*TASK QUARTERLY vol. 25, No 4, 2021, pp. 433–471*

# DATA JOURNALS AND DATA PAPERS IN VARIOUS RESEARCH AREAS AND SCIENTIFIC DISCIPLINES – BIBLIOMETRIC ANALYSIS BASED ON INCITES

BEATA ADAMCZAK

*Gdańsk University of Technology, Library*
*11/12 Gabriela Narutowicza*
*80-233 Gdańsk, Poland*
beata.adamczak@pg.edu.pl

(received: 3.12.2021; revised: 8.12.2021;
accepted: 16.12.2021; published online: 30.12.2021)

**Abstract:** The main aim of this work is to provide insight into a bibliometric analysis of Data Journals and Data Papers in terms of research areas, disciplines, publication year and country. In particular, we calculated many bibliometric indicators, especially: the number of publications and citations. Furthermore, this work also investigated the top 20 journals in which scientists published the largest number of Data Papers. It was found that the first Data Paper in Medical and Health Science was published in 2006. Our results indicate that about 64% of Data Papers are cited and about 30% of Data Papers are articles that have international co-authors. We also found that a significant number of Data Papers were published in top quartile (Q1) Open Access journals. Furthermore, most Data Papers were published in Multidisciplinary Sciences, followed by Biological Sciences, Earth and related Environmental Sciences as well as Computer and Information Sciences. Our analyses show that the largest number of Data Papers were published in journals with an Impact Factor ranging from 5 to 6. Additionally, the largest number of Data Papers were published in the USA, China, the United Kingdom, Italy and France. From the point of view of scientists, this work is needed because scientists can check information about journals in which they may publish research data. Moreover, they can find Data Journals with a high Impact Factor and MEiN points. This analysis is also interesting because scientists/publishers can compare the disciplines and countries in which researchers have published the greatest number of Data Papers. Finally, it is worth noting that this work gives insights into the various bibliometric indicators related to citations and the number of publications. This information can be useful for people associated with bibliometric analyses.

**Keywords:** data journals, data papers, research areas, bibliometric analysis

**DOI:** https://doi.org/10.34808/tq2021/25.4/p

## 1. Introduction

Data Journals are journals that publish articles referred to as Data Papers, which are about datasets or quote datasets. Data Journals are also defined as ordinary journals that publish original articles with research data (taken from research data repositories or created/generated by the authors on their own). Most journals in which authors can publish articles with research data have a standard format, such as: Introduction, Methods, Results, Discussion and Conclusions. These articles contain research data, especially in the main text of the publications or in the Supplementary Materials. However, it should be highlighted that there are journals in which scientists can publish articles about research data only. These journals (so-called pure journals) do not contain the standard sections: Methods, Results and Discussion. These journals contain only detailed descriptions of research data (i.e. title of dataset, abstract, authors, keywords, DOI, references) as well as information about the used license, access data, language and the data [1–3]. A fundamental example of such a journal is Data in Brief.

It needs to be emphasised that journals in which authors can publish articles with research data are divided into two criteria [1]:

A) Type of journals:

- pure – journals in which authors can publish research data only, such as: Dataset Papers in Science, Earth System Science Data, Geoscience Data Journal, Scientific Data, Journal of Open Archaeology Data, Journal of Open Health Data, Journal of Open Humanities Data, Journal of Open Psychology Data;
- mixed – journals in which authors publish articles with research data (they can add research data, for example in the Supplement Materials)

B) Subject of journals:

- multidisciplinary sciences – Data in Brief, Scientific Data.
- detailed scientific disciplines – journals in which authors can publish articles within selected research areas/scientific disciplines, for example: BMC Genetics, Neuroinformatics, Human Genomics.

It is worth keeping in mind that Data Papers are one of the main types of articles that are specific for journals in which scientists can publish articles with research data. It is also worth mentioning that certain types of publications are characteristics of specific journals or publishers. It is worth noting that these information can be found on the journal pages or directly in the article. For example:

- Data note – descriptions of biomedical data; journals such as: BMC Research Notes, F1000 Research, BMC Genetics, EvoDevo
- Dataset paper: Dataset Papers in Science
- Original data article: Neuroinformatics

- Data reports – journals such as: Frontiers in Bioengineering and Biotechnology, Acta Crystallographica Section E, Human Genome Variation, Aerosol and Air Quality Research
- Genome database: Human Genomics
- Database article – descriptions of research databases; usually journals published by Biomed Central [1].

It is well known that the leading science funding organisations (i.e. National Institutes of Health, Wellcome Trust) indicate that data publishing has become increasingly important [3]. Some publishers (i.e. PLOS, BioMed Central) have data-sharing policies [4, 5] providing detailed information about sharing and access to scientific data that can be published in their Data Journals. However, it is worth noting that these sharing policies vary between Data Journals and publishers. Some works provide insights into the fundamental principles associated with sharing and analysing research data. For example, Penev et al. [3] present guidelines that contain a lot of information about research data, especially: why publish data, how to publish data, and how to cite data. Other scientists [6, 7] present a comprehensive description of Data Papers in terms of various categories: data analysis, data classification, data collection, data formatting, data identification, data sharing, data modification, data validation, data visualisation. Moreover, Jiao and Darch [4] compare two journals (Earth System Science Data, Journal of Physical and Chemical Reference Data) indicating that Data Journals representing different scientific disciplines very often have various principles regarding to sharing and reuse of research data. It should be highlighted that research data published in journals should be compatible with the FAIR rules (findable, accessible, interoperable, reusable) [3] and standard academic practices [1].

Some previous works report that bibliometric analysis of Data Journals has been studied by scientists in recent years [1, 2, 8, 9]. In particular, Candela et al. [1] show a bibliometric analysis of 116 Data Journals published by 15 various publishers. They perform a detailed overview of the selected journals in terms of some indicators: nature of papers (pure/mix), length of papers, indexed by databases, open access nature, subject classification. They also show how Data Journals describe datasets, promote dataset availability and open access, guarantee dataset quality, and support dataset citation. Seo and Kim [8] investigated nine Data Journals indexed in Web of Science (Data in Brief, Scientific Data, Human Genome Variation, Earth System Science Data, Geoscience Data Journals, Journal of Open Archaeology Data, Data, Gigascience, Biodiversity Data Journal), especially in terms of data and metadata quality. They checked some criteria: data availability, usefulness of data, identifiers of data, acceptable data formats, originality of science, consistency, validated data, integrity of data, accuracy of data descriptions, open licenses. Walters [9] and Penev et al. [10] provide insights into a detailed overview of 13 Data Journals in various scientific disciplines, while Schopfel [11] presents a description of Data Journals, indicating five

major aspects of Data Papers: relevance to specific disciplines, business models (gold Open Access (OA) with Article Processing Charges (APC), gold OA without APC, hybrid, subscription), indexing in databases and repositories, links between academic publishing and the production/dissemination of Data Papers as well as evaluation (key factors for future development, especially: publishers, licensing, selection, structure and length, metadata and identifiers).

Our study performs a bibliometric analysis of Data Journals and Data Papers that were published in the years 1980–2020. At this point, it is worth noting that InCites allow us to analyse articles published in this period of time. In particular, we focussed our research on the quantity and quality of bibliometric indicators, especially: the number of Web of Science documents, the percentage of Open Access documents, the percentage of documents in Q1 journals, Category Normalized Citation Impact (CNCI), the percentage of documents cited, the percentage of documents in the Top 10%. Moreover, for all countries around the world, we determine the publication research output and indicators related to citations as a function of the publication year. We also show an analysis of Data Journals and Data Papers in various Organisation for Economic Co-operation and Development (OECD) research areas and disciplines. Furthermore, this work also presents the top 20 journals in which scientists from all countries published the largest number of Data Papers. Finally, we assess the correlation between specific bibliometric indicators.

## 2. Research Methods

To perform the bibliometric analysis of the Data Journals and Data Papers, we used InCites [12, 13], which is the most frequently used analytical tool in this kind of analysis. Bibliometric analyses were carried out for 15,134 articles identified as Data Papers and published by scientists from 154 countries in the years 1980–2020. We analysed 121 journals and 33 publishers. In this place, it is worth mentioning that we identified articles as Data Papers based on the InCites filters associated with the document type. The data was collected from the database on 10th March 2021. We calculated many bibliometric indicators, including: the number of Web of Science documents, the number of Open Access documents, the number of citations [14], the number of highly cited Data Papers [15], the percentage of documents in the Top 10% (the percentage of publications in the Top 10% based on citations by category, year and document type), Category Normalized Citation Impact (CNCI, average number of citations per paper normalised for the subject, year and document type), the percentage of Data Papers cited at least once, the percentage of Data Papers with international co-authors, the percentage of Data Papers where the first/corresponding author is affiliated with the selected countries. The values of particular bibliometric indicators were shown based on the country and the publication year as well as the main OECD research areas and OECD scientific disciplines, Impact Factor, and quartile. It is worth noting that the percentage of documents in Q1 - Q4 journals

are defined as the percentage of documents that appear in a journal in a particular Journal Impact Factor quartile in a given year. For example, the percentage of documents in Q1 journals is the number of documents in Q1 journals divided by the number of documents in IF journals [12]. We also calculated the number of Data Journals in which scientists can publish Data Papers as a function of the publisher, Impact Factor, Ministry of Science and Higher Education (MEiN) points, the main OECD research area, and scientific discipline. In this place, it is worth noting that MEiN points are points which journals receive by the Ministry of Science and Higher Education in Poland. It is worth adding that we ascribed the relevant Journal Citation Reports (JCR) categories, fields of science, and scientific disciplines to every journal in accordance with the OECD classification. Furthermore, it should be added that Figure 13 and Figure 14 show only Data Papers that were published in Data Journals with IF.

## 3. Results and Conclusions

To determine the publishing houses with the greatest number of journals in which authors can publish articles with research data (so-called Data Papers), we first classified 121 Data Journals according to their publishers. In Figure 1, we present the publishing houses with at least 4 journal titles in which authors can publish articles identified as Data Papers.

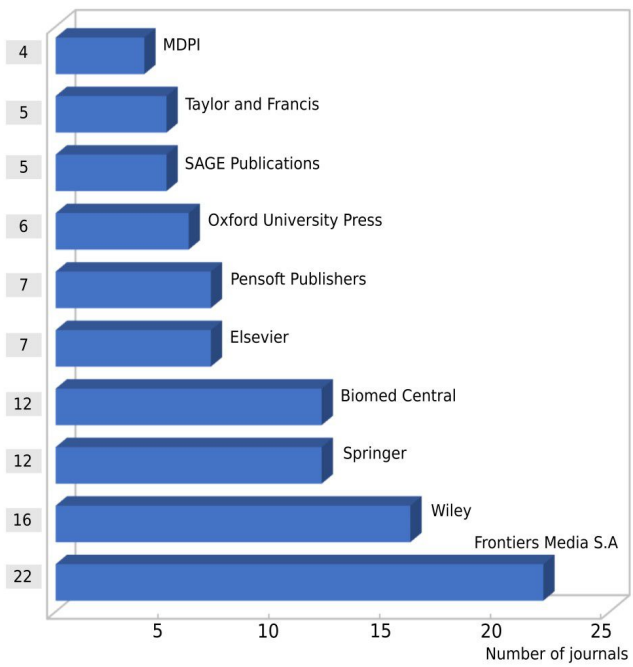


**Figure 1.** The number of Data Journals in which authors can publish Data Papers by publisher

Figure 1 shows that Frontiers Media S.A has as many as 22 Data Journals in which authors can publish Data Papers. Other publishers have fewer such journals, for example: Wiley – 16 journals, Springer and Biomed Central – 12 journals each, Elsevier and Pensoft Publishers – 7 journals each, Oxford University Press – 6 journals, Sage Publications and Taylor and Francis – 5 journals each, and MDPI – 4 journals. It is also worth highlighting that the largest number of Data Journals in which authors can publish Data Papers are Open Access journals, while the remaining ones are hybrid journals. Articles in the abovementioned Open Access Data Journals are usually published under the CC-BY 4.0 license. In this place, it is worth adding that hybrid journals are defined as non-gold Open Access journals which publish hybrid and subscription content. In particular, it should be noted that in these journals, scientists can pay an article publication charge and publish their article as Gold Open Access Data Paper with an open licence (Creative Commons) [12]. Furthermore, it is worth noting that 14480 Data Papers, which constitutes 96% of all the analysed Data Papers are Open Access articles.

