2022). Meanwhile, restricted access to scientific literature may impede researchers' ability to stay informed (Larivière and Sugimoto 2018; Piwowar et al. 2018). As alternatives to traditional closed-access journals (where readers must subscribe to gain access or pay per article), articles may be made OA through several avenues, with different levels of availability and durability.

There are four commonly defined types of OA: Gold, Hybrid, Bronze, and Green. Gold OA denotes articles published in journals that are entirely OA as a matter of journal policy. Hybrid OA denotes articles for which the authors have elected to pay a fee for publication as OA rather than as closed access. Bronze OA denotes articles that appear as OA after an embargo period of closed access or articles that appear available as OA despite lacking license information to guarantee OA in the long term. Green OA denotes articles that are self-archived by authors in OA repositories, which are often maintained, curated, and administered by universities or other institutions. The Hybrid and Bronze categories have been combined as Other Journal-Based OA in this report because of their similar structure as journal-hosted types of OA that allow only conditional—and potentially revocable—OA.

The number of articles published annually in closed-access journals increased by 112% between 2003 and 2022 (**Figure PBS-8**). Over the same period, annual publishing of Green OA articles increased by 228%, while Other Journal-Based OA articles (Hybrid and Bronze OA) increased by 198%. Gold OA articles (which are published in OA journals with no restrictions) had the largest percentage growth, from 19,089 articles in 2003 to 991,805 articles in 2022, an increase of over 5,000%. Hence, although the majority (77%) of S&E articles in 2003 whose access status is known were published in closed-access journals, fewer than half (49%) were in closed-access journals in 2022.

Figure PBS-8

S&E publications, by publication access type: 2003–22



OA = open access.

Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication. OA types are mutually exclusive. For articles published under multiple OA types, the article will be counted as part of only the first type it matches in this list: Gold OA, Other Journal-Based OA, or Green OA. Summing all OA and closed-access article counts results in a smaller number of articles than for all S&E because the access status of some articles (e.g., those without digital object identifiers) cannot be reliably ascertained. Green articles are published in toll-access journals but archived in an OA archive, or "repository." These repositories may be discipline specific (like arXiv) or institutional repositories operated by universities or other institutions. Green articles may be published versions or preprints and can have any license or no license. Bronze (Other Journal-Based OA) articles are free to read on the publisher's website, without a license that grants any other rights. There may be a delay between publication and availability to read, and often articles can be removed unilaterally by the publisher. Hybrid (Other Journal-Based OA) articles are free to read at the time of publication, with an open license. These are usually published in exchange for an article processing charge. Gold articles have all the same characteristics as Hybrid articles but are published in all-OA journals, which are in turn called "Gold journals" or just "OA journals."

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

To conclude this section, the findings of the output analysis reveal the growth in scientific publications over time, with upper-middle-income economies exhibiting particularly large percentage increases. Meanwhile, the distribution of publications across scientific fields shows that life sciences dominated in the United States, Europe, and Japan, whereas publications in engineering and computer sciences dominated in China and India. In OA, the dramatic growth of Gold OA publications and the steady growth of publications in other OA categories show an increased shift toward open science. However, OA can impede the dissemination of some scientific research. Publishing research as OA often requires authors to pay article processing fees, which may be prohibitive for scientists in less-developed nations or whose funders do not subsidize those fees.[4] The fees can be seen as shifting the costs of accessing research from readers and libraries to authors (Larivière and Sugimoto 2018).

Impact of Published Research

As a complement to measures of the total production of articles published by U.S.-based researchers, indicators gauging the scientific significance of those articles aid in assessing the vitality and competitiveness of the U.S. S&E enterprise. A key assertion in the study of scientific communication is that the author of a scientific journal article cites the previously published articles that influenced or informed the author's research (Merton 1973; Tahamtan and Bornmann 2018). Thus, when an article is referenced by many other articles, it is deemed to have exceptional scientific impact (Garfield 1973). Alternate explanations for citation patterns exist, such as the belief that they are a mechanism for the citing author to bestow credit on an earlier author (Small 2004). These interpretations of citation patterns suggest that authors who produce highly cited articles have greater scientific influence than those who do not. By extension, nations whose authors produce a disproportionate share of the most highly cited articles in a given field are regarded as generating research results with greater impact.

In general, most of the scientific articles published garner few or no citations, whereas for a small share, each article ends up with hundreds or thousands of citations (Van Noorden, Maher, and Nuzzo 2014). Recently published articles tend to have fewer citations than those published many years ago because articles are referenced only after they have been disseminated and read throughout the research community. Counting the cumulative citations earned by an article at least 2 years after publication provides the means to normalize the measure of impact for a given year. This 2-year lag provides sufficient time for high-impact articles to attract attention and therefore receive a distinguishing number of citations (Wang 2012). Those in the top 1% of the articles based on this measurement of citations are designated as highly cited articles (HCAs). Therefore, this report provides impact measures for articles up to those published in 2020, calculated using the citation counts for those articles at the end of 2022. Note that some articles may receive a much higher share of their citations many years after publication (Ke et al. 2015).

A method of representing scientific impact at the national level calculates the share of a specific nation's scientific output among the HCAs in a given year. To identify the HCAs for each year, publications in each scientific field are ranked based on their accumulated citations, generating a list of the top 1% of that distribution. A separate list is generated for each field to account for differences in citation practices and patterns among the various disciplines (Science-Metrix 2021b). Aggregating the lists for all fields produces the HCA list; that list is the compilation of the top 1% of articles in each field by citations rather than the top 1% overall. Each article on the list is attributed to a nation if at least one author on the publication is affiliated with an institution located in that region, country, or economy (meaning that one article will be attributed to every nation represented among its coauthors).

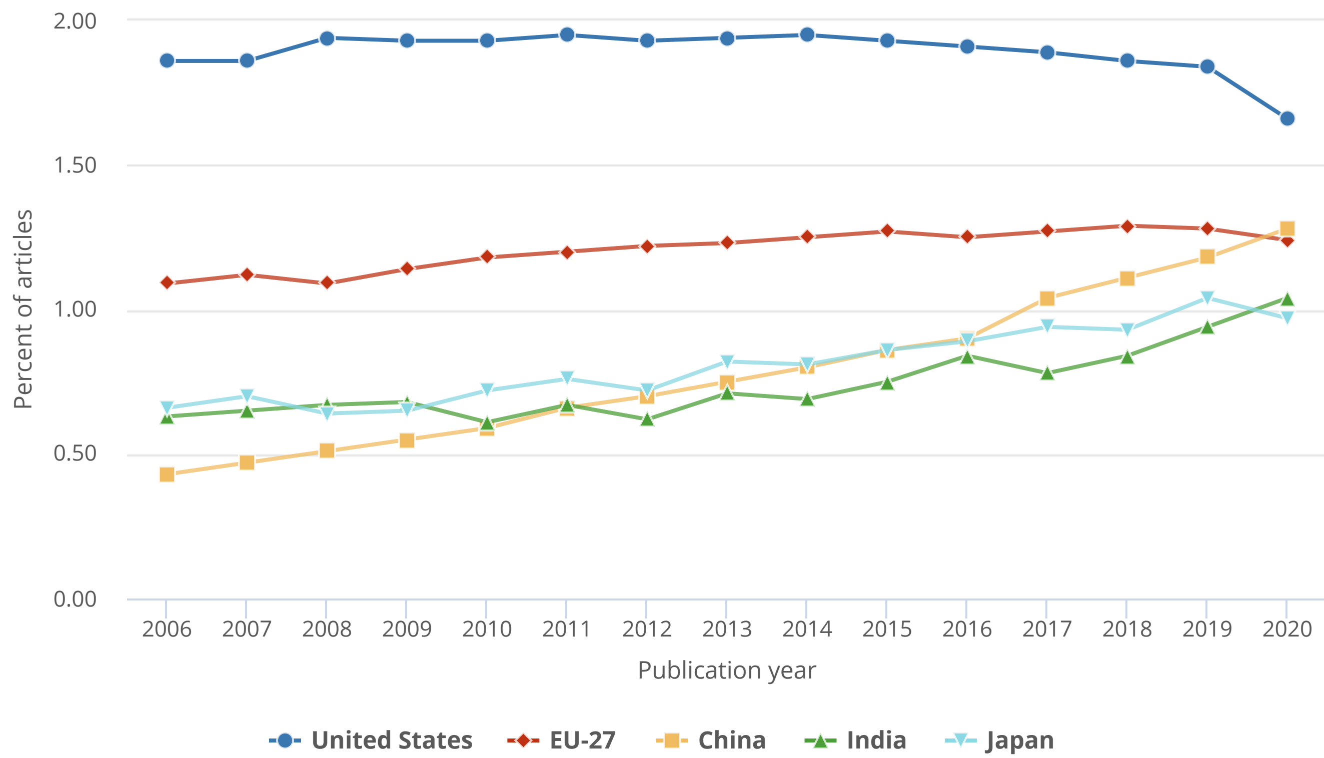
The estimate of scientific impact is based on the share of a specific region, country, or economy's articles designated as HCAs. For a region, country, or economy with impact on par with that of the entire global scientific community, 1% of that region, country, or economy's scientific articles would be found in the top 1%. An HCA share above 1.0 means that more than 1% of the nation's articles achieved HCA status, indicating that the nation's published research has disproportionately more impact relative to the volume of articles its researchers produce.[5] A consistent increase in a region, country, or economy's HCA ratio over time suggests that its S&E enterprise is rising in stature as a producer of notable research findings.

Impact by Region, Country, or Economy

For 2020, 1.7% of U.S. scholarly publications with U.S. authors were on the HCA list, showing that the United States generates a disproportionate share of those articles (**Figure PBS-9**). The HCA share for the United States had stayed above 1.9% in each year from 2008 through 2016. The nation's HCA share has declined each year since 2014 and stands at its lowest level since 2006.

Figure PBS-9

S&E publications in the top 1% most-cited journal articles as a share of all S&E journal articles, by selected region, country, or economy: 2006–20



EU-27 = European Union.

Note(s):
Articles refer to articles from peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Citation data are based on all citations made to articles in their publication year and all following years and are normalized by subfield and publication year to allow for comparisons across subfields and over time, resulting in the world level standing at 1.00 for each subfield and year. A minimum 2-year citation window is needed for a highly cited article (HCA) score to be computed. This results in scores regarding HCA not being computed after 2020 because the citation window for more recent years is not yet complete. The share of articles in the top 1% is computed as follows: $Sx = HCAx / Ax$, where Sx is the share of output from country x in the top 1% most-cited articles; $HCAx$ is the number of articles from country x that are among the top 1% of most-cited articles (using full counting, with the exception of articles at the limit of the top 1%, which are fractioned so the world average can stand at 1%); and Ax is the total number of articles from country x with a relative citation score, which excludes articles released after 2020 and unclassified publications. The world average is 1.00 for each year. For more details, see Table SPBS-70. See Table SPBS-1 for countries included in the EU; beginning in 2020, the United Kingdom was no longer a member of the EU.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

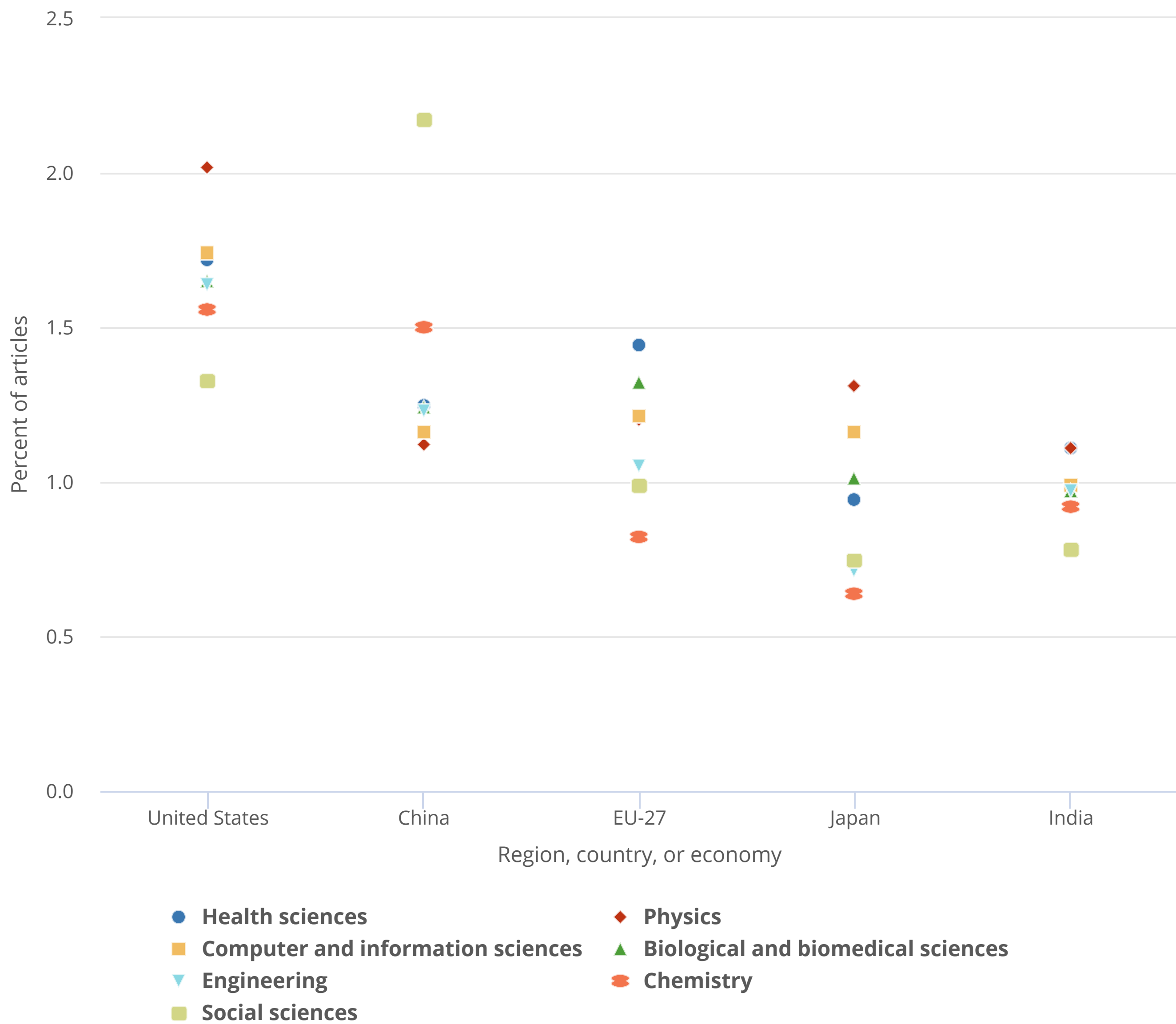
In contrast, China's HCA share has increased in each of the past 15 years, rising from 0.4% in 2006 to 1.3% in 2020 (**Figure PBS-9**). The disparity in HCA shares between the United States and China has decreased steadily since 2008. One factor that may affect this trend is the increase in the number of publications with U.S. and Chinese coauthors, discussed in the section **International Collaboration Patterns**. Any such articles in the 1% of the most highly cited publications in a given year would be credited toward the HCA shares of both countries. Given that on a whole-count basis, China produced more scientific articles than the United States in 2020, this convergence appears to represent a consistent increase in the scientific impact of publications with Chinese authors (Table SPBS-17).

China's HCA share in 2020 exceeded that of the EU-27 region for the first time (**Figure PBS-9**). The collective HCA share for the EU-27 has stayed at or above 1.2% since 2010. The HCA share of Japan, another major contributor to the world's scientific literature, has tended to lag behind the overall global HCA share. Japan's HCA share reached 1.0% in 2019 for the first time since at least 2006 and maintained that level in 2020. India's HCA share matched the global HCA share in 2020 for the first time, reaching 1.0%. Note that scientists who publish predominantly in non-English-language journals may receive lower citations in general because the language barrier could limit the readership of their articles in the international community (Di Bitetti and Ferreras 2016).

