DESIDOC Journal of Library & Information Technology, Vol. 45, No. 4, July 2025, pp. 350-366, DOI : 10.14429/djlit.21026

Open Data in Social Sciences: Growth, Impact, and Equity in Data Paper Publishing

Sanjoy Kar[~,*] and Durga Sankar Rath[#]

[~]*Institute of Development Studies Kolkata (IDSK), Kolkata - 700 064, India*
[#]*Vidyasagar University, Rangamati, Midnapore - 721 102, West Bengal, India*
[*]*E-mail: sanjoy@idsk.edu.in*

ABSTRACT

The rapid growth of data-driven research has elevated the prominence of data papers as a specialised scholarly publication format, which enhances data accessibility, transparency, and reproducibility in scientific research. This study provides a comprehensive analysis of peer-reviewed data papers in social science, examining their growth, scholarly impact, adoption trends, mandates, policies, and funding landscape across the globe. Results show a 36 % annual growth rate ($R^2 = 0.865$), with 83 % of data papers published after 2021, driven by open-access mandates, funding agency requirements, digital repositories and growing emphasis on open science. The United States and China dominate publication volume, while Switzerland and the UK lead in citation impact. Despite a weak but significant open-access citation advantage ($r = 0.052$, $p < 0.001$), 22.7 % of data papers remain uncited, reflecting a "citation paradox." Altmetric data highlights societal impact through media mentions (46 %), policy influence (36 %), patents (9 %) and engagement across social media platforms (X, Facebook, etc). Collaboration and funding patterns reveal entrenched Global North-South disparities, with 75 % of publications and 78 % of collaborative strength concentrated in the Global North. Only 42.5 % of journals enforce FAIR principles, and 35 % address CARE compliance, highlighting policy inconsistencies. To advance equitable open science, the study recommends standardised ethical frameworks, equitable funding models, and institutional support for global south scholars. These insights aim to strengthen data-sharing norms, promote research transparency, and foster inclusive collaboration.

Keywords: Data publication; Data journal; FAIR data; Open science; Data-sharing; Bibliometrics; Altmetrics; Social sciences

1. INTRODUCTION

Data serves as the foundation of social science research, empowering scholars to explore human behaviour, societal dynamics, and global phenomena. The social sciences have undergone a profound transformation with the advent of the data revolution, shifting from traditional qualitative methodologies to advanced computational and quantitative approaches[1]. The exponential growth of digital data, the availability of large-scale datasets, and the rise of open-access repositories have facilitated new ways of conducting research and policy analysis. Today, real-time data from social media, mobile devices, and satellite imagery[2] supports complex analyses, while qualitative and quantitative approaches provide complementary insights into cultural subtleties and population-level patterns[3,4]. This revolution has enhanced research accuracy, reproducibility, and the capacity to handle complex social phenomena with empirical rigor.

The increasing availability of research data has led to new forms of academic publication, one of which is the 'data paper'. Data papers serve as scholarly outputs that describe datasets, ensuring their discoverability, accessibility, reuse, and citation within the broader scientific community. Thus, this distinct publication format plays an instrumental role in advancing open science and improving the transparency and reproducibility of research in the age of scientific crisis[5-9]. As social science research has become more data-intensive[10], the formal recognition of datasets as standalone research contributions has gained traction. In social sciences, where qualitative and quantitative datasets play a crucial role, structured data papers enhance interdisciplinary research and evidence-based policymaking[11]. However, the growing emphasis on open data and data papers has led to the emergence of specialised data journals that prioritize dataset description and accessibility. A study identified 116 such journals that publish data papers, among which three focuses on social sciences[12]. However, a recent study[13] identified such 9 journals. Some of the most popular dedicated data journals include Data in Brief, Journal of Open Psychology Data and the (CODATA) Data Science Journal.

The origins of data papers are closely linked to the broader open science movement, which advocates for transparency and accessibility in research. The Bromley

Received : 28 February 2025, Revised : 12 April 2025
Accepted : 21 May 2025, Online published : 15 July 2025

Principles (1991) first emphasised the need for open access to primary scientific data, setting the stage for subsequent policies such as the Berlin Declaration on Open Access (2003) and the OECD Principles and Guidelines for Access to Research Data (2004). These initiatives underscored the importance of data availability for scientific progress. The formalisation of data papers as a distinct publication type emerged in biodiversity and environmental sciences, where large-scale datasets required structured dissemination. The Global Biodiversity Information Facility (GBIF) and Pensoft Publishers pioneered the concept by integrating data papers into their scholarly journals[14]. This model was later adopted in other disciplines, including social sciences, where data availability became essential for reproducibility and meta-analysis[15].

However, the growing reliance on data-driven methodologies has exposed vulnerabilities. During COVID-19 pandemic time, rapid data collection and analysis enabled timely insights into public health behaviours and policy impacts but also revealed instances of misconduct, including data manipulation and falsification. High-profile retractions, such as those in The Lancet due to flawed datasets, underscored the consequences of unethical behaviour[16]. Similarly, errors in widely cited papers in social science, such as coding mistakes in Excel spreadsheets leading to retractions in economics[17], highlighted the urgent need for transparency and accountability in data handling. Studies have documented research misconduct in social science, including data manipulation and falsification, using retraction watch data[18,19]. Such cases highlight the critical need for open data publishing to ensure research integrity and reproducibility. To address these challenges, journals, funding agencies and institutions have implemented stricter guidelines for data sharing, replication studies, and conflict-of-interest disclosures. Initiatives such as the Transparency and Openness Promotion (TOP) guidelines aim to promote responsible conduct by incentivizing transparency and accountability[20]. In this context, data papers have played as critical tools for promoting reproducibility and advancing open science[21]. Unlike traditional research articles, data papers prioritize dataset documentation, ensuring discoverability, accessibility, reuse, and citation within the broader scientific community. These publications play an instrumental role in improving transparency and reproducibility, particularly amid the replication crisis in social science[22]. Despite their potential, the adoption of data papers in social science lags behind natural sciences due to disciplinary norms, ethical complexities, and inconsistent adherence to FAIR principles (Findable, Accessible, Interoperable, and Reusable)[23-24]. Only a few of social research datasets meet all FAIR criteria[25], underscoring gaps in methodological rigor and knowledge dissemination. As social science research becomes increasingly data-intensive[10], structured data papers enhance interdisciplinary collaboration and evidence-based policymaking[11]. By formally recognising datasets as standalone contributions, data papers advance open science and address the pressing need for transparency in an age of scientific crises[7-9].

Despite the growing prominence of data papers, there is limited empirical research on their role and impact in social science. Existing studies tend to focus on specific disciplines or case studies, leaving gaps in our understanding of broader trends and challenges[26]. This study investigates the role of data papers in advancing transparency and reproducibility in social science through a longitudinal analysis. By examining trends, thematic clusters, and collaboration networks, we aim to address gaps in understanding their impact and contribute to the growing emphasis on open science. Notably, this study presents the first systematic, data-driven assessment of the evolution, adoption, and scholarly impact of data papers across disciplines, offering critical insights into their role in advancing open science and research transparency.

2. OBJECTIVES

The primary objectives of this study are as follows:

- To examine the evolution and growth of data papers in social science research through a systematic and evidence-based analysis.
- To quantify publication trends and citation impact of data papers across disciplines, institutions, and geographic regions.
- To map collaborative networks among authors, institutions, and countries involved in data paper publishing, highlighting patterns of research cooperation.
- To evaluate the role of funding agencies and journal policies in influencing data-sharing practices, with a particular focus on adherence to the FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles.
- To investigate Global North-South disparities in data-sharing norms and identify systemic barriers and enabling factors influencing the adoption of data papers in the Global South.
- To assess the citation advantage of open-access data papers compared to paywalled counterparts using Pearson's correlation analysis.
- To propose strategic insights and policy recommendations aimed at enhancing the visibility, accessibility, and scholarly impact of data papers, thereby promoting an open, transparent, and equitable research ecosystem in the social sciences.

3. LITERATURE REVIEW

Data papers emerged in the early 2000s in genomics[27] to address the "data deluge"[28-29]. Journals such as Scientific Data and the Journal of Open Psychology Data formalised the format by emphasising structured metadata. In social sciences, data papers gained momentum after 2015, driven by funder mandates (e.g., NIH's Data Sharing Policy) and open-access advocacy[30]. Despite this growth, data papers

exhibit a "citation paradox," receiving fewer citations than traditional articles[31-32] but gaining higher Altmetric attention due to reuse in policy and education. Citation practices and data reuse cultures vary across disciplines, further complicating their adoption[33-35].

The adoption of data papers has been fuelled by institutional mandates, funding agency requirements (e.g., NSF, NIH), journal policies (e.g., PLOS ONE), technological advancements, and growing emphasis on open science[36]. Studies show that social science articles with shared datasets receive twice as many citations as those without open data[37]. Similarly, Zhang and Ma[38] confirmed this citation advantage in economics using a causal inference approach. However, despite these benefits, many researchers remain reluctant to share data due to insufficient incentives[39] and ethical concerns. Meyer[40] found that 66 % of qualitative researchers avoid data sharing due to re-identification risks, even with anonymised data. Moreover, Institutional Review Boards (IRBs) often lack consistent guidelines for data paper ethics, resulting in inconsistent practices[41].

Recent studies highlight disciplinary differences in data paper adoption. Chung[42] analysed 713 data papers, identifying Biotechnology and Physics as dominant fields with distinct citation patterns. A later study by Chung[43] analysed 6,086 data papers and 84,908 cited references, uncovering trends in authorship, analytical tools, and key research areas such as genome composition and environmental sciences.