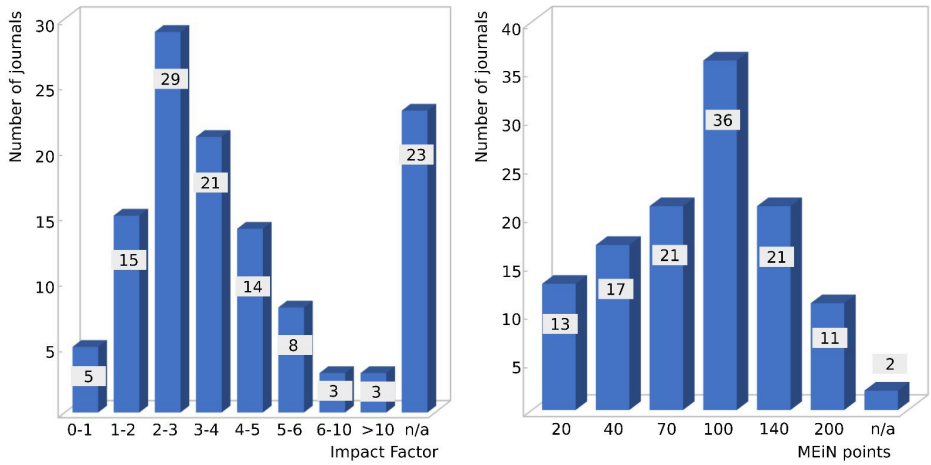


**Figure 2.** The number of Data Journals in which authors can publish Data Papers by their Impact Factor and the number of points awarded by MEiN

From the scientist's point of view, it is interesting to know whether Data Journals in which authors can publish Data Papers have a high Impact Factor and/or points awarded by MEiN. Therefore, in Figure 2 we present the number of Data Journals and the corresponding IF values and points awarded by MEiN. Based on this Figure, it can be observed that the largest number of Data Journals (N=29) in which authors can publish Data Papers have an Impact Factor of 2–3. Moreover, it is worth noting that 21 journals have an IF ranging from 3 to 4, 15 journals have an IF from 1 to 2, 14 journals have an IF from 4 to 5, 8 journals have an IF from 5 to 6 and 5 journals have an IF from 0 to 1. Interestingly, 3 journals in which authors can publish Data Papers have a very high IF, in particular: Chemical Reviews (IF=52.758), Nature Biotechnology (IF=36.558), and Blood

(IF=17.790). Figure 2 shows also that as many as 36 journals were awarded 100 points by MEiN, 21 journals each awarded 70 points and 140 points each, 17 journals were awarded 40 points, and 13 journals were awarded 20 points. It is worth noting that 11 journals in which authors can publish Data Papers were awarded as many as 200 points by MEiN. It also needs to be emphasised that 23 Data Journals (19% of all the analysed journals) do not have an IF and 2 Data Journals (1.7% of all analysed journals) were not awarded any points by MEiN.

It would be interesting to determine which publishing houses have the largest number of Data Journals with a high IF. Figure 3 presents the dependence between the number of Data Journals and the IF and points awarded by MEiN for the 6 publishing houses with the greatest number of journals in which authors can publish Data Papers. As further revealed by Figure 3, Biomed Central has 5 journals with an IF ranging from 3 to 4, and 4 journals have an IF from 2 to 3. It is also worth noting that Biomed Central has one journal with a smaller IF (BMC Musculoskeletal Disorders, IF=1.879). Interestingly, Biomed Central also has one journal (Journal of Cheminformatics) with a very high Impact Factor (IF = 5.326). The majority of journals published in Biomed Central were awarded 100 points by MEiN. Elsevier and Wiley have the greatest number of journals with 100 points awarded by MEiN. Moreover, it is worth emphasising that Elsevier has one journal in which authors can publish Data Papers that has a very high IF, namely Genomics Proteomics Bioinformatics (IF=7.051). Wiley also has one journal with an IF > 6.000. It is Global Ecology and Biogeography (IF=6.446). According to Figure 3, Frontiers Media has the greatest number of Data Journals with 100 points awarded by MEiN and an IF ranging from 3 to 4. In contrast, Springer has the most Data Journals that have 40 points awarded by MEiN and an IF ranging from 2 to 3. Pensoft Publishers have 6 journals that have an IF ranging from 1 to 2 and one journal with a slightly higher Impact Factor (Neobiota, IF=2.643). The greatest number of Data Journals (Comparative Cytogenetics, Zookeys, Nature Conservation-Bulgaria) published by Pensoft Publishers have 70 points awarded by MEiN. According to Figure S1, it is worth noting that for journals which have IF>10, 200 MEiN points were assigned them. On the other hand, for other journals (IF<10), we do not observe any correlation between IF and MEiN points.

To analyse the Data Journals in which authors can publish Data Papers more closely, we carried out bibliometric analyses with regard to specific fields of science in accordance with the OECD classification (Natural Sciences, Engineering and Technology, Medical and Health Sciences, Agricultural Sciences, Social Sciences, Humanities) and the disciplines of science corresponding to them.

By comparing the number of Data Journals in the different OECD research areas, based on Figure 4, it can be noted that the majority of Data Journals in which authors can publish Data Papers are devoted to Natural Sciences (69 journals, which constitutes 48% of all analysed journals), followed by Medical and Health Sciences (30 journals, which constitutes 21% of all analysed journals), Engineering and Technology (17 journals, which constitutes 12% of all analysed

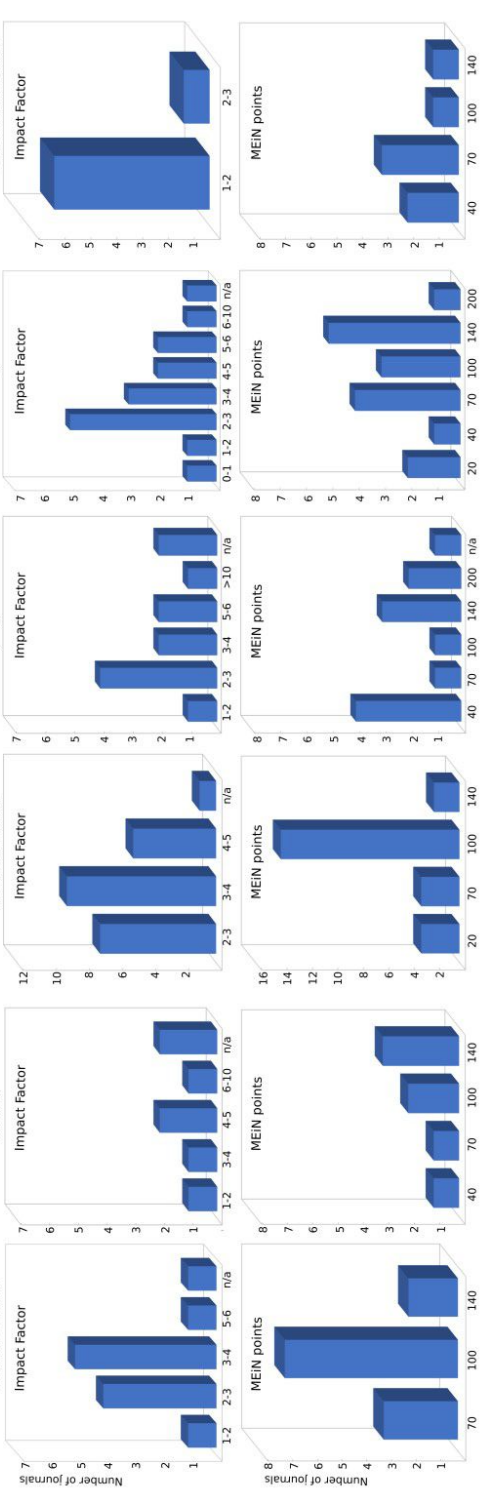


**Figure 3.** The number of Data Journals in which authors can publish Data Papers by the number of points awarded by MEiN and the Impact Factor of 6 publishing houses (Biomed Central, Elsevier, Frontiers Media, Springer, Wiley, Pensoft Publishers) that have the largest number of Data Journals

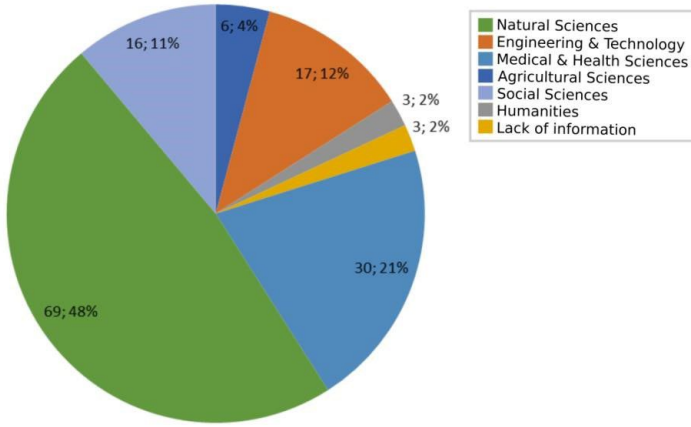


**Figure 4.** The number of Data Journals in which authors can publish Data Papers in relation to the main OECD research areas

journals), Social Sciences (16 journals, which constitutes 11% of all analysed journals), Agricultural Sciences (6 journals, which constitutes 4% of all analysed journals), and Humanities (3 journals, which constitutes 2% of all analysed journals).

In order to obtain a more detailed analysis, in Figure 5 we present the number of Data Journals in relation to the main OECD fields of science and the corresponding scientific disciplines. Based on this Figure, it can be seen that the majority of Data Journals are devoted to Biological Sciences (N=43), whereas significantly fewer are devoted to Earth and related Environmental Sciences (N=15), Basic Medical Research (N=11), Clinical Medicine (N=9), Health Sciences (N=8), and Computer and Information Sciences (N=8). In other fields of science, we observe a significantly smaller number of Data Journals in which authors can publish Data Papers. As revealed by an analysis of Figure 5, it can be seen that although there are only 2 journals in Multidisciplinary Sciences (Data in Brief, Scientific Data), the greatest number of Data Papers (N=7,506) were published in these journals. It is worth noting that a large number of Data Papers were also published in Biological Sciences (N=833), Earth and related Environmental Sciences (N=449), Computer and Information Sciences (N=208), Clinical Medicine (N=77), Chemical Sciences (N=59), Basic Medical Research (N=36), and Environmental Biotechnology (N=33). These journals are very popular among scientists because researchers can publish articles devoted to scientific data within all research areas. Furthermore, it is worth noting that although some disciplines (Mechanical Engineering, Chemical Engineering, Medical Engineering, Industrial Biotechnology, Nanotechnology, Animal and Dairy Science, Veterinary Science, Agricultural Biotechnology, Law, Social and Economic Geography, Media and Communications) do not have specialist Data Journals, these articles appear also in two mentioned Data Journals (Data in Brief, Scientific Data). Importantly, it should be noted that some journals can be ascribed to more than one field/disci-

pline of science, as a result of which the total number of Data Journals presented in Figure 4 and Figure 5 is higher than the total number of Data Journals in which authors can publish Data Papers.

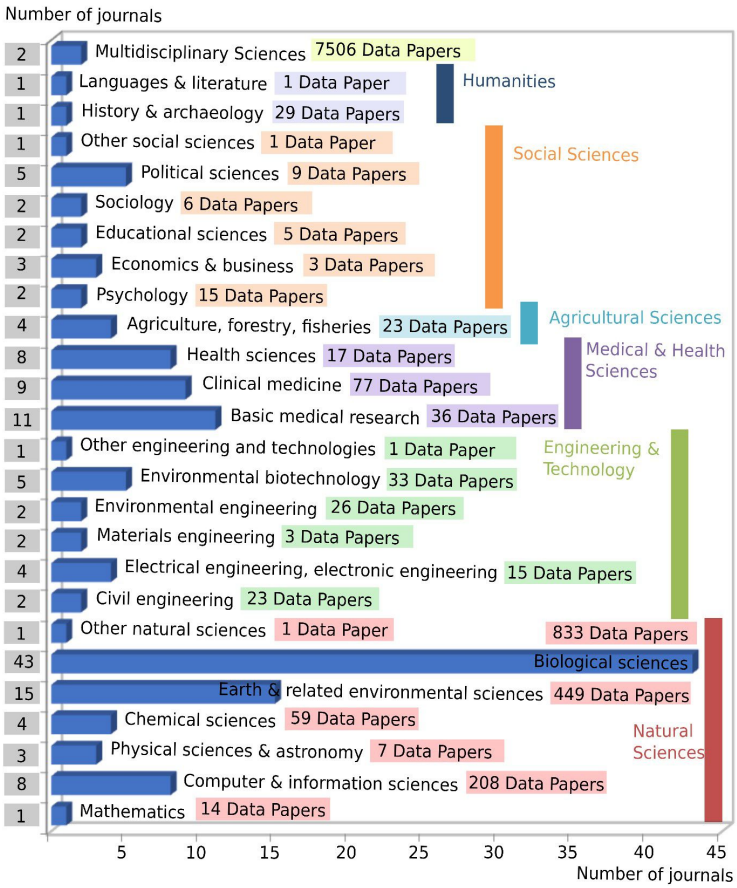


**Figure 5.** The number of Data Journals and the number of Data Papers published by scientists from all countries in 1980–2020 in relation to the main OECD research areas (Natural Sciences, Engineering and Technology, Medical and Health Sciences, Agricultural Sciences, Social Sciences, Humanities) and the corresponding OECD scientific disciplines

In order to determine the years and the OECD fields of science in which scientists published the greatest number of Data Papers, in Figure 6 we depict the dependence of the number of Data Papers as a function of the publication year for the main OECD research areas. It can be seen that the trend graphs show an increase in the number of Data Papers during the period from 2006 to 2020. As further revealed by Figure 6, the first Data Papers were written in the years 2006, 2008 and 2010. In 2011, scientists wrote 7 Data Papers and then 65 such articles in 2012, 123 articles in 2013, and 180 articles in 2014. Interestingly, a rapid increase in the number of Data Papers was observed in