Figure PBS-10 shows the field-specific HCA shares of the five regions, countries, or economies with the largest number of scientific publications in 2022, focusing on the fields representing the largest share of scientific articles in 2020 (Table SPBS-3 through Table SPBS-16). This figure illustrates how national scientific impact varies by discipline. HCA shares by whole count for all regions, countries, and economies across all fields are provided in Table SPBS-72 through Table SPBS-85.

Figure PBS-10

S&E publications in the top 1% most-cited journal articles as a share of all journal articles for selected S&E fields, by selected region, country, or economy: 2020



EU-27 = European Union.

Note(s):
Articles refer to articles from peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are allocated to a region or country on a whole count basis. Citation data are based on all citations made to articles in their publication year and all following years and are normalized by subfield and publication year to allow for comparisons across subfields and over time, resulting in the world level standing at 1.00 for each subfield and year. A minimum 2-year citation window is needed for a highly cited article (HCA) score to be computed. This results in scores regarding HCA not being computed after 2020 because the citation window for more recent years is not yet complete. The share of articles in the top 1% is computed as follows: *Sx* = *HCAx* / *Ax*, where *Sx* is the share of output from country *x* in the top 1% most-cited articles; *HCAx* is the number of articles from country *x* that are among the top 1% of most-cited articles (using full counting); and *Ax* is the total number of articles from country *x* with a relative citation score, which excludes articles released after 2020 and unclassified publications. The world average is 1.00 for each year. For more details, see Table SPBS-70 through Table SPBS-84. See Table SPBS-1 for countries included in the EU; beginning in 2020, the United Kingdom was no longer a member of the EU.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

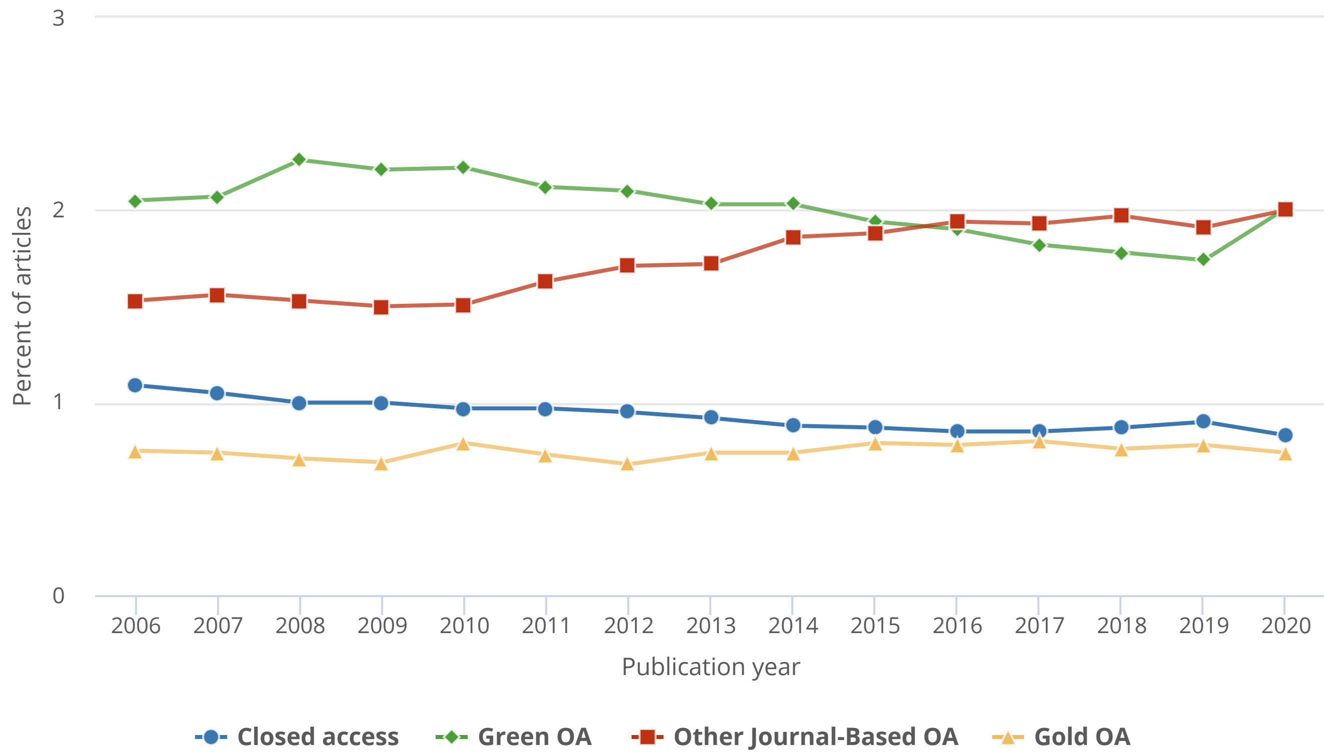
In the United States, physics had the highest share of publications in 2020 designated as HCAs (**Figure PBS-10**). Other fields in which the United States had a high HCA share were geosciences, atmospheric sciences, and ocean sciences and materials science (Table SPBS-77 and Table SPBS-79). Note that physics is not a dominant field in the United States in terms of number of publications (**Figure PBS-5**). For China, the social sciences had the highest share of articles designated as HCAs in 2020 (**Figure PBS-10**), although that field accounts for a relatively small share of China's total publications (see **Figure PBS-6**). The EU-27 region demonstrated higher relative impact in articles in health sciences, as well as geosciences (**Figure PBS-10**; Table SPBS-77). In Japan, publications in the fields of astronomy and astrophysics, geosciences, and physics showed high HCA shares (**Figure PBS-10**; Table SPBS-72 and Table SPBS-77). India showed relatively higher impact in physics than in other fields and had particularly notable impact in astronomy and astrophysics (**Figure PBS-10**; Table SPBS-72).

Impact and Open Access

As discussed earlier, the share of scientific articles published as OA has increased dramatically since 2012 (**Figure PBS-8**). Researchers have suggested that OA articles have a greater potential readership than articles published in closed-access journals and therefore are more likely to have higher citation rates. A recent review of such studies showed mixed evidence for that claim (Langham-Putrow et al. 2021). Worldwide analysis of highly cited articles by publication access type also shows mixed results (**Figure PBS-11**). The share of articles published under closed access in the 1% of most highly cited articles per year fell from 1.1% in 2006 to 0.8% in 2020. The HCA share for Green OA journals stayed at or above 2% for most years since 2006. In contrast, the HCA share for combined Hybrid and Bronze OA journals ("Other Journal-Based OA") rose from 1.5% in 2009 to 2% in 2020. As noted earlier, Gold OA journals now have the largest share of scientific publications by access type (**Figure PBS-8**). However, the HCA share for those journals has lagged behind that of the other categories for the past 15 years, staying in the 0.7%–0.8% range for most of that period. Significant publishers of Gold OA journals, such as the Multidisciplinary Digital Publishing Institute (MDPI), are increasing the number of articles published per year at a high rate, meaning that the number of articles required to meet the 1% HCA threshold is also growing rapidly.[6]

Figure PBS-11

S&E publications in the top 1% most-cited journal articles as a share of all articles, by publication access type: 2006–20



OA = open access.

Note(s):
Articles refer to articles from peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Citation data are based on all citations made to articles in their publication year and all following years and are normalized by subfield and publication year to allow for comparisons across subfields and over time, resulting in the world level standing at 1.00 for each subfield and year. A minimum 2-year citation window is needed for a highly cited article (HCA) score to be computed. This results in scores regarding HCA not being computed after 2020 because the citation window for more recent years is not yet complete. The share of articles in the top 1% is computed as follows: *Sx = HCAx / Ax*, where *Sx* is the share of output from country *x* in the top 1% most-cited articles; *HCAx* is the number of articles from country *x* that are among the top 1% of most-cited articles (using full counting, with the exception of articles at the limit of the top 1%, which are fractioned so the world average can stand at 1%); and *Ax* is the total number of articles from country *x* with a relative citation score, which excludes articles released after 2020 and unclassified publications. The world average is 1.00 for each year. For more details, see Table SPBS-86 through Table SPBS-89.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

To conclude this section, when looking at trends in the articles with the highest citation rates per year, the United States produces articles that gain significant recognition after publishing at a disproportionate rate compared with global publications each year. In contrast, although China's annual article output exceeds that of the United States, China's articles are not as well represented among the most highly cited articles. However, the impact of articles published by Chinese authors has seen significant annual increases since 2006; as of 2020, China's presence among the most highly cited articles exceeds that of the EU-27 and Japan.

International Collaboration and Citations

S&E research has become more global over the past decade—a trend evidenced cross-nationally (Glänzel and Schubert 2005; Luukkonen et al. 1993; Royal Society 2011). Researchers gather scientific expertise beyond their region, country, or economy's borders through collaboration (coauthoring articles) and citation (referencing articles from other regions, countries, or economies). Measured at the region, country, or economy level, international collaboration and citation are strongly influenced by the size and the policies of the region, country, or economy. For example, some regions, countries, or economies provide preferential funding for cross-national research within a region through programs such as the European Commission's *Horizon Europe* (European Commission 2021).

This section of the report examines trends in collaborations between researchers—as measured by coauthorships and citations—particularly among those involving international connections. In the sidebar **Artificial Intelligence Publication Output and International Collaboration**, this section also explores international collaboration in greater depth in the field of artificial intelligence using network analysis.

Researchers may collaborate for several reasons, including to develop a scientific relationship with another researcher or to gain access to costly or shared equipment. They may also work together to meet conditions attached to research funding that require international collaboration (Wagner 2018).

In general, national governments encourage international collaboration to achieve outcomes that exceed what they could achieve individually (although they may perceive risks in collaborating with regions, countries, or economies they regard as a potential threat). These positive outcomes include training a robust S&E workforce, partnering with researchers from developing countries, advancing domestic science excellence, increasing the impact of discoveries through better distribution of knowledge, strengthening scientific and diplomatic relations, and enhancing a sense of shared responsibility for future action (Lyons et al. 2016). Also, international collaborations increase the impact of research, as measured by citations (Chinchilla-Rodríguez, Sugimoto, and Larivière 2019; Narin, Stevens, and Whitlow 1991; Sugimoto et al. 2017). Domestic collaboration among researchers in the academic, government, and industry sectors also results in articles that receive higher citation rates when compared with articles from a single author or authors from only one sector (see the section Business Collaborations in Published Literature in *Indicators 2022* report "[2022] **Invention, Knowledge Transfer, and Innovation**").

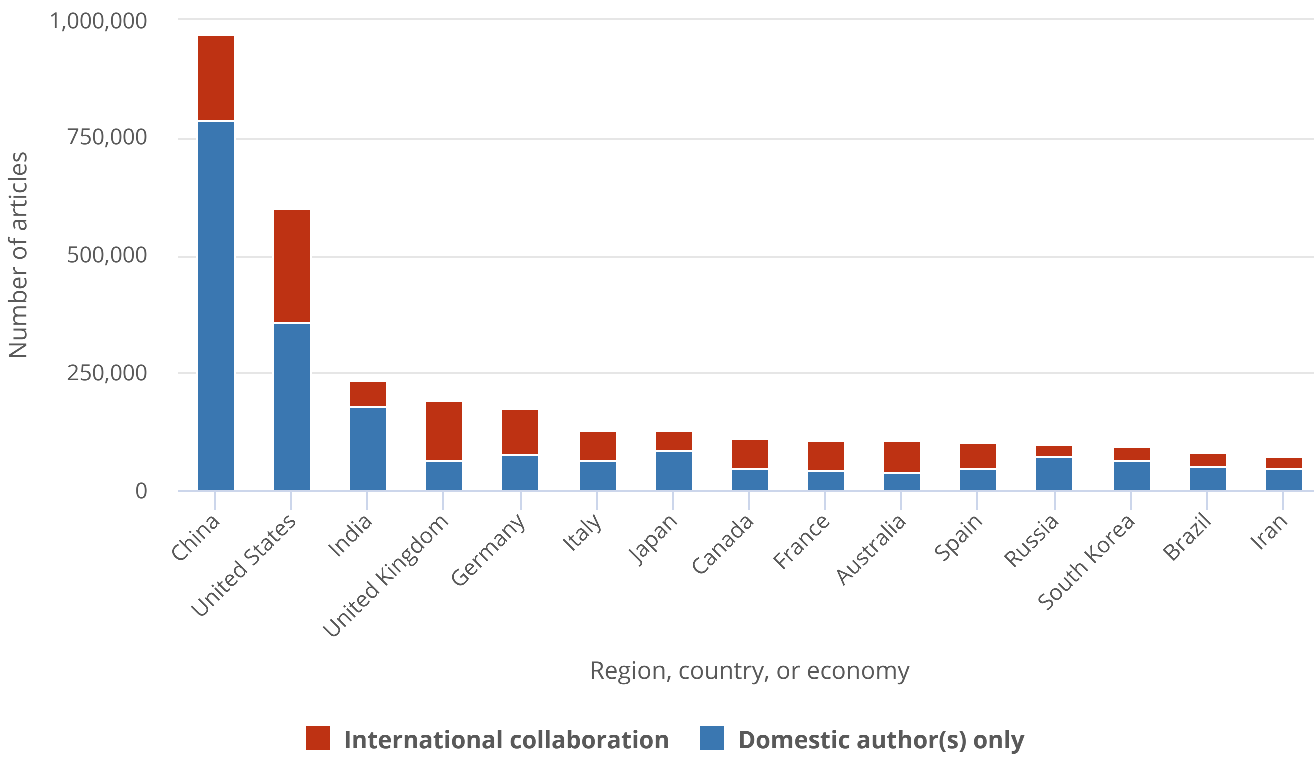
International Collaboration Patterns

An article is classified as an international collaboration if at least two author organizations are located in different regions, countries, or economies, as determined by their addresses on the article. In 2022, the United States contributed to the largest number of articles involving international coauthorship (241,823 articles, representing 32% of all internationally coauthored articles) (Table SPBS-33). The most frequent coauthorship partners for the United States were China, the United Kingdom, Canada, and Germany (Table SPBS-35).

In 2022, the global rate of international collaboration was 23%, but these rates varied by region, country, or economy. Researchers in the United States collaborated with international partners on 40% of their articles in 2022 (Table SPBS-33). Of the top 15 largest producers, the regions, countries, or economies that had higher international collaboration rates than the United States included the United Kingdom (67%), Australia (63%), France (60%), and Canada (60%). Conversely, regions, countries, or economies with rates of international collaboration lower than the U.S. rate included China (19%), India (24%), and Russia (25%) (**Figure PBS-12**). Beyond the 15 largest producers of publications, the 2022 international collaboration rates varied—Saudi Arabia (80%), Switzerland (74%), and Belgium (73%) had higher collaboration rates than the United States, whereas Turkey (29%) and Brazil (38%) had lower collaboration rates, albeit still higher than those of China, India, and Russia (Table SPBS-33).[7]

Figure PBS-12

International coauthorship of S&E publications for the 15 largest producing regions, countries, or economies of S&E publications: 2022



Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating region, country, or economy is credited with one count). Articles without international coauthorship are counts of articles with one or more institutional addresses all within a single region, country, or economy, which include single-author articles and articles coauthored under the same institutional address. International articles are articles with institutional addresses from more than one region, country, or economy. The numbers of articles from the international collaboration and domestic author(s) only categories may not sum to the total article number because some coauthored publications have incomplete address information in the Scopus database. These publications often cannot be reliably identified as international or domestic collaborations. For this reason, they are not included in either subcategory but are still counted toward the total number of articles. For more detail, see Table SPBS-37.