However, significant gaps remain in the adoption and impact of data papers in social sciences. Concerns over intellectual property, ethical considerations, and the absence of standardised guidelines hinder broader adoption. While STEM disciplines demonstrate established citation patterns and collaborative networks, the impact of data papers in social sciences remains underexplored. The variability in citation practices and data reuse cultures across disciplines further complicates the integration of data papers into mainstream academic discourse. This study addresses these gaps by conducting a longitudinal analysis to assess the evolution, scholarly impact, and disciplinary challenges of data papers in the social sciences, contributing to a deeper understanding of their role in the open science ecosystem.

4. METHODOLOGY

This study systematically investigates trends and visualises the landscape of peer-reviewed data publications in social science on a global scale, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[44] framework to ensure transparency, reproducibility, and rigor. Both quantitative methods and knowledge-mapping techniques were employed to analyze publication trends, collaboration networks, and citation impact, providing insights into how data papers advance transparency, reproducibility, and open science. The Scopus database was selected as the primary source due to its comprehensive coverage, reliability, and advanced search capabilities, including metadata fields essential for analysis[45-48]. To ensure consistency, datasets were retrieved during the first week of February 2025.

The search strategy used Scopus' All Science Journal Classification (ASJC) codes and SUBJAREA filters to address the interdisciplinary nature of social sciences and minimize terminology variability. The search string targeted publications classified under Business, Management, and Accounting (BUSI), Decision Sciences (DECI), Economics, Econometrics, and Finance (ECON), Psychology (PSYC), and Social Sciences (SOCI), excluding unrelated categories and non-English publications. The final dataset comprised 3,957 peer-reviewed data papers published up to 2024, after excluding non-empirical articles, non-English items, and incomplete records (Fig. 1).

Data cleaning involved resolving inconsistencies, missing values, and redundancies in the raw dataset, including standardising author names, institutional affiliations, and keywords. Citation data were retrieved from Google Scholar, Scopus, Web of Science, and Dimensions, with Altmetric scores obtained from Dimensions. Journal Impact Factors were collected from the Journal Citation Reports®. The cleaned dataset was analysed using Microsoft Excel, Python, VOSviewer[49] and the Bibliometrix package[50] applying descriptive statistics, fractional counting, and citation normalisation techniques. Growth trends were evaluated using polynomial regression, with statistical significance determined at $p < 0.05$. Visualisation tools facilitated robust insights into the dataset while maintaining analytical rigor.

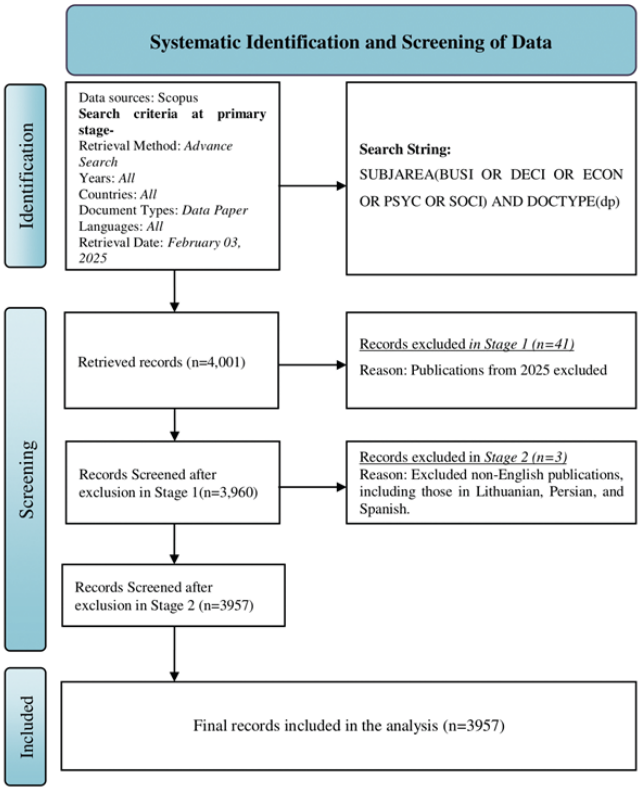


Figure 1. PRISMA workflow for data collection.

5. RESULTS

5.1 Data Papers and Temporal Growth Trends

The publication of Data Papers (DPs) in social sciences has grown significantly, with 3,957 DPs published between 2013 and 2024, reflecting an annual growth rate of 36 %. This surge, particularly after 2021 (Fig. 2A), is driven by data-sharing mandates, advancements in digital repositories, and widespread adoption of open-access policies. A linear regression model (R^2= 0.865, p < 0.0016) confirms a strong positive correlation between publication volume and time, highlighting the increasing academic recognition of DPs.

Geographically, the United States remains the dominant contributor (R^2 = 0.67, p < 0.045, τ = 0.82), while China has shown remarkable growth post-2018 (R^2=0.92, p < 0.002, τ = 0.1), reflecting national efforts to promote open science. Germany, the United Kingdom, and Italy demonstrate steady growth (R^2 between 0.72 and 0.85), while Spain and Switzerland exhibit moderate but consistent contributions. These trends signal a paradigm shift in scholarly communication, emphasising structured data documentation to enhance transparency and reproducibility. The sustained growth of DPs highlights their increasing importance in fostering open science and promoting responsible research practices across disciplines.

5.2 Geographic Trends and Research Impact

The geographic distribution reveals significant disparities in research productivity and impact. Fig. 2B and Table 1 show that among 160 contributing countries, the United States leads in publication volume (1,272), followed by China (1,092), Germany (570), and the United Kingdom (567). However, the Compound Annual Growth Rate (CAGR) analysis indicates that China (50.4 %) experienced the highest growth, followed by Italy (35.7 %) and Canada (35.5 %), while the United States shows slower growth, suggesting a plateau in DP output. Developed economies in the Global North dominate DP publications, contributing 75 % of the total output, whereas the Global South accounts for only 25 %, reflecting persistent research disparities (Fig. 2D and 2E).

In terms of research impact (Table 1), the US has the highest total citation count (29,671), followed by China (17,696), and the UK (16,573). However, mean citations per paper position Switzerland at the top (52.17), followed by the UK (29.23) and Australia (25.44). The h-index confirms the US' sustained influence (h = 75). In contrast, the Global South faces challenges in achieving similar citation impact, hindered by limited international collaboration (Fig. 3), access barriers, and lower visibility in high-impact journals. These disparities highlight the need for equitable funding and policy reforms to bridge the research gap.

Table 1. Top 10 countries in terms of data publications

Rank	Country	DP	CAGR%	Citations	MC	*h*-index
1	USA	1272	16.4	29671	23.33	75
2	China	1092	50.4	17696	16.21	58
3	Germany	570	32.0	12830	22.51	53
4	UK	567	27.1	16573	29.23	57
5	Italy	348	35.7	8513	24.46	46
6	France	334	27.9	7266	21.75	45
7	Australia	270	23.1	6870	25.44	36
8	Canada	264	29.3	5642	21.37	37
9	Spain	263	35.5	6667	25.35	40
10	Switzerland	212	26.2	11060	52.17	50

DP= Data paper, CAGR= Compound annual growth rate, MC= Mean citations

5.3 Collaboration Networks

Research collaboration enhances knowledge dissemination and research impact, analysed here at three levels: country, institution, and author.

5.3.1 Collaboration Network

The Country Collaboration Map (Fig. 3A) highlights global research partnerships, with the US leading in collaboration strength, followed by China, and the UK. European nations, such as France (523) and the Netherlands (486), maintain strong global ties. The US-China partnership (strength: 46) reinforces their dominance in global research productivity[51,52].

However, Global North (GN) countries dominate 78 % of collaborative strength, driven by dense intra-regional ties and high-impact GN-GN papers (mean citations: 28.6). In contrast, Global South (GS) nations exhibit weak integration, accounting for only 4 % of network strength, with limited South-South collaboration and lower citation impact (12.4 vs. 28.6 citations). Structural inequities hinder equitable knowledge production, leaving GS countries like India and Brazil with weaker ties to high-resource networks.

5.3.2 Institutional Collaboration Network

Figure 3B identifies the University of Chinese Academy of Sciences as the most influential institution (collaboration strength: 1,086), followed by Tsinghua University (895), and Peking University (752). Despite lower publication volumes, institutions such as the Finnish Meteorological Institute (h-index: 32.7) and Delft University of Technology (h-index: 29.8) contribute significantly to specialised research areas. The institutional network reflects an interdisciplinary approach, integrating diverse expertise to enhance research quality.