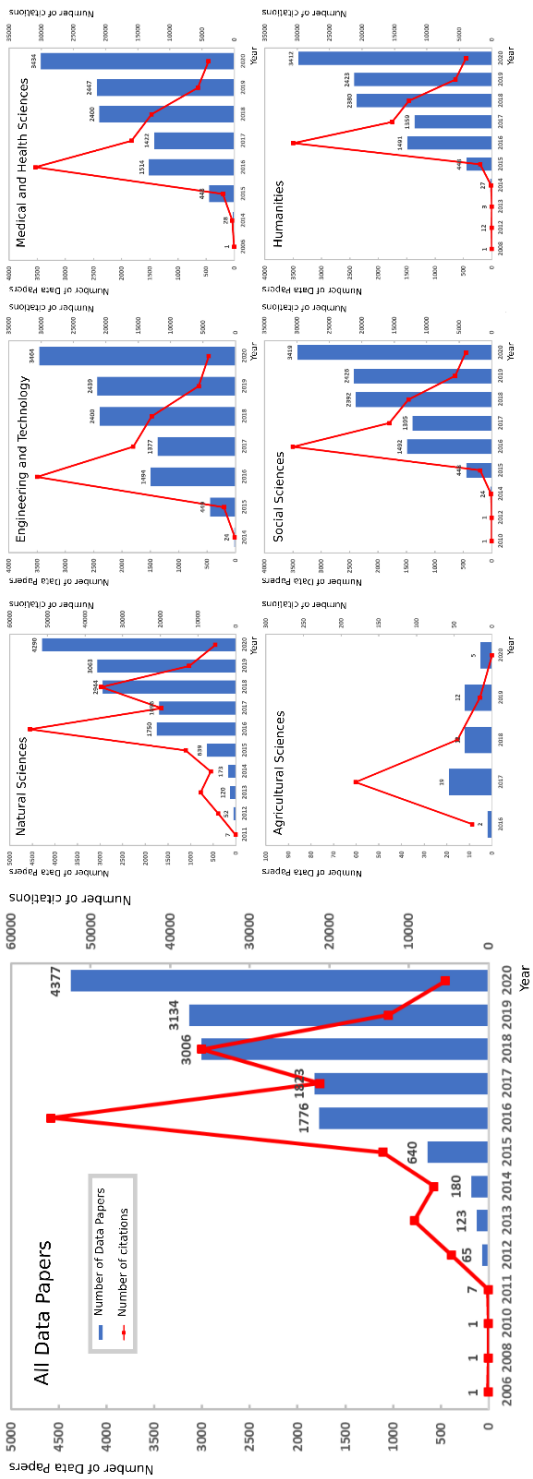


**Figure 6.** The total number of Data Papers published in 1980–2020 by scientists from all countries and the number of Data Papers published in the main OECD research areas (Natural Sciences, Engineering and Technology, Medical and Health Sciences, Agricultural Sciences, Social Sciences, Humanities) and the number of citations of these articles. Analyses were conducted using the InCites analytical tool

2015 (640 Data Papers). The trend of the number of Data Papers is upward, and in the years 2016–2017, scientists wrote almost 2000 Data Papers. However, the highest productivity was observed in the years 2018–2020, when scientists wrote about 3000 Data Papers. It is worth highlighting here that the greatest number of Data Papers are Open Access articles (Figure S2). We note that a significantly smaller number of Data Papers were published in hybrid journals or hybrid/transformative journals. Changes in the number of Data Papers as a function of publication year in particular research areas are shown in Figure 6. These results confirm that the first Data Paper was written in Medical and Health Sciences in 2006, followed by Humanities (2008), Social Sciences (2010), Natural Sciences (2011) as well as Engineering and Technology (2014). In particular, it is worth highlighting that the first article in Medical and Health Sciences is associated with dermatology, while the first article in Humanities is related to linguistics. Figure 6 shows that 7 articles were written in Natural Sciences in 2011, while in 2012 scientists wrote 52 Data Papers. Moreover, it is worth noting that most of the Data Papers were published in recent years, in particular in the years 2016–2017, when almost 2000 articles were written, in the years 2018–2019, when scientists wrote about 3000 articles and in 2020, when 4290 articles were written. In the field of Engineering and Technology, scientists wrote a significantly smaller number of Data Papers than in the field of Natural Sciences. However, it is worth noting that the first Data Papers in Engineering and Technology were written in 2014, whereas in 2020 scientists wrote 3464 such articles. Interestingly, even though in research areas such as Natural Sciences as well as Engineering and Technology, an increase in the number of publications can be observed over the years. In the field of Medical and Health Sciences, the largest number of Data Papers were written in 2020 (3434 articles). It should be noted that previous work reports that Data Papers have often been cited [4]. Therefore, as revealed by an analysis of the number of citations over the span of years (Figure 6), the most frequently cited Data Papers are articles that were published in Natural Sciences, followed by Medical and Health Sciences, Engineering and Technology, Social Sciences and Humanities. The number of citations in Agricultural Sciences is significantly smaller. Moreover, it is worth highlighting that the maximum number of citations was observed in Natural Science in 2016 (an increase of about 41,390 citations in comparison to the previous year). On the other hand, the maximum number of citations was observed in Engineering and Technology as well as in Medical and Health Sciences also in 2016 (increase of about 29000 citations, in comparison to 2015). In Social Sciences and Humanities, the maximum number of citations was also observed in 2016, while in Agricultural Sciences, the maximum number of citations was observed in 2017.

In order to obtain a more detailed picture about the publishing houses in which scientists from all countries published the greatest number of Data Papers in the selected scientific OECD disciplines, Figure 7 presents the number of Data Papers in relation to the OECD disciplines for the 5 publishing houses in which

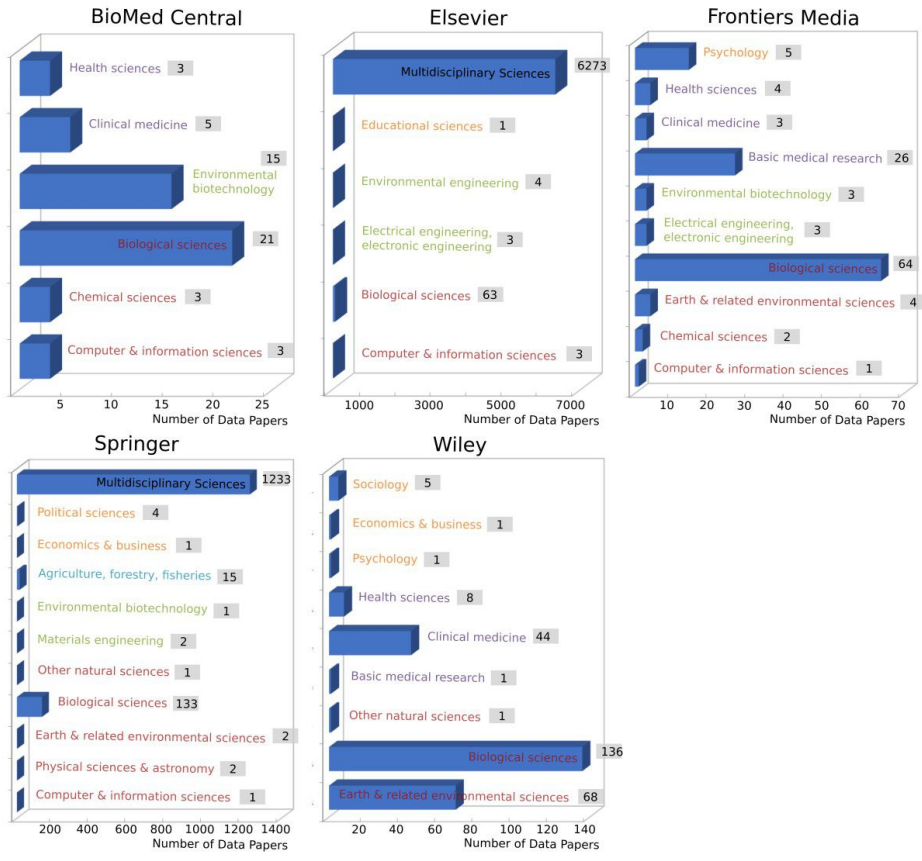


**Figure 7.** The number of Data Papers in relation to the OECD scientific disciplines for 5 publishing houses in which scientists published the greatest number of Data Papers (BioMed Central, Elsevier, Frontiers Media, Springer and Wiley)

scientists published the largest number of Data Papers. The results indicate that Biomed Central, Frontiers Media, and Wiley have the greatest number of Data Papers devoted to Biological Sciences. In Wiley, a significant number of Data Papers were also published in Earth and the related Environmental Science as well as Clinical Medicine. In contrast, Frontiers Media has also a significant number of Data Papers that were published in Basic Medical Research. It is also worth noting that Elsevier and Springer have the greatest number of Data Papers with Multidisciplinary Sciences (N=6,273 and N=1,233, respectively). Interestingly, it is worth highlighting that all Data Papers published in Pensoft Publishers are assigned to one OECD discipline (Biological Sciences). As we have mentioned earlier, it should be noted that the sum of Data Papers in particular scientific disciplines does not reflect the total number of articles in this discipline because one Data Paper can be published in a journal that is assigned to one or more WoS-defined journal categories.

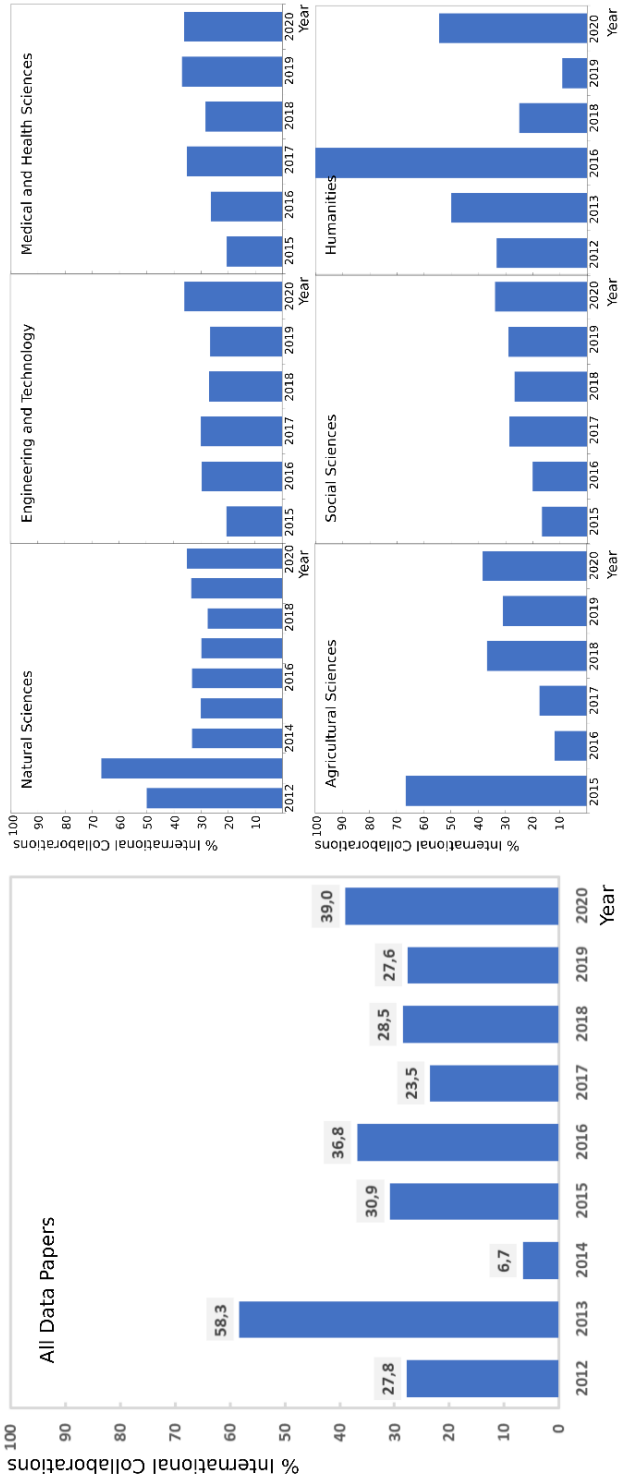


**Figure 8.** Percentage of Data Papers that have international co-authors by publication year and main OECD research area

It is well known that international collaboration can have an effect on the quality of the research publication output [16, 17]. In order to check which Data Papers were published by scientists from foreign countries, Figure 8 shows the percentage of Data Papers that have international co-authors as a function of the publication year and the main OECD research areas. In this place, it is worth adding that foreign countries means any other countries than the analyzed/selected country. Based on this Figure, it can be seen that about 30% of Data Papers are articles that have authors from at least two different countries. As is also evident from Figure 8, the distributions of the percentage of Data Papers that have international co-authors between various research areas are quite similar (with differences of a few years only). For example, in Natural Sciences in the years 2012–2013, we observed that approximately 50–65% of Data Papers were written by scientists from abroad. About 68% of Data Papers had international co-authors in Agricultural Sciences in 2015, while in the years 2016–2017, only 10–20% of Data Papers had authors from abroad. In Humanities, we observed that the percentage of Data Papers with international co-authors was less than 30% in the years 2018–2019, while in the years 2013, 2016 and 2020, these values were over 30%. However, it should be highlighted that scientists published a very low number of Data Papers in Humanities Sciences.

To reveal significant similarities and differences in the percentage of Data Papers that were published in Data Journals with particular quartiles (Q1, Q2, Q3 and Q4), we analysed trends in these indicators as a function of the publication year and the main OECD research areas, as shown in Figure 9. As is also evident from Figure 9, the largest number of Data Papers were published in Q1 Data Journals. Furthermore, we found that the percentage of Data Papers that were published in Q1 Data Journals in Natural Sciences decreased over 4 years from 2016 to 2019. A similar decrease in this indicator is also seen in other research areas, such as: Engineering and Technology, Medical and Health Sciences.