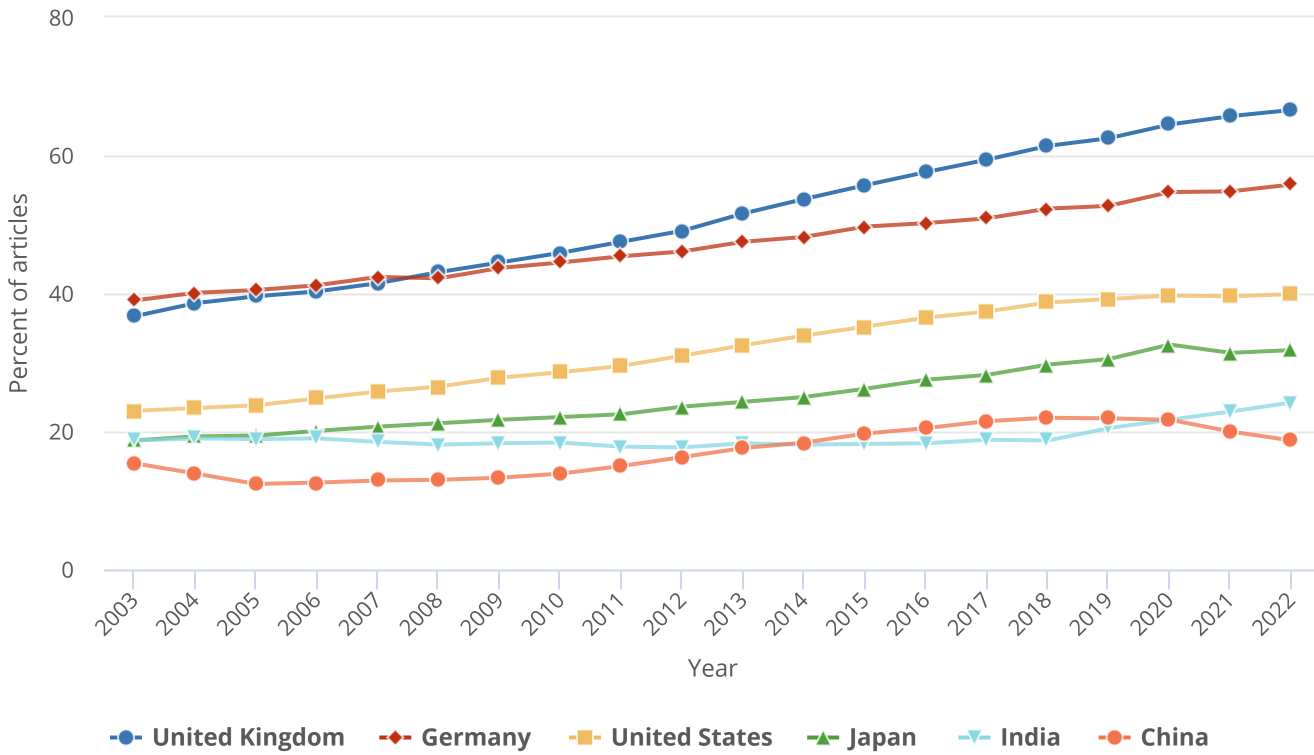
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

The percentage of worldwide S&E articles produced with international collaboration has grown over time, increasing from 19% in 2012 to 23% in 2022 (Table SPBS-33).[8] Each region, country, or economy leading in publications showed increases in international collaboration rates (**Figure PBS-13**). United Kingdom researchers had international coauthors on 37% of their articles in 2003, a percentage that had increased to 67% by 2022. Similarly, Germany's international collaboration rate increased from 39% to 56% over the same period. The United States and Japan both saw notable increases in international collaborations between 2003 and 2022 (from 23% to 40% and from 19% to 32%, respectively), whereas there was less change in the rates for China (from 15% to 19%) and India (from 19% to 24%).

Figure PBS-13

Selected leading region, countries, or economies with publications with international coauthors: 2003–22



Note(s):
Articles refer to publications from a selection of journals and conference proceedings in S&E from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating region, country, or economy is credited with one count). Articles with international institutions are counts of articles with institutional addresses from more than one region, country, or economy. For additional countries, see Table SPBS-33.

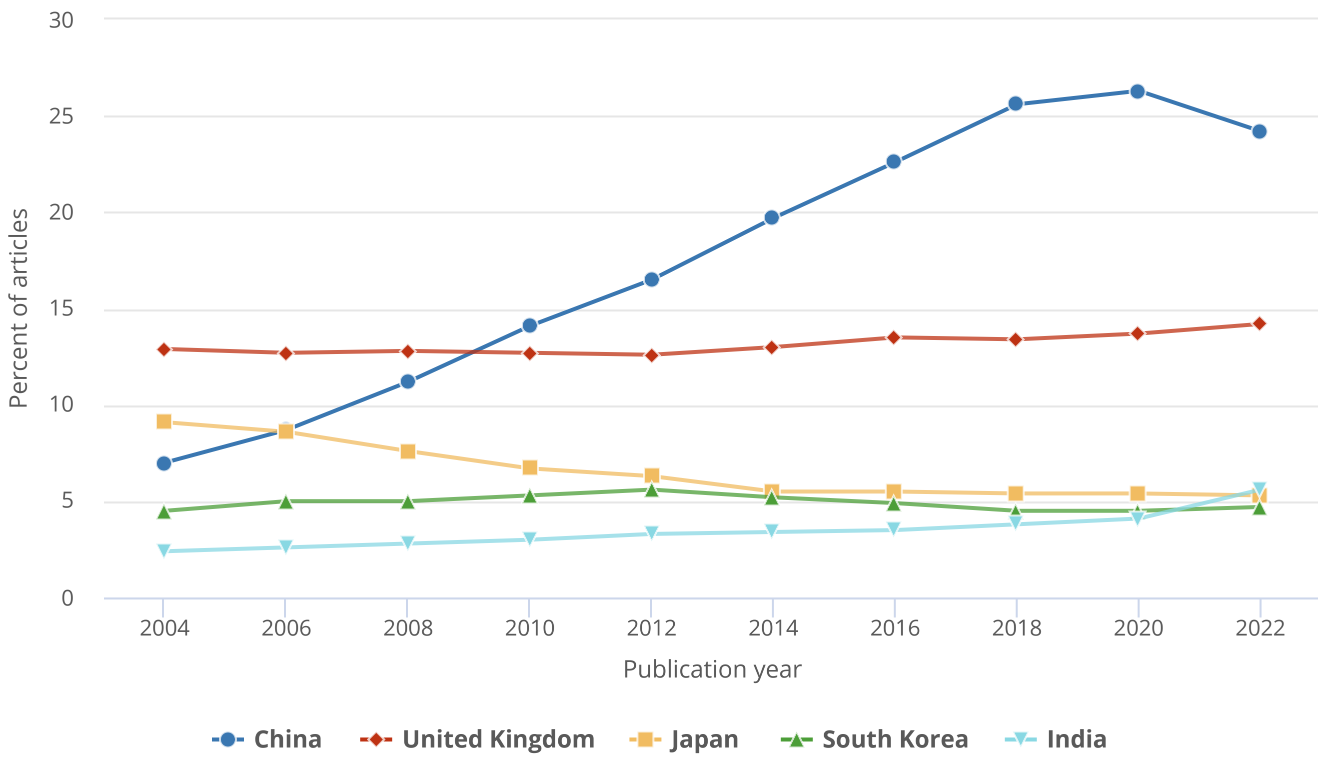
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Over time, the top collaborators of the United States have changed. In 2004, the United Kingdom accounted for 13% of articles that the United States coauthored internationally—the highest percentage of any partner region, country, or economy (**Figure PBS-14**). By 2022, China had become the largest collaborator with the United States, with 24% of internationally coauthored U.S. articles having a Chinese coauthor, although this represents a slight decline from 26% in 2020. Meanwhile, the percentage of U.S. internationally coauthored articles with the United Kingdom increased slightly over this period to 14% in 2022.

Figure PBS-14

U.S. international S&E publications with coauthor(s) from the United Kingdom and Asian countries: Selected years, 2004–22



Note(s):
Articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating region, country, or economy is credited with one count). Articles with international institutions are counts of articles with institutional addresses from more than one region, country, or economy. For more detail, see Table SPBS-36.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

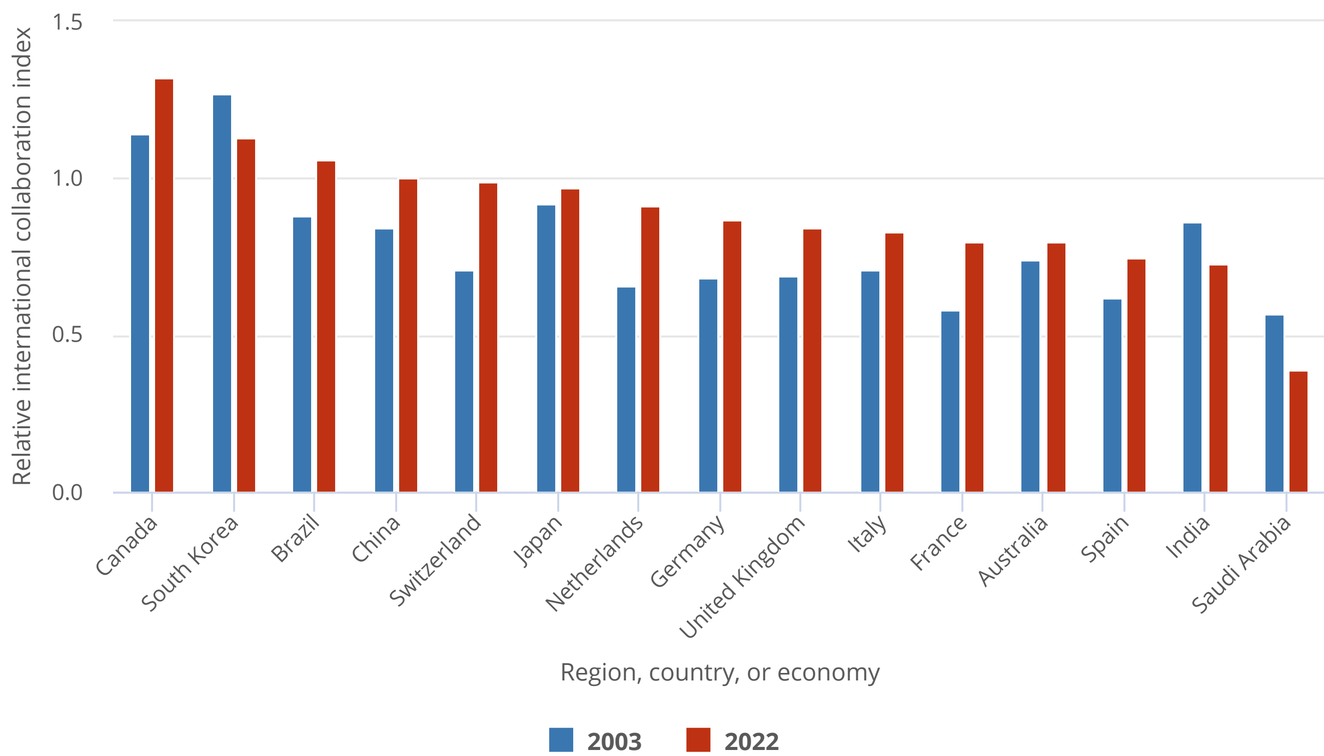
The U.S.-China collaboration on publications has benefited not only both countries but also global science through the amount of published research collaborations, funding agency support from the United States and China, and roles of collaborators on publications (Lee and Haupt 2020). The rapid growth in U.S.-China collaborations coincided with China's growing scientific and technological capabilities, such as rising R&D spending and university degree awards (see *Indicators 2022* report "[2022] **Higher Education in Science and Engineering**")—both of which may be contributing factors to the U.S.-China collaboration pattern.

Rates of U.S. collaboration with other Asian regions, countries, or economies have also changed over time. From 2004 to 2022, the percentage of U.S. internationally coauthored articles with Japan decreased (from 9.1% to 5.3%), whereas the share coauthored with India increased (from 2.4% to 5.6%). The share of U.S. international collaborations involving South Korea grew between 2004 and 2012 (from 4.5% to 5.6%), then declined through 2022 (from 5.6% to 4.7%) as South Korea increased its partnerships with other Asian regions, countries, or economies (**Figure PBS-14**; Table SPBS-35). Meanwhile, there was little change from 2003 to 2022 in the U.S. coauthorship percentages with Canada (12% in 2022) and with European regions, countries, or economies such as the United Kingdom (14% in 2022) and Germany (11% in 2022) (Table SPBS-35).

Normalizing international collaborations by a region, country, or economy's publication output enables comparison independent of its size. For example, the international collaboration index (ICI), adapted from He (2009), is obtained by dividing a region's, country's, or economy's share of collaboration with a partner by the partner's overall share of international collaborations with all regions, countries, or economies. An ICI value of 1.0 shows that the level of coauthorship between two regions, countries, or economies is proportional to the partners' overall rates of international coauthorship. ICI values above 1.0 indicate more extensive ties between two regions, countries, or economies, whereas values below 1.0 indicate weaker ties. In 2022, the United States had ICI values above 1.0 with Canada (1.3), South Korea (1.1), and Brazil (1.1). ICI values between the United States and most other major research-producing regions, countries, or economies increased between 2003 and 2022, except for South Korea (from 1.3 to 1.1), India (from 0.9 to 0.7), and Saudi Arabia (from 0.6 to 0.4) (**Figure PBS-15**).

