Business Ethics and Leadership, Volume 6, Issue 4, 2022
ISSN (online) – 2520-6311; ISSN (print) – 2520-6761

Open Research Data in the Open Science Ecosystem and Business Environment

Anton Boiko, https://orcid.org/0000-0002-1784-9364
Doctor of Economic Sciences, Associate Professor, Sumy State University, Ukraine
Olha Kramarenko, https://orcid.org/0000-0002-2616-0147
PhD student, Sumy State University, Ukraine
Tetyana Mayboroda, https://orcid.org/0000-0002-4547-5822
PhD, Associate Professor, Sumy State University, Ukraine
Corresponding author: Olha Kramarenko, o.kramarenko@ssu.edu.ua
Type of manuscript: research paper

Abstract: *Today, one can observe shifts in the research landscape, which is formed by digitization and open science principles. The open science movement continues to gain momentum, attention and debate. In parallel with the principle of unity, open science gives rise to a taxonomy of several related ideas, guidelines and concepts, such as open access, open replicable research and open data. Over the past fifteen years, research institutions have focused on open access to publications. However, recently the focus of attention has shifted to research data as a "new currency" in research activities and their distribution in open access, and the guiding principles of data management are becoming crucial for the wide implementation of open science practices and the effective use of data in research, industry, business and other sectors of the economy. In this context, it is relevant to carry out a thorough study of primary scientific works on open science issues and to study the role of the concept of "open research data" in the paradigm of a holistic ecosystem of open science and business ecosystem. In this work, it is proposed to use the methods of quantitative and qualitative bibliometric analysis, which allows to identify the main trends and form the basis for further research. The information base for this work was the international scientometric database Scopus, which enables to analyze bibliographic data using built-in tools and import them for external use in the VOSviewer software. The study revealed an increasing trend in the number of publications on the subject under study, with the highest annual growth rate in 2017 (76%) and 2019 (66%). Qualitative bibliographic analysis made it possible to analyze the most cited and, therefore, trending works on the selected topic. In terms of the number of citations per year, the results show that the studies with such directions in open science as open program code (open source); data/research reproducibility, research data management; open access to publications (open access) are most popular. In addition, a cluster analysis of the co-prevalence of keywords was conducted. It formed clusters dedicated to both institutional and infrastructural problems of the development of open science and research data. Separately, the results of the analysis create a scientific basis for further research into the key determinants of the effectiveness of the implementation of a proper research data management system at the micro, meso, and macro levels. It will improve the effectiveness of the implementation of scientific developments from one field of knowledge to another, while achieving increased interdisciplinary research. In parallel with this, interested persons of the real sector of the economy get the opportunity to analyze scientific results, determining the possibility of their adoption in their own activities.*
Keywords: technological progress, digitization, research data, open science, bibliometric analysis.
JEL Classification: I23, I25, I28, O1.

Received: 10 September 2022 **Accepted:** 31 October 2022 **Published:** 31 December 2022

Funding: This research was funded by the Ministry of Education and Science of Ukraine and performed the results of the projects «Business-Education-Science Coopetition: Institutional and Economic Models of Innovation Transfer for National Security and Sustainable Development» (№ 0122U000772).

Publisher: Sumy State University
Founder: Sumy State University

Cite as: Boiko, A., Kramarenko, O., & Mayboroda, T. (2022). Open Research Data in the Open Science Ecosystem and Business Environment. *Business Ethics and Leadership*, *6*(4), 79-91. http://doi.org/10.21272/bel.6(4).79-91.2022.

Business Ethics and Leadership, Volume 6, Issue 4, 2022
ISSN (online) – 2520-6311; ISSN (print) – 2520-6761

