

Who funds what: An assessment of research funding networks in data papers

Yurdagül Ünal 

Department of Information Management, Hacettepe University, Ankara, Türkiye

Müge Akbulut 

Department of Information Management, Ankara Yıldırım Beyazıt University, Ankara, Türkiye

Information Development
1–20

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/026666669251352185

journals.sagepub.com/home/idv



Abstract

This study examines the role of funding collaborations in shaping the production and dissemination of scientific information through data papers, a rapidly growing academic publication format. To the best of our knowledge, there are no studies investigating, and evaluating the data paper-funder relationship. The goal of this study was, therefore, to evaluate data papers and funder information in detail, extracted from the data papers themselves, in order to reveal the collaborative characteristics of funders, and to provide guidance to researchers and funding agencies. Data papers published between 2006–2017 were downloaded from the Web of Science database. The same papers were retrieved from Dimension, which offered more detailed category classifications. These classifications were then utilized for further analysis based on categories. The names of funders were standardized by matching them using the Crossref funder registry, and associated funding metadata. A statistical, and social network analysis were performed. The top funding country was the USA; the top funding institution was the *U.S. Department of Health and Human Services, National Institutes of Health*. The collaboration network among funders exhibited relatively low density. A collaboration network of 1197 links between 69 countries was created. The USA had connections with 62 countries. Our study is important because it standardizes the funding data for data papers by associating them with Crossref funding metadata. The widespread increase of data papers, and their relatively dispersed funding among a variety of funders points to the need for research evaluating collaborations between funders, as important both for the funded researchers, and for understanding and optimizing the shortcomings of current funding management.

Keywords

data papers, funders, social network analysis, betweenness centrality, funded research, funding networks, funded collaboration

Submitted: 21 March 2025; accepted: 8 June 2025

Introduction

Data has become a strategic asset in today's scientific world. Open science, and data-oriented, as well as data-intensive studies are shaping scientific research, and scientific communication. Recent developments in the scientific world have also brought about new academic publication types, such as *data journals* and *data papers*, the number of both data papers and journals continue to grow. From the open science perspective, data papers have an important role to play in increasing the visibility and reuse of research data.

However, while the proliferation and benefits of open data sharing are undeniable, significant challenges also threaten this trend. Specifically, current geopolitical tensions are diminishing the willingness of some countries or regions to share their data openly. The increasing importance of data as a

Corresponding author:

Yurdagül Ünal, Department of Information Management, Hacettepe University, 06800 Beytepe Ankara, Türkiye.

Email: yurdagul@hacettepe.edu.tr

strategic asset, combined with international competition and security concerns, can lead to protectionist approaches that may restrict the free flow of scientific data (Ghermandi et al., 2023). This situation holds the potential to endanger global scientific collaboration, and the fundamental principles of open science.

Regarding traditional research articles, there are various studies which evaluate the research impact of funded studies from multiple perspectives; investigate the relationship between funding, and scientific productivity; and analyze funded studies by country, subject, and funder. Studies on data papers generally compare data papers with traditional research articles. Some studies focus on research impact, the number of citations the articles receive, and why they are cited. Additionally, there are studies which examine subject distributions, and investigate whether data papers increase the visibility or impact of other research articles with which they are related. Although there are many studies in the literature linking the research impact of articles with their funding, there are no studies investigating, and analyzing this relationship from the perspective of the data papers themselves.

The focused goal of this study is to examine the funding of scientific collaborations within the specific context of data papers. Although it is assumed that all funding providers are correctly listed in the Web of Science database, incorrect entries may exist. This distinction will be made clearer in the manuscript of our paper. Our study focuses on data papers, because they directly present data, and project-oriented studies. For this reason, the evaluation of funding information extracted from the data papers themselves is critically important.

Data papers “main content is a description of published research datasets, along with contextual information about the production and the acquisition of the datasets” (Schöpfel et al., 2019: 635–636). Although they go through publication processes similar to traditional research articles, the aim of data papers is to focus on the creation, management, processing, access, and use of datasets, rather than the conceptual background, methodology, or presentation of findings (McGillivray et al., 2022: 2). There are pure data journals, which publish only data papers, as well as “mixed” journals, which publish data papers along with traditional articles. A data journal is defined as a “journal that advertised itself as primarily publishing articles about data in some form, whether or not that data was usually associated with other papers, or saved in a separate dataset” (Thelwall, 2020: 699).

Research funding is one of the most important factors that enable researchers to realize their goals (Rusu et al., 2022), therefore it is important for researchers to be able to identify the most appropriate funders for their projects. Thus, funding can have a significant impact on the individual productivity of researchers, and the trajectory of their careers (Ayoubi et al., 2019). It is equally important for funding institutions and organizations to ensure diversity and fairness in the distribution of funds, and to use resources efficiently. Given that the funding process has come under increasing scrutiny, and, in order to ensure the success of research investments (Alvarez-Bornstein and Bordons, 2021), data papers are also an important instrument for measuring the research effectiveness of funded projects.

Funding, increasing research effectiveness, and data sharing are related issues. It is necessary to investigate whether the funding support provided has achieved its purpose, and whether it has had an impact on the scientific output produced. Understanding the relationship between funding and its research impact will help researchers, funding agencies, and institutions make informed decisions about resource allocation, and contribute to the advancement of science (Ali and Nazim, 2025). Data sharing is an important and necessary phenomenon for research transparency, accountability, visibility, and reuse, and is one of the mandatory requirements of funding agencies, such as the *European Commission* and the *U.S. Department of Health and Human Services, National Institutes of Health*,¹ therefore funded research data should be shared in accordance with the data sharing standards of the respective funder.

Funding support for scientific projects can come from different sources, such as academic institutions, governments, private organizations, and foundations. The type and amount of funding can vary depending on the scope, objectives, and potential impact of the project. Researchers identify and select appropriate funding sources, while funders typically seek to support institutions and researchers whose profiles align with their specific research policies, expectations and priorities, regarding project types, scientific outputs, and geographical areas. Sharing data from funded research, in accordance with FAIR principles, is important, in order to increase research impact, to enable other researchers to use the same data, to verify results, and, ultimately, to use all varieties of resources more effectively and efficiently. Since *data papers* have become such an important means

of publishing research data, data paper publication practices should be addressed in more detail in the future development of open data policies (Li and Jiao, 2022: 835).

In this study, it was decided to analyze at the level of co-funder and co-country, in order to examine in more detail interactions and connections between funders. A method frequently used to monitor the development of scientific fields is visualization through social network analysis. Scientific networks have a variable structure, and, in the process, some variables are relatively milder and some are more dramatically expressed. Understanding the effects of these variables is important for monitoring the basic features and development of the discipline in question (Chen, 2004; Wu and Duan, 2015). The networks visualized by social network analysis are critical for both researchers and funders in terms of monitoring potential funders who may be brokers/intermediaries, and visually representing potential funder-institution collaborations. Observing symbiotic relationships, which cannot be observed using traditional statistical methods, with social network analysis, will also be valuable for researchers, especially in interdisciplinary studies.

The purpose of this study is to examine in detail the data papers indexed in the *Web of Science* (hereafter *WoS*) database; and, in particular, to analyze the funding information within the papers themselves. Furthermore, understanding the data paper funder relationship can shed light on funders' strategic practices, which often involve supporting particular institutions, or researchers, whose work aligns with the funders' interests and priorities, sometimes on an ongoing basis.

We address the following research questions:

- Which research categories had the highest likelihood of receiving funding?
- Which countries and funders played a more central role in encouraging research collaboration?
- What was the level of interdisciplinary collaboration by funders?

Literature review

Research analyzing funded studies is important for guiding researchers and funders. Many different variables come to the fore when examining funding effectiveness, which is multifaceted. For example, there are

differences in publications which receive local government funding, and those which receive international support (Zhou et al., 2020). Unfortunately, the opportunity to receive funding support is not the same for every country, or researcher. The number and type of supporting institutions can also vary from country to country, as well as the amount of support provided. In particular, there are differences between countries in the regulation of government funding sources. It has been found that national funding agencies² are more effective than non-focal agencies in increasing citation impact; international cooperation increases citation impact, and developing countries benefit more from this cooperation (Zhou et al., 2020).

Recent studies illustrate a shifting global landscape in which China and the United States dominate research funding, contributing to a duopoly in the global scientific ecosystem. China has surpassed the United States in terms of publications acknowledging domestic and international funding, while the United States remains the most significant global research partner (Miao et al., 2023). This duopoly highlights the strategic positioning of these two nations in global research ecosystems. According to Miao et al. (2023), China has not only surpassed the United States in terms of domestically funded research outputs, but has also increased its international collaboration rates, making it a central actor in the global science funding landscape. Meanwhile, the United States continues to maintain its role as the primary global research partner, especially in high-impact scientific collaborations.

The sustainability and success of scientific research largely depend on aligning with the strategic priorities set by funding agencies, and developing effective application strategies. Traditionally, the management and distribution of research funding have been shaped by the political priorities of funders. However, researchers are far from passive recipients in these processes. On the contrary, they strategically reshape these processes, taking effective steps to both preserve their own research priorities, and to align with funder policy goals.