Figure PBS-15

Relative international collaboration index of selected large-producing regions, countries, or economies with the United States: 2003 and 2022



Note(s):
Article counts for computing the index refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating region, country, or economy is credited with one count). Regions, countries, or economies that have contributed to less than 1% of all internationally coauthored articles in 2022 are omitted. The index of collaboration is calculated as follows: $ICxy = (Cxy / Cx) / (Cy / Cw)$, where $ICxy$ is the index of collaboration between country x and country y, Cxy is the number of publications coauthored between country x and country y, Cx is the total number of international coauthorships by country x, Cy is the total number of international coauthorships by country y, and Cw is the total number of international coauthorships in the database. For additional regions, countries, or economies, see Table SPBS-38.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

SIDEBAR

Artificial Intelligence Publication Output and International Collaboration

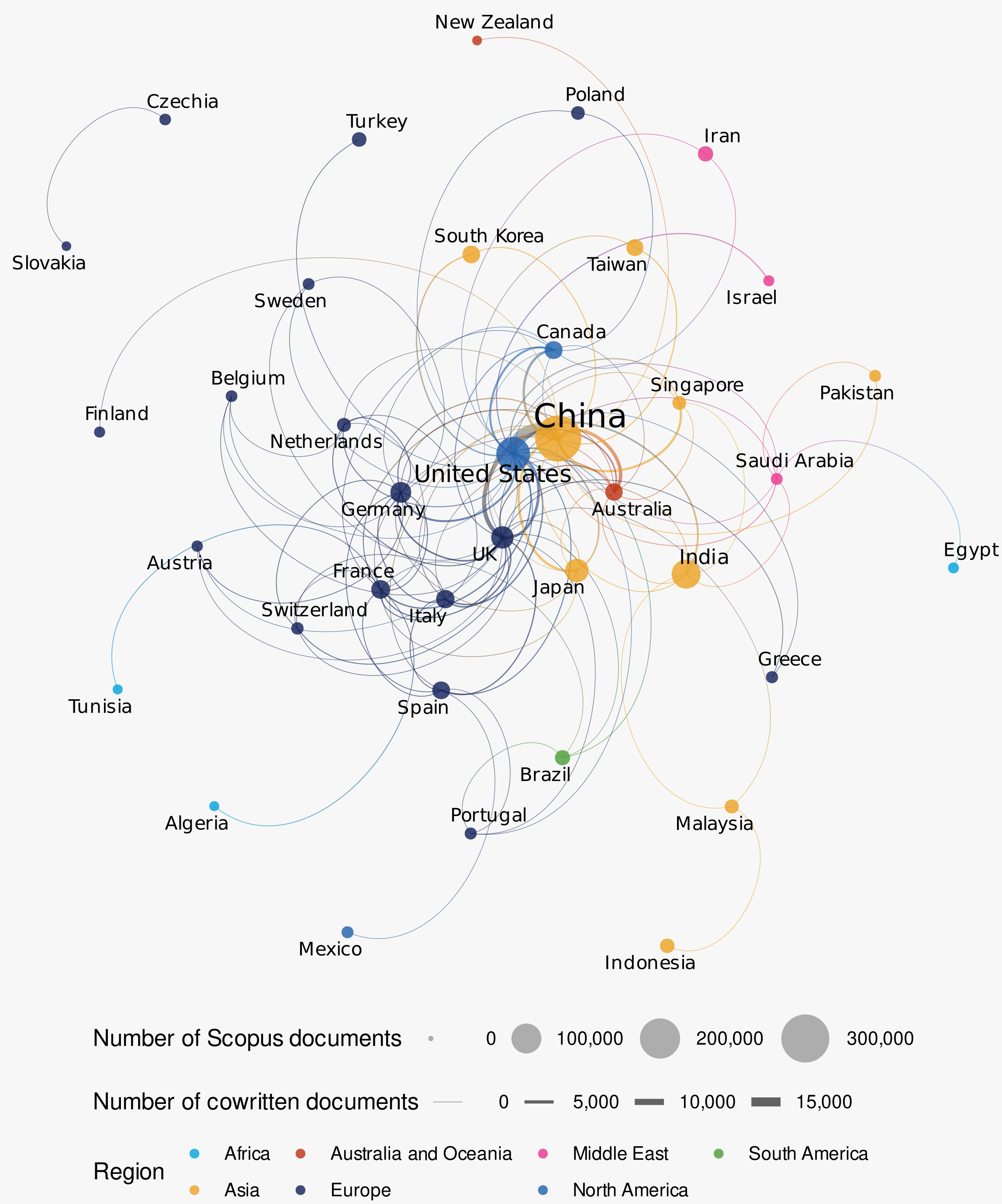
The National Science Board (NSB) highlighted the importance of international collaboration and work in technology-intensive industries in its *Vision 2030* report (NSB 2020). This work became important because of the increasing globalization of science and engineering and the significance of research in fields such as artificial intelligence (AI) playing a large role "to empower U.S. business and entrepreneurs to succeed globally" (NSB 2020). The National Artificial Intelligence Research Resource Task Force (NAIRRTF) has also drawn special attention to the researchers behind AI during the field's rapid growth through a strategic plan to increase access to resources and training for diverse AI researchers (NAIRRTF 2023). Research in the field of AI is an important technology-intensive area that has grown rapidly in recent years and in which international collaborations are crucial (Liu, Shapira, and Yue 2021). The nature of international partnerships between top-producing regions, countries, or economies, such as the United States and China, and other smaller, yet relatively important partnerships, such as Slovakia and Czechia, is an important influence on the AI advancements and priority topics of study.

The initial step to explore the state of international collaboration in the field of AI is determining the inclusion of articles from scientific journals and conference proceedings relevant for a network analysis of AI international collaboration.[*] Scopus has a series of classification groups that help isolate relevant AI work; this subset of the Scopus data set was used to produce two network figures. The first represents the global collaboration network of AI and displays the number of coauthored articles between regions, countries, or economies to highlight the most prolific collaborations and collaborators. The second network focuses on the relative importance of collaborations by normalizing for overall publication output in AI. This helps highlight smaller, but relatively important, collaborative relationships. Together, the two figures allow for a more complete snapshot of the global collaboration network of AI.[†]

Figure PBS-B shows the network of AI research collaboration. The largest contributors from 2003 to 2022, in terms of total documents produced and the largest collaborative hubs, were China (274,096 total articles as a whole count) and the United States (133,601) (Table SPBS-91). Indeed, this country pairing was responsible for the most coauthored articles of any pair (13,631 articles) (**Figure PBS-B**). Further, of the top 10 most prolific pairs, all feature the United States or China. The first pairing that does not is the 11th most prolific pair of the United Kingdom and Germany (2,166). Other important contributors in the network include Australia, Japan, Canada, and Singapore. The centrality of the United States and China may follow from their high publication output (see the section **Output by Region, Country, or Economy**), which enables a greater number of network connections and echoes previously observed patterns of collaboration at the region, country, or economy level (Leydesdorff and Wagner 2008).

Figure PBS-B

AI collaboration network, by region, country, or economy: 2003–2022



AI = artificial intelligence.

Note(s):
This network diagram shows the number of cowritten articles by all pairs of regions, countries, or economies within the top 60 producers of AI-related research based on whole counting for those pairs that cowrote 400 articles or more. AI article counts refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus that were classified as AI in the All Science Journal Classification. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Links are only shown in a single direction, dictated by alphabetical order. The size of the nodes is proportional to the total number of AI-related articles written by each region, country, or economy. The width of the links between nodes is proportional to the quantity of articles both regions, countries, or economies have cowritten. Positioning of nodes is defined using the Kamada-Kawai algorithm. For the list of regions, countries, and economies and their respective geographic regions in this figure, see Table SPBS-91.

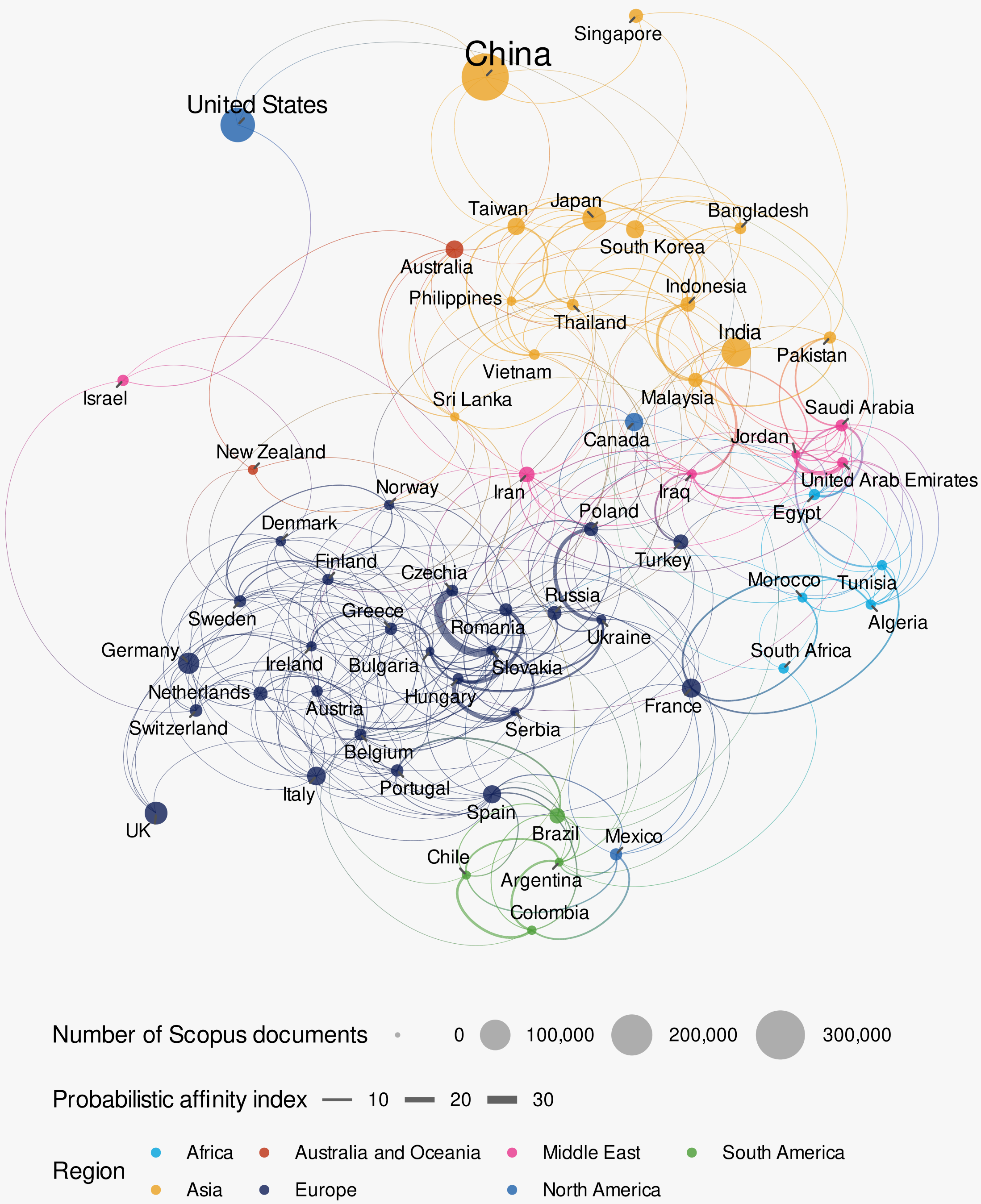
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Although whole-counted documents produced by region, country, or economy pairs is an effective way to describe the overall hub-centered structure of collaboration in the field of AI, in **Figure PBS-C**, the focus on relationships relative to overall output highlights more regional collaboration patterns and partnerships. Instead of line thickness representing the number of coauthored documents, it now shows the *index of collaboration*. This index is best interpreted as the propensity of a region, country, or economy to collaborate with another, given their total collaborations. The baseline of this measure is equal to 1.00, where values greater than 1.00 show a preferential collaborative relationship. This helps highlight important relationships of regions, countries, or economies not significantly featured in **Figure PBS-B**. Specifically, the values of collaborative pairs Slovakia and Czechia (30.42), Serbia and Hungary (15.79), and the United Arab Emirates and Jordan (15.03) were notably large, given the number of coauthored publications that these countries produced. Unsurprisingly, geography and shared history play significant roles in these relationships, but they help illustrate smaller, and yet important, collaborative relationships in AI.

Figure PBS-C

AI index of collaboration, by region, country, or economy pairs: 2003–2022



AI = artificial intelligence.

Note(s):
The index of collaboration is calculated as follows: $ICxy = (Cxy / Cx) / (Cy / Cw)$, where $ICxy$ = index of collaboration between country x and country y, Cxy = number of publications coauthored between country x and country y in the relevant field, Cx = total number of international coauthorships by country x in the relevant field, Cy = total number of international coauthorships by country y in the relevant field, and Cw = total number of international coauthorships in the relevant field overall. This network diagram shows indices of collaboration between all pairs of regions, countries, or economies within the top 60 producers of AI-related research, based on whole counting, for pairs having an index of collaboration higher than 1.00. AI articles refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus that were classified as AI in the All Science Journal Classification. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Links are only shown in a single direction, dictated by alphabetical order. The size of the nodes is proportional to the total number of AI-related articles written by each region, country, or economy. The width of the links between nodes is proportional to the index of collaboration between both regions, countries, or economies. Positioning of nodes is defined using the Fruchterman-Reingold algorithm. For the list of regions, countries, and economies and their respective geographic regions in this figure, see Table SPBS-91.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

In conclusion, looking at international collaboration in the field of AI through a network analysis provides some helpful conclusions about the current state of an increasingly important field. Although the United States and China were the most prominent actors in the global collaboration network, looking at relative relationships provides additional perspective for important partnerships like that of Slovakia and Czechia, which would be otherwise obscured. Going forward, AI research will likely remain an influential field affecting multiple facets of society in the coming years. The nature of the networks producing forthcoming research will be just as important in shaping new advancements as the international partnerships that produce them.

* *Network analysis* refers to a broad range of visualization, mathematical, and statistical techniques centered around the conceptualization and depiction of entities and the relationships between them as nodes and edges (linkages between nodes). The analyses here are descriptive.

† Although these figures display work on AI specifically, they also show general trends in international collaboration. For example, the large collaboration between the United States and China is not unique to AI but is true for most subject areas. Further, the relative collaborative relationships and propensity scores are also likely true of other areas as they highlight geographically or culturally close ties where AI research is one part of the larger collaborative relationship. The focus on AI as an important area of study serves as a useful lens to look at these relationships and is not shared with the purpose of distinguishing AI networks from other fields of study.

International Citation Patterns

Another indicator of international collaboration is cross-national citations, which are citations to a region, country, or economy's publications that come from publications authored outside that region, country, or economy. This section outlines an international citation measure called the *relative citation index* (RCI). The RCI is computed by dividing the share of the citing region's, country's, or economy's outgoing citations going to the cited region, country, or economy, then dividing this number by the share of publications attributed to the cited region, country, or economy. The measure is normalized by the publication output of each region, country, or economy. Otherwise, regions, countries, or economies such as the United States, the United Kingdom, Germany, and China would be favored, simply because they produce so many articles and are therefore more likely to attract citations. Findings based on the RCI provide additional evidence of the globalization of S&E research.

An RCI value above 1.00 shows that the citation rate between two regions, countries, or economies is higher than the baseline (and a value below 1.00 shows a citation rate lower than the baseline), taking into account their relative publication outputs.[9] U.S. authors tended to cite English-speaking regions, countries, or economies disproportionately, with RCI values above 1.00 for the United Kingdom (1.37), Canada (1.29), and Australia (1.07) (**Table PBS-2**). This finding may also be a function of the many scientific journals being published in English (Di Bitetti and Ferreras 2016). Some other notable European countries with high research activity had moderate RCI values based on citations from the United States, such as Germany (0.96), France (0.93), and Italy (0.87). U.S. authors cite China (0.48), Iran (0.32), and India (0.25) less frequently than the baseline, given the number of articles produced by those countries.

Table PBS-2

Relative citation index for 15 largest producing regions, countries, or economies: 2020

(Index)

	Cited region, country, or economy														
Citing region, country, or economy	China	United States	India	United Kingdom	Germany	Italy	Japan	Canada	France	Australia	Spain	Russia	South Korea	Brazil	Iran
China	2.43	0.74	0.48	0.69	0.56	0.56	0.54	0.69	0.54	0.95	0.52	0.13	0.96	0.32	0.73
United States	0.48	3.24	0.25	1.37	0.96	0.87	0.52	1.29	0.93	1.07	0.64	0.11	0.56	0.34	0.32
India	0.90	0.62	5.43	0.78	0.53	0.79	0.39	0.68	0.57	0.82	0.65	0.17	0.93	0.58	1.52
United Kingdom	0.52	1.34	0.35	6.52	1.19	1.15	0.51	1.29	1.14	1.60	0.91	0.13	0.54	0.42	0.41
Germany	0.46	1.27	0.26	1.65	6.80	1.20	0.62	1.05	1.31	1.06	0.92	0.21	0.56	0.37	0.31
Italy	0.56	1.08	0.36	1.44	1.05	8.69	0.51	0.96	1.29	0.92	1.36	0.15	0.60	0.55	0.60
Japan	0.64	1.13	0.33	1.14	1.04	0.92	8.64	0.86	1.01	0.93	0.70	0.16	0.94	0.32	0.33
Canada	0.60	1.55	0.36	1.61	0.93	0.95	0.47	8.91	1.00	1.47	0.77	0.12	0.62	0.46	0.65
France	0.53	1.26	0.33	1.63	1.37	1.47	0.66	1.16	8.66	1.08	1.13	0.20	0.57	0.51	0.40
Australia	0.67	1.21	0.40	1.83	0.88	0.83	0.46	1.39	0.82	11.02	0.77	0.11	0.63	0.45	0.61
Spain	0.58	1.03	0.42	1.49	1.04	1.78	0.48	1.04	1.21	1.16	8.84	0.16	0.63	0.75	0.66
Russia	0.61	0.75	0.50	0.90	0.96	1.01	0.55	0.70	0.91	0.73	0.79	10.15	0.60	0.47	0.70
South Korea	1.02	0.97	0.63	0.89	0.74	0.84	0.76	0.84	0.67	0.95	0.67	0.14	8.99	0.38	0.78
Brazil	0.62	0.86	0.68	1.10	0.71	1.23	0.39	0.94	0.89	1.03	1.23	0.15	0.59	10.51	1.00
Iran	1.04	0.62	1.06	0.70	0.46	1.00	0.33	0.78	0.56	0.85	0.74	0.17	0.80	0.63	11.93

Note(s):

Citations refer to publications from a selection of journals, books, and conference proceedings in S&E from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) listed in the article. Articles are credited on a fractional count basis (i.e., for articles with collaborating institutions from multiple regions, countries, or economies, each region, country, or economy receives fractional credit on the basis of the proportion of its participating institutions). Citation counts are based on all citations made to articles in their publication year and in the following 2 years (i.e., 3-year citation window; scores in 2020 are based on citations to articles published in 2020 that were made in articles published in 2020–22). The relative citation index (RCI) normalizes cross-national citation data for variations in relative size of publication output. RCI is computed by dividing the share of the citing region, country, or economy's outgoing citations attributed to the cited region, country, or economy, then dividing that amount by the share of publications attributed to the cited region, country, or economy. An RCI of 1.00 means that the citing region, country, or economy cites publications from the cited region, country, or economy as much as would be expected to happen randomly, showing no particular affinity between the regions, countries, or economies. Scores higher than 1.00 mean that the citing region, country, or economy has a higher-than-expected tendency to cite the cited region, country, or economy's S&E literature. For more detail, see Table SPBS-39. Cells in which the region, country, or economy collaborates at or above the world average for that year are shaded green.