Open Research Data in the Open Science Ecosystem and Business Environment

Introduction

Humanity continues to face "grandiose" problems on the way to its sustainable development. Global challenges such as climate change, food security, health and well-being across the lifespan require concerted efforts in environmental science, education, policy and business. Today, science is an essential factor in the development of the population's education, business, culture and welfare. A significant role of science as a component of sustainable development, competitiveness, innovation, infrastructure, and quality education is assigned in the Sustainable Development Strategy of Ukraine until 2030 (Ministry of economic development and trade of Ukraine, 2017). Currently, state support for international scientific and scientific and technical cooperation, as well as the integration of the scientific and innovative system of Ukraine into the European Research Area, is one of the key priorities and is defined in the following regulatory documents "On approval of the road map for the integration of the scientific and innovative system of Ukraine into the European research of space" (Pro zatverdzhennia dorozhnoi karty z intehratsii naukovo-innovatsiinoi systemy Ukrainy do yevropeiskoho doslidnytskoho prestorio, 2021), Law of Ukraine "On scientific and scientific and technical activities" (Pro naukovu i naukovo-tekhnichnu diialnist, Zakon Ukrainy, 2021), Cabinet resolution of Ministers of Ukraine "On the implementation of the Association Agreement between Ukraine, on the one hand, and the European Union, the European Atomic Energy Community and their member states, on the other hand" (Postanova Kabinetu Ministriv Ukrainy No. 1106, 2020). In turn, one of the important elements of the Single Digital Market of Europe (Digital Single Market), as well as a component of the "Open Innovations − Open Science − Openness to the World" paradigm, which is developing within the framework of the European Research and Innovation Space, is the development of the European Cloud of Open Science (European Open Science Cloud) and the European data infrastructure aimed at the development of open scientific practices driven by research data.

Literature Review

Foreign scientists have carried out many scientific studies devoted to open science, open data and application practices. Scholarly research (Baack, 2015; Kattge et al., 2014; McKiernan et al., 2016; Piwowar & Vision, 2013) demonstrates both the advantages and disadvantages of open science and the need to share and document data. Open access and open science are a path to faster, more collaborative, and more inclusive science (Jamali et al., 2020). Importantly, open science is supported by open data, models and also by open publications (Björk, 2017), and open software code (Hey & Payne, 2015; Easterbrook, 2014). At the same time, considerable attention is focused on open practices' reproducibility and quality indicators (Hey & Payne, 2015; Easterbrook, 2014; Koltay, 2020; Class et al., 2021; Cai & Zhu, 2015). In the study (Hajduk et al., 2019), scientists emphasize that data exchange should be simple, feasible and accessible. Therefore, ensuring quality, reproducibility and interoperability is possible by applying FAIR principles to the open dissemination of scientific knowledge (Hauschke et al., 2021; De Smedt et al., 2020). Applying FAIR principles is a component of the research data management (Wilkinson et al., 2016). Some domestic and foreign scientists investigate the issues of open science and research data in the paradigm of digital transformations, e-infrastructures, grid-infrastructures, and cyberinfrastructures (Svystunov, S.Ya., Shevchenko, A.Yu., 2014; Hey, 2005); in the paradigm of creating a single research space in Ukraine and developing innovative ecosystems (Matyushenko et al., 2017).

The Aim of the Study

The aim of the article is to study the foundations of the concept of "open research data" in the paradigm of a holistic ecosystem of open science and the business environment. A systematic analysis of the retrospective development of the open research data was carried out in the following sequence:

- The essence of "open research data" concept was analyzed and the prerequisites for its development were considered.
- A bibliometric analysis was carried out for a comprehensive understanding of the researched problem, in-depth analysis of the features of the development of scientific research and identification of the main scientific directions.

Presentation of the Main Research Material

From a historical perspective, the open science development as we know it today, and those components that laid the foundation for many of the principles that guide science today, is believed to have its roots in the 17th century. It was a period of the first scientific revolution, when research societies science was made public for the first time, and the publication of scientific works became the predominant means of spreading scientific knowledge (Bartling, S., Friesike, S., 2014). Although, science is becoming open, the invention in the 1990s of the World Wide Web, the primary purpose of which was to rapidly exchange information between scientists around the world (CERN, n.d.), is driving technological progress that in turn, it critically accelerates the open science movement, changes its landscape, and transforms the very concept of "open" as a whole. Thus, rapid technological progress and innovations caused by digitalization contribute to the historical development of the period of the second scientific revolution, which is characterized by the emergence of such trends in science as "Science 2.0", "Open Science", "eScience", "Mode2" and "Open Research" (Bartling, S., Friesike, S., 2014).

As you can see, many terms are used to denote transformations of scientific practices, and open science is one of them. It is worth noting that it was the European Commission that contributed to the generalization of all concepts under the single term "Open Science" by choosing it from among the alternative "Science 2.0" by voting in 2014 (Burgelman et al., 2019). Considering the above, it becomes evident that digitalization, innovation and technological progress can be attributed to the key drivers of changing views on the nature of knowledge dissemination and the formation of open science. The combination of all these components leads to a rise of new technologies of scientific cooperation and as a result, contributes to the democratization of science and research in terms of the free dissemination of knowledge, the exchange of ideas, software and data, the creation of new environments for joint scientific research, computing (Friesike et al., 2014).