The study conducted by Morris and Rip (2006) examined how life science researchers strategically adapted to changing research policies. The findings indicated that researchers did not merely comply with policy changes; instead, they developed various strategies to maintain and sustain their research priorities. These strategies included diversifying funding

sources, aligning research topics with policy priorities, and establishing flexible collaborations. This demonstrates that researchers are not strictly bound by the frameworks established by funders; rather, they actively reinterpret these frameworks according to their own research objectives. Similarly, Shove (2003) demonstrated that researchers did not simply adapt to the political priorities of funding agencies, but also integrated these priorities into their own research strategies. Shove's study shows that researchers are not passive responders to the demands of funders; rather, they reshape research processes, in order to align with funder policy goals. In doing so, they not only comply with the policies, but also develop strategies to extend the boundaries of funding frameworks, and preserve their own research agendas.

Furthermore, recent studies examining North–South research collaborations in the context of Research for Development (R4D) projects emphasize that Southern researchers actively adapt to the rigid project frameworks defined by Northern funding agencies. Rather than merely complying with predefined structures, they strategically reinterpret project requirements to align both donor expectations and local research priorities. This process fosters a co-creation of project frameworks that incorporate contextual realities and local expertise, enhancing the sustainability and relevance of development-oriented research. Additionally, findings indicate that Southern researchers are capable of negotiating project terms and influencing final research content, suggesting a shift from passive participation to proactive engagement. This capacity to reshape project structures reflects Southern researchers' strategic agency in navigating asymmetric North–South funding relationships, challenging traditional assumptions of unidirectional control (Alom Bartrolí, 2023).

These findings reveal that research funding processes are not solely determined by the strategic priorities set by funding agencies. Rather, they are actively reinterpreted and reshaped by researchers through strategic interventions. The studies by Morris and Rip (2006), Shove (2003), and Alom Bartrolí (2023) collectively demonstrate that researchers are not merely passive recipients in funding processes, but are strategic actors. Therefore, to achieve a deeper understanding of research funding mechanisms, it is essential to consider the researcher perspective, as a critical component of these processes.

The relationships between research funding and funding organizations which contribute to published

work are important for identifying potential field-specific funding opportunities. In a study analyzing nearly 6 million articles from WoS, it was found that the main sponsors of funded articles were government agencies. The top three funding organizations were predominantly local, and a significant portion of the funds were given to local research projects. In the same study, China was found to have the highest proportion of funded articles in its country's total scientific output, while Italy had the lowest proportion (Huang and Huang, 2018).

In another study examining the impact of government funding on research productivity, it was similarly found that approximately 70% of research articles in China received funding (Wang et al., 2012). From a discipline perspective, papers in the life sciences had the highest proportion of funded papers from that field's total paper output, while natural sciences had the highest proportion of papers of all funded research in every country (Huang and Huang, 2018).

Impact evaluation research on funded research is used as a guide for science and technology policies, and is crucial for the proper use of resources. Measuring the return on investment of research projects is not always easy, but is not impossible. Measuring the research impact of the academic outputs of funded studies, such as comparing the number of citations received by funded and non-funded studies, is important in this context.

There are many different variables involved in examining investments and their outcomes. For example, the type of funding institution and the structure of the funding are related to the impact of the output. In this context, while there is no relationship between funding and first citation, there is a significant relationship between the number of citations and top percentile citation impact. Citation impact is positively related to funding variety, and negatively related to funding intensity (Gök et al., 2016).

Whether the research is funded or unfunded impacts its influence within scientific communication. In all disciplines, funded research is published in more prestigious journals, has higher citation rates, and is developed by larger teams (Alvarez-Bornstein and Bordons, 2021). Similarly, publications that receive funding have a higher impact, both in terms of citations and journal ranking, than those which do not (Wang and Shapira, 2015). More specifically, the Matthew effect is higher for funded articles, than for unfunded articles (Roshani et al., 2021). We

conjecture that these findings are influenced by the fact that research which applies for funding goes through an evaluation process, and the more promising or important projects are selected through that process. It is not surprising that the scientific output of research projects which have gone through a pre-evaluation and screening process have a higher research impact. This is described as a “positive cycle” (Wang and Shapira, 2015). Research or researchers who fail to enter this cycle have a low chance of success, and are likely to face difficulties in a competitive research environment.

Financial support is also important for the visibility and recognition of research. In a study of 10 core journals in the field of Library and Information Science, it was reported that the proportion of funded articles was 14%, and the average citation counts of these articles (24.56) was higher than unfunded articles (20.49). In addition, funded open-access articles had a higher scholarly impact, than funded subscription-based articles (Ali and Nazim, 2025). Unlike other studies, Zhao et al. (2018) investigated the research impact of funded studies through the WoS Usage Count, and concluded that there is a positive relationship between usage and funding, which varies by discipline.

The rise of data-intensive research has also been fueled by the strategic funding mechanisms that prioritize open data and reproducibility in scientific outputs. This transition is evident in the proliferation of data papers, which play a critical role in the visibility and impact of funded research. Data papers not only facilitate data reuse, but also represent a new form of research visibility, especially in the context of funded projects where transparency and accessibility are prioritized. Data papers play a dual role in enhancing research visibility and supporting reproducibility, particularly in projects backed by major funding agencies.

Data papers have a positive impact on both data reuse, and metrics related to research articles, such as Altmetrics, citation, views, downloads, and tweets (McGillivray et al., 2022). Similarly, it has been observed that papers published in *Data in Brief*, which has published the highest number of data papers in recent years, attracted more than five Mendeley readers within a year of publication, and received a significant number of citations. Overall, *Data in Brief* makes a positive contribution to science by enabling access to data of various types, although its contribution to data reuse is limited (Thelwall, 2020). Another study on *Data in Brief* (Fu et al., 2023) investigated whether publishing a

data paper in an open access data journal had an impact on the number of citations of the related research paper. The findings showed that research articles related to data papers had a higher number of citations, than other articles published in the same issue of the same journal. Content analysis was used to investigate how data papers increased the citations of related research articles. It was found that authors cited the data paper and the related research article together, in order to reuse the underlying data, or to better understand the underlying data and related research articles. In 14.2% of the co-cited studies, this association was due to data reuse.

In a study (Candela et al., 2015) conducted on journals publishing data papers, it was determined that there are seven journals which exclusively publish data papers, and 109 journals publishing data papers along with other types of papers. The subject distribution of these journals, according to Scopus classification, most of which are peer-reviewed and open access, is represented by *Medicine* (52.67%), *Biochemistry, Genomics and Molecular Biology* (25.89%) and *Agricultural and Biological Sciences* (16.07%). Schöpfel et al. (2021) conducted a subject analysis of data papers published in WoS up to March 2020, and stated that more than 85% of the papers were in the *Multidisciplinary Science* subject area, which makes the determination of subject distribution difficult.³

In another study investigating the geographical dynamics behind data paper production (Chen et al., 2022), 6821 data papers published in the *Scientific Data* and *Data in Brief* journals between 2014 and 2020 were analyzed to see how researchers from different countries collaborate. It was found that the majority of data papers (67.7%) were based on local collaboration (i.e., co-authors from the same country), 28.6% were international collaborations, and 3.7% were single-author papers. The same collaboration model largely reflected the model for research articles.

Data sources and methodology

Within the scope of the research, in addition to the bibliometric analysis of data papers, funder information was also examined in detail. The WoS and the Crossref funder registry were used to create the dataset. The collection and verification of funder names was conducted with great care and meticulous precision, in order to ensure the complete accuracy of name uniformity. The funder information was extracted from the Web of Science (WoS) database. But, due to inconsistencies

and errors in the raw data, further standardization was necessary. The extracted funder names were matched and standardized using the Crossref Funder Registry, which maintains an open and unique registry of grant-giving organizations worldwide. This process involved the automatic matching of funder names with high similarity scores, followed by manual verification to correct any discrepancies. The manual verification step was crucial to ensure that all funder names were accurately represented, avoiding duplication and misclassification.

Figure 1 shows how the dataset was created.⁴ The number of citations received by each data paper downloaded from WoS was also examined. In this context, since the publication year of the oldest dated data paper was 2006, and a certain period of elapsed time was allowed for a study to be cited, the year range was determined to be 2006–2017. In the first stage of the research, the data papers and the metadata were accessed and downloaded on April 23, 2023, using the query *DT=(Data Paper) AND PY=(2006–2017)* (Figure 1, Step #1).

While including more recent publications is often desirable, the 2006–2017 date range for this study was deliberately chosen for several key reasons. Firstly, this period encompassed the initial emergence and early, yet substantial, growth phase of the data paper, as a recognized document type within the WoS database. This enabled an analysis of its establishment within the scholarly communication landscape. Secondly, ending the timeframe in 2017 provided a sufficient window (minimum 5–6 year by the time of analysis) for publications within the set to circulate, be indexed, and accrue at least some initial citations. This allowed for a more stable cohort for analysis, compared to including very recent papers, whose citation records would have been highly incomplete, even if citation impact had not been the primary focus of this study. Thirdly, given the exponential increase in data paper publications after 2017 (as shown in Figure 2), and the considerable manual effort required for standardizing funder information, limiting the dataset to this period ensured the feasibility and methodological consistency of data processing and analysis. Analyzing the distinct dynamics of the post-2017 period, with its vastly larger volume, would have demanded a separate, larger-scale undertaking, which would have required different resource allocations, and potentially different methodologies. This timeframe focus is acknowledged as a limitation in the Discussion section, and the

findings reflect the specific characteristics of this foundational period for data papers in WoS.