Source(s):

National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

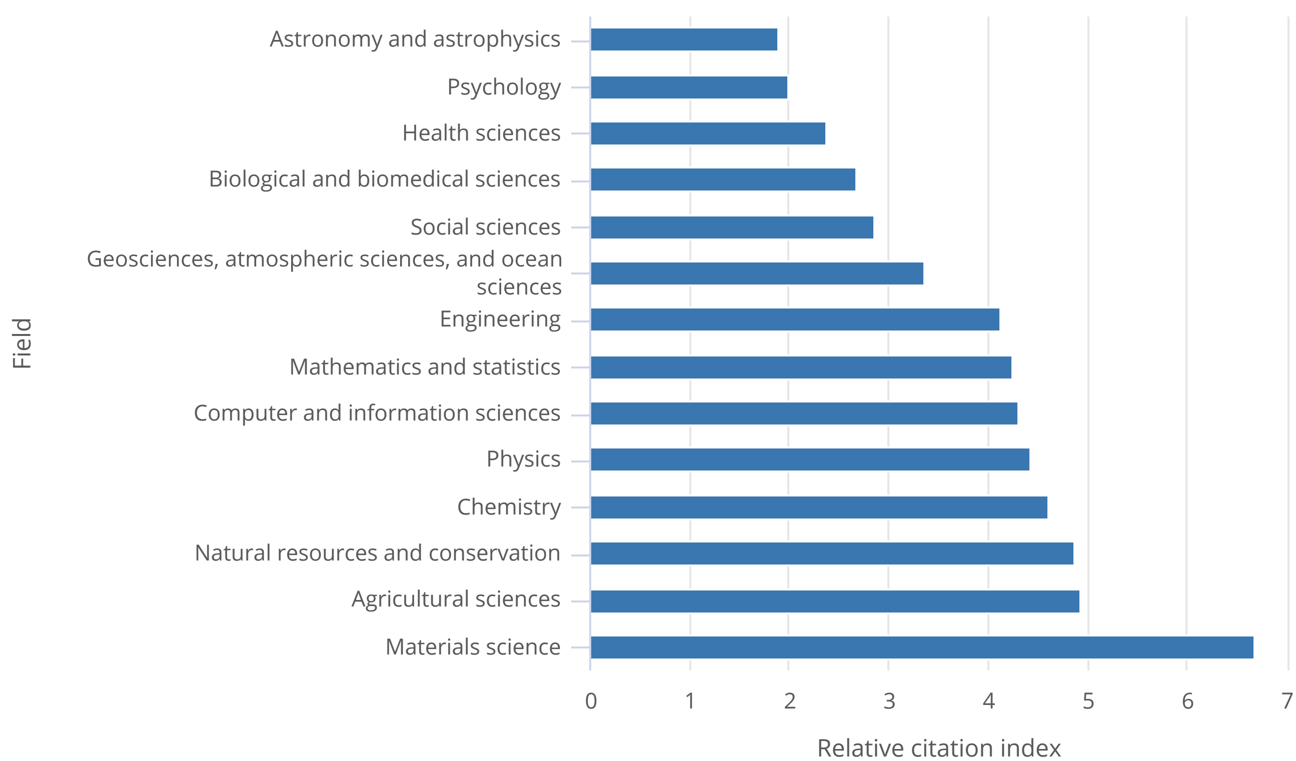
Science and Engineering Indicators

The RCI also provides a view into the international nature of the scientific research in each region, country, or economy by determining the number of other regions, countries, or economies from which the region, country, or economy draws its research (i.e., citations). For example, authors from China predominantly cited articles from China (2.43) and cite articles from only South Korea and Australia (among the leading publication producers) near the baseline rate (**Table PBS-2**). Iran (11.93) and Russia (10.15) tended to cite their own publications extensively. Conversely, France frequently cited research from numerous other regions, countries, or economies, such as the United Kingdom (1.63), Italy (1.47), Germany (1.37), and the United States (1.26). Many other European regions, countries, or economies displayed a similar pattern of extensive cross-national citations.

RCI results at the scientific field level provide insights into the extent to which regions, countries, or economies build on their own research within a given field rather than on research gleaned from elsewhere (Table SPBS-40 through Table SPBS-53). These results are thus a proxy measure for the degree to which researchers in a given field are insular or international in their citation behavior. For example, in 2022, the United States had an overall domestic RCI of 3.24 (**Table PBS-2**). This means that U.S. articles had more than three times as many citations to earlier U.S. publications than the baseline, given the number of these earlier U.S. publications. In materials science, the United States had a domestic RCI of 6.68 in 2022 (**Figure PBS-16**), which means that there are more than six times as many citations to U.S. articles as the baseline. This suggests that U.S. materials science researchers built particularly extensively on earlier U.S. research. The same holds true in agricultural sciences (4.92), natural resources and conservation (4.86), and chemistry (4.60). Meanwhile, the United States had a lower domestic RCI in astronomy and astrophysics (1.89), psychology (2.00), and health sciences (2.37), suggesting that U.S. researchers in these fields had a relatively stronger connection to international research. That said, these RCIs remain above 1.00 because researchers tend to cite domestic publications disproportionately.

Figure PBS-16

U.S. domestic relative citation index across fields: 2022



Note(s):
Article counts for computing the index refer to publications from a selection of conference proceedings and peer-reviewed journals in S&E fields from Scopus. Articles are classified by their year of publication and are assigned to a region, country, or economy on the basis of the institutional address(es) of the author(s) listed in the article. Articles are credited on a whole count basis (i.e., each collaborating region, country, or economy is credited with one count). Regions, countries, or economies that have contributed to less than 1% of all internationally coauthored articles in 2022 are omitted. The domestic relative citation index is calculated as follows: $ICxy = (Cxy / Cx) / (Cy / Cw)$, where $ICxy$ is the index of collaboration between country x and country y, Cxy is the number of publications coauthored between country x and country y, Cx is the total number of international coauthorships by country x, Cy is the total number of international coauthorships by country y, and Cw is the total number of international coauthorships in the database.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

To conclude this section, the findings from the analysis of international collaborations and citations reveal that regions, countries, or economies had wide variations in their propensity to collaborate with international partners, although all major regions, countries, or economies increased their collaboration rate over time. Researchers in the United States collaborated increasingly with Asian partners—most notably from China. Meanwhile, international citation patterns reveal that the United States disproportionately cited publications from established research communities, including its own, with its international connections varying widely across scientific fields.

Conclusion

Based on the overall size of the U.S. contribution to S&E research publication output and its relative impact, as measured by citations to its S&E publications, the United States remains a highly influential nation. The publication outputs of the United States, Japan, the United Kingdom, and the EU-27 are concentrated in health sciences, whereas publications from China and India focused more on engineering and computer and information sciences, respectively. In terms of S&E publication quantity, China's output has grown rapidly and is now nearly double that of the United States. In terms of impact among S&E publications, China has increased rapidly in the last decade. The high-income economies (including the United States, the EU-27, and Japan) have slowly increased their large base of S&E publications, whereas middle-income and upper-middle-income economies have rapidly increased their production, collaboration, and impact despite beginning from a smaller S&E publications base.

International research collaboration is increasing, reflecting traditional ties across regions, countries, or economies and new relationships that stem from growing capabilities in the middle-income economies. Greater publication output—with greater and more diverse collaborations—means more regions, countries, or economies are contributing, and many are doing so with U.S. authors. Finally, OA articles showed considerable growth in terms of output and impact. The growth of open science should continue to impact the way research is produced, consumed, and cited in coming years.

Glossary

Definitions

Citations: Citations, generally at the end of each article, provide researchers with the list of the prior research relied on for the article. Citations of S&E publications by other S&E publications provide an indication of the impact of publications and of the flow of knowledge or linkage between sectors or geographic locations.

Coauthorship: Coauthorship refers to cases in which more than one author is listed on a publication. Data on coauthorship can be used to measure collaboration across regions, countries, economies, and institutional sectors. Publication counts of coauthorship use whole counting, so each region, country, or economy contributing to the article receives credit for that article. An article is considered to contain an international coauthorship when institutional addresses for its authors are located in two or more different regions, countries, or economies. Table SPBS-36 shows international coauthorship from 2003 to 2022.

European Union (EU-27): The EU comprises 27 member nations: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden. Bulgaria and Romania joined the EU in 2007, and Croatia joined in 2013—these nations are included in the EU grouping for all years analyzed in this report. In 2020, the United Kingdom left the EU, and data covering the United Kingdom are excluded from the EU-27 in the text of the report but are available in the supplemental tables as “EU-27 and United Kingdom.”

Fractional counting: A method of counting S&E publications in which credit for coauthored publications is divided among the collaborating institutions or regions, countries, or economies based on the proportion of their participating authors. Fractional counting allocates the publication count based on the proportion of the coauthors named on the article with institutional addresses from each region, country, or economy. Fractional counting enables the counts to sum up to the number of total articles. For example, if a publication were authored by two researchers from the University of Oslo, one from University College London, and one from the University of Washington, half of the publication would be attributed to Norway, and a quarter each to the United Kingdom and the United States when the fractions are calculated at the level of researchers. For this report, fractions were calculated at the level of researchers. If an author provides multiple institutions, and those institutions are in different regions, countries, or economies, then each region, country, or economy receives an appropriate fraction of the count.

Highly cited article (HCA): An HCA ratio provides an indication of scientific impact (Waltman, van Eck, and Wouters 2013). The HCA ratio for a region, country, or economy is calculated as the share of all articles published in a given year by authors with institutional addresses within that region, country, or economy that fall within the top 1% by citation count of all articles published that year, measured for each research field. The HCA ratio is indexed to 1.00, so a region, country, or economy whose authors produce highly cited articles at the expected (i.e., global average) rate has an HCA ratio of 1.00—that is, 1% of the region’s, country’s, or economy’s articles are among the top 1% of the world’s highly cited articles. A region, country, or economy with an HCA ratio greater than 1.00 is producing a disproportionately high level of articles with exceptional scientific impact, whereas a region, country, or economy whose authors produce relatively fewer influential articles will have an HCA ratio below 1.00.

International collaboration index (ICI): The ICI helps identify the propensity of collaboration between two regions, countries, or economies. The ICI is calculated as follows: $Icxy = (Cxy / Cx) / (Cy / Cw)$, where $Icxy$ is the index of collaboration between country x and country y, Cxy is the number of publications coauthored between country x and country y, Cx is the total number of international coauthorships by country x, Cy is the total number of international coauthorships by country y, and Cw is the total number of international coauthorships in the database. An index greater than 1.0 means that a country-country pair has a stronger-than-expected tendency to collaborate; an index less than 1.0 indicates a weaker-than-expected tendency to collaborate (Table SPBS-38).

Open access (OA): OA refers to peer-reviewed publications that are accessible online to any reader without requiring a journal subscription or other fees from readers (Piwowar et al. 2018). Several commonly defined types of OA have been adopted for the purposes of this analysis. Gold OA denotes articles published in journals that are entirely OA as a matter of journal policy. Hybrid OA refers to articles appearing in closed-access journals where the authors have paid a fee to make the article OA. Bronze OA denotes articles in closed-access journals that become OA after an embargo period of closed access or articles that appear available as OA despite lacking the license information to guarantee OA in the long term. Green OA denotes articles that are self-archived by authors in OA repositories, which are often maintained and administered by universities or other institutions.

Relative citation (RC): The RC is a normalization of the relative scientific impact of publications produced by a given region, country, or economy that takes into consideration variations in citation behavior between fields and years of publication. For a publication in a given scientific field and publication year, the citation count is divided by the average count of all publications in the relevant field and publication year.

Relative citation index (RCI): The RCI normalizes cross-national citation data for variations in relative size of publication output. It is computed by dividing the share of the citing region's, country's, or economy's outgoing citations going to the cited region, country, or economy, then dividing this number by the share of publications attributed to the cited region, country, or economy.

Whole counting: This measure (also called *full* or *integer* counting) assigns one count to each region, country, or economy or institutional sector involved in coauthoring the article, irrespective of its proportionate involvement in authorship. Although fractional counting aims to assess the proportionate contributions of regions, countries, or economies or sectors, whole counting aims to assess the participation of regions, countries, or economies or sectors. One result of this difference is that the sum of articles from regions, countries, or economies or institutional sectors will exceed the total number of articles when whole counting is used. For the United States in 2022, there were 458,000 publications in the Scopus database as measured on a fractional-count basis and 608,000 as measured on a whole-count basis (Table SPBS-2 and Table SPBS-17). In the full-counting method, each publication is counted once for each entity listed in the address field. For example, if a publication were authored by two researchers from the University of Oslo, one from University College London, and one from the University of Washington, the publication would be counted once for Norway, once for the United Kingdom, and once for the United States. When it comes to aggregating groups of institutions (e.g., research consortia) or groups of countries (e.g., the EU-27), double counting is avoided. This means that if authors from Croatia and France co-published an article, this publication would be credited only once when counting publications for the EU-27, although each country had been credited with one publication count.

Key to Acronyms and Abbreviations

AI: artificial intelligence

APC: article processing charge

EU-27: European Union

HCA: highly cited article

ICI: international collaboration index

NCSES: National Center for Science and Engineering Statistics

NSB: National Science Board

OA: open access

R&D: research and development

RCI: relative citation index

S&E: science and engineering

TOD: Taxonomy of Disciplines

References

Bornmann L. 2013. What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey. *Journal of the American Society for Information Science and Technology* 64:217–33. Available at **https://doi.org/10.1002/asi.22803**. Accessed 24 March 2023.

Brainard J, Kaiser J. 2022. White House Requires Immediate Public Access to All U.S.-Funded Research Papers by 2025. *Science* 377(6610):1026–27. Available at **https://doi.org/10.1126/science.ade6577**. Accessed 17 April 2023.

Chinchilla-Rodríguez Z, Sugimoto CR, Larivière V. 2019. Follow the Leader: On the Relationship between Leadership and Scholarly Impact in International Collaborations. *PLOS ONE* 14(6):e0218309. Available at **https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0218309**. Accessed 28 August 2023.