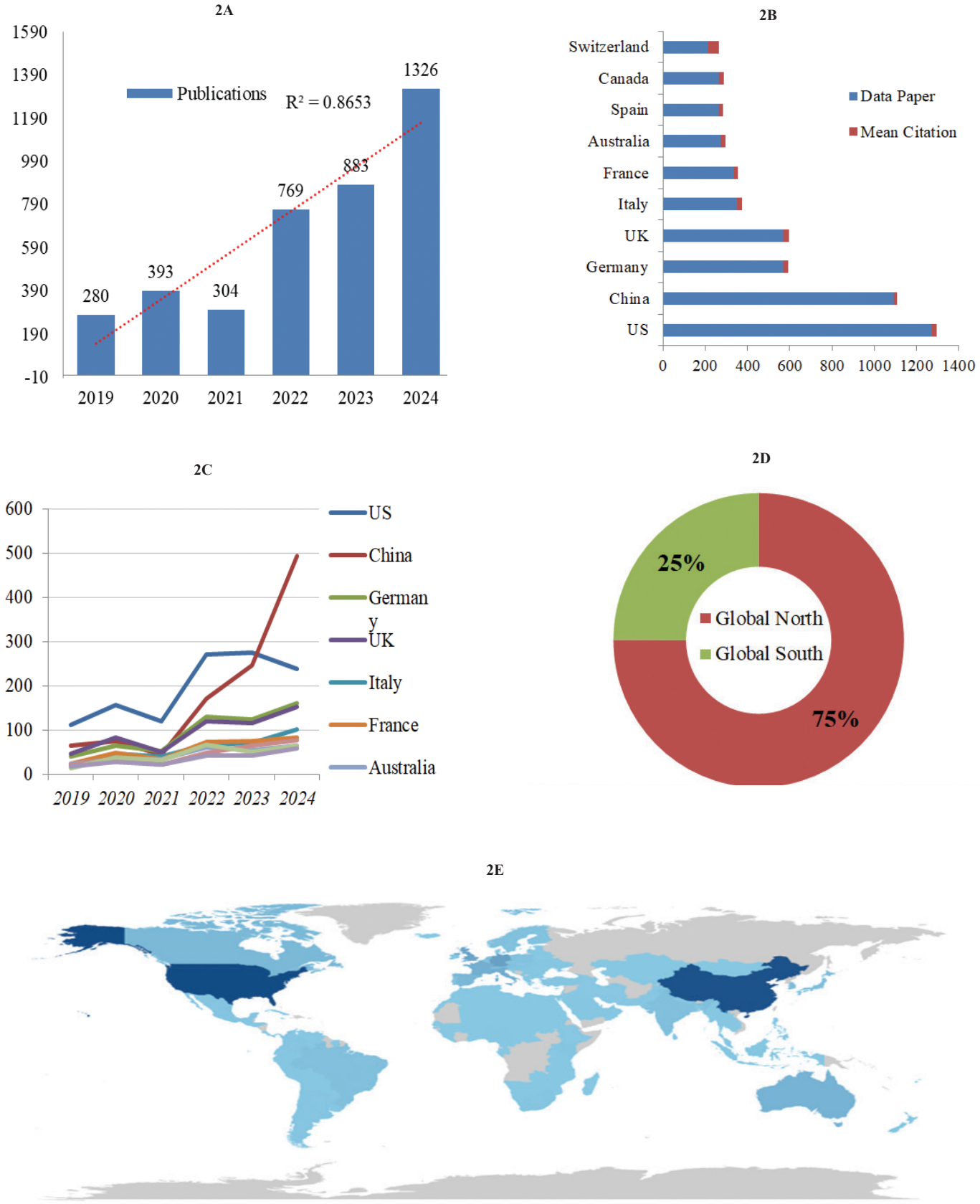


Figure 2. Growth and geographic influence of data papers in social sciences: (A) Annual growth of data papers (2019-2024). (B) DP outputs of top 10 productive countries and their research impact (Citation Mean). (C) DP publication trend in top 10 countries and territories over the years. (D) Global north vs. Global south. (E) Distribution of DP across the globe (Choropleth Map).

5.3.3 Author Collaboration Network

The Author Density Visualisation (Fig. 3C) highlights a highly interconnected co-authorship network, with influential researchers such as Liu, Zhu (citations: 341), Ciais, Philippe (citations: 398), and Shan, Yuli (citations: 1,523) occupying central positions. The density of author clusters suggests that social science data research is primarily conducted through collaborative efforts. The Author Network Map underscores strong link strengths, positioning researchers like Chen, Lei, Jian Jianbo, and Shi Qiong as central figures in the collaborative network.

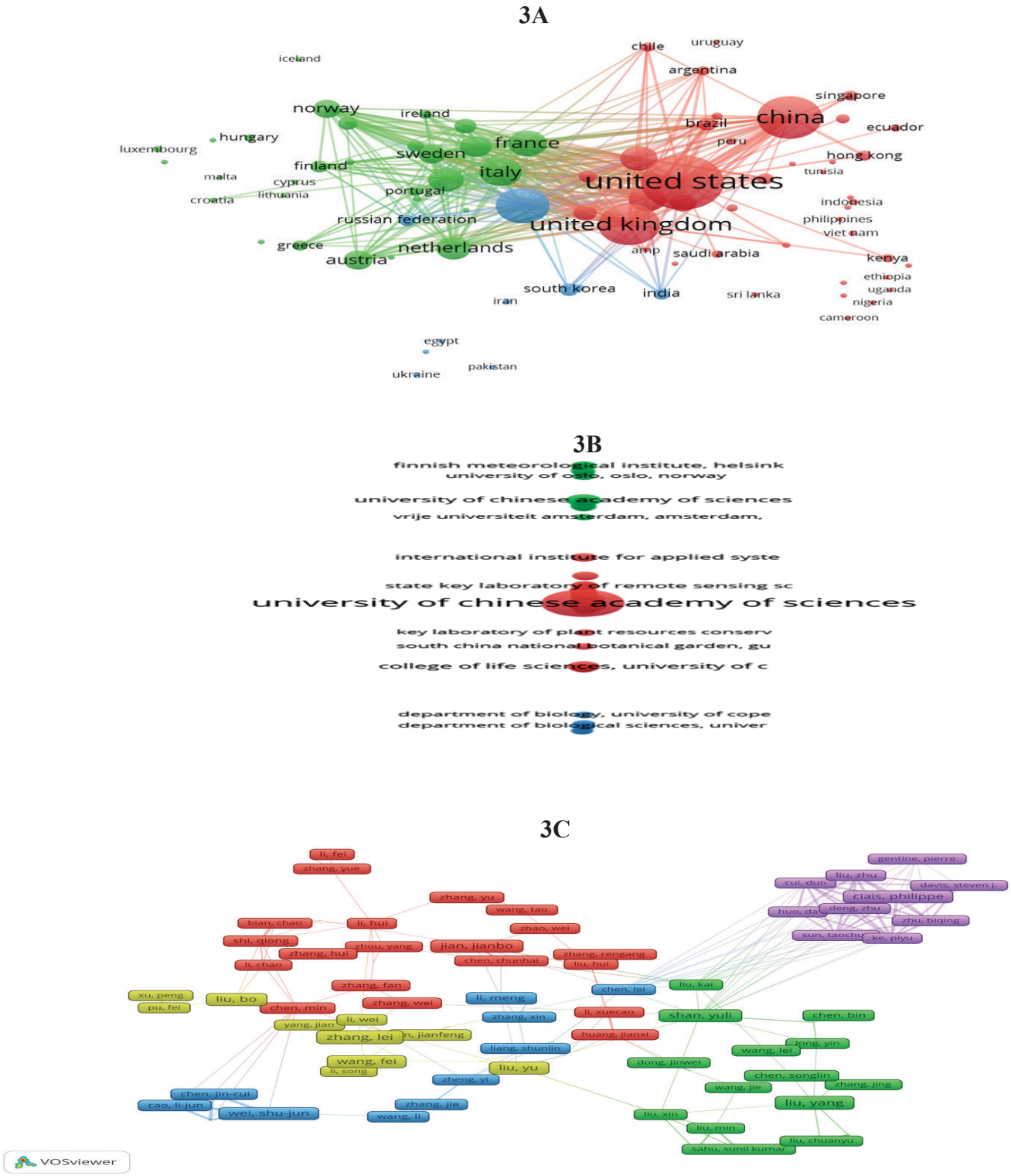


Figure 3. Research collaboration network: (A) Country collaboration (network visualisation), (B) Organisation (network visualisation), (C) Author collaboration (density visualisation).

5.4 Most Productive Institutions

Table 2 identifies the top 10 institutions contributing to data paper, emphasizing their publication volume, citation impact, and h-index. Among 20,991 institutions worldwide, 51 published at least five DPs during the study period. The Chinese Academy of Sciences leads globally (281 papers, 6,868 citations, h-index: 34), followed by CNRS (173 papers, 4,687 citations, h-index: 35). Government agencies such as the Ministry of Education of China (123 papers, mean citation: 33.55) and the Ministry of Agriculture of China (82 papers) play a pivotal role. Notable universities include Tsinghua University (85 papers, mean citation: 45.76) and leading European institutions such as INRAE, the University of Oxford, and the University of Cambridge.

5.5 Most Productive Authors

Table 3 highlights the top 10 most prolific authors in data publications, highlighting their document count, citation impact, mean citations per document, and h-index. Of 33,868 contributing authors, 68 have authored at least five DPs. China (5 authors) and the UK (3 authors) dominate this list. Yuli Shan (UK) leads with 11 papers, 1,523 citations (mean: 138.45), and an h-index of 9. Philippe Ciais (France) follows with 10 papers, 399 citations (mean: 39.90), and an h-index of 8. Jacqueline M. Cole (UK) and Zhenyu Liu (China) also exhibit notable contributions. These findings underscore the pivotal roles of UK, China, and France-based scholars in advancing data-driven research and open science in the social sciences.

Table 2. Top 10 institutions by data publications

Rank	Organisations	DP	Citations	MC	*h*-index
1	Chinese academy of sciences	281	6868	24.44	34
2	Cnrs centre national de la recherche scientifique	173	4687	27.09	35
3	University of chinese academy of sciences	158	3419	21.64	22
4	Ministry of education of the people's republic of china	123	4127	33.55	24
5	Inrae	88	2515	28.58	22
6	Tsinghua university	85	3890	45.76	24
7	Ministry of agriculture of the people's republic of china	82	752	9.17	12
8	University of oxford	80	1827	22.84	21
9	Consiglio nazionale dellericerche	72	1855	25.76	17
10	University of cambridge	67	1683	25.12	22

5.6 Analysis of Data Publication Venues and Citation Impact

The publication landscape of data papers has expanded significantly, with 39 journals publishing 3,957 papers between 2013 and 2024. Scientific Data (ISSN: 2052-4463) dominates the field, contributing 3,600 papers with 61,325 citations, a mean citation rate of 17.03, an h-index of 120, and an impact factor (5.8 in 2023). Data (ISSN: 2306-5729) ranks second with 107 papers (h-index: 38), followed by Journal of Open Humanities Data (n = 62) and Journal of Open Archaeology Data (n = 40), both by WPUP (Table 4).