Between 2006 and 2017, journals indexed in the WoS database published a total of 2714 data papers. The analysis of funders was conducted on 2044 studies, since 670 of the data papers did not include information on funders. In the next three steps (#2 - #4), standardization was made at the funder level. For this purpose, first, the article-level data were reduced to the funder level by preserving the article connection information (Step #2). Funders were standardized using the Crossref funder registry, which maintains an open and unique registry of persistent identifiers of grant-giving organizations worldwide (Lammey, 2020). In the second step of the standardization phase, we downloaded all data on funding organizations identified in Crossref (Step #3), and categories identified in Dimensions (Step #4). In the next step, the data downloaded from WoS, and reduced to the funder level, were matched with the Crossref and Dimensions data (Step #5). In this step, python code was written for automatic matching. 62% of the list was matched automatically, including duplicate records, but not including individual funders. For the remaining unmatched funders, a similarity algorithm was run, and funders with 90% similarity in WoS and Crossref data were merged. Since there is the probability of a high error rate when using automatic matching, the matching records were checked manually. It was observed that the data in the funder field of the records in WoS was quite dirty, and that there were many incorrect entries. Finally, records that did not match automatically, and were mentioned at least three times, were manually checked.⁵ The standardized funders were then brought back to the article level (Step #6). For the social network analysis, the dataset had to be converted to the WoS format. Using the software VOSviewer and CiteSpace, analyses, such as country and funder, were performed at the article level. In our study, we chose to evaluate at the level of co-funder and co-country, in order to gain a more detailed understanding of the interactions and connections between funders. When converting the dataset to the WoS format, standardized funders were used instead of author names (without losing the article relationship). The same process was carried out for country analysis. That is, instead of author names, the country equivalents of the standardized funders in Crossref were used. In this way, a structure was created in which symbiotic connections at the funder level could be tracked (Step #7).

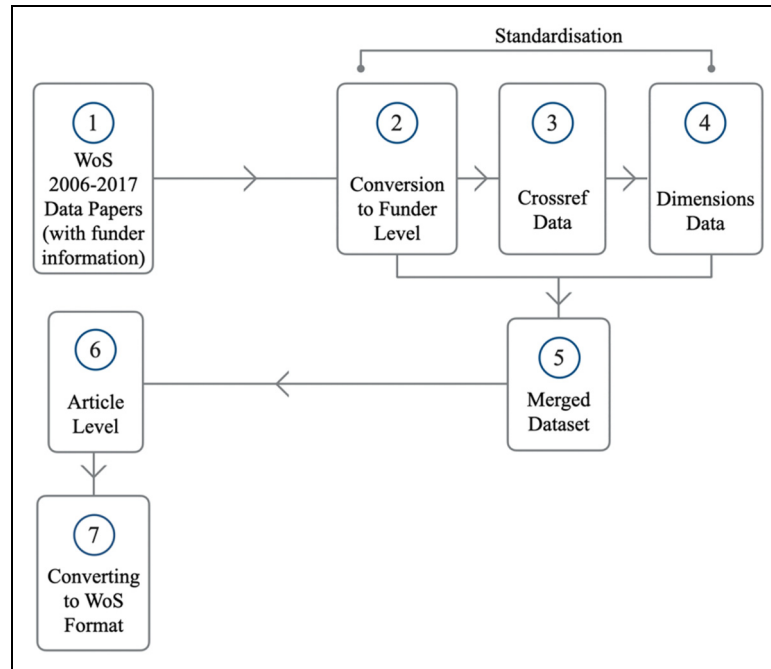


Figure 1. The six steps followed in the making data suitable for analysis.

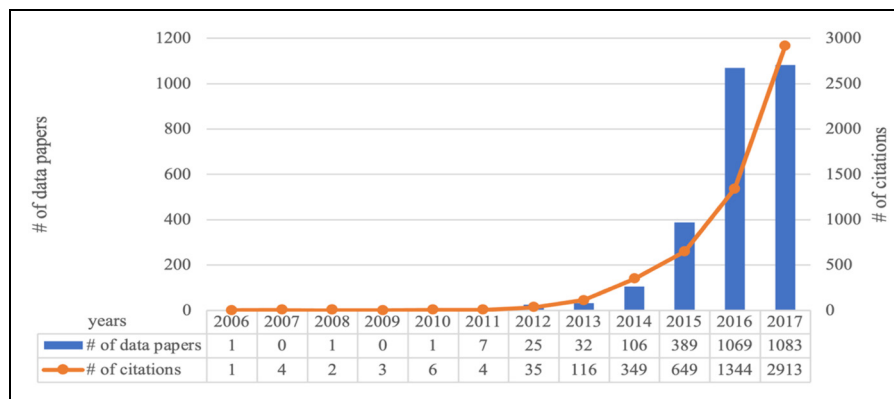


Figure 2. Distribution of data papers, and citations by years.

Note: The number of citations and the number of papers is independent of each other at the year level. For example, while the number of data papers published in 2009 was 0, the total number of citations received by data papers in WoS in that same year was 3.

Findings

Data paper level descriptive statistics

Between 2006–2017, a total of 2714 data papers were published in journals indexed in WoS. The total number of citations of these publications was 5426. The distribution of these data papers and citations by years is shown in Figure 2.

Although the dataset covers a 12-year period, it was observed that approximately 80% of the total data papers were produced in the last two years of that

period (2016–2017). Annual data paper production, which was very rare until 2011, increased continuously every year after 2011. In 2014, it showed a very sharp increase. By 2016, it had increased by a factor of 10, compared with two years earlier.⁶

As expected, as the number of data papers increased, the citations of these papers also increased steadily. One of the reasons for this increase may be that 98.7% of the data papers were open access. The fact that most of the studies (68.8%) were published in the journal *Data in Brief* (see Table 1) may have

Table 1. Top 10 journals with the highest number of data papers.¹¹

Journal Name	# of Data Papers	%
Data in Brief	1867	68.8
Scientific Data	294	10.8
Earth System Science Data	99	3.6
Biodiversity Data Journal	61	2.2
Acta Crystallographica Section E	47	1.7
Crystallographic Communications		
Genomics Data	45	1.7
Gigascience	41	1.5
Human Genome Variation	36	1.3
Oncologist	31	1.1
Data	23	0.8

contributed to this high open access rate. In addition, studies in the literature (Fu et al., 2023) show that publishing an article in an open access data journal such as *Data in Brief* increases the number of citations, not only for the data paper, but also for the research paper to which the data relates.

As mentioned earlier, approximately 25% (N = 670) of the data papers published from 2006 to 2017 did not have any funder information. In terms of categories, it was seen that the data papers published in the journals indexed in Dimensions, were from 22 different categories.⁷ Table 2 shows in more detail, by category, the distribution of data papers which were either supported, or not supported by funder(s). The most intensive data paper production during this period was in “31 Biological Sciences” (N = 1133, 41.7%) (see Table 2). 58.7% of the data papers were published in “31 Biological Sciences” and “32 Biomedical and Clinical Sciences” categories. This was expected given the data-intensive nature of the studies in these categories.

66% of all data papers received support. More specifically, more than half of the data papers were funded in 18 of 22 categories (82%). In terms of funding, while the proportion of all data papers in the “31 Biological Sciences” category was 30.5%, the proportion of supported data papers in the “31 Biological Sciences” category is higher (32.3%). In other words, 909 of the 1133 (80.2%) studies in this category were funded. The average funding rate for the “31 Biological Sciences” and “32 Biomedical and Clinical Sciences” categories, which account for about half (48.6%) of all data papers, was 79.9%. Looking at the overall distribution, these two

Table 2. Distribution of data papers according to the published category.

Dimensions Categories	Total N	Funded		Unfunded	
		N	%	N	%
31 Biological Sciences	1133	909	80.2	224	19.8
32 Biomedical and Clinical Sciences	657	523	79.6	134	20.4
34 Chemical Sciences	393	311	79.1	82	20.9
40 Engineering	347	245	70.6	102	29.4
37 Earth Sciences	240	181	75.4	59	24.6
41 Environmental Sciences	157	104	66.2	53	33.8
30 Agricultural, Veterinary and Food Sciences	145	115	79.3	30	20.7
46 Information and Computing Sciences	137	103	75.2	34	24.8
52 Psychology	86	67	77.9	19	22.1
42 Health Sciences	85	53	62.4	32	37.6
51 Physical Sciences	67	56	83.6	11	16.4
44 Human Society	53	34	64.2	19	35.8
33 Built Environment and Design	48	24	50.0	24	50.0
35 Commerce, Management, Tourism and Services	47	24	51.1	23	48.9
38 Economics	39	15	38.5	24	61.5
43 History, Heritage and Archaeology	26	16	61.5	10	38.5
49 Mathematical Sciences	14	12	85.7	2	14.3
48 Law and Legal Studies	13	5	38.5	8	61.5
47 Language, Communication and Culture	11	6	54.5	5	45.5
39 Education	9	6	66.7	3	33.3
50 Philosophy and Religious Studies	8	5	62.5	3	37.5
36 Creative Arts and Writing	4	2	50.0	2	50.0
Total	3719	2816	66.0	903	34.0

Note: Some data papers are multi-classified, so the total number adds up to more than 2714.

categories were among the categories with the highest support rates.