Di Bitetti MS, Ferreras JA. 2016. Publish (in English) or Perish: The Effect on Citation Rate of Using Languages Other Than English in Scientific Publications. *Ambio* 46:121–27. Available at **https://doi.org/10.1007%2Fs13280-016-0820-7**. Accessed 24 March 2023.

Ebadi A, Schiffauerova A. 2016. How to Boost Scientific Production? A Statistical Analysis of Research Funding and Other Influencing Factors. *Scientometrics* 106(3):1093–1116. Available at **https://doi.org/10.1007/s11192-015-1825-x**. Accessed 24 March 2023.

European Commission, Directorate-General for Research and Innovation. 2021. *Horizon Europe, the EU Research and Innovation Programme (2021–27): For a Green, Healthy, Digital and Inclusive Europe*. Luxembourg: Publications Office of the European Union. Available at **https://data.europa.eu/doi/10.2777/052084**. Accessed 28 August 2023.

Franzoni C, Scellato G, Stephan P. 2011. Changing Incentives to Publish. *Science* 6043:702–3. Available at **https://science.sciencemag.org/content/333/6043/702.full**. Accessed 28 August 2023.

Garfield E. 1973. Citation Frequency as a Measure of Research Activity and Performance. *Current Contents* 5:406–8. Available at **http://www.garfield.library.upenn.edu/essays/V1p406y1962-73.pdf**. Accessed 24 March 2023.

Glänzel W, Schubert A. 2005. Domesticity and Internationality in Co-Authorship, References and Citations. *Scientometrics* 65(3):323–42.

Grudniewicz A, Moher D, Cobey KD, Bryson GL, Cukier S, Allen K, Ardern C, Balcom L, Barros T, Berger M, Ciro JB, Cugusi L, Donaldson MR, Egger M, Graham ID, Hodgkinson M, Khan KM, Mabizela M, Manca A, Milzow K, Mouton J, Muchenje M, Olijhoek T, Ommaya A, Patwardhan B, Poff D, Proulx L, Rodger M, Severin A, Strinzel M, Sylos-Labini M, Tamblyn R, van Niekerk M, Wicherts JM, Lalu MM. 2019. Predatory Journals: No Definition, No Defence. *Nature* 576:210–12. Available at **https://www.nature.com/articles/d41586-019-03759-y?fbclid=IwAR30GTZSaMn_D7bPyeQtk5hHdDXlcD4-hq8Mcl37TXX0PpjB1Tw4sYpSxjA**. Accessed 28 August 2023.

He T. 2009. International Scientific Collaboration of China with the G7 Countries. *Scientometrics* 80(3):571–82.

Ke Q, Ferrara E, Radicchi F, Flammini A. 2015. Defining and Identifying Sleeping Beauties in Science. *Proceedings of the National Academy of Science* 112(24):7426–31. Available at **https://doi.org/10.1073/pnas.1424329112**. Accessed 28 August 2023.

Langham-Putrow A, Bakker C, Riegelman A. 2021. Is the Open Access Citation Advantage Real? A Systematic Review of the Citation of Open Access and Subscription-Based Articles. *PLOS One* 16(6): e0253129. Available at **https://doi.org/10.1371/journal.pone.0253129**. Accessed 28 August 2023.

Larivière V, Sugimoto CR. 2018. Do Authors Comply with Mandates for Open Access? [Commentary] *Nature* 562:483–86. Available at **https://www.nature.com/articles/d41586-018-07101-w**. Accessed 17 April 2023.

Lee JJ, Haupt JP. 2020. Winners and Losers in U.S.-China Scientific Research Collaborations. *Higher Education* 80:57–74. Available at **https://doi.org/10.1007/s10734-019-00464-7**. Accessed 28 August 2023.

Leydesdorff L, Wagner C. 2008. International Collaboration in Science and the Formation of a Core Group. *Journal of Informetrics* 2(4):317–25. Available at **https://doi.org/10.1016/j.joi.2008.07.003**. Accessed 24 March 2023.

Leydesdorff L, Wagner C. 2009. Macro-Level Indicators of the Relations between Research Funding and Research Output. *Journal of Informetrics* 3(4):353–62. Available at **https://doi.org/10.1016/j.joi.2009.05.005**. Accessed 24 March 2023.

Liu N, Shapira P, Yue X. 2021. Tracking Developments in Artificial Intelligence Research: Constructing and Applying a New Search Strategy. *Scientometrics* 126(4):3153–92. Available at **https://doi.org/10.1007/s11192-021-03868-4**. Accessed 24 March 2023.

Luukkonen T, Tijssen RJW, Persson O, Sivertsen G. 1993. The Measurement of International Scientific Collaboration. *Scientometrics* 28:15–36. Available at **https://doi.org/10.1007/BF02016282**. Accessed 28 August 2023.

Lyons E, Colglazier EW, Wagner CS, Börner K, Dooley DM, Mote CD Jr, Roco MC. 2016. How Collaborating in International Science Helps America. *Science & Diplomacy* 5(2).

Merton RK. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.

Narin F, Stevens K, Whitlow ES. 1991. Scientific Co-Operation in Europe and the Citation of Multinationally Authored Papers. *Scientometrics* 21:313–23. Available at **https://link.springer.com/article/10.1007%2FBF02093973**. Accessed 28 August 2023.

National Artificial Intelligence Research Resource Task Force (NAIRRTF). 2023. *Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource.* Washington, DC: National Artificial Intelligence Initiative Office. Available at **https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf**. Accessed 24 March 2023.

National Science Board (NSB), National Science Foundation. 2018. *Science and Engineering Indicators 2018* (*Indicators 2018*). NSB-2018-1. Alexandria, VA. Available at **https://www.nsf.gov/statistics/2018/nsb20181/**.

National Science Board (NSB), National Science Foundation. 2020. *Vision 2030*. NSB-2020-15. Alexandria, VA. Available at **https://www.nsf.gov/nsb/publications/2020/nsb202015.pdf**.

National Science Board (NSB), National Science Foundation. 2021a. Academic Science and Engineering Article Output per 1,000 Science, Engineering, and Health Doctorate Holders in Academia. *Science and Engineering Indicators: State Indicators*. Alexandria, VA. **https://ncses.nsf.gov/indicators/states/indicator/academic-se-articles-per-1000-seh-doctorate-holders-in-academia**.

National Science Board (NSB), National Science Foundation. 2021b. Academic Science and Engineering Article Output per $1 Million of Academic Science and Engineering R&D. *Science and Engineering Indicators: State Indicators*. Alexandria, VA. **https://ncses.nsf.gov/indicators/states/indicator/academic-se-articles-per-1-million-in-academic-rd**.

National Science Board (NSB), National Science Foundation. 2022a. Higher Education in Science and Engineering. *Science and Engineering Indicators 2022* (*Indicators 2022*). NSB-2022-3. Alexandria, VA. Available at **https://ncses.nsf.gov/pubs/nsb20223**.

National Science Board (NSB), National Science Foundation. 2022b. Invention, Knowledge Transfer, and Innovation. *Science and Engineering Indicators 2022* (*Indicators 2022*). NSB-2022-4. Alexandria, VA. Available at **https://ncses.nsf.gov/pubs/nsb20224/**.

Petrou C. 2020. Guest Post—MDPI's Remarkable Growth. *Scholarly Kitchen.* August 10. Available at **https://scholarlykitchen.sspnet.org/2020/08/10/guest-post-mdpis-remarkable-growth/**. Accessed 11 July 2023.

Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S. 2018. The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access articles. *PeerJ* 6:e4375. Available at **https://doi.org/10.7717/peerj.4375**. Accessed 24 March 2023.

Royal Society. 2011. *Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century*. RS Policy Document 03/11. London: Royal Society.

Science-Metrix. 2019. *Bibliometric Indicators for the* SEI 2020. *Technical Documentation*. Montreal, Canada: Science-Metrix. Available at **https://www.science-metrix.com/bibliometric-indicators-for-the-sei-2020-technical-documentation/**. Accessed 28 August 2023.

Science-Metrix. 2021a. *Bibliometric Indicators for the* Science and Engineering Indicators 2022. *Technical Documentation.* Available at **https://science-metrix.com/bibliometrics-indicators-for-the-science-and-engineering-indicators-2022-technical-documentation/**. Accessed 28 August 2023.

Science-Metrix. 2021b. *Special Tabulations of Elsevier Scopus Abstract and Citation Database.* Montreal, Canada: Science-Metrix.

Small H. 2004. On the Shoulders of Robert Merton: Towards a Normative Theory of Citation. *Scientometrics* 60(1):71–79. Available at **https://doi.org/10.1023/B:SCIE.0000027310.68393.bc**. Accessed 24 March 2023.

Stephan PE. 2012. *How Economics Shapes Science.* Cambridge, MA: Harvard University Press.

Sugimoto CR, Larivière V. 2018. *Measuring Research: What Everyone Needs to Know*. Oxford, United Kingdom: Oxford University Press.

Sugimoto CR, Robinson-Garcia N, Murray DS, Yegros-Yegros A, Costas R, Larivière V. 2017. Scientists Have Most Impact When They're Free to Move. *Nature* 550(7674):29–31. Available at **https://www.nature.com/news/scientists-have-most-impact-when-they-re-free-to-move-1.22730**. Accessed 28 August 2023.

Tahamtan I, Bornmann L. 2018. Core Elements in the Process of Citing Publications: Conceptual Overview of the Literature. *Journal of Informetrics* 12:203–16. Available at **https://doi.org/10.1016/j.joi.2018.01.002**. Accessed 24 March 2023.

Van Noorden R, Maher B, Nuzzo R. 2014. The Top 100 Papers. *Nature.* 514:550–53. Available at **https://doi.org/10.1038/514550a**. Accessed 24 March 2023.

Wagner CS. 2018. *The Collaborative Era in Science: Governing the Network.* London: Palgrave Macmillan.

Waltman L, van Eck NJ, Wouters P. 2013. Counting Publications and Citations: Is More Always Better? *Journal of Informetrics* 7(3):635–41. Available at **https://www.sciencedirect.com/science/article/abs/pii/S1751157713000357**. Accessed 28 August 2023.

Wang J. 2012. Citation Time Window Choice for Research Impact Evaluation. *Scientometrics* 94:851–72. Available at **https://doi.org/10.1007/s11192-012-0775-9**. Accessed 24 March 2023.

Yin Y, Dong Y, Wang K, Wang D, Jones BF. 2022. Public Use and Public Funding of Science. *Nature Human Behavior* 6:1344–50. Available at **https://doi.org/10.1038/s41562-022-01397-5**. Accessed 24 March 2023.

Notes

1 Publication output includes only those indexed in the Scopus database. The publication output discussion uses fractional counting, which credits coauthored publications according to the collaborating institutions or regions, countries, or economies based on the proportion of their participating authors. Country assignments refer to the institutional address of authors, with partial credit given for each international coauthorship. As part of the data analysis, filters were employed on the raw Scopus S&E publication data to remove publications with questionable quality, which appear in what are sometimes called predatory journals (NSB ***Indicators 2018***: **Bibliometric Data Filters**).

2 The proportion of output attributable to the large producers is consistent whether using fractional counting, as in **Figure PBS-2** and **Table PBS-1**, or whole counting, as in Table SPBS-17. There is a slight difference between the United States and China when looking at the whole-counting total production numbers. Using whole counting for 2022, the United States had 605,633 articles, whereas China had 976,141. A whole-counting measure allocates one full count to each region, country, or economy with an author contributing to the article; in fractional counting, each region, country, or economy receives a proportion of the count based on the number of authors from that region, country, or economy. For example, if an article had four authors—two from the United States, one from China, and one from Brazil—the fractional scores would be half for the United States, a quarter for China, and a quarter for Brazil. The difference between whole and fractional counting indicates that the United States has more authors working with Chinese authors than China has working with U.S. authors.

3 The use of whole counting or fractional counting to tally the publication output of nations can change the calculated publication count based on the degree to which a region, country, or economy is involved in international collaborations. Under whole counting, a nation receives credit for any publication with an author from that nation. Under fractional counting, the nation's credit for a publication is prorated based on the share of the publication's coauthors who are located in that nation (Table SPBS-17 through Table SPBS-31).

4 Many publishers make their article processing charges (APCs) known publicly. For example, a list of Elsevier APCs can be found at **https://www.elsevier.com/about/policies/pricing**. Wiley APCs are at **https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/article-publication-charges.html**. Springer Nature APCs are at **https://www.springernature.com/gp/open-research/journals-books/journals**.

5 In mathematical terms, a region, country, or economy's HCA ratio (the share of the articles ranked in the highest 1% based on citations) is computed as follows: $Sx = HCAx / Ax$, where Sx is the share of output from country x in the top 1% most-cited articles, $HCAx$ is the number of articles from country x that are among the top 1% of most-cited articles, and Ax is the total number of articles from country x with a relative citation (RC) score. $HCAx$ and Ax are based on whole counting. The RC score is a normalized citation score assigned to a publication and is used to rank articles into the top 1%. The RC score takes into consideration the citation behavior between fields and years of publication. Based on the observation that distinct differences in the citation rates of articles appear 2 years or more after publication, the HCA ratio is calculated with a time lag of at least 2 years (meaning that in 2021, the HCA ratios were calculated based on publications issued in 2019).

6 For an analysis of the rapid growth in articles published each year by MPDI, see Petrou (2020).

7 Regions, countries, or economies contributing less than 1% of all internationally coauthored publications in 2020 were not included in the analysis.

8 The total international collaboration rate differs from the rate of co-publications of individual regions, countries, or economies. Individual scores use whole counting, in which each region, country, or economy with an institutional address on the publication receives 1 point. Therefore, the basis count for individual regions, countries, or economies will be larger than the number of publications that have authors with institutional addresses from two or more regions, countries, or economies, which is the basis for computing the total international collaboration rate.

9 RCI is a citation-based measure of scientific influence. It is calculated from a region, country, or economy's share of all cited S&E publications divided by the other region, country, or economy's share of all cited S&E publications; an index less than 1.00 means a lower-than-expected tendency to cite the other region, country, or economy's publications.

10 Because the bibliometric database is constantly updated, the National Center for Science and Engineering Statistics (NCSES) does not recommend comparing bibliometric data across different editions of *Indicators*. For each edition of *Indicators*, NCSES uses a fixed snapshot of the database. This means that although trends are comparable, the exact number of articles, citations, and other data will vary across editions. For more information about comparing fixed versus dynamic journal data sets, see Schneider et al. (2019).

11 Bibliometric databases such as Dimensions and Crossref are larger than Scopus and do not provide the same level of curation.

12 More information about the selection of journals and conference papers is available at **https://www.elsevier.com/online-tools/scopus/content-overview** and **https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection**.

13 For articles on low-quality publications, see Beall (2012), Bohannon (2013), Carey (2016), and Kolata (2013).

14 In DOAJ, journals can be flagged for the following reasons: (1) suspected editorial misconduct by the publisher or society, (2) invalid International Standard Serial Number (ISSN) or an ISSN not registered or in ISSN database, (3) invalid contact information, or (4) no editorial board. For the DOAJ list of excluded journals, see **https://docs.google.com/spreadsheets/d/183mRBRqs2jOyP0qZWXN8dUd02D4vL0Mov_kgYF8HORM/edit#gid=0**. Note that DOAJ also flags serials that are no longer available in open access; although an important and evolving phenomenon in the research landscape, open access status is not associated here with any specific demarcation of quality, whether low or high. Thus, NCSES does not filter the titles flagged by DOAJ solely for open access–related reasons out of the *Indicators* database.