By comparing the publication research output in particular countries, we analysed the number of Data Papers that were published by scientists from all countries in the years 1980–2020 and the corresponding number of citations of these articles, as shown in Figure 10. The Figure shows that the largest number of Data Papers were published in the USA (N=2,153 articles). Scientists from other countries wrote significantly fewer Data Papers. For example, researchers from China wrote 933 Data Papers, which constitutes 43% of articles written by scientists from the USA. Scientists from Germany and the United Kingdom wrote 862 articles each, scientists from Italy wrote 697 articles, and scientists from France wrote 560 articles. As can be seen in Figure 10, a significant number of Data Papers (>100) were also published in the following European countries: Spain (N=471), Russia (N=351), the Netherlands (N=297), Switzerland (N=276), Sweden (N=236), Belgium (N=214), Denmark (N=184), Norway (N=172), Portugal (N=159), Austria (N=145), Finland (N=133), Poland (N=129), and Greece (N=115). A significant number of Data Papers were also published in countries in

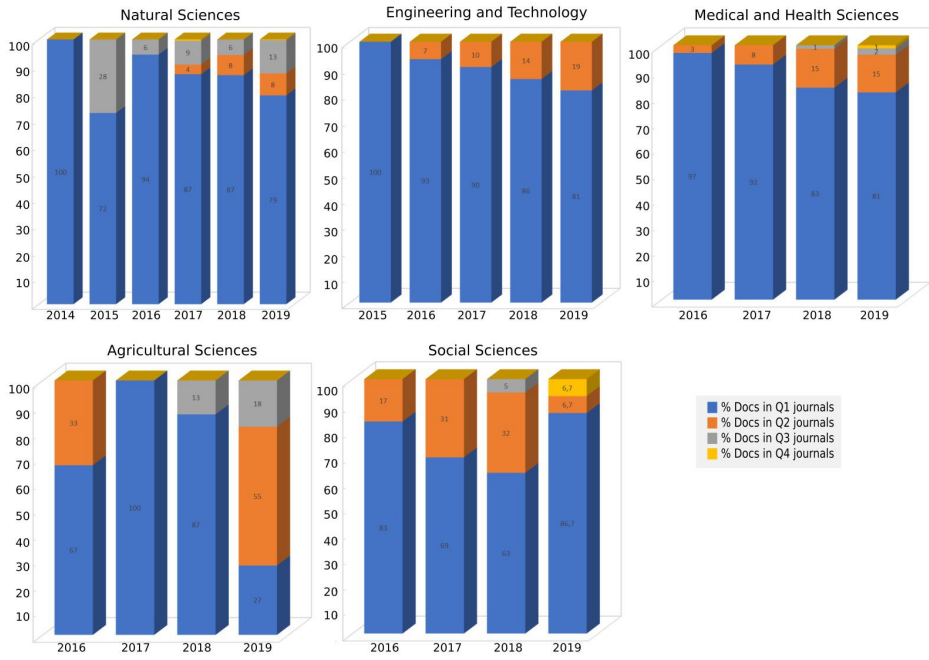


**Figure 9.** Percentage of Data Papers that were published in Data Journals with particular quartiles (Q1, Q2, Q3 and Q4) as a function of the publication year and research area

other continents: in Asia (Japan, India, Iran, South Korea, Indonesia, Malaysia), in America (Canada, Brazil, Colombia, Mexico) as well as in Australia, Africa (Nigeria) and New Zealand. It is also worth highlighting that Data Papers written by scientists from the USA are cited the most often (N=22,204 citations). The second place with regard to the largest number of cited Data Papers is occupied by Germany (N=12,546 citations), followed by the United Kingdom (N=12,066 citations), and Switzerland (N=8,773 citations). However, if we take into account the average number of citations per publication, the highest values of this indicator were observed for countries in which scientists published a smaller number of Data Papers, for example: Bolivia (4 articles, average number of citations per publication: 123), Bermuda (9 articles, average number of citations per publication: 50) and Oman (4 articles, average number of citations per publication: 27). It is worth noting that if we take into account only the countries in which scientists published more than 100 Data Papers, the greatest number of citations per publication was observed for Norway (172 articles, average number of citations per publication: 34), Switzerland (276 articles, average number of citations per publication: 32), Belgium (212 articles, average number of citations per publication: 29), the Netherlands (295 articles, average number of citations per publication: 28), New Zealand (103 articles, average number of citations per publication: 28), and Sweden (236 articles, average number of citations per publication: 23). On the other hand, in countries in which scientists published the largest number of Data

Papers, the average number of citations per publication was in the range from 7 to 15 (Italy – 7.09, China – 7.88, the USA – 10.31, France – 13.63, the UK – 13.92, Germany – 14.40). In order to establish the countries in which the average number of citations is higher than the average world values, Figure 11 shows the average values of the Category Normalized Citation Impact (CNCI) for all countries around the world. It should be clarified that the average global CNCI value is equal to 1.0. As is also evident from Figure 11, the highest CNCI values were observed for countries in which scientists published a lower number of Data Papers (2–4), in particular: Libya (CNCI=9.07), Bolivia (CNCI=5.41), and Malta (CNCI=4.58). In the case of countries in which scientists published the greatest number of Data Papers (for example: the USA, China, the UK, Germany, Italy, France), the CNCI values are in the range from 1.30 to 1.56, which means that the average number of citations per publication is slightly higher than the average global value. It is also worth highlighting that among countries in which scientists published a significant number of Data Papers (>100), the highest CNCI values were observed for Brazil (CNCI=1.69), Austria (CNCI=1.73), Finland (CNCI=1.77), Norway (CNCI=1.98), the Netherlands (CNCI=1.99), New Zealand (CNCI=2.01), Sweden (CNCI=2.07), Switzerland (CNCI=2.26), and Belgium (CNCI=2.42). Interestingly, it should be noted that although there are countries in which scientists published a significant number of Data Papers (128–555), the CNCI values are slightly lower than the average global values, especially: Japan (CNCI=0.97), Russia (CNCI=0.91), Nigeria and South Korea (CNCI=0.93), Malaysia (CNCI=0.76), Colombia (0.78), Portugal (0.84), Indonesia (0.82), and Mexico (0.76).

To obtain a more detailed picture of the role of scientists in creating/writing their Data Papers, in Figure 12 we show the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries. Figure 12 shows that for 128 countries (which constitutes 83% of all the analysed countries), the percentage of Data Papers that have the first author affiliated with the selected countries is slightly greater than or equal to the percentage of Data Papers that have the corresponding author affiliated with the selected countries. It is also worth adding that in countries in which scientists published a significant number of Data Papers (>100), the average values of these indicators are equal to about 64%. Interestingly, in the case of some countries (Iran, Nigeria, Indonesia, Colombia, Malaysia, India) it was observed that the percentage of Data Papers with the first/corresponding author affiliated with the selected countries was over 80%. On the other hand, in the case of some other countries (Norway, New Zealand, Denmark, the Netherlands, Switzerland, Austria, Sweden, Belgium) the average values of these indicators are below 50%. Additionally, in this place, it is worth highlighting that if the article has only one author, the author is either the first and the corresponding author. Interestingly, 339 Data Papers (2.2% of all analyzed articles) have 1 author only. Furthermore, it is worth mentioning that 100% of the percentage of first/corresponding authors means that in the selected country, all articles have the first or the corresponding author affiliated in the

mentioned country. Interestingly, it should be noted that authors from some countries published 1-2 Data Papers only, while authors from other countries published a significantly greater number of Data Papers.

Moreover, in order to determine the Data Journals with the highest Impact Factor, in Figure 13 we present the number of Data Papers in relation to the Impact Factor of particular Data Journals for the 5 countries in which scientists published the largest number of Data Papers: the USA, the United Kingdom, Germany, China, and Italy. This Figure shows that for all the analysed countries, the largest number of Data Papers were published in Data Journals with an Impact Factor of 5–6. Interestingly, the Data Journals in which scientists from the abovementioned countries published the largest number of Data Papers are Scientific Data (the USA – 518 articles, the United Kingdom – 228 articles, China – 201 articles, Germany – 185 articles, Italy – 102 articles) and Gigascience (China – 71 articles, the USA – 61 articles, the United Kingdom – 24 articles, Germany – 18 articles, Italy – 9 articles). Scientists from the abovementioned countries wrote slightly fewer Data Papers in Data Journals with a slightly higher IF, i.e. 6–10: Earth System Science Data (IF=9.197), Genomics Proteomics Bioinformatics (IF=7.051), and Global Ecology  Biogeography (IF=6.446). It is worth emphasising that Earth System Science Data published 135 articles from the USA, 124 articles from Germany, 89 articles from the United Kingdom, 46 articles from Italy, and 32 articles from China. Interestingly, scientists from the USA published one article in Chemical Reviews (IF=52.758) and one article in Nature Biotechnology (IF=36.558).

Moreover, the number of Data Papers as a function of the Impact Factor for all countries is shown in Figure S3. It should be added that if we take into account only Data Papers that were published in Data Journals with IF, the largest number of Data Papers were published in journals with an IF in the range of 5–6. It is also evident from Figure S3 that the greatest number of Data Papers were published by scientists from Europe, North America, and Asia. In Figure 14 and Figure S4, we compare the number of Data Papers in relation to the Impact Factor and the publication year. Figure 14 also presents the number of Data Papers in relation to the OECD research areas. Based on these Figures, it can be noted that the number of Data Papers increased every year. Moreover, it is worth noting that in the years 2012–2015/2016, scientists wrote Data Papers that were published in Data Journals with a specific IF. For example, in the years 2012–2014, the majority of Data Papers in Natural Sciences were published in Data Journals with an IF ranging from 6 to 10. On the other hand, in the initial years 2015–2016, the largest number of Data Papers in Engineering and Technology were published in Data Journals with IFs ranging 3–4 and 6–10. In Medical and Health Sciences, the first articles referred to as Data Papers were published in Data Journals with an IF ranging 5–6. Starting from 2017, we observed an increase in the number of Data Papers that were published in journals with a varying IF. Interestingly, in 2018 Data Journals in Engineering and Technology that published the greatest

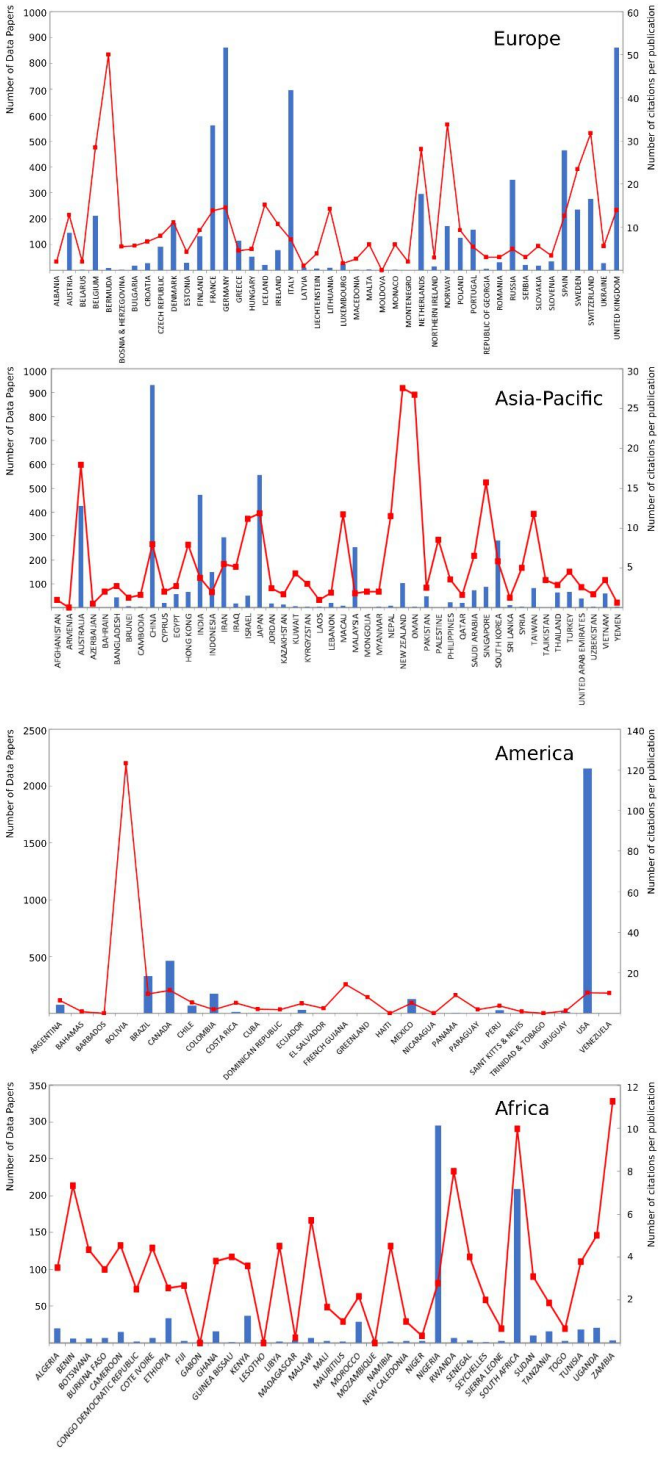


**Figure 10.** The number of articles indexed in the Web of Science and defined as Data Papers (1980–2020) and the corresponding number of citations of these articles. Results are shown for all countries

number of Data Papers had an IF of 3–4, whereas in 2017 Data Journals in Medical and Health Sciences that published the greatest number of Data Papers had an IF of 5–6.