A similarly high support rate was observed for the categories “49 Mathematical Sciences” and “51 Physical Sciences”. This may have been due to the relatively low number of data papers in these two categories (14 and 67 respectively). “38 Economics” and “48 Law and Legal Studies” categories had a higher rate of unsupported data papers (61.5% unsupported data papers for both categories).

Dimensions subject categories included both general and detailed subject classifications. For this reason, the category “31 Biological Sciences”, which contained the most data papers, was also examined according to its sub-categories (See Table 3). There were nine sub-categories under the general category of “31 Biological Sciences”. Approximately three quarters (72.4%) of the data papers were in three sub-categories, namely “3101 Biochemistry and Cell Biology”, “3105 Genetics”, and “3102 Bioinformatics and Computational Biology”. 85.4% of the data papers in these sub-categories were funded.

When the sub-topic categories were analyzed according to their funding status, a varied distribution was observed. There were five sub-topic categories (“3103 Ecology”, “3104 Evolutionary Biology”, “3106 Industrial Biotechnology”, “3108 Plant Biology”, “3109 Zoology”) with a funded data paper rate below the overall category average (80.2%). The reason why more than half of the sub-topic categories were like this is that the data papers with 85% or more support were distributed in a smaller number of sub-topic categories. The category with the highest number of supported to unsupported data papers was “3109 Zoology”; of the 24 data papers in this category, 13 received funding support.

Funder level analysis

After matching and manually checking WoS and Crossref data, 2051 unique funders were identified in the funder fields of the data papers. The top 10 funders and their country of origin (according to Crossref data) are presented in Table 4. In terms of funder information (FU field in WoS), the most frequently mentioned funder in data papers was U.S. Department of Health and Human Services; National Institutes of Health. In this WoS example, what appears as two separate funders, separated by semicolons, in our analysis became a single funder. Therefore, 289 data papers (14.1%), were corrected to *U.S. Department of Health and Human Services, National Institutes of Health* and defined as a single funder.⁸

Our research shows (see Table 4 above) that more than half of the articles (50.7%) were funded by 10 different organizations. This data reveals that the USA was in a leading position, making the largest contribution to scientific research, especially through research funding provided by the *U.S. Department of Health and Human Services, National Institutes of Health*. Three of the top 10 funders were of USA

Table 3. Distribution of data papers for the category “31 Biological Sciences”.

Dimensions Categories	Total N	Funded		Unfunded	
		N	%	N	%
3101 Biochemistry and Cell Biology	340	287	84.4	53	15.6
3102 Bioinformatics and Computational Biology	307	263	85.7	44	14.3
3103 Ecology	142	88	62.0	54	38.0
3104 Evolutionary Biology	36	22	61.1	14	38.9
3105 Genetics	328	282	86.0	46	14.0
3106 Industrial Biotechnology	28	21	75.0	7	25.0
3107 Microbiology	111	95	85.6	16	14.4
3108 Plant Biology	31	21	67.7	10	32.3
3109 Zoology	24	13	54.2	11	45.8
Total	1347	1092	73.5	255	26.5

Note: Some data papers are multi-classified, so the total number adds up to more than 1133 (# of data papers in category “31 Biological Sciences”).

origin. The USA invests heavily in scientific research, and these investments contribute to important scientific advances worldwide. At the same time, Japan’s two major organizations, the *Ministry of Education, Culture, Sports, Science and Technology* and the *Japan Society for the Promotion of Science*, also provided significant research funding (6.1%, and 4.8% respectively). Another noteworthy country was China, which ranked third. With the *National Natural Science Foundation of China* (5.3%), China stood out as a major financial contributor to scientific research. *The European Union (EU)* ranked fourth, and played an important role in promoting and supporting scientific research. The EU funded a wide range of research; a significant proportion of these data papers were the result of collaboration.

It is noteworthy that during the period covered by our study, policies promoting or mandating data sharing were increasingly implemented by major funders, particularly influential bodies like the National Institutes of Health (NIH) in the USA, which has had a data sharing policy since 2003; the *European Commission*, which under Horizon 2020, has progressively required data sharing since 2017; and Natural Environment Research Council in the UK, which has been a long-standing proponent of the practice. These policy shifts likely contributed significantly to the rise in data paper publications

Table 4. Top 10 funding organizations and their countries.

Funder	Country	# of Data Papers	%
U.S. Department of Health and Human Services, National Institutes of Health	USA	289	14.1
Ministry of Education, Culture, Sports, Science and Technology	Japan	125	6.1
National Natural Science Foundation of China	China	120	5.9
European Commission	European Union	108	5.3
Japan Society for the Promotion of Science	Japan	99	4.8
National Institute of General Medical Sciences	USA	76	3.7
Deutsche Forschungsgemeinschaft	Germany	62	3.0
National Cancer Institute	USA	61	3.0
Natural Environment Research Council	United Kingdom	52	2.5
Conselho Nacional de Desenvolvimento Científico e Tecnológico	Brazil	48	2.3
Others	—	1011	49.3
Total	—	2051	100.0

observed towards the end of the period (2016–2017) of our study. The composition of the top 10 funders highlights the dominant role of established scientific nations from the Global North (USA, Japan, Germany, UK) and the European Union. While China represents a major emerging economy, Brazil's CNPq is the sole representative from Latin America. No African nation appears in this top tier, potentially reflecting broader global disparities in large-scale research funding relevant to data paper production, indexed in WoS during this period.

The distribution and support rates by year also provided insights into the process by which data papers gained their current prominence. Table 5 shows the number of data papers by year, and the number of the data papers receiving funding. The data papers that received funding were relatively more recent. Since the number of data papers in the first years of the dataset (2006–2010) was very low, it may be misleading to comment from the funder's point of view. However, since 2011, when the number of data papers started to increase, it was immediately noticeable that more than half of the data papers were funded articles. Both the number of data papers and the proportion of funded articles within these data papers began increasing. In 2016 (76.5%), and 2017 (77.3%), the data papers indexed in WoS were supported much more than in previous years by at least one funder. A significant contributing factor to this trend, especially in 2016 and 2017, was the increasing implementation of data sharing policies by major funding agencies, such as the NIH and the European Commission, requiring or strongly encouraging funded researchers to make their data openly available. These requirements likely spurred the increase

Table 5. Number of data articles and funding rates by year.

Year	Total N	Funded		Unfunded	
		N	%	N	%
2006	1	1	100.0	0	n/a
2007	0	0	n/a	0	n/a
2008	1	1	100.0	0	n/a
2009	0	0	n/a	0	n/a
2010	1	0	n/a	1	100.0
2011	7	4	57.1	3	42.9
2012	25	18	72.0	7	28.0
2013	32	17	53.1	15	46.9
2014	106	72	67.9	34	32.1
2015	389	276	71.0	113	29.0
2016	1069	818	76.5	251	23.5
2017	1083	837	77.3	246	22.7
Total	2714	2044	75.3	670	24.7

in the publication of data papers, as a means to comply with the new policies, and to disseminate data.

We also looked at how many different funders supported each data paper. The data paper entitled *Global Carbon Budget 2015* (Le Quéré et al., 2015) was the study supported by the largest number of individual funders with 75 funders. Figure 3 shows the distribution of data papers and funders up to a maximum of 16 funders per paper. 670 (24.7%) of the data papers did not contain any funder information. The number of data papers supported by a single funder was 589 (21.7%). More than half of the articles (52.8%) were supported by between 1 and 4 different funders. As was easily predicted, the paper frequency decreased as the number of funders increased.

We analyzed whether the number of funders varied by category. Of the 2044 data papers supported by at

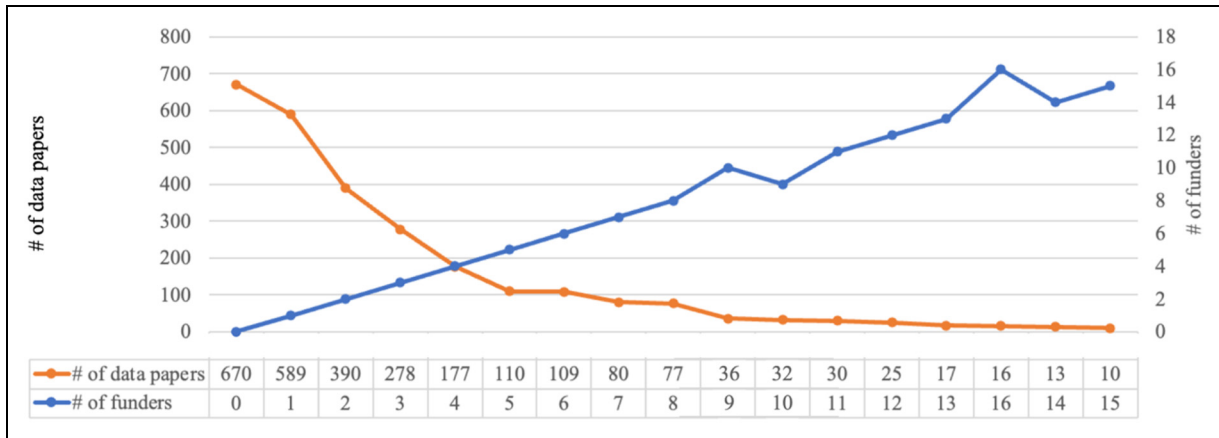


Figure 3. Number of funders supporting data papers up to a maximum 16 funders per data paper.

least one funder, the top 10 categories are shown in Figure 4.⁹ When we look at the number of funders of the studies in the field of “31 Biological Sciences”, where 41.7% of the data papers were categorized, it is seen that the median (50%) value is 2. When the first 10 categories were analyzed, it was revealed that the median values were mostly 2, while the median values were less (1) for “40 Engineering”, “41 Environmental Sciences”, “42 Health Sciences”, and “46 Information and Computing Sciences”. Generally speaking, we can say that studies in these disciplines were supported by fewer funders.