Journals beyond the data-publishing domain also contribute significantly. Frontiers in Psychology (n=26) and the Morbidity and Mortality Weekly Report (n=17) have integrated data papers, reflecting the broader adoption of data dissemination. Other contributors include Review of International Organisations, Data Intelligence, and the Australian Economic Review.

Co-citation analysis (Fig. 4B) reveals strong linkages between Scientific Data, Frontiers in Psychology, and Data Intelligence. Bibliographic coupling analysis (Fig. 4C) positions Scientific Data and Data as central nodes in global research networks. Citation growth trends (Fig. 4D) confirm rising academic engagement with data papers, driven by journals such as Scientific Data and Review of International Organisations.

5.7 Citation Impact Analysis

Table 5 shows that 3,957 data papers accumulated 59,580 citations between 2019 and 2024, with an h-index of 98, reflecting significant academic impact. However, citation accumulation varies across publication years, necessitating the use of Time-Adjusted Normalized Citation Score (TANCS) and Citation Mean-Adjusted TANCS (CM-TANCS) to ensure fair comparisons.

TANCS normalises citation performance by adjusting for publication year differences:

$$\text{TANS} = \frac{C}{P \times A}$$

Table 3. Top 10 productive authors

Rank	Author	DP	Citations	MC	*h*-index
1	Jianbo Jian, China	11	67	06.09	5
2	Yuli Shan, UK	11	1523	138.45	9
3	Philippe Ciais, France	10	399	39.90	8
4	Jacqueline M. Cole., UK	10	307	30.70	8
5	Li-Jun Cao, China	8	6	00.75	1
6	Jin-Cui Chen, China	8	6	00.75	1
7	Zhenyu Liu, China	8	341	42.63	6
8	Qiong Shi, China	8	50	06.25	3
9	Dabo Guan, UK	7	959	137	6
10	Linda See, Austria	7	227	32.43	5

Table 4. Top 10 data journals by publications and citation metrics

Rank	Journal	DP	Citations	MC	*h*-index	IF (2023)	Cite score	SJR (2023)	ETD year	Publisher
1	Scientific data	3600	61325	17.03	120	5.8	11.2	1.94	2014	Springer
2	Data	107	203	1.90	38	2.2	4.3	0.5	2016	MDPI
3	Journal of open humanities data	62	138	2.23	3	NA	0.6	0.6	2015	WPUP
4	Journal of open archaeology data	40	68	1.70	6	NA	2.2	0.24	2012	WPUP
5	Journal of open psychology data	28	56	2.00	NA	NA	0.8	NA	2013	WPUP
6	Frontiers in psychology	26	160	6.15	184	2.6	5.3	0.8	2010	Frontiers
7	Morbidity and mortality weekly report	17	22	1.29	260	21	65.4	13.08	1981	CDC
8	Review of international organisations	10	1471	147.10	49	4.5	8.1	1.56	2006	Springer
9	Data intelligence	8	14	1.75	22	1.3	6.6	0.75	2019	CNPIEC
10	Australian economic review	6	14	2.33	35	1	1.9	0.36	1968	WILEY

IF= Impact Factor, SJR= SCImago Journal Rank, RIS= Research Impact Score

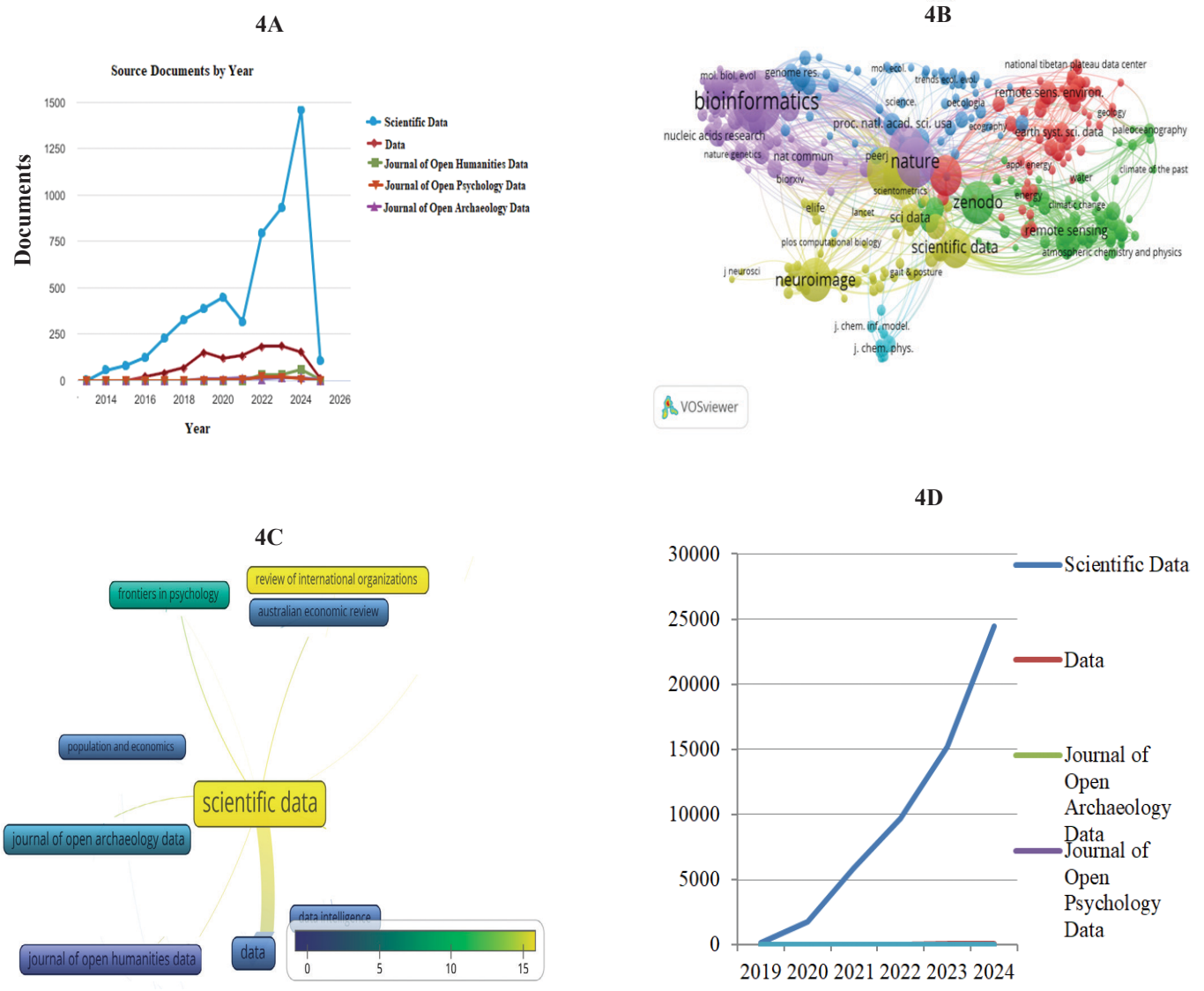


Figure 4. Publication venues for data papers (data journals): (A) Source production over time. (B) Co-citation analysis of the source journals. (C) Bibliographic coupling (source overlay). (D) Citation growth of top 5 journals.

Where C is total citations, P is the number of papers, and A is the citation age factor.

CM-TANCS refines TANCS by incorporating the citations mean (M) for each year relative to the overall mean $(\bar{M})$, adjusting for variations in citation behavior:

$$\text{CM-TANCS} = \frac{C}{P \times A} \times \frac{M}{\bar{M}}$$

Table 5 highlights that older papers accumulate more citations, with the 2020 cohort emerging as the most influential (21,730 citations, CM-TANCS: 25.24). In contrast, 2023 and 2024 papers display lower CM-TANCS values (1.04 and 0.03), reflecting limited exposure time. Papers from 2021 and 2022 show moderate citation impact (CM-TANCS: 6.79 and 2.64), underscoring the influence of research visibility, data accessibility, and disciplinary engagement in shaping citation trajectories.

5.8 Citation Impact of Open Access vs. Paywalled Data Papers

Open Access (OA) publishing enhances scientific visibility and knowledge dissemination, yet its impact on citation rates for data papers remains underexplored. A scatter plot (Fig. 5) assesses the relationship between OA status and citation impact, with color-coded bubbles indicating citation ranges. Pearson's correlation coefficient (r = 0.052, p = 0.00096) reveals a weak but statistically significant positive relationship, suggesting a slight citation advantage for OA data papers.

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sqrt{(Y - \bar{Y})^2}}}$$

Where X denotes the Open Access status, Y represents the citation count (from the "Cited by" field), $\bar{X}$ is the mean value of X, and $\bar{Y}$ is the mean value of Y.

Figure 5 shows that highly cited papers (blue and purple bubbles) cluster under OA, while paywalled papers tend to have lower citation counts. This finding supports the Open Access Citation Advantage (OACA) hypothesis, where OA articles attract ~18 % more citations[53-54], aligning with previous studies by Tennant[55], *et al.* and Archambault[56], *et al*. However, the presence of low-citation papers in both OA and paywalled categories, coupled with the near-flat trend line, indicates that factors beyond accessibility-such as journal prestige, research domain, and dataset utility-primarily drive citation counts.