To identify the top 20 Data Journals in terms of the number of publications, Table 1 shows the Data Journals in which scientists from all countries published the greatest number of Data Papers. In Table 1 we also analysed the values of other bibliometric indicators: the number of Data Papers (WoS), the number of citations, the percentage of documents cited at least once, the Category Normalized Citation Impact (CNCI), the Impact Factor, and the quartiles. For every journal, we ascribed the JCR categories and the corresponding OECD main research areas and OECD disciplines. Based on this Table, it can be seen that the Data Journals with Multidisciplinary Sciences (Data in Brief and Scientific Data) are those in which scientists published the significantly greatest number of Data Papers. It is also worth highlighting that the Data Journals on the Top 20 list have the high values of the Impact Factor and the percentage of documents cited at least once. The following Data Journals have the highest values of the Impact Factor: Earth System Science Data (IF=9.197), Genomics Proteomics and Bioinformatics (IF=7.051), GigaScience (IF=5.993), Scientific Data (IF=5.541), Circulation Cardiovascular Quality and Outcomes (IF=5.071), and Oncologist (IF=5.025). It is also worth noting that the Data Papers published in Data in Brief, Scientific Data, and Earth System Science Data were cited over 10,000 times. In order to obtain a more detailed picture about the Data Journals in which scientists from around the world published Data Papers, Figure S5 shows the name of the Data Journals and the corresponding number of Data Papers and the Impact Factor for the five countries (USA, China, Germany, United Kingdom, Italy) in which scientists published the greatest number of Data Papers. Based on this Figure, it can be noted that in the mentioned countries, the greatest number of Data Papers were published by scientists in the following Data Journals: Data in Brief, Scientific Data, and Earth System Science Data. The only exception is China, in which a slightly higher number of Data Papers were published in GigaScience than in Earth System Science Data. Additionally, in Table S1, we analyse the countries in which scientists published the greatest number of Data Papers in 4 top Data Journals. We do not observe any correlation between the publication source country and affiliation of authors. Furthermore, it is worth noting that although the publication source country of Data in Brief are the Netherlands, the greatest number of Data Papers in this journal were published by scientists from USA (1117 articles), China (484 articles) and Italy (441 articles). Similar results were observed for Scientific Data (England - publication source country) and Data (Switzerland - publication source country). In the case of Scientific Data, the greatest number of Data Papers were published by scientists from the USA (552 articles), the United Kingdom (246 articles) and China (210 articles), while in the case of Data, the greatest number of Data Papers were published by scientists from the USA (1896 articles), the United Kingdom (786

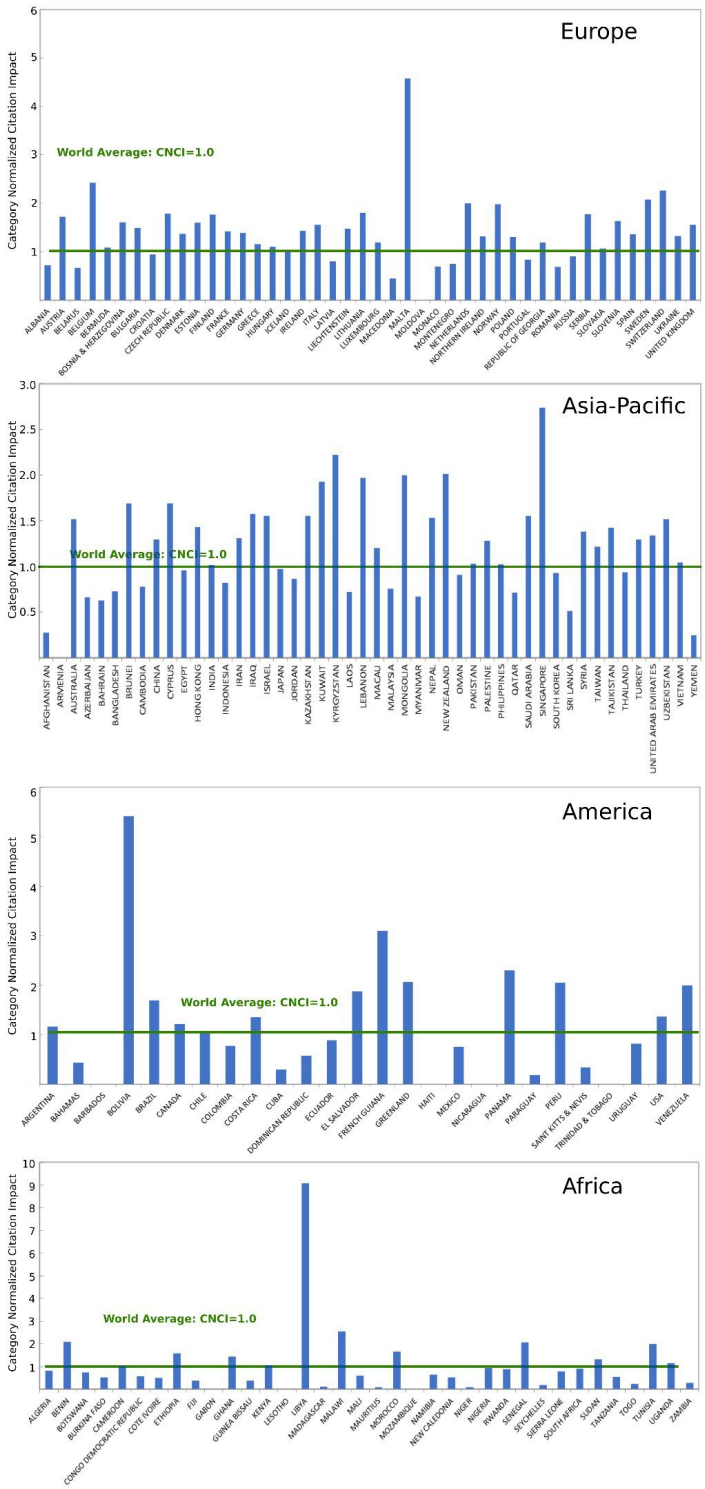


**Figure 11.** Average values of Category Normalized Citation Impact (CNCI) for all countries. Results were calculated for Data Papers published in the years 1980–2020

articles) and Germany (782 articles). Switzerland takes the 17th place in this ranking. Interestingly, Germany is the publication source country in Earth System Science Data and takes the second place in the ranking of the most popular countries which published in this journal.

Finally, to understand the relationship between particular bibliometric indicators, in Figure 15 we analysed the correlation between the percentage of Data Papers that have international co-authors as a function of some bibliometric indicators, especially: CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, the percentage of Data Papers cited at least once, the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries, the percentage of Data Papers in the Top 10%, and the percentage of Data Papers in Q1 Data Journals. Figure 16 and Figure S6 present the dependence of selected bibliometric indicators as a function of the percentage of Data Papers that have the first/corresponding author affiliated with the mentioned countries, respectively. Further, Figure S7 reflects the dependence of the selected bibliometric indicators as a function of the percentage of Data Papers in Q1 Data Journals. The results are shown for the countries in which scientists published at least 100 Data Papers.

As further revealed by Figure 15, there is a strong linear correlation between the selected bibliometric indicators (R-squared value on the charts > 50% in the case of most relations). We found that if the percentage of the Data Papers that have international co-authors was higher, the values of particular indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, and the percentage of Data Papers in the Top 10%) were also higher. It is worth highlighting that there are very strong correlations between the percentage of Data Papers that have international co-authors and the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries (R-squared value on the charts > 89– 92%). In particular, it can be seen that if the percentage of Data Papers that have international co-authors is higher, the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries is lower. According to Figure 16, we also observed that if the percentage of Data Papers that have international co-authors was higher, the values of certain indicators (percentage of Data Papers in Q1 Data Journals, and the percentage of Data Papers cited at least once) were also higher. However, it should be mentioned here that the correlation between these indicators is weak (R-squared value on the charts=30% and 18%, respectively). In contrast, Figure 16 and Figure S6 show a significant correlation (R-squared value on charts=59–78%) between the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries and other bibliometric indicators. In particular, Figure 16 and Figure S6 indicate that if the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries is higher, the values of particular indicators (CNCI, the average number of citations per publication, the percentage of highly

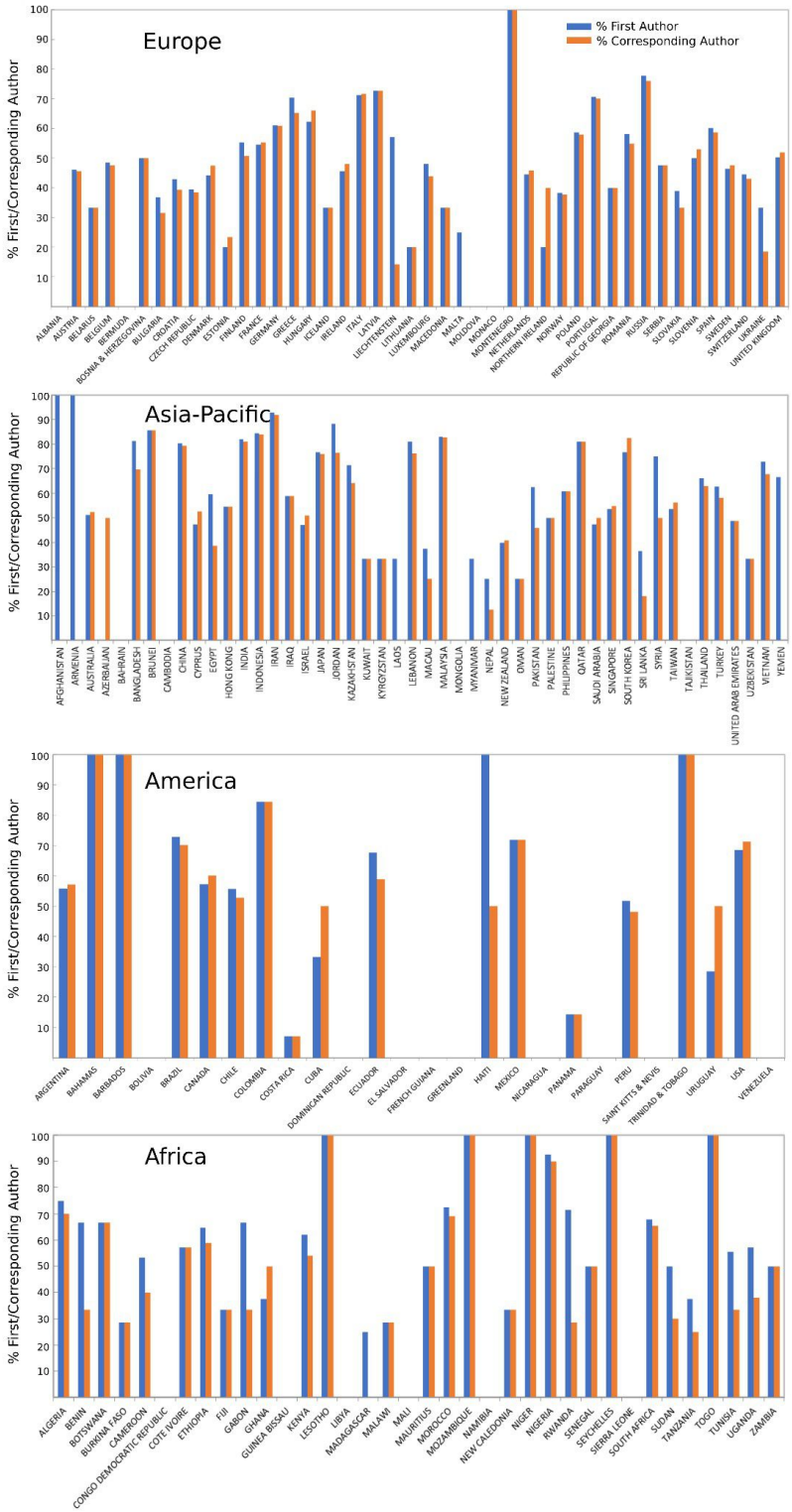


**Figure 12.** Percentage of Data Papers that have the first/corresponding author affiliated with the selected countries. Results are shown for articles published in the years 2008–2020

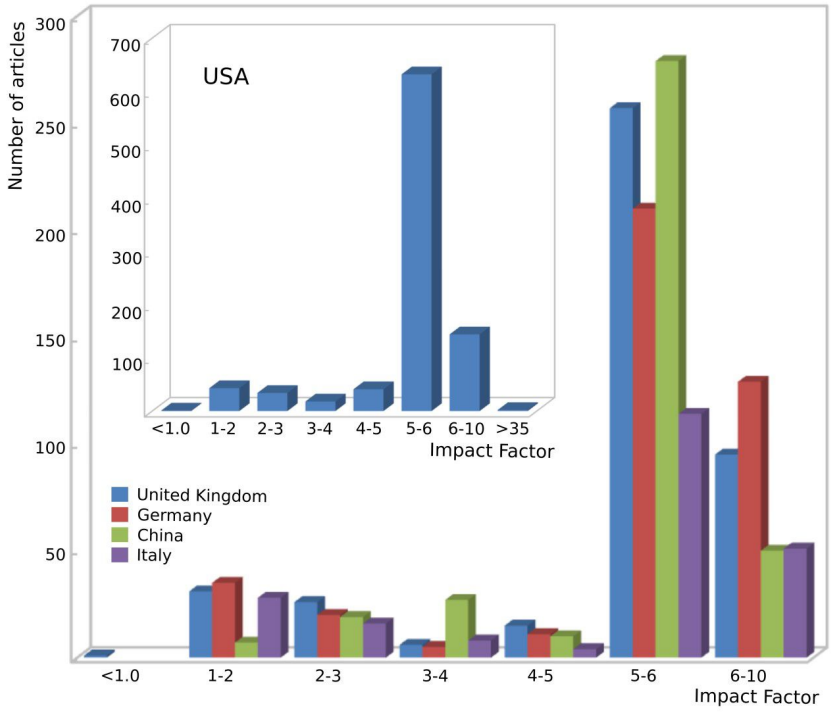


**Figure 13.** The number of articles indexed in the Web of Science and defined as Data Papers (1980–2020) in relation to the Impact Factor for the 5 countries in which the largest number of Data Papers were published

cited Data Papers, and the percentage of Data Papers in the Top 10%) are lower. Further, we also observed that if the percentage of Data Papers with the first/corresponding author affiliated with the selected countries was higher, the values of certain indicators (the percentage of Data Papers in Q1 Data Journals, and the percentage of Data Papers cited at least once) were lower. However, the correlation between these indicators is weak (R-squared value on the charts=29% and 37%, respectively). Moreover, based on Figure S7, it can be seen that if the percentage of Data Papers in Q1 Data Journals is higher, the values of certain indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, and the percentage of Data Papers in the Top 10%) are also higher. However, it should be noted that there is a very low correlation between these indicators (R-squared value on the charts =20–26%).