Considering the quartile values, the value of 1 was also prominent in the top 25% of categories, which received the least amount of funder support. Since the data papers were relatively evenly distributed across categories, it was expected that the median values and the top 25% would be similar. It was revealed that in the top 10 fields with the largest number of data papers, “40 Engineering”, “41 Environmental Sciences” and “42 Health Sciences” received less funding support than the other categories.

When it comes to the 75% quartile value level, which had a relatively higher number of funders per data paper, it is observed that the general distribution was again mostly even. Considering all categories (http://www.mugeakbulut.com/data_papers/dimensions_full_boxplot.html), it should be noticed that the number of funders in the field of “51 Physical Sciences” was higher than in other fields. Although the number of data papers in this category was relatively low (67), the number of funders was higher than in other categories. This is directly related to the nature of physical science studies. Since it is a science that also forms the basis of technology, developments in

physics lead to the development of new technologies, and the production of new products. It would be expected for these studies to receive funding support from more organizations, since they have a significant potential for economic growth and development.

The collaboration network of funders

In the dataset downloaded from WoS, the names of the funder(s) of the data papers were used, instead of the author’s name(s). In this way, a collaboration network of funding institutions was obtained (codes can be accessed at <https://github.com/...>). The collaboration network, grouped according to the top three categories, based on their density, is shown in Figure 5.

One of the most important issues regarding the structure of a collaboration network is density. The greater the number of connections between the nodes, the denser the network is. When all nodes are connected to each other, the network density is 1. The network resulting from our analysis was not dense, because the density was 0.0619. The high betweenness centrality of the *U.S. Department of Health and Human Services, National Institutes of Health* allows us to comment on intra-country dynamics, particularly in the fields of “31 Biological Sciences” and “32 Biomedical and Clinical Sciences”, where *U.S. Department of Health and Human Services, National Institutes of Health* is a key player. At the country level, the United States of America is also a key node for these fields, because the *U.S. Department of Health and Human Services, National Institutes of Health*, is a national

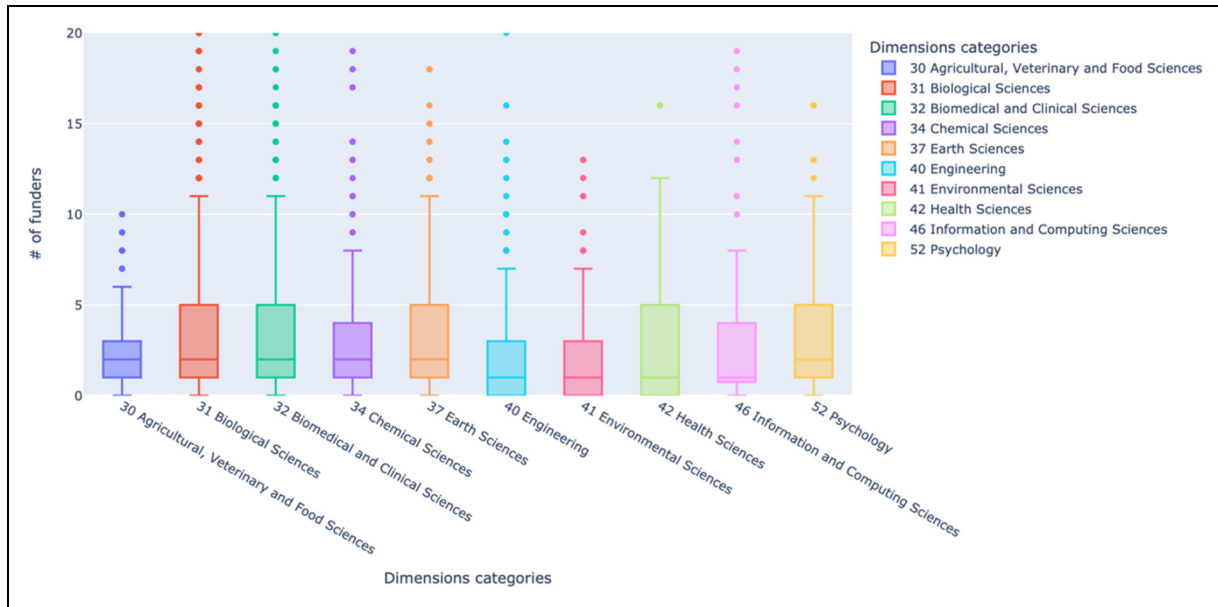


Figure 4. Comparison of the number of funders by category (for an interactive version see: https://mugeakbulut.com/data_papers/10_cat_box_plot.html).

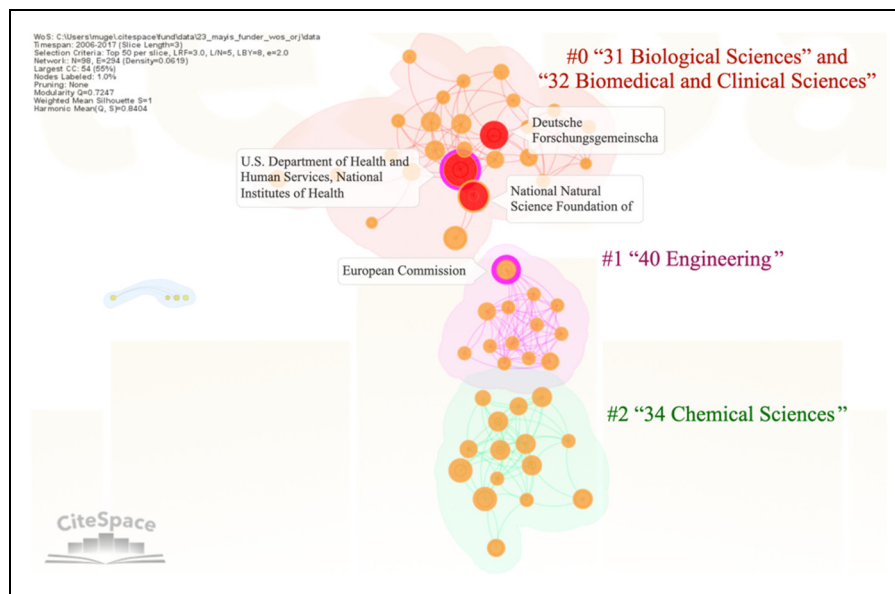


Figure 5. Collaboration network of funders of data papers.

organization with the highest betweenness centrality in the overall network. This finding is similar to Verma et al.'s study (2023: 4311).

There was also an isolated part of the network which was not connected to the other clusters. At the cluster level, cluster #0 ("31 Biological Sciences" and "32 Biomedical and Clinical Sciences"), which was the cluster with the most funders, and cluster #2 ("34 Chemical Sciences")

had looser connections. In cluster #1 ("40 Engineering"), on the other hand, the connections were tighter, with density decreasing from the center outwards. The most important reason for the high density in cluster #1 is that the funders in that cluster were linked to the *European Commission*, which had a high betweenness centrality score. Studies supported by the *European Commission* are often supported by other funders. The fact

that the *European Commission* both encourages scientific cooperation between member states, and supports projects in a wide range of disciplines enhances scientific collaboration by increasing the likelihood that various funders will contribute to the same project.

The distribution of data papers according to the most frequently published categories is given in Table 2. Since the largest number of data papers was in “31 Biological Sciences”, and “32 Biomedical and Clinical Sciences”, it is not surprising that that cluster had both the highest density, and the greatest number of nodes. On the other hand, the “40 Engineering” cluster, which ranked seventh among the funded data papers, and second among the non-funded papers, was the most concentrated network in terms of relationships. In other words, research within this category was supported by multiple funders who also supported studies in that same category. This may be due to the nature of the “40 Engineering” category. Figure 4 shows that although studies in the category of “40 Engineering” were not supported by multiple funders (Median 2), there were more intense relationships between those funders. Thus, a study supported by one funder in this category was very likely to be supported by other funders also interested in this category.

Among network metrics, the centrality metric is related to the position of funders within the same network. One of the network metric’s most important objectives is to identify the main funders. Betweenness centrality, which is one of the measures related to the concept of centrality, is a measure of the level of connection of a unit with other units with which it is not directly connected (Chen, 2016). A betweenness centrality score is calculated using the number of shortest paths between two nodes divided by the total number of paths (Akça and Akbulut, 2023). Funders with a high betweenness centrality score act as a bridge between different clusters. The funder with the highest betweenness centrality in the network we created was the *European Commission* (0.13). Looking at the category information of the data papers, it is seen that the studies supported by the *European Commission* were from nine different categories. The betweenness centrality value for the *U.S. Department of Health and Human Services, National Institutes of Health*, which supported the research with the highest number of data papers (289), was 0.07. Although it supported studies in 5 different categories, most of these studies (86,5%) were in “31

Biological Sciences”. Since only 39 of the supported studies were from the remaining 4 categories, the betweenness centrality value was accordingly lower.