Data papers function more as infrastructure for data reuse rather than direct citation sources[57]. These findings underscore the multifaceted nature of citation dynamics and suggest that while OA enhances visibility, its direct influence on scholarly impact remains modest. To advance open science, policies should promote OA for accessibility while incorporating complementary metrics (e.g., Altmetrics and policy impact) to capture broader societal engagement.

5.9 Global Funding Landscape for Data Papers in Social Sciences

Table 6 shows that 159 funding agencies have supported 3,333 papers across 7,205 instances. China leads this space, with the National Natural Science Foundation of China (NSFC) funding 620 papers (10,046 citations, h-index: 44), followed by the Ministry of Science and Technology (529 papers, 7,671 citations). The European Commission also demonstrates high impact (392 papers, mean citation: 18.12, h-index: 40). In the U.S., the NIH and NSF each supported 266 papers, contributing significantly to open data. So the global funding landscape for data papers reflects a strong commitment to open science and advancing data-driven research practices[58]. Despite strong investment from the Global North, funding in the Global South remains limited, revealing inequities in infrastructure and access. Greater collaboration and inclusive funding mechanisms are essential to advancing equitable participation in data-driven open science research[59].

Table 5. Citation impact of data papers in social sciences (2019-2024)

Pub year	DP	2019	2020	2021	2022	2023	2024	Citations	TANCS	CM-TANCS
1	280	151	1285	2282	2601	2677	2912	11908	7.09	12.44
2	393	***	650	3502	5131	5843	6604	21730	11.06	25.24
3	304		***	410	1796	2561	3032	7799	6.41	6.79
4	769			***	648	3712	6302	10662	4.62	2.64
5	883				***	1025	5239	6264	3.55	1.04
6	1326					***	1217	1217	0.92	0.03

TANCS= Time-Adjusted Normalised Citation Score, CM-TANCS= Citation Mean-Adjusted TANCS

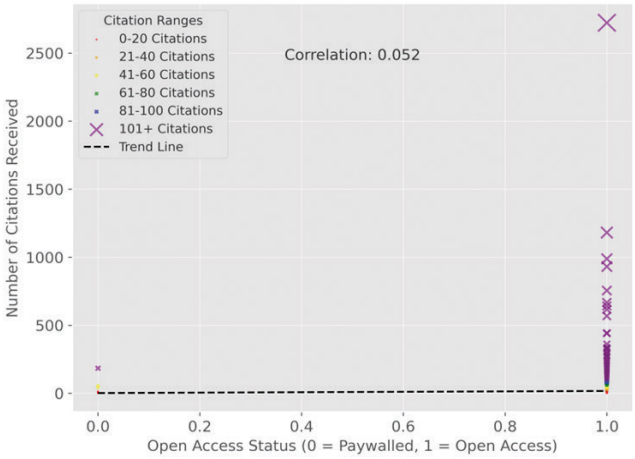


Figure 5. Citation impact of open access vs. paywalled data papers.

5.10 Citation Performance and Societal Impact of High-Impact Data Papers

The ten most cited data papers (Table 7) demonstrate strong scholarly impact, all being Gold Open Access and funded by agencies. Nine were published in Scientific Data (Springer Nature), with six from 2020, three from 2019, and one from 2023, highlighting rapid uptake. Harris, *et al.* leads with 3,582 Google Scholar citations, followed by Gygli, *et al.* with 2,073. Beyond citations, these papers exhibit high societal engagement, with notable Altmetric Attention Scores reflecting influence across policy, media, and social platforms.

5.11 Correlation Between Citations and Altmetrics

Figure 6A shows a moderate positive correlation (r=0.62) between citation counts and Altmetric Attention Scores for the top 100 cited data papers, suggesting that highly cited papers often achieve broader societal visibility. However, discrepancies arise, as some moderately cited papers (<500 citations) attain high Altmetric Scores (>200), while highly cited papers (>2000 citations) exhibit lower altmetric engagement. This highlights that altmetrics reflect public discourse and immediacy rather than long-term scholarly influence[60-64]. Thelwall[61], *et al.* previously noted that while social media and policy sources contribute to visibility, they do not necessarily drive citation accumulation, reinforcing the divergence between public engagement and scholarly uptake.

Figures 6B and 6C reveal that 100 % of top-cited papers were shared on X (Twitter), 36 % were cited in policy documents, but only 9 % appeared in patents, indicating limited technological translation. Altmetric engagement remains skewed toward dominant platforms, amplifying Global North visibility.

5.12 Data Sharing Mandates and Policies of Journals

This section examines how journal policies influence data-sharing practices, particularly adherence to FAIR and CARE principles65-69. Fig. 7 highlights inconsistencies in data-sharing mandates and enforcement mechanisms across journals. While 57.5 % of journals outline data-sharing policies, only 42.5 % enforce FAIR compliance, and just 35 % address CARE principles, raising concerns about data interoperability and ethical standards.

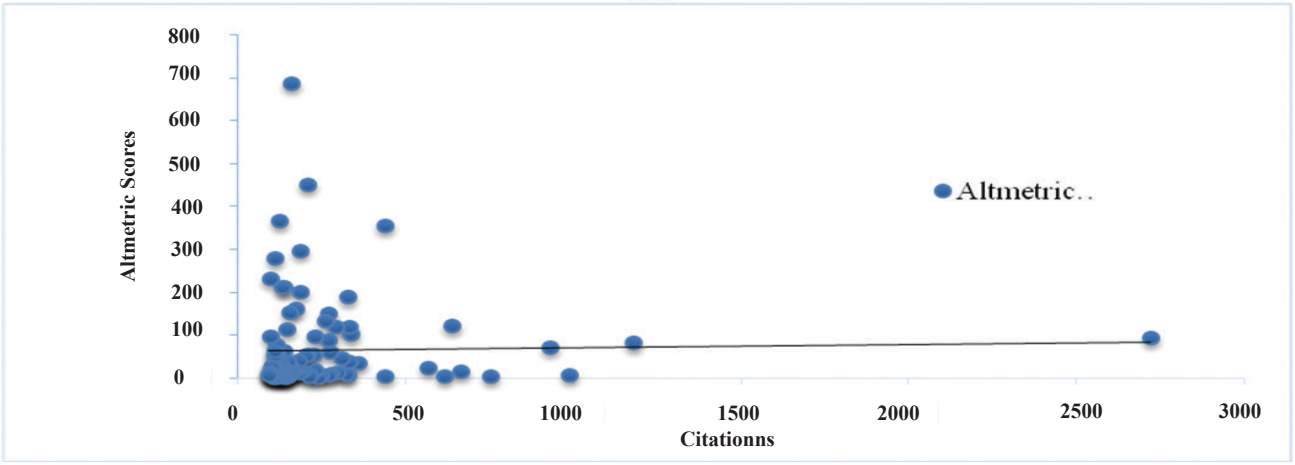
Moreover, 32.5 % mandate data sharing, while 67.5 % adopt weak or optional guidelines. Only 40 % require repository storage, and 47.5 % enforce data citation practices. Compliance checks are rare, with only 35 % implementing verification mechanisms. Although 67.5 % promote cross-disciplinary data sharing and 62.5 % operate under an open-access model, gaps in policy standardisation and enforcement persist. Strengthening regulatory frameworks is essential to ensure transparency, equity, and responsible data stewardship.

Table 6. Top 10 funding agencies

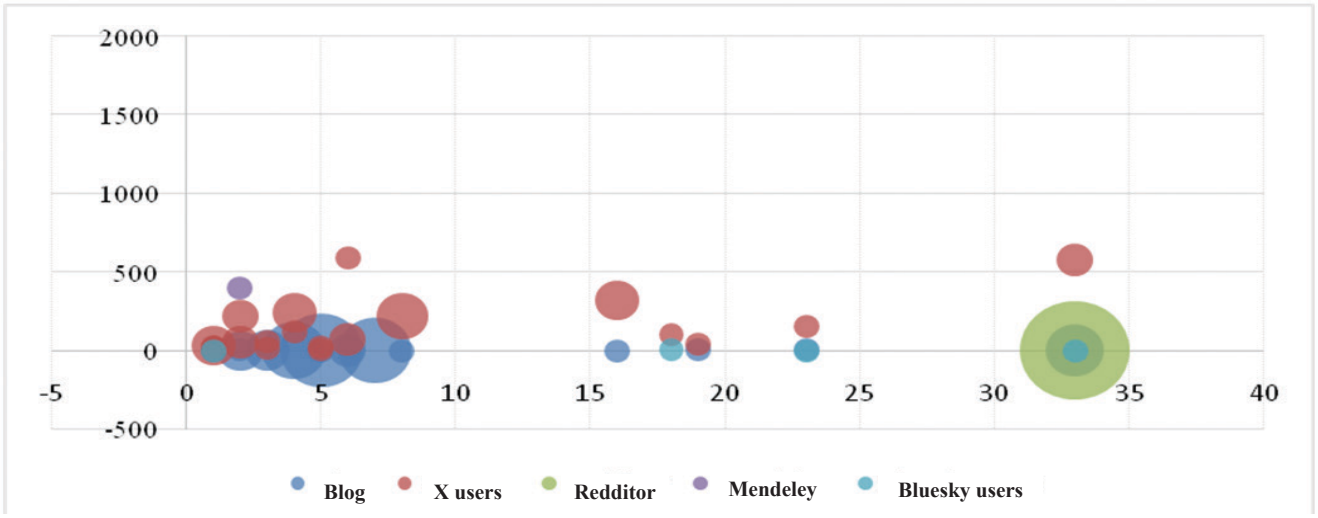
Rank	Sponsors	DP	Citations	MC	*h*-index
1	National Natural Science Foundation of China (NSFC)	620	10046	16.20	44
2	Ministry of Science and Technology of the People's Republic of China (MOST P.R. China)	529	7671	14.50	39
3	European Commission (EC)	392	7102	18.12	40
4	National Key Research and Development Program of China (NKRDPC)	269	5132	19.08	30
5	National Institutes of Health (NIH)	266	7422	27.90	37
6	National Science Foundation (NSF)	266	4872	18.32	34
7	Horizon 2020 Framework Programme (H2020)	235	8427	35.86	36
8	U.S. Department of Health and Human Services (HHS)	218	6055	27.78	35
9	UK Research and Innovation (UKRI)	164	5183	31.60	32
10	U.S. Department of Energy(DOE)	147	6161	41.91	30