## 4. Conclusions

This paper provides insight into a comprehensive overview of Data Journals and the Data Papers that were published in 1980–2020. In particular, we calculated some bibliometric indicators related to publication activity and impact of

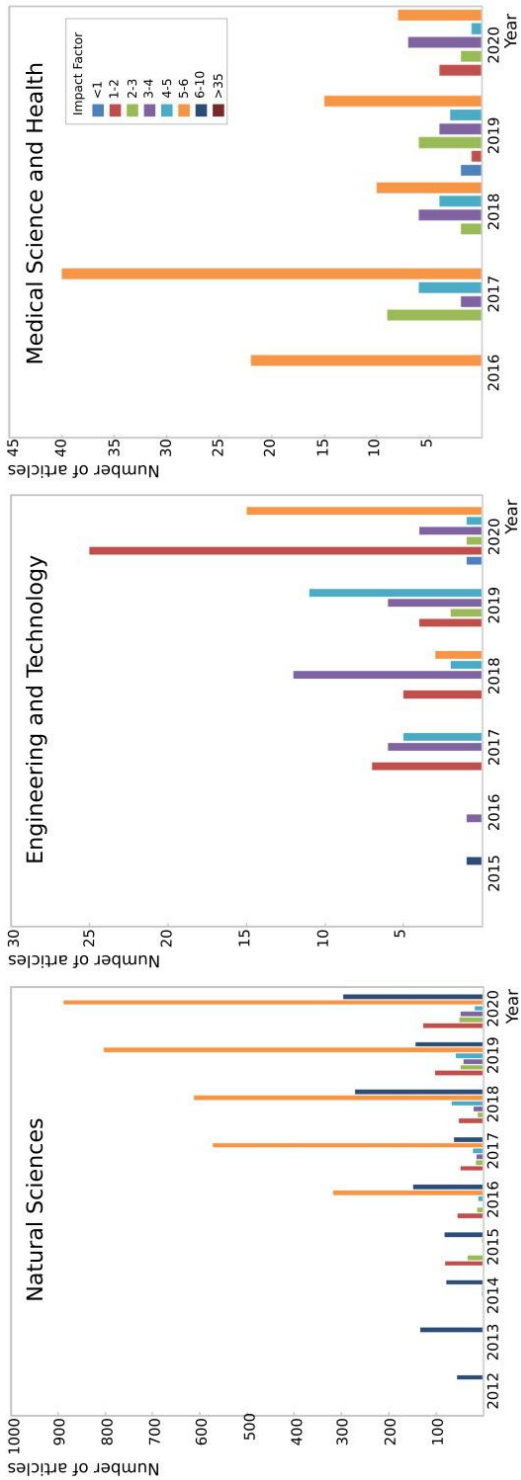


**Figure 14.** The number of articles identified as Data Papers by the Impact Factor and publication year in Natural Sciences, Engineering and Technology, and Medical and Health Sciences. Analyses were conducted using the InCites analytical tool and carried out for countries in which at least 30 Data Papers were written in the years 1980–2020

**Table 1.** Top 20 journals in which scientists published the greatest number of Data Papers in the years 1980–2020 and the corresponding number of citations, percentage of documents cited at least once, Category Normalized Citation Impact (CNCI), Impact Factor (IF), quartile, JCR categories, and main OECD research areas and OECD disciplines

| No. | Name of journal | WoS | Number of citations | % of docs cited | CNCI | Impact Factor | Quartile | JCR categories (OECD research areas and OECD disciplines) |
|---|---|---|---|---|---|---|---|---|
| 1 | DATA IN BRIEF | 6,273 | 11,234 | 54.25 | 0.69 | n/a | n/a | Multidisciplinary Sciences |
| 2 | SCIENTIFIC DATA | 1,233 | 15,325 | 79.08 | 2.31 | 5.541 | Q1 | Multidisciplinary Sciences |
| 3 | EARTH SYSTEM SCIENCE DATA | 338 | 10,195 | 87.57 | 1.10 | 9.197 | Q1 | Meteorology and Atmospheric Sciences; Geosciences, Multi-disciplinary (Earth and related Environmental Sciences; Natural Sciences) |
| 4 | DATA | 184 | 508 | 63.59 | 0.96 | n/a | n/a | Computer Science, Information Systems (Computer and Information Sciences; Natural Sciences) |
| 5 | BIODIVERSITY DATA JOURNAL | 170 | 384 | 52.35 | 0.88 | 1.331 | Q3 | Biodiversity Conservation (Biological Sciences; Natural Sciences) |
| 6 | GIGASCIENCE | 158 | 1,700 | 89.87 | 1.02 | 5.993 | Q1 | Biology (Biological Sciences; Natural Sciences) |
| 7 | HUMAN GENOME VARIATION | 130 | 331 | 77.69 | 0.76 | n/a | n/a | Genetics and Heredity (Biological Sciences; Natural Sciences) |
| 8 | ECOLOGY | 78 | 539 | 74.36 | 1.00 | 4.700 | Q1 | Ecology (Biology Sciences; Natural Sciences) |
| 9 | ACTA CRYSTALLO-GRAPHICA SECTION E-CRYSTALL-OGRAPHIC COMM-UNICATIONS | 53 | 40 | 47.17 | 1.00 | n/a | n/a | Crystallography (Chemical Sciences; Natural Sciences) |
| 10 | GEOSCIENCE DATA JOURNAL | 50 | 290 | 54.00 | 0.29 | 2.714 | Q2 | Meteorology and Atmospheric Sciences; Geosciences Multi-disciplinary (Earth and related Environmental Sciences; Natural Sciences) |
| 11 | GENOMICS DATA | 45 | 261 | 93.33 | 0.88 | n/a | n/a | Genetics and Heredity (Biological Sciences; Natural Sciences) |
| 12 | ONCOLOGIST | 44 | 429 | 90.91 | 1.27 | 5.025 | Q2 | Oncology (Clinical Medicine; Medical and Health Sciences) |

**Table 1 – continued.**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 13 | ZOOKEYS | 33 | 82 | 72.73 | 1.04 | 1.137 | Q3 | Zoology (Biological Sciences; Natural Sciences) |
| 14 | JOURNAL OF OPEN ARCHAEOLOGY DATA | 29 | 36 | 37.93 | 0.87 | n/a | n/a | Archaeology (History and Archaeology; Humanities) |
| 15 | ECOLOGICAL RESEARCH | 27 | 67 | 70.37 | 0.33 | 1.580 | Q3 | Ecology (Biological Sciences; Natural Sciences) |
| 16 | CIRCULATION CARDIOVASCULAR QUALITY AND OUTCOMES | 24 | 129 | 75.00 | 1.71 | 5.071 | Q1 | Cardiac and Cardiovascular Systems (Clinical Medicine; Medical and Health Sciences) |
| 17 | FRONTIERS IN GENETICS | 23 | 59 | 52.17 | 0.85 | 3.260 | Q2 | Genetics and Heredity (Biological Sciences; Natural Sciences) |
| 18 | EARTHQUAKE SPECTRA | 22 | 84 | 68.18 | 1.06 | 1.930 | Q2 | Engineering Civil; Engineering Geological (Civil Engineering; Environmental Engineering; Engineering and Technology) |
| 19 | FRONTIERS IN MARINE SCIENCE | 20 | 50 | 60.00 | 1.41 | 3.661 | Q1 | Environmental Sciences; Marine and Freshwater Biology (Earth and related Environmental Sciences, Biological Sciences, Natural Sciences) |
| 20 | GENOMICS PROTEOMICS & BIOINFORMATICS | 17 | 152 | 76.47 | 2.67 | 7.051 | Q1 | Genetics and Heredity (Biological Sciences; Natural Sciences) |

citations in terms of OECD research areas, scientific disciplines, publication year, and country.

The main conclusions of our study are as follows:

- If the percentage of Data Papers that have international co-authors is higher, the values of particular indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, and the percentage of Data Papers in the Top 10%) are higher (R-squared value on the charts = 53–65%). It means that international collaboration can have an effect on the greater number of citations, highly cited Data Papers as well as the percentage of Data Papers in the Top 10% of the best journals.
- If the percentage of Data Papers that have international co-authors is higher, the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries is lower (R-squared value on the charts = 89–92%).
- If the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries is higher, the values of particular indicators (CNCI, the average number of citations per publication, the

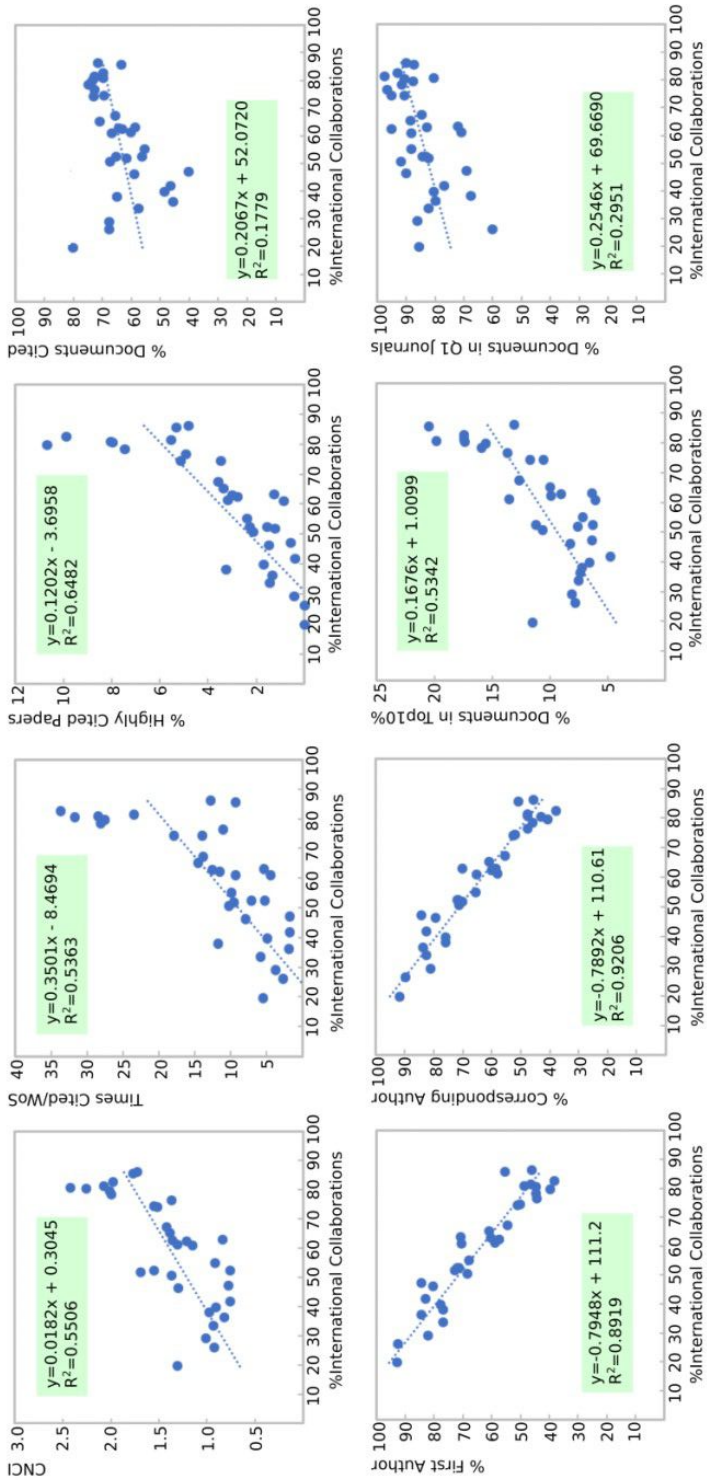


**Figure 15.** Dependence of selected bibliometric indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, the percentage of Data Papers cited at least once, the percentage of Data Papers that have the first/corresponding author affiliated with the selected countries, the percentage of documents in the Top 10%, and the percentage of Data Papers in Q1 Data Journals) as a function of Data Papers that have international co-authors. The results are shown for the countries in which scientists published at least 100 Data Papers