15 For Elsevier's principles of quality, see **https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection**. During its periodic reevaluation of items flagged for follow-up, the Scopus Content Selection and Advisory Board elected to remove 670 titles as of 2021. NCSES retroactively removed these 670 titles from the *Indicators* database to create a valid time series for bibliometric analysis, although Elsevier does not claim that these titles were necessarily of low quality before 2021.

16 Computation uses fractional counting of articles.

Acknowledgments and Citation

Acknowledgments

The National Science Board (NSB) extends its appreciation to the staff of the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation and to the many others, too numerous to list individually, who contributed to the preparation of this report.

This report was produced under the leadership of Emilda B. Rivers, Director, NCSES; Christina Freyman, Deputy Director, NCSES; and John Finamore, Chief Statistician, NCSES. The report benefited from extensive contributions from NCSES staff, including Christina Freyman and Amy Burke. Alexandria Hughes provided advice on statistical issues. Clara Boothby (Oak Ridge Institute for Science and Engineering) also provided subject-matter expertise and drafting support.

RTI International assisted with report preparation under the leadership of Dr. Susan Rotermund. Alison Bean de Hernandez and Dr. Adams Bailey performed analyses and generated visualizations from the publications data. Robin Henke led the work on data quality assurance; August Gering led the editing team, with composition support from Alex Cone. Rossana Zetina-Beale and Deborah Kulik provided project management support. Staff at Penobscot Bay Media, LLC (PenBay Media), created the report site. The following persons and agencies reviewed this report:

James Dunham, Georgetown University

Francis Narin, CHI Research (Retired)

National Institutes of Health

National Science Foundation

Office of Science and Technology Policy

NSB is especially grateful to the Committee on National Science and Engineering Policy for overseeing preparation of the volume and to the National Science Board Office, under the direction of John Veysey, which provided vital coordination throughout the project. Nadine Lymn led the outreach and dissemination efforts. Amanda Vernon served as Board Office Liaison to the committee. Carol Robbins, Steve Deitz, May Aydin, and Anne Emig were the Executive Secretaries.

Citation

National Science Board, National Science Foundation. 2023. Publications Output: U.S. Trends and International Comparisons. *Science and Engineering Indicators 2024.* NSB-2023-33. Alexandria, VA. Available at **https://ncses.nsf.gov/pubs/nsb202333/**.

Technical Appendix

The *Science and Engineering Indicators 2024* report "Publications Output: U.S. Trends and International Comparisons" (PBS) uses a large database of publication records as a source of bibliometric data. Bibliometric data include each article's title, author(s), authors' institution(s), references, journal title, unique article-identifying information (journal volume, issue, and page numbers or digital object identifier), and year or date of publication. The PBS report uses Scopus, a bibliometric database owned by Elsevier and containing scientific literature with English titles and abstracts, to examine national and global scientific publication–related activity.[10] This appendix discusses the Scopus data and data limitations. More detailed documentation of the Scopus data and computation of bibliometric indicators is available in the report *Bibliometric Indicators for the Science and Engineering Indicators 2024. Technical Documentation* (Science-Metrix 2023).

Data

The counts, coauthorships, and citations presented in the PBS report are derived from information about research articles and conference papers (hereafter referred to collectively as *articles*) published in conference proceedings and peer-reviewed scientific and technical journals. All journals and conference proceedings must pass the standards for inclusion in Scopus, which includes reaching criteria for relevance and peer review. The articles exclude editorials, commentaries, errata, letters, and other material that do not present or discuss scientific data, theories, methods, apparatuses, or experiments. The articles also exclude working papers, which are not generally peer reviewed. The bibliometric data undergo review and processing to create the data presented in the PBS report (Science-Metrix 2023).

Beginning in the *Indicators 2016* report, the PBS report's analysis shifted from using Web of Science by Clarivate (previously Thompson-Reuters) to the Scopus database by Elsevier. In 2016, an examination of the two databases found expanded data coverage in Scopus of internationally recognized peer-reviewed scientific journals (NSB ***Indicators 2016*****: New Data Source for Indicators Expands Global Coverage**). Since 2016, both databases have continued to expand coverage. A recent study comparing the databases found 27 million documents in Scopus and 23 million in Web of Science, with an overlap of 18 million (Visser, van Eck, and Waltman 2020). The *Indicators 2024* PBS report uses the Scopus database to ensure the broadest coverage of a curated database.[11]

This section of the **Technical Appendix** continues with a brief overview of the database composition, followed by an explanation of potential biases in the data, such as exclusion of non-peer-reviewed articles, English-language bias, and the reasoning behind removing conference papers from the highly cited article (HCA) ratio.

Database Composition

Journal selection. Elsevier selects journals for the Scopus database based on evaluation by an international group of subject-matter experts who examine a candidate journal's editorial policy, content quality, peer-review policies, peer-review process and capacity, citations by other publications, editor standing, regularity of publication, and content availability.

Conference selection. Elsevier selects conference materials for the Scopus database by subject field based on quality and relevancy, including the reputations of the sponsoring organization and the publisher of the proceedings.[12]

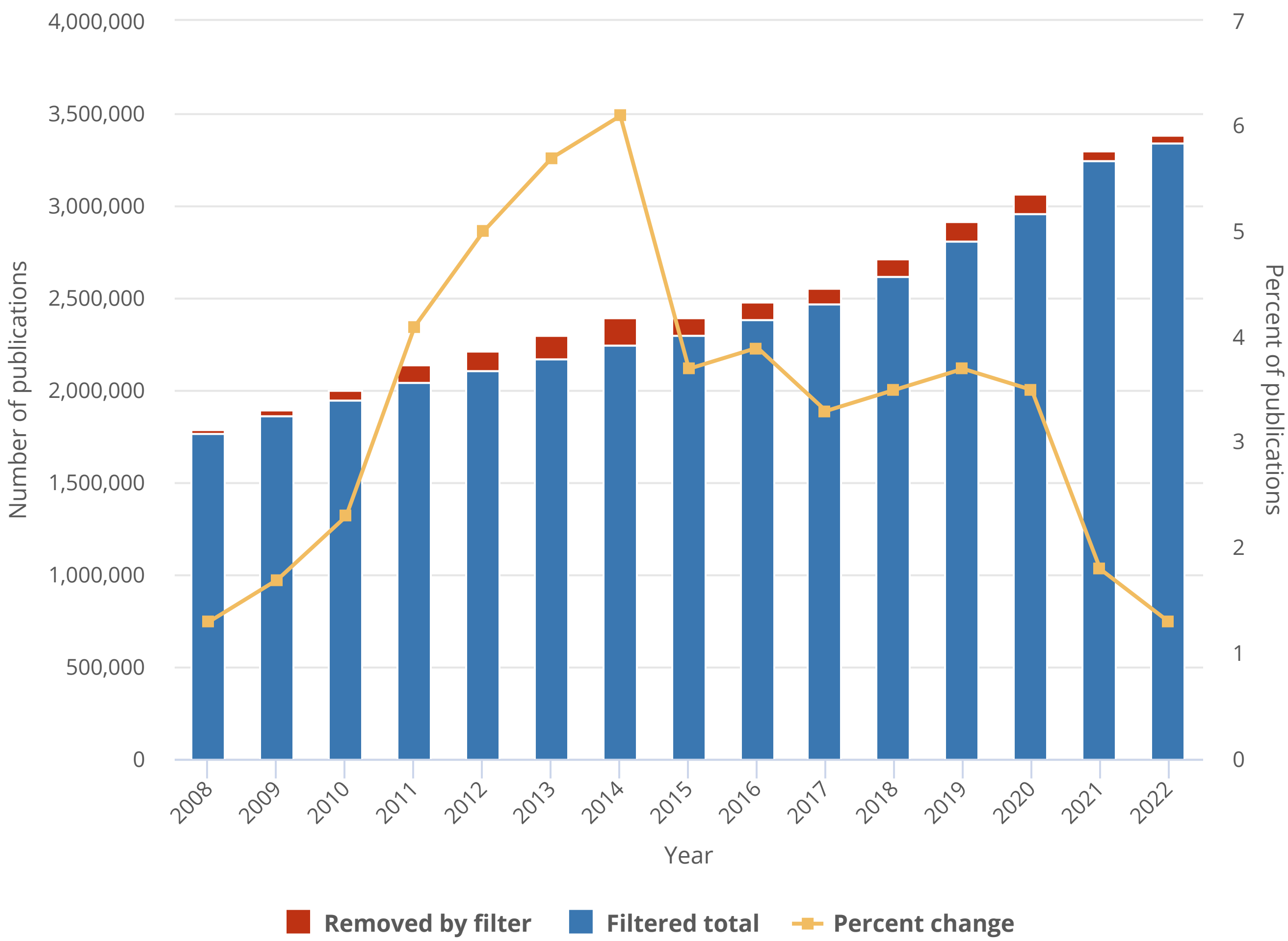
Database Filtering

The National Center for Science and Engineering Statistics (NCSES) undertakes additional filtering of the Scopus data to ensure that the statistics presented in *Indicators* measure original and high-quality research publications (Science-Metrix 2023). Around 2011, librarians and bibliometric experts began to note an increase in articles in the database from electronic journals and conference proceedings lacking substantive peer review.[13] To exclude these publications from the bibliometric data used in this report, NCSES removed journals and conference papers flagged by the Directory of Open Access Journals (DOAJ) for failing to adhere to its list of best practices or being suspected of editorial misconduct.[14] Titles removed by Elsevier from the Scopus database beginning in 2014 were removed retroactively from the *Indicators* database for all publication years (Science-Metrix 2023).[15]

As a result, NCSES removed 2% or fewer articles from the Scopus database prior to 2011, followed by about 4% (more than 88,000 articles) in 2011, then peaked with 5% to 6% (about 111,000 to 145,000 articles) each year from 2012 to 2014 (**Figure SAPBS-1**).[16] Since then, the removal rate has declined.

Figure SAPBS-1

Filtered and unfiltered publications in Scopus, by year: 2008–22



Note(s):
Percent change is computed as the difference in number of publications between the filtered and the unfiltered approaches divided by the number of publications in the unfiltered approach.

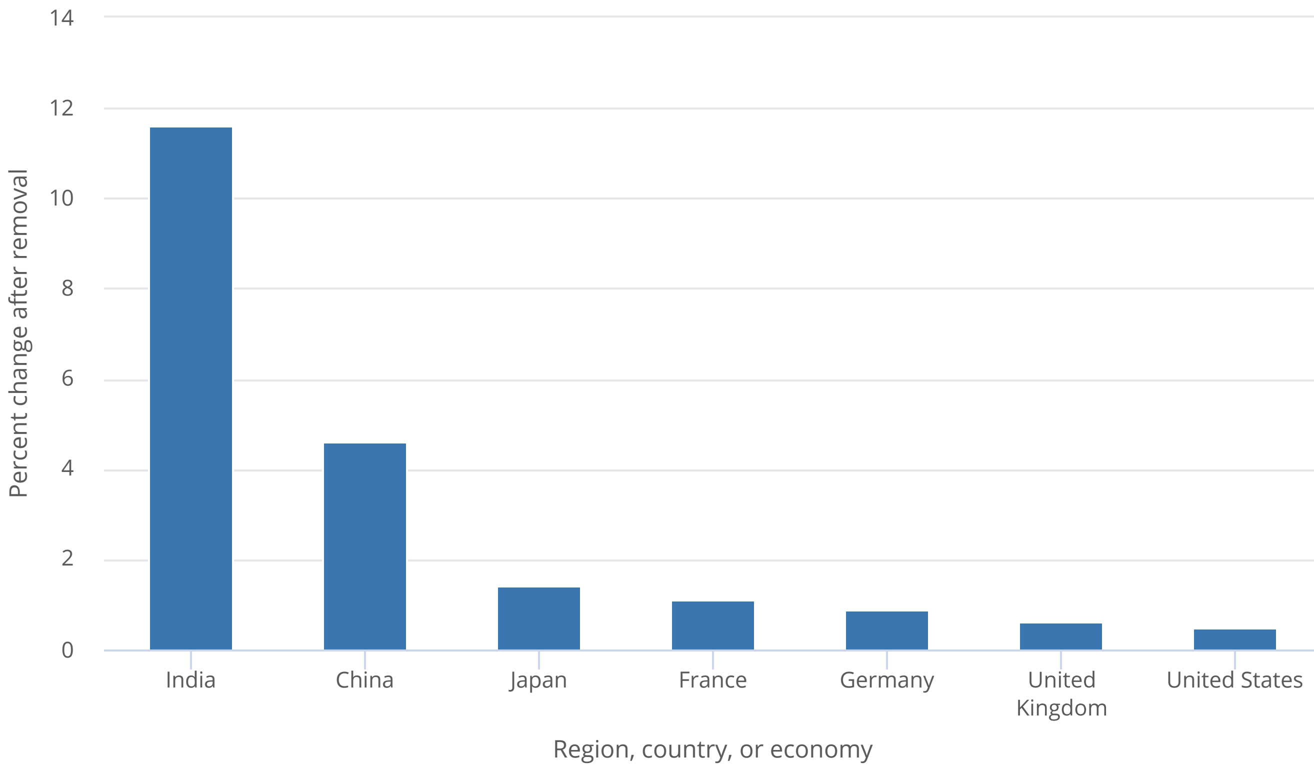
Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

The filtering has different impacts by region, country, or economy and field of science. NCSES has examined the effect of this filtering to better understand any potential bias. **Figure SAPBS-2** shows the numerical impact of the filters by region, country, or economy. From 2008 to 2022, India had the largest percentage of articles removed (approximately 12% of India's total unfiltered article count), followed by China (5%) (**Figure SAPBS-2**). By subject field, materials science (18% of articles filtered out) and computer and information sciences (9%) were the fields with the most filtered articles (**Figure SAPBS-3**).

Figure SAPBS-2

Impact of removing low-quality publications from Scopus, by selected region, country, or economy: 2008–22



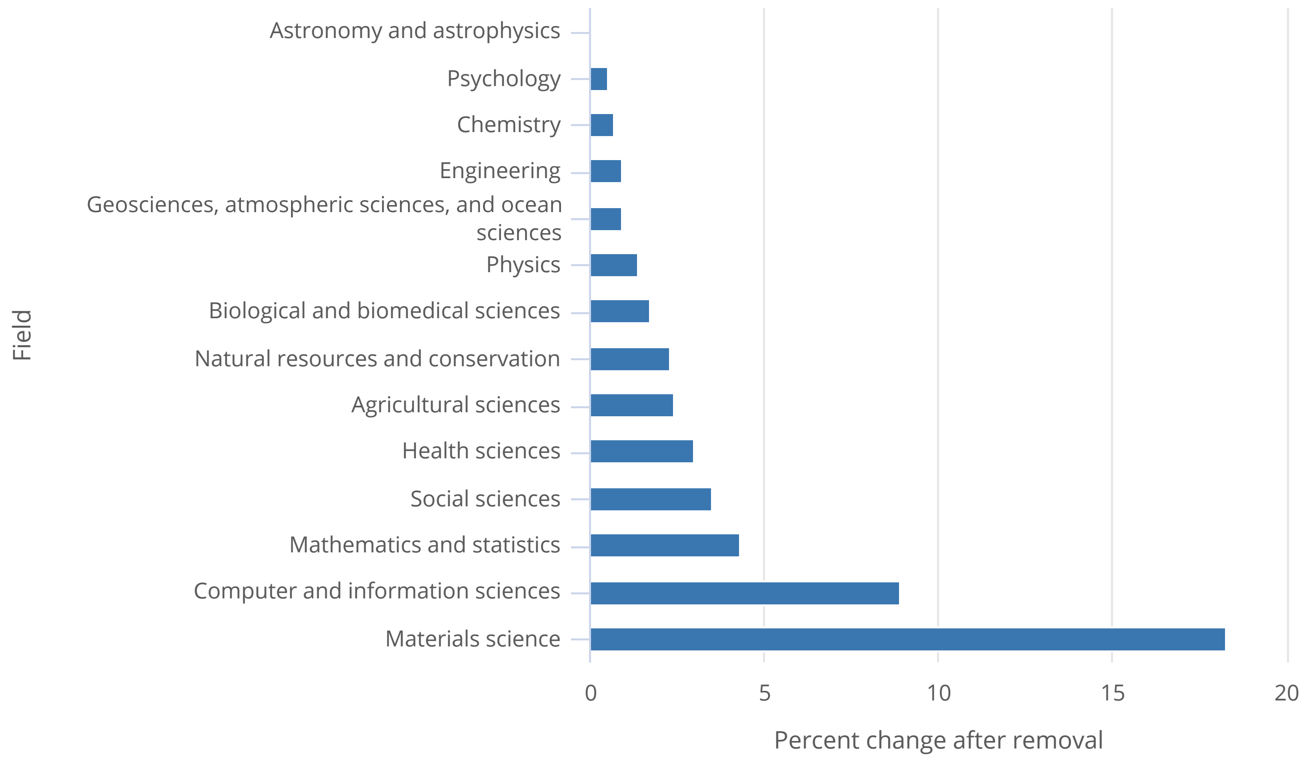
Note(s):
Articles are credited on a fractional count basis (i.e., for articles from multiple regions, countries, or economies, each region, country, or economy receives fractional credit on the basis of the proportion of its participating authors). Percent change is computed as the difference in number of publications between the filtered and the unfiltered approaches divided by the number of publications in the unfiltered approach.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

Figure SAPBS-3

Impact of removing low-quality publications from Scopus, by field of science: 2008–22



Note(s):
Percent change is computed as the difference in number of publications between the filtered and the unfiltered approaches divided by the number of publications in the unfiltered approach.