Rest 149 sponsoring agencies supported 4049 times with a total of 7205 instances

6A



6B



6C

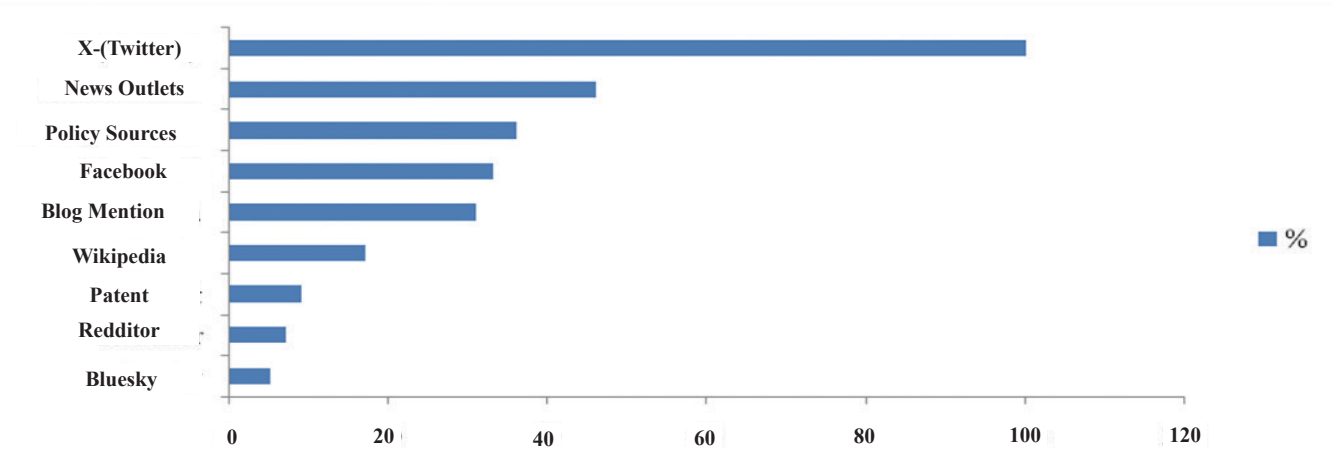


Figure 6. Correlation between citations and altmetric attention scores for the top 100 cited papers: (A) Traditional citations vs. altmetric attention. (B) Distribution of online sources contributing to altmetric attention. (C)Breakdown of altmetric contributions across different platforms.

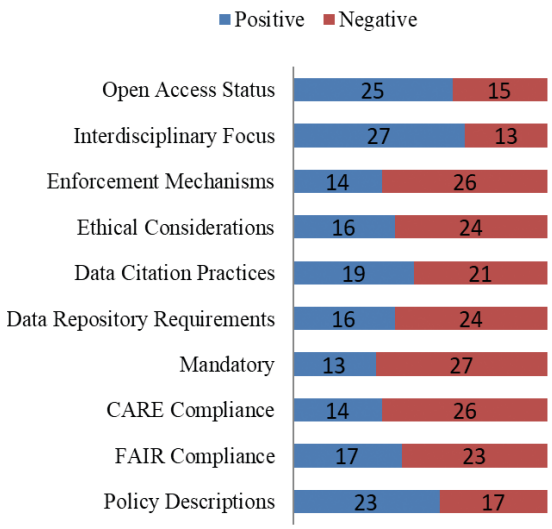


Figure 7. Data-sharing policies and compliance among journals.

6. DISCUSSION

This study highlights the growing role of data papers in social science research, with a 36 % annual growth rate and 83 % published after 2021, reflecting the shift toward transparency and reproducibility[70,71]. This growth reflects evolving academic policies, raising awareness of data-sharing benefits, and the establishment of data journals[72,73]. China dominates data paper growth (50.4 %), driven by national open science policies, while the U.S. maintains the highest publication volume despite slower growth (16.4 %). Notably, 20 % of data journals account for 98.31 % of all data papers, underscoring the concentration of data paper dissemination.

6.1 Collaboration and Geographic Disparities

Collaboration networks reveal entrenched geopolitical hierarchies, with Global North (GN) nations contributing 78 % of collaborative strength and maintaining citation dominance (GN-GN: 28.6 vs. GS-GS: 12.4) reflecting Collyer's[75] critique of epistemic hegemony. Asymmetric GN-GS partnerships (e.g., US-China) exemplify Becker's[76] concept of interdependence, where Global South (GS) nations act as data providers rather than equal partners[77,78]. Sparse South-South collaborations (4 % of network strength) reflect systemic inequities and support Ishengoma's[79] assertions that underfunding and colonial legacies marginalize Southern agency, calling for equitable funding, GS-centric infrastructures, and policies to dismantle extractive dynamics[80].

6.2 Citation Impact and Open Access Advantage

Although data papers contribute significantly to research, they generally receive fewer citations than traditional articles, reflecting a "citation paradox." The Pareto principle is evident, with 20 % of data papers generating 80 % of total citations. OA papers exhibit a weak but statistically significant citation advantage (r = 0.052, p = 0.00096), consistent with Piwowar[81], *et al.* findings on the Open Access Citation Advantage. However, factors such as journal prestige and research domain primarily shape citation patterns, reinforcing Mooney and Newton's[82] observations on selective advantages for high-impact datasets.

6.3 Funding Disparities and Global Imbalances

Funding agencies play a crucial role in data paper production, with 159 agencies supporting 3,333 papers. China's NSFC leads globally, followed by the European Commission and NIH. This distribution reflects Fecher and Friesike's[83] observations on institutional commitment to open science. Despite these investments, funding disparities persist in the GS, where limited resources hinder data-sharing initiatives[88]. Collaborative funding models and capacity-building efforts are essential to bridge these gaps and promote equitable participation in open science.

6.4 Journal Policies and Ethical Considerations

While Scientific Data leads in enforcing robust data-sharing requirements, inconsistencies in adherence to FAIR and CARE principles persist[84]. Only 42.5 % of journals mandate FAIR compliance, and 35 % incorporate CARE principles, highlighting weak enforcement mechanisms. Addressing these gaps requires methodological fairness[85], standardised policies, researcher incentives, and ethical training[55,86].

6.5 Alternative Metrics and Societal Impact

Given their role in supporting research infrastructure rather than advancing theoretical frameworks, traditional citation metrics may not fully capture the impact of data papers. Altmetric Attention Scores, policy mentions, and media engagement offer a more comprehensive evaluation of societal influence[87]. Highly cited papers often demonstrate strong societal engagement, reinforcing the need to integrate alternative metrics into research evaluation frameworks. Highly cited papers often demonstrate strong societal engagement, reinforcing the need to integrate alternative metrics into research evaluation frameworks.

Data papers are emerging as critical infrastructure for reproducible research, with notable growth in publication, collaboration, and policy implementation. However, persistent disparities in funding, collaboration, and ethical considerations necessitate stronger frameworks for equitable data sharing and reuse. Strengthening regulatory policies and fostering inclusive funding mechanisms can ensure that data papers continue to advance transparency, reproducibility, and societal impact across disciplines.

Table 7. Top 10 frequently cited data papers and their altmetric scores

S. No.	Data paper	Times cited				Alt-metrics
		GS	Scopus	WoS	Dim.	
1	Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Scientific data, 7(1), 109. https://doi.org/10.1038/s41597-020-0453-3	3582	2723	2799	3074	94
2	Gygli, S., Haelg, F., Potrafke, N., & Sturm, J. E. (2019). The KOF globalisation index–revisited. The Review of International Organizations, 14, 543-574. https://doi.org/10.1007/s11558-019-09344-2	2073	1181	1124	1511	82
3	He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., & Li, X. (2020). The first high-resolution meteorological forcing dataset for land process studies over China. Scientific data, 7(1), 25. https://doi.org/10.1038/s41597-020-0369-y	1045	988	1017	971	7
4	Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., ... & Law, B. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Scientific data, 7(1), 225. https://doi.org/10.1038/s41597-020-0534-3	1093	931	978	1065	72
5	Shan, Y., Huang, Q., Guan, D., & Hubacek, K. (2020). China CO2 emission accounts 2016–2017. Scientific data, 7(1), 54. https://doi.org/10.1038/s41597-020-0393-y	840	755	752	769	5
6	Johnson, A. E., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C. Y., ... & Horng, S. (2019). MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data, 6(1), 317. https://doi.org/10.1038/s41597-019-0322-0	1229	667	562	935	16
7	Johnson, A. E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., ... & Mark, R. G. (2023). MIMIC-IV, a freely accessible electronic health record dataset. Scientific data, 10(1), 1. https://doi.org/10.1038/s41597-022-01899-x	1131	642	600	848	123
8	Chen, J., Gao, M., Cheng, S., Hou, W., Song, M., Liu, X., ... & Shan, Y. (2020). County-level CO2 emissions and sequestration in China during 1997–2017. Scientific data, 7(1), 391. https://doi.org/10.1038/s41597-020-00736-3	703	616	641	605	4
9	Wagner, P., Strodthoff, N., Bousseljot, R. D., Kreiseler, D., Lunze, F. I., Samek, W., &Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. Scientific data, 7(1), 1-15. https://doi.org/10.1038/s41597-020-0495-6	958	568	431	627	24
10	Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., ... & Reichstein, M. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. Scientific data, 6(1), 74. https://doi.org/10.1038/s41597-019-0076-8	541	442	457	527	5

GS= Google Scholar, WoS= Web of Science, Dim= Dimensions

7. CONCLUSIONS

This study provides a systematic evaluation of data papers in social sciences, highlighting their growing prominence, citation impact, and role in advancing open science and research transparency. Despite this progress, challenges persist, including funding inequities, ethical concerns, and inconsistent data-sharing mandates. Adoption of data papers in social sciences remains slower than in STEM disciplines due to concerns around data privacy, informed consent, and anonymisation[55]. Weak compliance mechanisms further hinder adherence to FAIR and CARE principles, reinforcing call for methodological fairness[85].