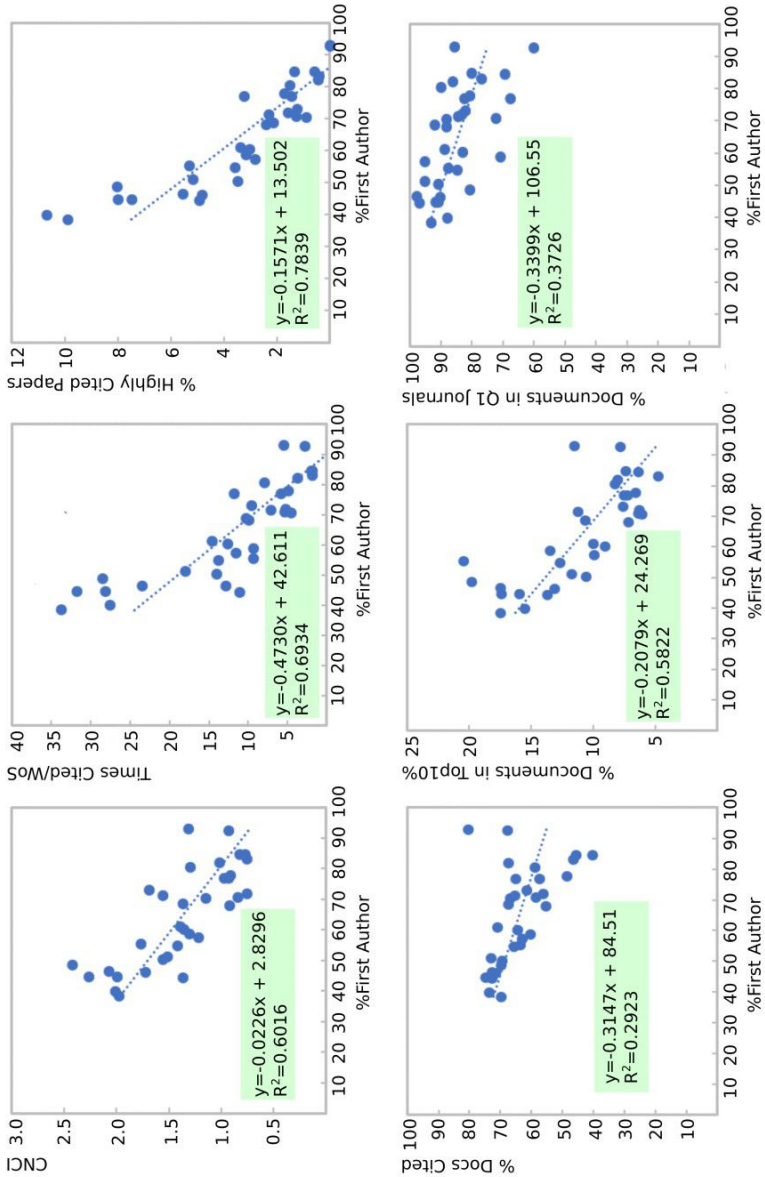


**Figure 16.** Dependence of selected bibliometric indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, the percentage of Data Papers cited at least once, the percentage of Data Papers in the Top 10%, the percentage of Data Papers in Q1 Data Journals) as a function of the percentage of Data Papers that have the first author affiliated with the selected countries. The results were shown for the countries in which scientists published at least 100 Data Papers

percentage of highly cited Data Papers, and the percentage of Data Papers in the Top 10%) are lower (R-squared value on the charts = 58–78%).
- If the percentage of Data Papers published in Q1 Data Journals is higher, the values of particular indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, and the percentage of Data Papers in the Top 10%) are higher (R-squared value on the charts =20–26%).

We also derive a lot of detailed conclusions associated with the bibliometric analysis of the Data Journals and Data Papers.

Bibliometric analysis of Data Journals:

- The majority of Data Journals in which scientists can publish Data Papers are owned by Frontiers Media S.A, followed by Wiley, Springer, Biomed Central, Elsevier, and Pensoft Publishers.
- Open access journals constitute 65% of all the analysed Data Journals.
- Most Data Journals in which authors can publish Data Papers are journals with an Impact Factor of 2–3.
- Most Data Journals in which authors can publish Data Papers are journals that were awarded 100 points by the Ministry of Science and Higher Education (MEiN).
- Most Data Journals in which authors can publish Data Papers are devoted to Natural Sciences, followed by Medical and Health Sciences as well as Engineering and Technology.
- Most Data Journals in which authors can publish Data Papers are devoted to disciplines such as: Biological Sciences, Earth and related Environmental Sciences, Basic Medical Research, Clinical Medicine, Health Sciences, Computer and Information Sciences.
- The Data Journals about Multidisciplinary Sciences (Data in Brief, and Scientific Data) are the journals in which scientists published the largest number of Data Papers.
- Data Journals (Data in Brief, Scientific Data, and Earth System Science Data) are the journals in which Data Papers were more cited (over 10,000 times).
- Three Data Journals in which authors can publish Data Papers have a very high Impact Factor: Chemical Reviews (IF=52.758), Nature Biotechnology (IF=36.558), and Blood (IF=17.794).

Bibliometric analysis of Data Papers:

- The first Data Paper in Medical and Health Science was published in 2006.
- The number of Data Papers is rapidly growing over the years.
- Over 3,000 Data Papers were published each year between 2018–2020.
- The largest number of Data Papers were published in Q1 Data Journals.
- About 64% of Data Papers are cited at least once.

- About 30% of Data Papers are articles that have international co-authors.
- For 128 countries (which constitutes 83% of all analysed countries), the percentage of Data Papers that have the first author affiliated with the selected country is slightly greater than or equal to the percentage of Data Papers that have the corresponding author affiliated with the selected country.
- The average percentage of Data Papers that have the first/corresponding author affiliated with the selected country is equal to about 64% (The assumption was to take into account only those countries in which at least 100 Data Papers were published in the years 1980–2020).
- 339 Data Papers (2.2% of all analyzed articles) have only 1 author.
- The largest number of Data Papers were published in Multidisciplinary Sciences (Elsevier, Springer), followed by Biological Sciences (Biomed Central, Frontiers Media, Wiley), Earth and related Environmental Sciences as well as Computer and Information Sciences.
- The most frequently cited Data Papers are articles that were written in disciplines such as: Natural Sciences, Medical and Health Sciences as well as Engineering and Technology.
- The largest number of Data Papers were published in Data Journals with an Impact Factor ranging from 5 to 6 (The assumption was to take into account only articles that were published in journals with an Impact Factor).
- Scientists from Europe, North America and Asia published the largest number of Data Papers with an Impact Factor in the range of 5–6 and 6–10.
- The largest number of articles identified as Data Papers were published in the USA, China, the United Kingdom, Italy, and France.

## *References*

[1] Candela L., Castelli D., Manghi P. and Tani A.. 2015 *Data Journals: A Survey*, *Journal of the Association for Information Science and Technology* **66** 1747 doi: doi.org/10.1002/asi.23358

[2] Callaghan S., Donegan S., Pepler S., Thorley M., Cunningham N., Kirsch P., Ault L., Bell P., Bowie R., Leadbetter A., Lowry R., Moncoiffé G., Harrison K., SmithHaddon B.,Weatherby A. and Wright D. 2012 *Making data a First class scientific output: Data citation and publication by NERCs environmental data centres*, *International Journal of Digital Curation* **7** (1) 107 doi: 10.2218/ijdc.v7i1.218

[3] Penev L., Mietchen D., Chavan V.S., Hagedorn G., Smith V.S., Shotton D., Tuama E., Senderov V., Georgiev T. and Stoev P., Groom Q.J., Remsen D., Edmunds S. C. 2017 *Strategies and guidelines for scholarly publishing of biodiversity data*, *Research Ideas and Outcomes* **3** doi: 10.3897/rio.3.e12431

[4] Jiao C. and Darch P.T. 2020 *The role of the data paper in scholarly communication*, *Proceedings of the Association for Information Science and Technology* **57** doi: 10.1002/pra2.316

[5] Silva L., Bloom T., Ganley E. and Winker M. 2014 *PLoS new data policy: Public access to data*, *PLoS ONE*

[6] Li K., Greenberg J. and Dunic J. 2020 *Data objects and documenting scientific processes: an analysis of data events in biodiversity Data Papers*, *Journal of the Association for Information Science and Technology* **71** (2) 172 doi: 10.1002/asi.24226
[7] Chavan V. and Penev L. 2011 *The Data Paper: a mechanism to incentivize data publishing in biodiversity science*, *BMC Bioinformatics* **12** (15) 1 doi: 10.1186/1471-2105-12-S15-S2
[8] Seo S. and Kim J. 2020 *Data Journals: Types of peer review, review criteria, and editorial committee members' positions*, *Science Editing* **7** (2) 130 doi: 10.6087/kcse.207
[9] Walters W. H. 2020 *Data Journals: Incentivizing data access and documentation within the scholarly communication system*, *Insights* **33** doi: 10.1629/uksg.510
[10] Penev L., Chavan W., Georgiev T. and Stoev P. 2012 *Data Papers as Incentives for Opening Biodiversity Data: One Year of Experience and Perspectives for the Future*, *Insights*
[11] Schöpfel J., Farace D., Prost H. and Zane A. 2019 *Data Papers as a New Form of Knowledge Organization in the Field of Research Data*, *Knowledge Organization* **46** (8) 622 doi: 10.5771/0943-7444-2019-8-622
[12] Clarivate Analytics 2018 *InCites Indicator Handbook*
[13] Clarivate Analytics 2018 *InCites Benchmark Analytics*, https://clarivate.libguides.com/incites_ba
[14] Garfield E. 1979 *Is citation analysis a legitimate evaluation tool?*, *Scientometrics* **1** (4) 359 doi: 10.1007/BF02019306
[15] Aksnes D.W. 2003 *Characteristics of highly cited papers*, *Research Evaluation* **12** (3) 159 doi: 10.3152/147154403781776645
[16] Schubert A. and Schubert G. 2020 *Internationality at university level*, *Scientometrics* **123** 1341 doi: 10.1007/s11192-020-03443-3
[17] LunaMorales M. E., LunaMorales E. and Perez-Angon M. A. 2020 *Influence of the international collaboration in the field of metric studies of science and technology: the case of Mexico*, *Scientometrics* **126** 2485 doi: 10.1007/s11192-02003522-5

# Supplementary Information

**Table 2.** Top 21 countries in which scientists published the greatest number of Data Papers in top 4 Data journals (Data in Brief, Scientific Data, Earth System Science Data and Data) and the corresponding number of Data Papers

| No | Country | DATA IN BRIEF | Country | Scientific Data | Country | EARTH SYSTEM SCIENCE DATA | Country | DATA |
|---|---|---|---|---|---|---|---|---|
| 1 | USA | 1117 | USA | 552 | USA | 142 | USA | 1896 |
| 2 | China | 484 | United Kingdom | 246 | Germany | 127 | United Kingdom | 786 |
| 3 | Italy | 441 | China | 210 | United Kingdom | 95 | Germany | 782 |
| 4 | India | 397 | Germany | 200 | France | 87 | China | 751 |
| 5 | Germany | 395 | Italy | 111 | Australia | 51 | Italy | 635 |
| 6 | United Kingdom | 360 | Australia | 110 | Italy | 49 | France | 509 |
| 7 | Nigeria | 289 | France | 102 | Netherlands | 46 | India | 442 |
| 8 | France | 287 | Canada | 96 | Norway | 46 | Spain | 421 |
| 9 | Iran | 285 | Switzerland | 95 | Switzerland | 45 | Canada | 411 |
| 10 | Japan | 284 | Spain | 84 | Spain | 44 | Japan | 396 |
| 11 | Spain | 261 | Sweden | 79 | Canada | 44 | Australia | 384 |
| 12 | Canada | 252 | Netherlands | 75 | China | 41 | Russia | 328 |
| 13 | Russia | 251 | Japan | 61 | Japan | 35 | Iran | 293 |
| 14 | Malaysia | 227 | Denmark | 54 | Sweden | 26 | Nigeria | 292 |
| 15 | Australia | 206 | Belgium | 48 | Belgium | 24 | Brazil | 284 |
| 16 | Brazil | 205 | Austria | 42 | Austria | 18 | Netherlands | 268 |
| 17 | South Korea | 188 | Brazil | 40 | Denmark | 16 | Switzerland | 253 |
| 18 | South Africa | 159 | South Korea | 39 | Poland | 14 | Malaysia | 241 |
| 19 | Colombia | 138 | Norway | 39 | Russia | 14 | South Korea | 239 |
| 20 | Indonesia | 134 | Finland | 38 | New Zealand | 14 | Sweden | 216 |
| 21 | Netherlands | 126 | Singapore | 31 | South Africa | 9 | South Africa | 195 |

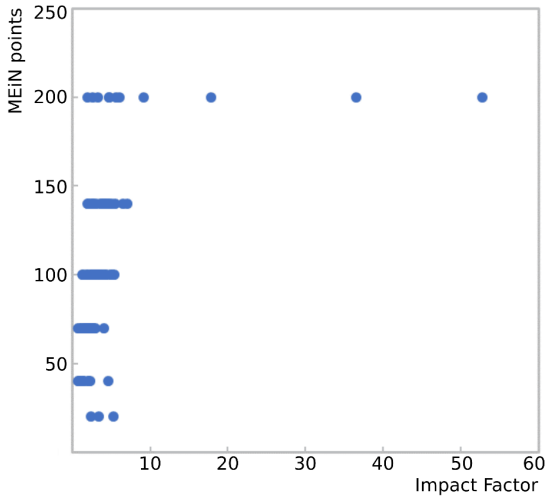


**Figure 17.** MEiN points as a function of IF for Data Journals which have assigned IF and MEiN points

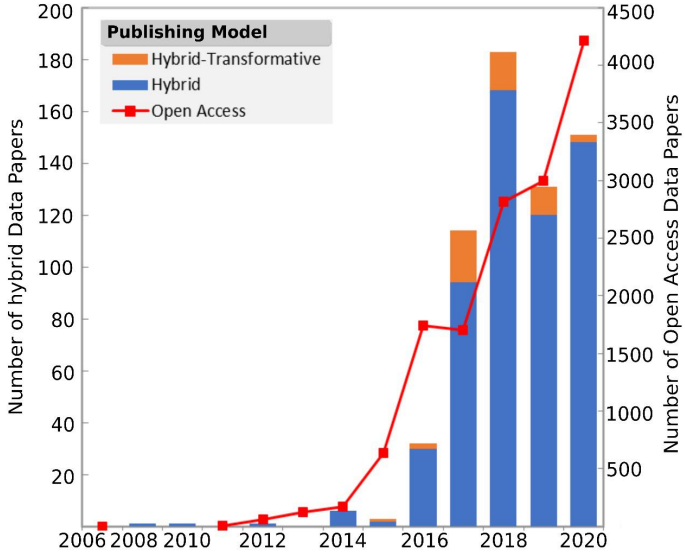


**Figure 18.** The number of Data Papers that were published in Open Access, hybrid and hybrid/transformative journals as a function of publication year