In general, the degree of centrality of the funders in the network was low (the highest was 0.13). Therefore, we can define the network of funders in the data papers as a loose network. This can be interpreted as the absence of a centralized funder for all three clusters. However, since clusters #1 and #2 contained interdisciplinary studies, it is expected that the degree of centrality would also be low. This is because the degree of centrality tends to be higher only for specific categories.

The small network in blue to the left side of the main collaboration network, positioned as a different component of the overall collaboration network, is a *robotics* cluster consisting of the following funders *Deutscher Akademischer Austausch Dienst Kairo, Engineering and Physical Sciences Research Council, Comisión Nacional de Investigación Científica y Tecnológica* and *Clearpath Robotics Partnerbot Program*. These funders are positioned as a small world with no links to other clusters.

When the data paper network is examined holistically, it can be said that the interdisciplinary categories (clusters #0 and #2) and *40 Engineering* (#1) exhibited similar behavior in terms of co-funding. All three groups were heavily co-funded within themselves. In “31 Biological Sciences” and “32 Biomedical and Clinical Sciences” citation burst was observed in three funders. Studies supported by the *U.S. Department of Health and Human Services, National Institutes of Health, Deutsche Forschungsgemeinschaft* and the *National Natural Science Foundation of China* experienced a citation burst soon after they were published. Citation burst is referred to in the literature as “smart girls”. The studies funded by the *European Commission* and the *U.S. Department of Health and Human Services, National Institutes of Health* were also supported by other funders. Unlike the other funders in the dataset, the studies supported by both of these key funders were from diverse fields. Studies supported by the key funders also received funding support from funders in other fields.

The funder level collaboration network between countries

Understanding the collaboration network between countries can also help researchers. For this purpose (Figure 6), a collaboration network of 1197

connections among 69 countries was created. Since we did not have information on the research funding budgets of the funders related to data papers, the full count method was used for preparing the network (The minimum number of funders from a country was 5). The USA, Germany, China and India were at the forefront of the network. The creation of the collaboration network resulted in 7 clusters, one of which contained only Spain and Algeria.¹⁰ Studies supported by funders from Algeria were only supported by Spain. On the other hand, Spain in the same cluster provided joint funding support along with 59 other countries. While Network Visualization (Figure 6) demonstrates prominent nodes for developing or emerging economies such as, China, India, and Brazil. Further analysis could explore whether these connections predominantly link to Global North hubs (like the USA or European countries), or if significant South-South collaboration networks are also evident within the data paper funding landscape. A preliminary observation suggests strong ties between the USA and many countries regardless of economic status, but the density of connections amongst Global South countries appears lower, warranting further investigation.

The USA had connections with 62 other countries, the highest number of such connections. Since the *U.S. Department of Health and Human Services, National Institutes of Health* is the organization which gives the most funding for data papers, it is not surprising that it stands out in the network. *The Ministry of Education, Culture, Sports, Science and Technology* and the *National Natural Science Foundation of China*, the second and third largest funders, are based in Japan and China, and were therefore also major nodes in the network. However, not many studies received support from German funders, but those which did receive support from German funders also often received support from other countries.

Discussion and conclusion

From the perspective of open science, data papers play a crucial role in enhancing the visibility and reuse of research data. Data papers also serve as vital tools for evaluating the research effectiveness of funded projects. It is therefore essential to investigate whether the provided funding achieves its intended objectives, and generates an impact on scientific output.

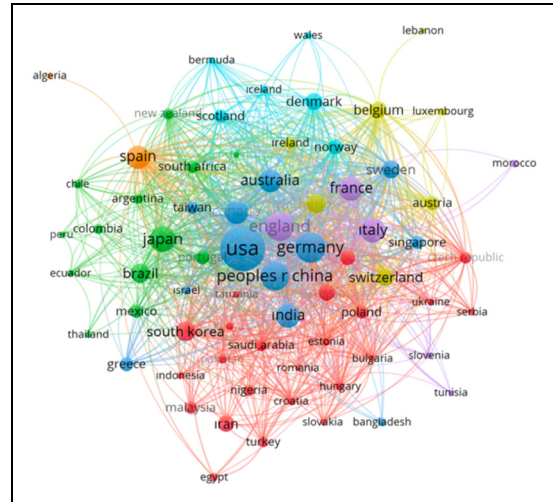


Figure 6. Collaboration network for data papers among countries.

Receiving funding from multiple sources does not necessarily indicate that the funders are part of a cohesive funding network. This paper was not framed primarily as an investigation distinguishing between direct, formal collaborations among funders and the more common scenario of coincidental co-funding arising from principal investigators (PIs) securing grants from multiple sources. However recent evaluations suggest that while our initial perspective emphasized the latter, we acknowledge that direct agency co-funding of individual projects might be less frequent than PI-driven efforts. This perspective requires further consideration. Indeed, emerging studies highlight the growing trend of funders increasingly collaborating through strategic alliances and joint initiatives, particularly around specific large-scale research challenges by pooling resources to improve coordination and efficiency (European Commission, n.d.). These cooperative funding mechanisms, especially in global research agendas, illustrate that funders are not only acting independently but are also forming synergistic partnerships to enhance research impact and resource optimization. This evolving landscape undoubtedly adds complexity to the international funding environment. Given the nature of our dataset (based on co-occurrence in publication acknowledgements) our analysis focused on interpreting these observed co-funding patterns, as a source of insight for researchers, navigating the funding landscape; rather than as an attempt to map the structure of formal funder alliances. Such a research project would necessitate different data and methodologies

specifically designed to capture the structural and strategic interconnections between funding bodies. Our approach allows for the identification of valuable information with which to guide researchers in identifying potential funding opportunities and understanding the dynamics of research support. We highlight the role of principal investigators as strategic agents in securing multi-source funding, aligning various grants with project goals, and navigating the complex web of international research financing.

In this context, in our study, 2051 unique funders were analyzed in detail for a total of 2714 data papers in journals indexed in WoS between 2006 and 2017. Although the first data paper indexed in WoS was published in 2006, the number of data papers published before 2016 was quite low, with a significant increase in 2016 and 2017. 80% (N = 2152) of the papers subjected to analysis were published in these two years. It should be noted that the ratio of data papers to the number of articles indexed in WoS in the same years (N = 3,538,561) was quite low (0.061). There are different reasons for the low number of data papers published before 2016. One of these reasons is the fact that it was a relatively new type of publication that had not yet been fully accepted by researchers and publishers. Although the *Journal of Chemical and Engineering Data* first appeared in 1956, most data journals were established only in the last ten years. While the number of data journals has increased, many researchers are still unaware of this type of journal and publication (Walters, 2020: 4–5).

Data paper publishing behavior also varies across disciplines. In this study, the majority of the papers were published in *Data in Brief* (68.8%), and classified under the *Multidisciplinary Sciences* category in WoS (79.6%). For this reason, the more detailed *Dimensions* subject categories were used for a detailed discipline-by-discipline analysis. A subject-based analysis of data papers showed that certain topics such as *Biological Sciences*, *Biomedical & Clinical Sciences*, and *Chemical Sciences* stood out, with 80% of data articles on these topics being funded. Funding research in relevant fields can make a major contribution to human and public health, and scientific progress. However, more efforts are needed to make these funding trends more balanced and equitable. Our study revealed that research within the subject categories of *Physical Sciences* typically attracted a higher number of funders compared to other fields. This trend may be attributed to the prioritization of

these subjects in contemporary scientific discourse, leading funders to allocate more resources to these areas. Additionally, the nature of research in *Physical Sciences* often involves large-scale collaborative projects, further contributing to the increased number of funders. Moreover, the formulation of funding strategies that promote cross-disciplinary collaboration could foster a more holistic approach to research funding, facilitating contributions to a wider array of research domains. The dense relationships in the *Engineering* category are likely due to the data-intensive nature of research in this field and its support by multiple and diverse funding sources. Data-driven research requires a large variety of data types and sources, which can lead researchers to seek support from multiple funding agencies. This can, in turn, foster greater collaboration and knowledge sharing among researchers in the engineering field.

Studies in the literature on data sharing, and data papers (McGillivray et al., 2022; Walters, 2020) showed that such publications were more common in STEM, and Health Sciences fields, than in the Social Sciences and Humanities. The interdisciplinary and heterogeneous nature of the Humanities causes data to vary in terms of types and characteristics. In the sub-disciplines of the Social Sciences and Humanities, the concept of “data”, and the criteria for the publishability of these data, may be approached differently (McGillivray et al., 2022). Looking at the top 10 WoS categories in our study in terms of the subject distribution of data papers, the fact that all categories except Multidisciplinary Sciences fall within the scope of the Science, Technology, Engineering, Mathematics, and Health Sciences fields may be considered as an indicator of the variability of the concept of “data”. The attitudes, thoughts, knowledge, and skills of researchers regarding “data sharing” also differ by discipline, and even country (Schmidt et al., 2016; Unal et al., 2019), which impacts the publication of data papers. The value of data papers in academia, and their impact on academic success is an important issue for researchers. For example, whether there is a professional reward for publishing data papers, or whether these potential publications meet the criteria for “publication” at all, including the circumstance that some data is not shared unless it is mandatory, affects the choices researchers make, and therefore has an impact on the quantity of data papers, and data sharing (Chavan and Penev, 2011).