The study identifies a weak yet statistically significant correlation between open access and citation impact, indicating that although increased accessibility enhances research visibility, citation influence is predominantly shaped by factors such as journal prestige and the specific research domain. Addressing funding disparities, particularly between the Global North and South, and strengthening regulatory frameworks are essential for

fostering equitable data-sharing practices. Only 30 % of journals implement robust enforcement mechanisms, underscoring the need for standardised policies that balance ethical considerations with open data accessibility.

Future research should investigate disciplinary variations in data-sharing behaviours, assess the effectiveness of different policy frameworks, and develop alternative impact metrics to capture the diverse contributions of data papers. Comparative analyses between social sciences and STEM fields can identify cross-disciplinary differences, while long-term evaluations of open data initiatives can inform strategies to maximize scientific and societal benefits. Building technological infrastructure and capacity in underrepresented regions will ensure inclusive participation in open science practices.

DATA AVAILABILITY

The supplementary data[89] supporting this study are openly accessible via Mendeley at https://doi.org/10.17632/bw4khm8zjd.1 and are available for free use, sharing, and reuse in accordance with open data principles.

REFERENCES

1. Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz J, *et al.* Manifesto of computational social science. Eur Phys J Spec Top. 2012;214:325-46. doi: 10.1140/epjst/e2012-01697-8
2. Cukier K, Mayer-Schoenberger V. The rise of big data. Foreign Aff. 2013;92(3):28-40.
3. Denzin NK, Lincoln YS, editors. The Sage handbook of qualitative research. Thousand Oaks (CA): Sage; 2011.
4. Salganik MJ. Bit by bit: Social research in the digital age. Princeton (NJ): Princeton University Press; 2019.
5. Fidler F, Wilcox J. Reproducibility of scientific results. In: Zalta EN, editor. The Stanford encyclopedia of philosophy. Stanford (CA): Metaphysics research lab, stanford university; 2018. Available from: https://plato.stanford.edu/archives/win2018/entries/scientific-reproducibility/
6. Hardwicke TE, Wallach JD, Kidwell MC, Bendixen T, Crüwell S, Ioannidis JPA. An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014-2017). R Soc Open Sci. 2020;7(2):190806. doi: 10.1098/rsos.190806
7. Bausell RB. The problem with science: The reproducibility crisis and what to do about it. Oxford: Oxford university press; 2021.
8. Moody JW, Keister LA, Ramos MC. Reproducibility in the social sciences. Annu Rev Sociol. 2022;48(1):65-85.
9. Freese J, Rauf T, Voelkel JG. Advances in transparency and reproducibility in the social sciences. Soc Sci Res. 2022;107:102770. Available from: doi: 10.1016/j.ssresearch.2022.102770
10. Kowalczyk S, Shankar K. Data sharing in the sciences. Annu Rev Inf Sci Technol. 2011;45(1):247-94. doi: 10.1002/aris.2011.1440450113
11. Fecher B, Friesike S, Hebing M. What drives academic data sharing? PLoS One. 2015;10(2):e0118053. doi: 10.1371/journal.pone.0118053
12. Candela L, Castelli D, Manghi P, Tani A. Data journals: A survey. J Assoc Inf Sci Technol. 2015;66(9):1747-62. doi: 10.1002/asi.23358
13. McGillivray B, Marongiu P, Pedrazzini N, Ribary M, Wigdorowitz M, Zordan E. Deep impact: A study on the impact of data papers and datasets in the humanities and social sciences. Publications. 2022;10(4):39. doi: 10.3390/publications10040039
14. Chavan V, Penev L. The data paper: A mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics. 2011;12(Suppl 15):S2. doi: 10.1186/1471-2105-12-S15-S2
15. Gherghina S, Katsanidou A. Data availability in political science journals. Eur Polit Sci. 2013;12(3):333-49. doi: 10.1057/eps.2013.8
16. Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet. 2023;401(10393):e21-33.
17. Herndon T, Ash M, Pollin R. Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. Camb J Econ. 2014;38(2):257–79. doi: 10.1093/cje/bet075
18. Craig R, Cox A, Tourish D, Thorpe A. Using retracted journal articles in psychology to understand research misconduct in the social sciences: What is to be done? Res Policy. 2020;49(4):103930. doi: 10.1016/j.respol.2020.103930
19. Basumatary B, Verma MK. Unearthing the misconduct and inaccuracies in social science research: a trend analysis of retracted articles based on retraction watch database. Glob Knowl Mem Commun. 2024. doi: 10.1108/GKMC-04-2024-0228
20. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, *et al.* Promoting an open research culture. Science. 2015;348(6242):1422-5. doi: 10.1126/science.aab2374
21. Borgman CL. Big data, little data, no data: Scholarship in the networked world. Cambridge (MA): MIT Press; 2017.
22. Open science collaboration. Estimating the reproducibility of psychological science. Science. 2015;349(6251):aac4716.
23. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, *et al.* The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3(1):160018. doi: 10.1038/sdata.2016.18
24. Kim Y, Adler M. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. Int J Inf Manag. 2015;35(4):408–18. doi: 10.1016/j.ijinfomgt.2015.04.007
25. Leonelli S, Lovell R, Wheeler BW, Fleming L,

Williams H. From FAIR data to fair data use: Methodological data fairness in health-related social media research. Big Data Soc. 2021;8(1). doi: 10.1177/20539517211010310
26. Pasquetto IV, Randles BM, Borgman CL. On the reuse of scientific data. Data Sci J. 2017;16(8):1-9. doi: 10.5334/dsj-2017-008
27. Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. Peer J. 2013;1:e175. doi: 10.7717/peerj.175
28. Hey T, Trefethen A. The data deluge: An e-science perspective. In: Berman F, Fox G, Hey T, editors. Grid computing: Making the global infrastructure a reality. Chichester: John Wiley & Sons. 2003;809-24. doi: 10.1002/0470867167.ch36
29. Hey T, Tansley S, Tolle K, editors. The fourth paradigm: Data-intensive scientific discovery. Vol. 1. Redmond (WA): Microsoft Research; 2009. https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf
30. Vicente-Saez R, Martinez-Fuentes C. Open Science now: A systematic literature review for an integrated definition. J Bus Res. 2018;88:428-36. doi: 10.1016/j.jbusres.2017.12.043
31. Mooney H. Citing data sources in the social sciences: do authors do it? Learn Publ. 2011;24(2):99-108. doi: 10.1087/20110204
32. Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz J. Research data explored: an extended analysis of citations and altmetrics. Scientometrics. 2016;107:723-44. doi: 10.1007/s11192-016-1887-4
33. Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. J Libr Schol Commun. 2012;1(1):eP1035. doi: 10.7710/2162-3309.1035
34. Park H, Wolfram D. An examination of research data sharing and re-use: Implications for data citation practice. Scientometrics. 2017;111:443-61. doi: 10.1007/s11192-017-2240-2
35. Jacoby J. Share and share alike? Data-sharing practices in different disciplinary domains. Soc Sci Libr Interdiscip Collect Serv Netw. 2010;144:79-94.
36. Munafò MR, Nosek BA, Bishop DV, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. Nat Hum Behav. 2017;1(1):1-9. doi: 10.1038/s41562-016-0021
37. Gleditsch NP, Metelits C, Strand H. Posting your data: Will you be scooped or will you be famous. Int Stud Perspect. 2003;4(1):89-97.
38. Zhang L, Ma L. Does open data boost journal impact: evidence from Chinese economics. Scientometrics. 2021;126(4):3393-419. doi: 10.1007/s11192-021-03897-z
39. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PLoS One. 2011;6(6):e21101. doi: 10.1371/journal.pone.0021101
40. Meyer MN. Practical tips for ethical data sharing. Adv Methods Pract Psychol Sci. 2018;1(1):131-44. doi: 10.1177/25152459177476
41. Vanpaemel W, Vermorgen M, Deriemaecker L, Storms G. Are we wasting a good crisis? The availability of psychological research data after the storm. Collabra. 2015;1(1):1–5. doi: 10.1525/collabra.13
42. Chung E. An investigation on scientific data for data journal and data paper. J Korean Soc Inf Manag. 2019;36(1):117-35.
43. Chung E. An investigation of intellectual structure on data papers published in data journals in Web of Science. J Korean Soc Inf Manag. 2020;37(1):153-77.
44. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372. doi: 10.1136/bmj.n71
45. Martín-Martín A, Orduna-Malea E, Thelwall M, Delgado-López-Cózar E. Google Scholar, Web of Science, and Scopus: which is best for me? Impact Soc Sci Blog. 2019. https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/03/google-scholar-web-of-science-and-scopus-which-is-best-for-me/
46. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics. 2016;106:213-28. doi: 10.1007/s11192-015-1765-5
47. Aksnes DW, Sivertsen G. A criteria-based assessment of the coverage of Scopus and Web of Science. J Data Inf Sci. 2019;4(1):1-21. Available from: doi: 10.2478/jdis-2019-0001
48. Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics. 2021;126:5113-42. doi: 10.1007/s11192-021-03948-5
49. Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84(2):523-38. doi: 10.1007/s11192-009-0146-3
50. Aria M, Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetr. 2017;11(4):959–75. doi: 10.1016/j.joi.2017.08.007
51. Haug S, Waisbich LT. Comprehensive power shifts in the making: China's policy transfer partnerships with the United Nations. Glob Policy. 2024;15:62-73. doi: 10.1111/1758-5899.13360
52. Zhang X, Jing Y. A mixed funding pattern: China's exercise of power within the United Nations. Glob policy. 2024;15:121-34. doi: 10.1111/1758-5899.13376
53. Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, *et al.*,The state of OA: A large-scale analysis of the prevalence and impact of open access articles. PeerJ. 2018;6:e4375.