Source(s):
National Center for Science and Engineering Statistics; Science-Metrix; Elsevier, Scopus abstract and citation database, accessed April 2023.

Science and Engineering Indicators

English-Language Bias

The Scopus database is constructed from articles and conference proceedings with an English-language title and abstract; therefore, the database contains an unmeasurable bias because not all science and engineering (S&E) articles and conference proceedings meet the English language requirement (Elsevier 2020). Scopus uses English because of its widespread acceptance as the assumed global language of science, and research shows that a representative plurality of researchers publish in English (Amano, González-Varo, and Sutherland 2016). Most notably, this restriction may undercount contributions from non-English-speaking regions, countries, or economies, but multiple factors influence global researchers' decisions to publish in English that may also affect the prominence of non-English-speaking regions, countries, or economies in Scopus.

Publication output data for non-English-speaking regions, countries, or economies also underestimate the S&E research output for China, Japan, and others (Amano, González-Varo, and Sutherland 2016; Xie and Freeman 2019). Therefore, when comparing English-speaking and non-English-speaking regions, countries, or economies (e.g., **Figure PBS-3**; **Table PBS-1**), there is a bias such that the non-English-speaking ones will be undercounted in the numbers of publications and citations. The bias varies by scientific field. Bibliometric researchers have found a native-language preference in citations (Liang, Rousseau, and Zhong 2012), and the social sciences exhibit more substantial linguistic bias than physical sciences, engineering, and mathematics (Mongeon and Paul-Hus 2015). One solution undertaken by Elsevier is to increase publications from non-English-speaking regions, countries, or economies.

Conference Papers Removed from the Highly Cited Article Ratio

Conference papers are included in the data analyzed in the report for output but not for HCA computation. Conference papers may bias HCA because of uneven inclusion in Scopus and widely different citation patterns compared with those of journal articles.

Generally, conference papers are cited less frequently than journal articles. Thus, a greater proportion of conference papers in the total article count for a region, country, or economy would tend to reduce impact based on HCA computation. Depending on the normalization approach, the score of regions, countries, or economies could be heavily impacted compared with others because conference papers represent a larger proportion of their output. The issue is demonstrated in a simplified two-country example. Both countries publish 1,000 journal articles, but one publishes 10 conference papers, and the other publishes 200 conference papers. Assume that, based on the 1,000 journal articles, both countries have the same impact. If scores are lower for conference papers, adding conference papers into the computation will reduce the combined HCA score of the country with 200 conference papers. Therefore, in this case, two countries with similar impact in research published in journals may have a different overall impact because of the propensity of one to also send people to conferences.

In other cases, conference papers can increase the HCA for a region, country, or economy. Some fields of science publish and cite conference proceedings at different rates. A field with a low average number of citations for each article can have an increase in HCA by including conference papers, even those with relatively few citations. For example, if the country's average number of citations is one and a conference paper receives two citations, the normalized impact of the conference paper will be 2.0, which is quite high. Adding these *high-impact* conference papers could boost the score of a country that specializes in a field with a below-average number of article citations and whose researchers participate in an above-average number of conferences.

The *Indicators 2024* PBS report removes conference proceedings from the HCA but keeps conference proceedings in the total output and collaboration analysis because conference proceedings for some fields and regions, countries, or economies are an important component of their output. The exception to this rule is for computer and information sciences. The HCAs for computer and information sciences for journal articles alone and for journal articles plus conference proceedings show similar HCA trends for the United States and China (Table SPBS-69). For both countries, including conference proceedings boosts the HCA from 2014 through 2018. Computer and information sciences was selected for this comparison because conference proceedings are close to 20% of the references in that field (Lisée, Larivière, and Archambault 2008).

Field Categorization

Articles were categorized by S&E fields corresponding to the 14 fields of science in the NCSES Taxonomy of Disciplines (TOD) (Science-Metrix 2023). This categorization first assigns the journal to one of the 176 subfields in the Science-Metrix classification, then to the TOD. This approach works well for most journals and fields—all of *dentistry* is assigned to *health sciences*, for example. Challenges arise for subfields that are more general, such as *energy*, and for multidisciplinary journals, such as *Science* or *Nature*. For these fields and journals, classification occurs at the article level based on an algorithm that uses author affiliations, the names of journals referenced in the bibliography, the titles of the references, the publication's abstract, the publication's author-defined keywords, the publication's title, and the scientific field of references.

Fractional and Whole Counting

Collaboration, including international collaboration, is increasingly common in scientific research, resulting in papers that are co-written by two or more coauthors. Co-written articles can pose challenges for appropriately allocating paper credit, by individual authors and by region, country, or economy, in the case of international collaboration. Articles with authors working in multiple regions, countries, or economies were accounted for in determining publication output by region, country, or economy and international collaborations, but different strategies were required for allocating credit. First, each article's affiliated region, country, or economy was determined by the institutional address of each author as listed in the article. Fractional counting was employed to determine *country output*, because each country received a fractional contribution to each paper based on the number of coauthors from that country. *Country output* aims to measure each country's contribution to worldwide publication, and fractional counting prevents double-counting collaborative papers and helps normalize the output of highly collaborative countries. Full counting was employed in measures of *international collaboration* because each region, country, or economy represented by one or more authors was counted once. The *international collaboration* analyses positioned each paper as an instance of international collaboration, such that whole counting was appropriate. Because whole counting was used for international collaboration and fractional counting for publication output, those values were not directly comparable. Finally, it should be noted that although the listed affiliation is generally reflective of the locations where research was conducted, authors may have honorary affiliations, have moved, or have experienced other circumstances preventing their affiliations from being an exact corollary to the research environment.

Network Analysis of Artificial Intelligence Papers

The sidebar **Artificial Intelligence Publication Output and International Collaboration** uses a network graph to show the interrelatedness of the international research effort. Visualization of the network uses a structure of nodes (representing regions, countries, or economies) and edges (representing connections between the regions, countries, or economies). A force-directed layout algorithm defines optimal bidimensional positions for the graph's nodes. The force-directed layout algorithms used are nondeterministic, such that even from identical starting positions and layout parameters, the final network layout may be slightly different. However, groups of nodes that are strongly linked together (clusters) will tend to be visually close to each other. As such, the position of a single node carries no meaning on its own; it must be compared with that of the others to obtain insights into the collaboration ecosystem. The table below provides terminology and definitions for an in-depth understanding of the network analysis (**Figure PBS-B** and **Figure PBS-C**; Table SPBS-91).

Indicator	Description
Degree	The degree of a node is the number of edges connected to the node. In the context of an international collaboration network, this corresponds to the number of other regions, countries, or economies with which the region, country, or economy has collaborated. In this case, the maximum value of this indicator is the number of nodes in the network minus one because the node that has its degree computed cannot have collaborated with itself.
Node strength	The node strength is the sum of the weights of edges connected to the node. For international collaboration, a single paper can generate multiple collaboration links. For example, if one author from the United States cowrote an article with two authors from France and one author from Canada, this article generates three collaboration links: United States–France, United States–Canada, and Canada-France, each with a weight of one, regardless of the number of authors.

Betweenness centrality	Betweenness centrality measures how often a given node in a network lies along the shortest paths between two other nodes that are not directly connected to one another. For example, this indicator would highlight entities that play an important *brokering* role, acting as a connecting link between entities that do not co-publish with one another directly. Nodes with a high betweenness centrality score are the bridges that connect relatively isolated islands of research communities within the overall topography. These entities play an important role in the interconnection of subgroups within the network as a whole.
Closeness centrality	Closeness centrality assesses the degrees of separation between one node and other nodes within a network. That is, it assesses the length of the chains that connect a given node to the rest of its community. Although, for example, betweenness centrality highlights entities that play an interconnecting role for their community, closeness centrality measures the level of access that a given entity has to its surrounding community. It highlights those that can tap into a large section of a network without passing through many degrees of separation or through distant and mediated connections. When calculating closeness centrality, a node directly connected to every other node in the network would score 1, the highest possible closeness centrality score.
Weighted eigenvector centrality	Weighted eigenvector centrality is a measure of the level of integration of a node in a collaboration network. The level of integration of nodes within a collaboration network is reflected by the number of nodes to which they are connected and the quality of their collaborations (i.e., the strength of the ties measured by the number of coauthored publications and the importance of the nodes to which they are connected in the network). The mathematical definition of eigenvector centrality is such that the centrality score of a node in a network is proportional to the sum of the centrality scores of all nodes connected to it. Thus, this indicator offers a good appreciation of the number and quality of an entity's collaborations because connections to high-scoring nodes contribute more to the score of that entity than equal connections to low-scoring nodes. A node scoring high for this indicator operates closer to the core of the network than a low-scoring node. High-scoring nodes are central and highly important to the network's structure. Eigenvector centrality provides a good appreciation of the integration of individual entities within a network; that is, the higher the score, the more integrated the entity. The weighted version of the indicator accounts for the size of the tie between nodes. Centrality scores are typically normalized between 1 (most central node) and 0 (least central node).
Weighted PageRank	PageRank, made famous through its use by the Google search engine, is a variant of eigenvector centrality. It can be thought of as the result of a *random walk*, meaning that the PageRank score of a given node corresponds to the probability that someone starting on a random node of the network and randomly following edges will end the walk on a particular node. The weighted version of the algorithm makes stronger links more likely to be followed than weaker links. PageRanks are shown as percentages to clearly indicate the share of random walks in which the end point was the given node. All scores sum to 100%. In undirected networks, weighted PageRank yields results very similar to node strength.

Key to Acronyms and Abbreviations

DOAJ: Directory of Open Access Journals

HCA: highly cited article

NCSES: National Center for Science and Engineering Statistics

PBS: Publications Output: U.S. Trends and International Comparisons

S&E: science and engineering

TOD: Taxonomy of Disciplines

References

Amano T, González-Varo JP, Sutherland WJ. 2016. Languages Are Still a Major Barrier to Global Science. *PLOS Biology* 14(12):e2000933. Available at **http://journals.plos.org/plosbiology/article?id=10.1371%2Fjournal.pbio.2000933**. Accessed 26 August 2023.

Beall J. 2012. Predatory Publishers Are Corrupting Open Access. *Nature* 489(179). Available at **https://www.nature.com/articles/489179a**. Accessed 26 August 2023.

Bohannon J. 2013. Who's Afraid of Peer Review? *Science* 342(6154):60–65. Available at **https://www.science.org/doi/10.1126/science.342.6154.60**. Accessed 26 August 2023.

Carey K. 2016. A Peek Inside the Strange World of Fake Academia. *New York Times* December 29:3. Available at **https://www.nytimes.com/2016/12/29/upshot/fake-academe-looking-much-like-the-real-thing.html?_r=0**. Accessed 26 August 2023.

Elsevier. 2020. *Scopus Content Coverage Guide*. Available at **https://www.elsevier.com/?a=69451**. Accessed 26 August 2023.

Kolata G. 2013. Scientific Articles Accepted (Personal Checks, Too). *New York Times* April 8:1. Available at **https://www.nytimes.com/2013/04/08/health/for-scientists-an-exploding-world-of-pseudo-academia.html**. Accessed 26 August 2023.

Liang L, Rousseau R, Zhong Z. 2012. Non-English Journals and Papers in Physics: Bias in Citations? *Scientometrics* 95(1):333–50. Available at **https://www.issi-society.org/proceedings/issi_2011/ISSI_2011_Proceedings_Vol1_49.pdf**. Accessed 26 August 2023.

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* 59:1776–84. Available at **https://doi.org/10.1002/asi.20888**. Accessed 26 August 2023.

Mongeon P, Paul-Hus A. 2015. The Journal Coverage of Web of Science and Scopus: A Comparative Analysis. *Scientometrics* 106:213–28. Available at **https://doi.org/10.1007/s11192-015-1765-5**. Accessed 26 August 2023.

National Science Board (NSB), National Science Foundation. 2016. New Data Source for Indicators Expands Global Coverage. *Science and Engineering Indicators 2016* (*Indicators 2016*). NSB-2016-2. Alexandria, VA. Available at **https://www.nsf.gov/statistics/2016/nsb20161/#/sidebar/chapter-5/new-data-source-for-indicators-expands-global-coverage**.

Schneider JW, van Leeuwen T, Visser M, Aagaard K. 2019. Examining National Citation Impact by Comparing Developments in a Fixed and Dynamic Journal Set. *Scientometrics* 119(2):973–85. Available at **https://doi.org/10.1007/s11192-019-03082-3**. Accessed 26 August 2023.

Science-Metrix. 2023. *Bibliometric Indicators for the Science and Engineering Indicators 2024. Technical Documentation.* Available at **https://science-metrix.com/bibliometrics-indicators-for-the-science-and-engineering-indicators-2024-technical-documentation/**. Accessed 26 August 2023.

Visser M, van Eck NJ, Waltman L. 2020. *Large-Scale Comparison of Bibliographic Data Sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic.* Available at **https://arxiv.org/abs/2005.10732**. Accessed 26 August 2023.

Xie Q, Freeman RB. 2019. Bigger Than You Thought: China's Contribution to Scientific Publications and Its Impact on the Global Economy. *China & World Economy* 27:1–27. Available at **https://doi.org/10.1111/cwe.12265**. Accessed 26 August 2023.

Contact Us

To report an issue with the website, please e-mail ncsesweb@nsf.gov. For questions about the National Science Foundation (NSF), please visit the NSF help page at **https://nsf.gov/help/**. To see more from the National Science Board, please visit **https://nsf.gov/nsb/**.

Report Authors

Benjamin Schneider
Interdisciplinary Science Analyst
National Center for Science and Engineering Statistics (NCSES)
beschnei@nsf.gov

Jeffrey Alexander
Director, Innovation Policy
RTI International, under contract to NCSES

Patrick Thomas
Founding Partner
1790 Analytics LLC, under contract to RTI International

NCSES

National Center for Science and Engineering Statistics
Directorate for Social, Behavioral and Economic Sciences
National Science Foundation
2415 Eisenhower Avenue, Suite W14200
Alexandria, VA 22314
Tel: (703) 292-8780
FIRS: (800) 877-8339
TDD: (800) 281-8749
ncsesweb@nsf.gov