Beyond the varying conceptions of data and the heterogeneous nature of the SSH (Social Sciences

and Humanities) fields previously mentioned (McGillivray et al., 2022), several other factors intrinsic to SSH, such as sensitive data concerning human participants, necessitating stringent anonymity protocols and careful ethical considerations regarding participant vulnerability, inherently complicate straightforward open data sharing (Crow and Wiles, 2008). Furthermore, SSH data are frequently highly contextual and qualitative, making them less standardized and potentially harder to replicate or reuse meaningfully outside of their original context. This leads to valid researcher concerns about the risks of misinterpretation or harm, stemming from decontextualization (Mannheimer et al., 2018). Prevailing disciplinary cultures within many SSH fields may also place greater emphasis on narrative analysis, interpretation, and theoretical contributions concerning the dissemination of raw data, thus influencing publication priorities and the perceived value of data papers compared with traditional articles (Schwartz-Shea, 2019). These combined factors further shape the landscape of disciplinary attitudes towards data sharing, and the acceptance of data papers.

Governments, the private sector, and foundations support research on globally important topics such as climate change, and natural disasters through various programs. Such support enables researchers to conduct data-driven, or big data-driven studies. Calls issued by major funders, such as the EU 7th Framework Programme or NSF's Division of Astronomical Sciences (AST) Programs, are effective in increasing the number of data-driven research, and data papers in these fields.

Citations of data papers should be distinguished from citations of other types of publications. The main purpose of data papers is to provide information about the relevant data, and to help reuse the data, or create new datasets from it. Therefore, citations to data papers are an important indicator of whether these papers have achieved their purpose. The papers analyzed in this study were cited 5426 times in total. Detailed content analysis or bibliometric analysis of the citations was not performed. However, in previous studies (Fu et al., 2023), where content analysis of the citations of these papers was performed, some of these citations were related to data reuse. Some of the citations to the data papers also increased the number of citations of the research article to which the data papers were related.

The funding analysis of data papers is important in terms of determining the topic addressed, the number

and diversity of funders which provided support, and the relationship between those funders. In 25% of the papers analyzed within the scope of our study, there was no funding information. However, it is not possible to say with certainty that these papers did not receive any funding support. For example, funding information may not have been included in cases in which the funders did not require that their financial support be recognized by the authors of the paper. Additionally, it was observed that the funding text or acknowledgement field in citation indexes, which generally contain the funding information for a paper, was not always accurate and complete. For example, it was observed in some records that the same funder appeared under different or incorrect spellings. Some studies that analyzed the information presented in databases such as WoS and Scopus, using fields such as "Funding text" or "Acknowledgement", revealed obvious errors and omissions regarding funding sources (Liu, 2020; Liu et al., 2020). 2044 papers with funding information were supported by 2051 unique funders. 29% of these papers were supported by only one funder. While we do not know the specific details of the support provided, it can be assumed that about one-third of data-driven research is relatively small-scale or conducted by small research groups. It was concluded that the papers from the period 2015–2017, during which most of the data papers were published, received support from three different funders on average. The papers published during this period are probably the products of larger-scale and more collaborative research. The number of funders is not sufficient information alone from which to draw definitive conclusions about funded research. The amount and type of funding provided, as well as information such as the country of origin, should also be taken into consideration. This lack of accurate funding source information limited the scope of our study. Nevertheless, using the standardization of funding information provided by Crossref Open Funder Registry (OFR) was invaluable for the conduct of our study.

It is important to acknowledge the specific limitations inherent to the data sources and methodology employed in this study. Firstly, our reliance on the WoS database introduces potential biases, particularly in terms of geographical and linguistic representation. WoS primarily indexes journal literature and has historically exhibited biases towards English-language publications, and research originating from North America and Europe. This indexing preference may

lead to the underrepresentation of contributions from other areas of the world, such as Latin America, Africa, or parts of Asia. Furthermore, WoS may overlook research published in alternative formats (e.g., books, reports) and certain disciplinary fields, particularly within SSH (Mongeon and Paul-Hus, 2016). We also acknowledge that estimates or counts of the total number of publications, classified as data papers during this period, may vary significantly, depending on the specific definition employed, and the database(s) consulted. Consequently, the funding landscape presented, including the rankings of countries and institutions, reflects the particular coverage and limitations of WoS during the period 2006–2017. Incorporating data from other databases, with distinct geographical or disciplinary strengths, could provide a more balanced global perspective. Secondly, while citation data were collected for context, this study focused on funding networks rather than citation impact analysis, thereby avoiding the direct limitations associated with using citation counts as a sole measure of impact. The accuracy and completeness of the funding information extracted from WoS, despite our rigorous standardization efforts using the Crossref Funder Registry, remains a potential limitation. Inaccuracies and omissions in acknowledged funding sources within indexed publications are known issues (Liu, 2020; Liu et al., 2020). The findings concerning funding networks and collaborations should be interpreted within the context of these methodological constraints.

Although it is assumed that all funding providers are correctly listed in the WoS database, incorrect entries may exist. To address this, we manually standardized the funding provider names using the Crossref Open Funder Registry (OFR), a process that was both data-intensive and time-consuming, involving extensive manual verification and correction.

In our study, 75% of the papers had funding support. Most of the data-driven or data-intensive research received support from different institutions (N = 2051). It can be said that data-intensive or data-driven research has become much more important in recent years, and is difficult to carry out without reliable funding support. Funders wish that the research topic of the data papers they consider supporting will make a valuable contribution to science. Research projects that apply for funding go through an evaluation process, in order to determine whether they are suitable to the interests of the funder. Therefore, the academic outputs of funded research, which a funder considers to have high research value,

or contribution to science, are likely to be published as an article or data paper. Funder policies, such as the obligation to share or open the data of the research they have supported, or the publication of that research output, are therefore important considerations for researchers, as stated in Schmidt et al.'s (2016) study.

The USA and *U.S. Department of Health and Human Services, National Institutes of Health* (N = 289) were the leading country and institution providing the most funding. *U.S. Department of Health and Human Services, National Institutes of Health* is a funder which encourages data sharing. It can be said that this data sharing policy has contributed to an increase in the publication of data papers. However, the betweenness centrality value of *U.S. Department of Health and Human Services, National Institutes of Health* is low (0.07), because the majority of the studies it supported were in the *Biological Sciences* and *Biomedical and Clinical Sciences* subject categories. In general, the collaboration network among funders was not dense (0.0619). It can be said that there was no dominant funder in the network, and that papers received support from many different funders (N = 2051). Studies supported by the *U.S. Department of Health and Human Services, National Institutes of Health*, the *Deutsche Forschungsgemeinschaft*, and the *National Natural Science Foundation of China* were rapidly cited immediately after publication. At the country level, the collaboration network for data-driven research was centered on the USA, Germany, China and India. The USA had the most connections with other countries. When examining the structural features of the collaboration network, the position of certain funders at the center of the network and their role as a bridge between research areas is noteworthy. In particular, given the importance of the *European Commission* in this field, its support for projects covering various research disciplines and its efforts to promote scientific collaboration is important. The European Commission is, and should continue to be, an important funder in terms of the level of support it provides, as well as its potential to encourage collaboration between various stakeholders from different countries, sectors and disciplines.

This study, which was conducted to examine the data paper funder relationship using statistical, and social network analysis, has shown that data papers, which are a relatively new type of publication, but whose number has increased rapidly in recent years, are widespread and extensively funded. In future research, in order to examine the relationship between funding support and data papers, there will

be a need for more detailed overall analysis, such as subject, content, citation, financial support, and reuse of data. Furthermore, the standardized funder dataset compiled for this study provides a valuable resource for future investigations specifically focused on exploring North-South dynamics, potential funding inequities, and collaborative patterns in the funding of data-intensive research and data papers.

Given their increasing importance to open science, more research is needed, in order to better understand the role of data papers in scholarly communication, to analyze the institutions that provide funding support, to determine which topics should receive the most support, and to reveal the possibilities of research, and financial cooperation among funders. Such research pursuits would be valuable for funders, both in terms of evaluating their past investments, and in planning their future investments. It is also important to examine in detail the results of data-driven research conducted by both small and large research groups. Small-scale research groups can be fast and flexible in providing innovative approaches or creative solutions, while large-scale research can achieve more comprehensive results by providing more resources and infrastructure.

In conclusion, the distribution of data papers, funding trends, and the structure of the collaboration network provide an important source for understanding trends, and funding distribution in scientific research. These findings provide important guidance for research policy makers and funders to effectively allocate research resources, and encourage multidisciplinary collaboration. It is also important for researchers to apply for appropriate calls for support, and to increase their chances of receiving support, in order to realize their work.