doi: 10.7717/peerj.4375
54. Fu J, Tian L, Zhang C, Li J. Opening research data contributes to the citations of related research articles: Evidence from data in brief. Learn Publ. 2023;36(3):426-38. doi: 10.1002/leap.1551
55. Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CH. The academic, economic and societal impacts of open access: An evidence-based review. F1000Res. 2016;5:632. doi: 10.12688/f1000research.8460.3
56. Archambault É, Côté G, Struck B, Voorons M. Research impact of paywalled versus open access papers. Copyr Fair Use Schol Commun. 2016;29:1-5. http://digitalcommons.unl.edu/scholcom/29
57. McGillivray B, Marongiu P, Pedrazzini N, Ribary M, Wigdorowitz M, Zordan E. Deep impact: A study on the impact of data papers and datasets in the humanities and social sciences. Publications. 2022;10(4):39. doi: 10.3390/publications10040039
58. Fecher B, Friesike S. Opening science. In: Bartling S, Friesike S, editors. Open science: One term, five schools of thought. Cham: Springer Int Publ; 2014; 17–47. doi: 10.1007/978-3-319-00026-8_2
59. Turba R, Thoré ESJ, Bertram MG, Bridg H, Sabet SS, Gamboa M, *et al.* Global north-south science inequalities due to language and funding barriers. Zenodo. 2015. doi: 10.5281/zenodo.14902147
60. Priem J, Taraborelli D, Groth P, Neylon C. Altmetrics: A manifesto. Copyr Fair Use Schol Commun. 2015;185. https://digitalcommons.unl.edu/scholcom/185
61. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do altmetrics work? Twitter and ten other social web services. PLoS One. 2013;8(5):e64841. doi: 10.1371/journal.pone.0064841
62. Bornmann L. Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. J Informetr. 2014;8(4):895-903. doi: 10.1016/j.joi.2014.09.005
63. Haustein S, Bowman TD, Costas R. Interpreting "altmetrics": viewing acts on social media through the lens of citation and social theories. In: Sugimoto CR, editor. Theories of informetrics and scholarly communication. Berlin: De Gruyter. 2016;372-406.
64. Sugimoto CR, Work S, Larivière V, Haustein S. Scholarly use of social media and altmetrics: A review of the literature. J Assoc Inf Sci Technol. 2017;68(9):2037–62. doi: 10.1002/asi.23833
65. Piwowar H, Chapman W. A review of journal policies for sharing research data. Nat Prec. 2008;1:1. doi: 10.1038/npre.2008.1700.1
66. Sturges P, Bamkin M, Anders JH, Hubbard B, Hussain A, Heeley M. Research data sharing: Developing a stakeholder-driven model for journal policies. J Assoc Inf Sci Technol. 2015;66(12):2445-55. doi: 10.1002/asi.23336
67. Herndon J, O'Reilly R. Data sharing policies in social sciences academic journals: Evolving expectations of data sharing as a form of scholarly communication. In: Kellam L, Thompson K, editors. Data brarianship. Chicago: American library association. 2016;219-43. doi: 10.15139/S3/12157
68. Vasilevsky NA, Minnier J, Haendel MA, Champieux RE. Reproducible and reusable research: Are journal data sharing policies meeting the mark? PeerJ. 2017;5:e3208. doi: 10.7717/peerj.3208
69. Huh S. Recent trends in medical journals' data sharing policies and statements of data availability. Arch Plast Surg. 2019;46(6):493-7. doi: 10.5999/aps.2019.01515
70. Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz J, *et al.* Manifesto of computational social science. Eur Phys J Spec Top. 2012;214:325-46. doi: 10.1140/epjst/e2012-01697-8
71. Hardwicke TE, Wallach JD, Kidwell MC, Bendixen T, Crüwell S, Ioannidis JPA. An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014-2017). R Soc Open Sci. 2020;7(2):190806. doi: 10.1098/rsos.190806
72. Michener WK. Ecological data sharing. Ecol Inform. 2015;29:33-44. doi: 10.1016/j.ecoinf.2015.06.010
73. Chawinga WD, Zinn S. Global perspectives of research data sharing: A systematic literature review. Libr Inf Sci Res. 2019;41(2):109-22. doi: 10.1016/j.lisr.2019.04.004
74. Zhang L, Downs RR, Li J, Wen L, Li C. A review of open research data policies and practices in China. Data Sci J. 2021;20(3):1-17. doi: 10.5334/dsj-2021-003
75. Collyer FM. Global patterns in the publishing of academic knowledge: Global north, global south. Curr Sociol. 2018;66(1):56-73. doi: 10.1177/0011392116680020
76. Becker B. International inequality and demand for redistribution in the global south. Polit Sci Res Methods. 2024;12(2):407-15. doi: 10.1017/psrm.2023.36
77. Chandiwana S, Ornbjerg N. Review of North–South and South–South cooperation and conditions necessary to sustain research capability in developing countries. J Health Popul Nutr. 2003;21(3):288-97. https://www.jstor.org/stable/23499226
78. Binka F. North-south research collaborations: A move towards a true partnership? Trop Med Int Health. 2005;10(3):207-9. doi: 10.1111/j.1365-3156.2004.01373.x
79. Ishengoma JM. North–South research collaborations and their impact on capacity building: A Southern perspective. In: Halvorsen T, Nossum J, editors. North–South knowledge networks: Towards equitable collaboration between academics, donors and universities.

Cape Town: African Minds.2016;149-86.
80. Boampong MS, Boakye-Danquah J, Boafo YA, Appiagyei K, Tweneboah-Koduah D, Antwi EK. Experiences of inequality in international collaborative research–Perspectives from environmental and sustainability scholars from Ghana, West Africa. Environ Sci Policy. 2024;152:103661. doi: 10.1016/j.envsci.2023.103661
81. Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, et al. The state of OA: A large-scale analysis of the prevalence and impact of open access articles. PeerJ. 2018;6:e4375. doi: 10.7717/peerj.4375
82. Mooney H, Newton MP. The anatomy of a data citation: Discovery, reuse, and credit. J Libr Schol Commun. 2012;1(1):eP1035. doi: 10.7710/2162-3309.1035
83. Fecher B, Friesike S. Open science: One term, five schools of thought. In: Bartling S, Friesike S, editors. Opening science. Cham: Springer. 2014;17-47. doi: 10.1007/978-3-319-00026-8_2
84. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, *et al.* The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3(1):1-9. doi: 10.1038/sdata.2016.18
85. Leonelli S, Lovell R, Wheeler BW, Fleming L, Williams H. From FAIR data to fair data use: Methodological data fairness in health-related social media research. Big Data Soc. 2021;8(1):1-14. doi: 10.1177/20539517211010310
86. Resnik DB, Morales M, Landrum R, Shi M, Minnier J, Vasilevsky NA, *et al.* Effect of impact factor and discipline on journal data sharing policies. Account Res. 2019;26(3):139-56. doi: 10.1080/08989621.2019.1591277
87. Kassab O, Bornmann L, Haunschild R. Can altmetrics reflect societal impact considerations? Exploring the potential of altmetrics in the context of a sustainability science research center. Quant Sci Stud. 2020;1(2):792-809. doi: 10.1162/qss_a_00032
88. Kar S. Designing a framework of open-data repository for social science research in India [PhD thesis]. Midnapore: Vidyasagar University; 2025.
89. Kar S, Rath DS. Open data in social sciences. Mendeley Data. 2025; V1. doi: 10.17632/bw4khm8zjd.1

CONTRIBUTORS

Mr. Sanjoy Kar is an academician at the Institute of Development Studies Kolkata, India. With over 10 years of experience leading modern information services, he is actively engaged in research, teaching, and professional development. He is also a Visiting Fellow at the University of California, Los Angeles, USA. His research interests Span meta science, Responsible research assessment, Open science, and Research data curation & management. The current paper is the outcome of his doctoral research. His contributions to this study include conceptualisation, data collection, analysis, visualisation, and writing the original draft.

Prof. Durga Sankar Rath is a prolific researcher and distinguished faculty member in Information Science at Vidyasagar University, West Bengal, India. He is a proud alumnus of the Indian Statistical Institute (ISI), Bangalore, and Vidyasagar University. With 25 years of teaching and research experience, he has mentored numerous Ph.D. scholars across India. His research interests include: Artificial intelligence (AI), ICT applications, Research evaluation metrics, and Knowledge organisation.
His contribution to this study includes conceptualisation, supervision, review and editing of the final draft.