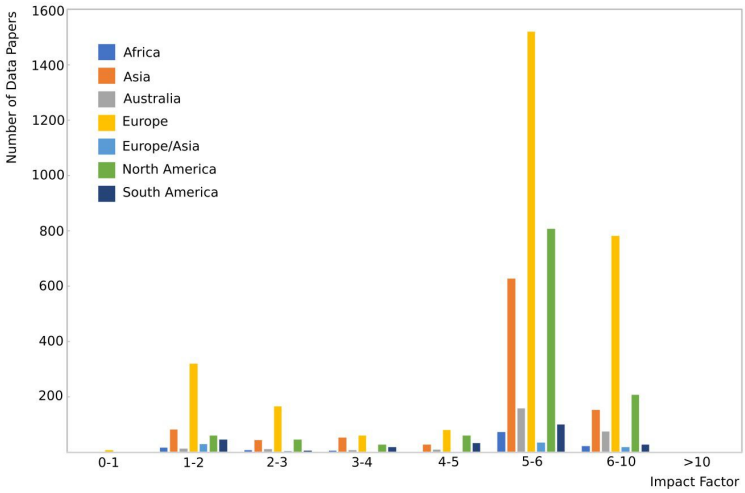


**Figure 19.** The number of Data Papers that were published in journals with an Impact Factor as a function of Impact Factor for all countries. The number of Data Papers was calculated for particular continents (Europe, Asia, North America, South America, Australia, Africa)

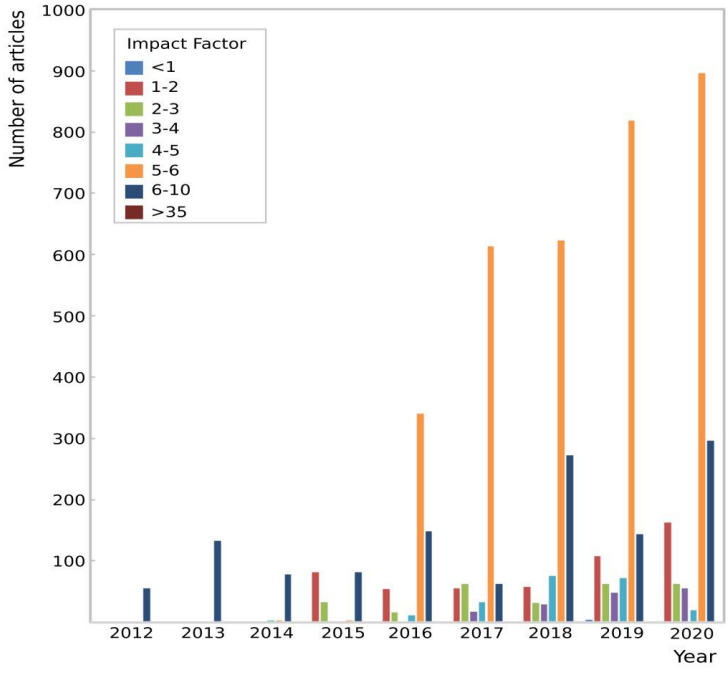


**Figure 20.** The number of Data Papers that were published in journals with an Impact Factor as a function of the publication year and the Impact Factor. The results are shown for the countries in which scientists published at least 30 Data Papers in the years 1980–2020

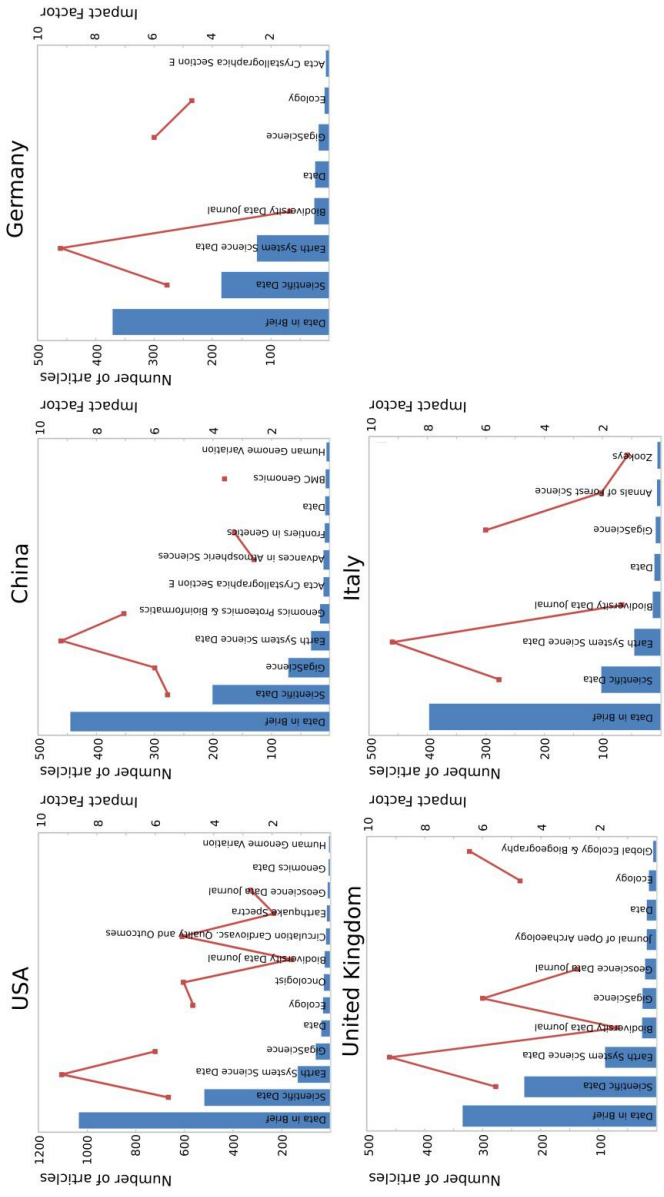


**Figure 21.** The number of Data Papers that were published in journals with the greatest number of Data Papers as a function of Impact Factor. The number of Data Papers is shown for the 5 countries in which scientists published the greatest number of Data Papers: USA, China, Germany, UK, and Italy

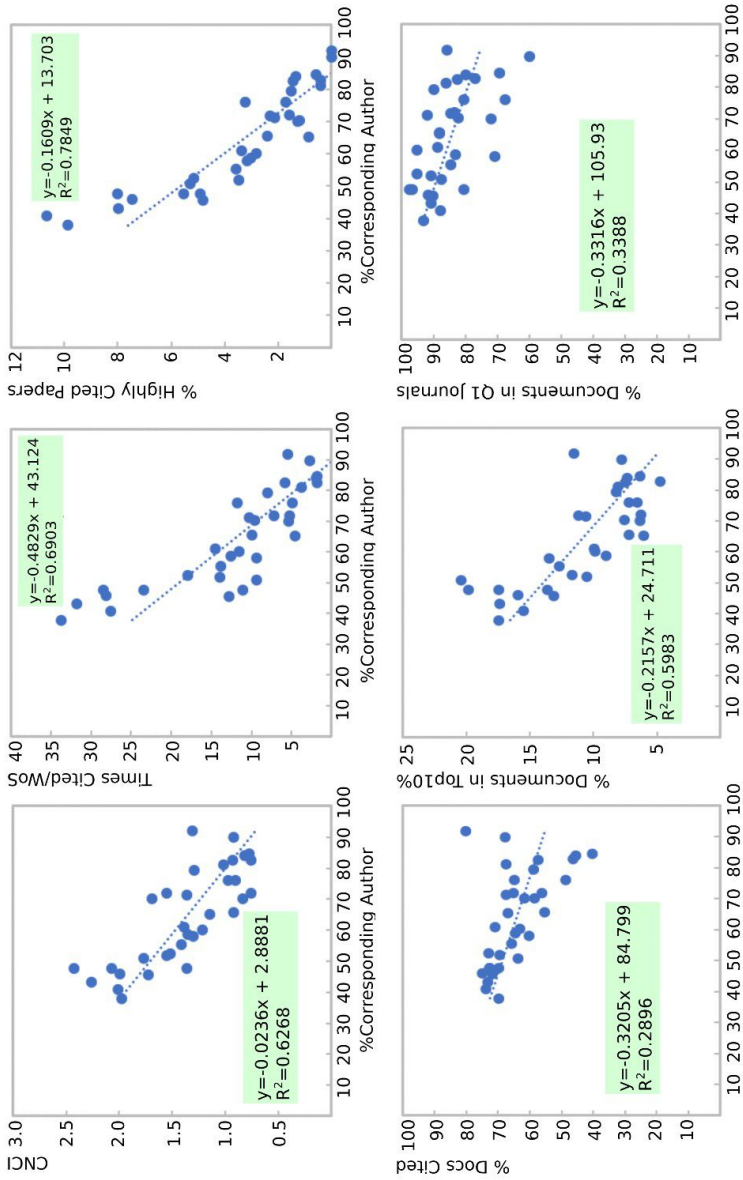


**Figure 22.** Dependence of selected bibliometric indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, the percentage of Data Papers cited at least once, the percentage of documents in the Top 10%, the percentage of Data Papers in Q1 Data Journals) as a function of the percentage of Data Papers that have the corresponding author affiliated with the selected countries. The results are shown for the countries in which scientists published at least 100 Data Papers

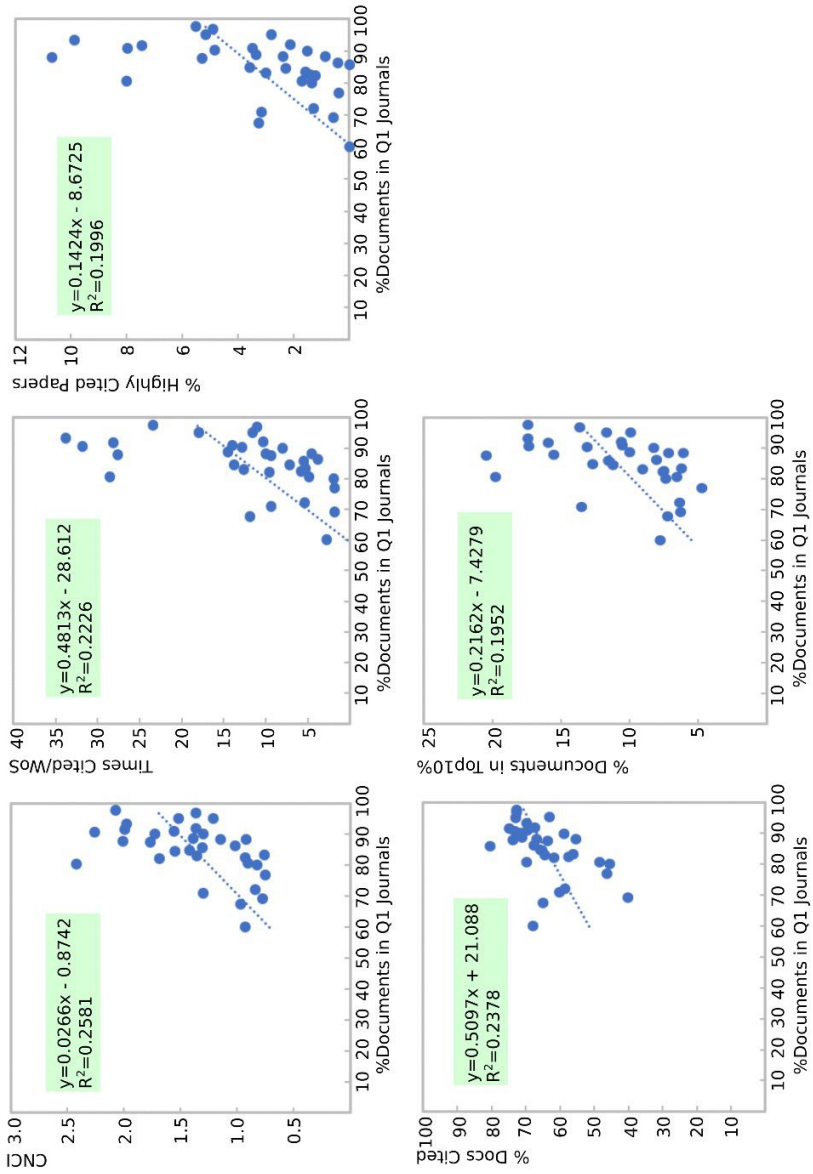


**Figure 23.** Dependence of selected bibliometric indicators (CNCI, the average number of citations per publication, the percentage of highly cited Data Papers, the percentage of Data Papers cited at least once, the percentage of documents in the Top 10%) as a function of the percentage of Data Papers in Q1 Data Journals. The results are shown for countries in which scientists published at least 100 Data Papers

**Beata Adamczak** is a librarian at the Gdańsk University of Technology. She received her PhD in Physical Chemistry at the Gdańsk University of Technology. Her dissertation was associated with molecular dynamic simulations. Her research interests focus on large data set analysis using Microsoft Excel, especially pivot tables and charts. She has experience in analytical reports and bibliometric analysis in InCites and SciVal. In particular, she performed bibliometric analysis related to insights into the publishing activity of the Gdańsk University of Technology compared to other world universities (1945-2019), Open Access publishing among the Visegrad Group countries. She also depicted the bibliometric analysis of the Polish Maritime Research journal in terms of publications and citations in order to celebrate the 27th anniversary since the publication of its first issue in 1994. Moreover, she is responsible for the institutional repository The Bridge of Knowledge. She checks the Open Access policy of journals and publishers.