Acknowledgments

We sincerely like to thank Professor Chaomei Chen, whom we consulted during our study, and who patiently answered our questions.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Yurdagül Ünal  <https://orcid.org/0000-0002-6519-9845>
Müge Akbulut  <https://orcid.org/0000-0003-0026-6485>

Notes

1. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>
2. The agencies focused on were: National Natural Science Foundation of China (NSFC), US National Science Foundation (NSF), German Research Foundation, Netherlands Organization for Scientific Research (NWO), National Research Foundation of South Africa (NRF), and National Council for Scientific and Technological Development of Brazil (CNPq).
3. In Schöpfel et al.'s (2021) study, the subject distribution of journals was based on WoS subject categories, whereas in Candela et al.'s (2015) study, Scopus subject categories were used for more detailed classification.
4. All codes and data sets prepared for the study can be accessed at <https://github.com/mugeakbulut/datapapers>
5. Records which were mentioned once or twice did not affect the analysis results, since they were in the queue. The majority of these matches were also manually checked.
6. As of March 2025, there were 17,412 data papers in WoS. The number of data papers gradually increased every year from 2006 to the present date with the exception of the pandemic period 2021/2022.
7. While analyzing the findings regarding the categories, it should be kept in mind that a publication may belong to more than one category.
8. Since *U.S. Department of Health and Human Services, National Institutes of Health* has different programs, FU fields are mentioned 687 times. Therefore, it can be said that an average of two or more *U.S. Department of Health and Human Services, National Institutes of Health* programs are mentioned in a data paper.
9. See the interactive box plot <https://...> where all fields are shown, where it can be seen that “51 Physical Sciences” received support from different funders.
10. Orange dots in the upper left corner.
11. There are 58 unique journals in the dataset. Table 1 shows the top 10 journals containing about 94% of the data papers.

References

- Akça S and Akbulut M (2023) Social network analysis of mythology field. *Library Hi Tech* 41(3): 903–920.
- Ali A and Nazim M (2025) Does funded research have a greater scholarly impact? A study of funded and non-funded research published in high-impact library and

- information science journals. *Global Knowledge, Memory and Communication* 74(3/4): 1295–1318.
- Alom Bartroli M (2023) North–south research funding dynamics of collaborative projects: Researchers’ appropriation strategies of agencies’ project frameworks. *Science and Public Policy* 50(5): 935–946.
- Alvarez-Bornstein B and Bordons M (2021) Is funding related to higher research impact? Exploring its relationship and the mediating role of collaboration in several disciplines. *Journal of Informetrics* 15(1): 101102.
- Ayoubi C, Pezzoni M and Visentin F (2019) The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? *Research Policy* 48(1): 84–97.
- Candela L, Castelli D, Manghi P, et al. (2015) Data journals: A survey. *Journal of the Association for Information Science and Technology* 66: 1747–1762.
- Chavan V and Penev L (2011) The data paper: A mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12: 1–12.
- Chen C (2004) Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences* 101(Supp. 1): 5303–5310.
- Chen C (2016) *CiteSpace: A Practical Guide for Mapping Scientific Literature*. Hauppauge, NY: Nova Science Publishers.
- Chen PY, Li K and Jiao C (2022) A preliminary analysis of geography of collaboration in data papers by S&T capacity Index. *Proceedings of the Association for Information Science and Technology* 59: 642–644.
- Crow G and Wiles R (2008) Managing anonymity and confidentiality in social research: The case of visual data in community research. NCRM Working Paper, ESRC National Centre for Research Methods. (Unpublished). Available at: eprints.ncrm.ac.uk/id/eprint/459/ (accessed 30 April 2025).
- European Commission (n.d.) Joint Programming Initiatives. Available at: research-and-innovation.ec.europa.eu/research-area/health/joint-programming-initiatives_en (accessed 30 April 2025).
- Fu J, Tian L, Zhang C, et al. (2023) Opening research data contributes to the citations of related research articles: Evidence from Data in Brief. *Learned Publishing* 36: 426–438.
- Ghermandi A, Langemeyer J, Van Berkel D, et al. (2023) Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use. *One Earth* 6: 236–250.
- Gök A, Rigby J and Shapira P (2016) The impact of research funding on scientific outputs: Evidence from six smaller European countries. *Journal of the Association for Information Science and Technology* 67: 715–730.
- Huang MH and Huang MJ (2018) An analysis of global research funding from subject field and funding agencies perspectives in the G9 countries. *Scientometrics* 115: 833–847.
- Lammy R (2020) Funder registry – Crossref. Available at: www.crossref.org/services/funder-registry/ (accessed 20 May 2023).
- Le Quéré C, Moriarty R, Andrew RM, et al. (2015) Global carbon budget 2015. *Earth System Science Data* 7(2): 349–396.
- Li K and Jiao C (2022) The data paper as a sociolinguistic epistemic object: A content analysis on the rhetorical moves used in data paper abstracts. *Journal of the Association for Information Science and Technology* 73(6): 834–846.
- Liu WS (2020) Accuracy of funding information in Scopus: A comparative case study. *Scientometrics* 124(1): 803–811.
- Liu WS, Tang L and Hu GY (2020) Funding information in Web of Science: An updated overview. *Scientometrics* 122(3): 1509–1524.
- Mannheimer S, Pienta A, Kirilova D, et al. (2018) Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist* 63(5): 43.
- McGillivray B, Marongiu P, Pedrazzini N, et al. (2022) Deep impact: A study on the impact of data papers and datasets in the humanities and social sciences. *Publications* 10(4): 39.
- Miao L, Larivière V, Wang F, et al. (2023) Cooperation and interdependence in global science funding. *Papers* 2308.08630, arXiv.org, Revised Feb 2024. Available at: ideas.repec.org/p/arx/papers/2308.08630.html (accessed 30 April 2025).
- Mongeon P and Paul-Hus A (2016) The journal coverage of web of science and Scopus: A comparative analysis. *Scientometrics* 106(1): 213–228.
- Morris N and Rip A (2006) Scientists’ coping strategies in an evolving research system: The case of life scientists in the UK. *Science & Public Policy* 33: 253–263.
- Roshani S, Bagheryllooie MR, Mosleh M, et al. (2021) What is the relationship between research funding and citation-based performance? A comparative analysis between critical disciplines. *Scientometrics* 126: 7859–7874.
- Rusu VD, Mocanu M and Bibiri AD (2022) Determining factors of participation and success rates in research funding competitions: Case study. *PLoS ONE* 17(7): e0272292.
- Schmidt B, Gemeinholzer B and Treloar A (2016) Open data in global environmental research: The Belmont forum’s open data survey. *PLoS ONE* 11(1): e0146695.

- Schöpfel J, Farace D, Prost H, et al. (2019) Data papers as a new form of knowledge organization in the field of research data. *Knowledge Organization* 46(8): 622–638.
- Schöpfel J, Farace DJ, Prost H, et al. (2021) Data documents. *Knowledge Organization* 48(4): 307–328.
- Schwartz-Shea P (2019) Under threat? Methodological pluralism in public administration. *Public Performance & Management Review* 44(5): 75.
- Shove E (2003) Principals, agents and research programmes. *Science & Public Policy* 30: 371–381.
- Thelwall M (2020) Data in brief: Can a mega-journal for data be useful? *Scientometrics* 124: 697–709.
- Unal Y, Chowdhury G, Kurbanoglu S, et al. (2019) Research data management and data sharing behaviour of university researchers. *Information Research* 24(1): paper isic1818.
- Verma MK, Khan D and Yuvaraj M (2023) Scientometric assessment of funded scientometrics and bibliometrics research (2011–2021). *Scientometrics* 128: 4305–4320.
- Walters WH (2020) Data journals: Incentivizing data access and documentation within the scholarly communication system. *Insights: The UKSG Journal* 33(1): 18.
- Wang J and Shapira P (2015) Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers. *PLoS ONE* 10(2): e0117727.
- Wang X, Liu D, Ding K, et al. (2012) Science funding and research output: A study on 10 countries. *Scientometrics* 91(2): 591–599.
- Wu Y and Duan Z (2015) Social network analysis of international scientific collaboration on psychiatry research. *International Journal of Mental Health Systems* 9(1): 1–10.
- Zhao SX, Lou W, Tan AM, et al. (2018) Do funded papers attract more usage? *Scientometrics* 115(1): 153–168.
- Zhou P, Cai X and Lyu X (2020) An in-depth analysis of government funding and international collaboration in scientific research. *Scientometrics* 125: 1331–1347. DOI: 10.1007/s11192-020-03595-2.

About the authors

Yurdagül Ünal is an associate professor at the Department of Information Management at Hacettepe University. She received her PhD from the Department of Information Management at Hacettepe University, Turkey. She also held a visiting scholar position at the University of North Carolina at Chapel Hill. Her major areas of teaching and research include bibliometrics, collection development, research data management, and cultural heritage. Dr Ünal has actively participated in various projects supported by the European Commission and TUBITAK (Scientific and Technological Research Council of Turkey). Some notable EU projects she has been involved in include “Local content in a Europeana cloud (LoCloud),” “AccessIT: Accelerating the circulation of culture through the exchange of skills in information technology,” “Re-designing and co-creating innovative cultural heritage services through libraries (iTSELF),” and “Enhancing Key Civic Competences for the Post-truth Era: News Literacy and Critical Thinking (ESSENTIAL)”.

Müge Akbulut is a Faculty Member at the Department of Information Management at the University of Ankara Yıldırım Beyazıt in Ankara, Turkey. She received her PhD from the Department of Information Management at Hacettepe University, Turkey. She carries out research on information retrieval, bibliometrics and research methods. Ms. Akbulut received the certificates of Google Qualified Developer Programme, Software and Algorithms (from Microsoft Virtual Academy), Computer Operating (from Ministry of National Education); and she worked in the department of software at Informascope, Mikro Bilgi and ANKAREF in the past. Ms. Akbulut is the member of International Society for Scientometrics and Informetrics (member no: 00296), the Informatics Association of Turkey (member no: 99008214) and University and Research Librarians Association.