Innovative methods of distribution and exchange of knowledge, first, transform publishing activity and radically change the form of access and distribution of peer-reviewed scientific literature. Therefore, the support of the open science strategy by the world's scientific publishing houses and the change of views on the nature of research data form a new publishing model – "open access journal" and "data journal". Depending on the organization of open access, such publishing houses allow scientists to distribute by self-archiving not only research results (postprints), but also manuscripts of their research (preprints) (Suber, P., 2012). On the other hand, the emphasis on the importance of data sharing and reuse, which has been developing in recent years, makes data journals a new channel for realizing this goal by facilitating the dissemination by scientists beyond the article of primary scientific products generated during research, such as datasets, software, program code, experiments, etc. (Candela et al., 2015).

As one can see, there are other sources of knowledge dissemination than publication in a journal. Alternative sources of dissemination of scientific information, including scientific blogs, repositories (academic, thematic, and research data), scientific social networks, are developing as integral components of open science. These sources are primarily popular among scientists since non-journal sources contribute to greater visibility of research results; faster use and greater citation of research that may be relevant where the peer review and editorial publication process takes too much time; involvement in debate and discussion; involvement of the public in scientific processes.

In turn, digitization, technological progress and new forms of knowledge exchange cause the growth of the volume and diversity of scientific information and give rise to the interdisciplinary nature of scientific issues. Undoubtedly, all these processes could not but contribute to developing and improving technological means of data collection, processing, and analysis. Thus, there is a rapid increase in the volume, complexity, and variety of data, which makes science more dependent on joint scientific and technical research environments, and complex computing platforms. Therefore, there is a development of e-infrastructures and grid-infrastructures, better known as cyber-infrastructures, which provide researchers with access to better computing power, and data storage (Hey, 2005).

Open science encompasses concepts and understandings of knowledge creation and dissemination, which are highly context-, perspective-, and interest-dependent. All transformation processes in science caused by technological development are structured by scientists Fecher & Friesike in terms of five different schools of thought, each of which is based on different assumptions and pursues different goals (Table 1) (Fecher B., Friesike S., 2014).

Table 1. Five Schools of Open Science

School of Thought	Background	Involved groups	Essence
State school	Science should be accessible to the public.	Scholars and citizens	Ensuring the accessibility of the research (production) process and the comprehensibility of the research result (product) for citizens. This school strives to engage the public in collaborative research through citizen science, and to make science more clear by lay people
Democratic school	Access to knowledge is unevenly distributed.	Scholars Politicians Citizens	Any research products, especially those that have been funded by state, should be publicly available. Thus, it is necessary to ensure unimpeded open access to scientific knowledge (including publications and data) for all citizens, inter alia by ensuring copyright through licenses for free distribution.
Pragmatic school	Scientific activity could be more effective due to the cooperation of scientists.	Scholars	Making the research more effective by revealing the methods and processes of knowledge creation.
School of infrastructure	The effectiveness of the research depends on the available tools and applications.	Scholars and platform providers.	Creation of open access platforms, tools and services for scientists for the purpose of spreading knowledge or distributed computing.
School of measurements	Today, scientific contributions require traditional and alternative approaches to impact measurement.	Scholars and politicians	Development of already existing and new alternative methods of the metric system of scientific influence.

Source: Fecher B., Friesike S., 2014

Undoubtedly, digitalization has accelerated the movement of open science, but its definition still needs to be defined. It is fair to note that although open science is a general term, its main components are accessibility, transparency and inclusiveness. Therefore, the Organization for Economic Cooperation and Development (OECD) invests in the concept of openness in science access to research results funded by the state in a digital format without or with minimal restrictions (OECD, 2015). According to the European Commission, open science is the transformation, discovery and democratization of scientific research and innovation through information and communication technologies (Ramjoué, 2015). According to the UNESCO recommendation, open science is an inclusive construct combining various movements and practices to make scientific knowledge searchable, accessible, compatible and reusable (UNESDOC, 2021). Scientists themselves, for the most part, include in its content the improvement of efficiency, control and reproducibility of results, the reuse of scientific work and the increase of transparency of all scientific processes, but scientists (Vicente-Saez & Martinez-Fuentes, 2018) consider open science as transparent and accessible knowledge shared and extended by users through shared networks.

Today, according to some researchers, available research data refers to research data that is included in a scientific article published in open access. However, for the most part, more than the data presented in the research material is needed to reproduce and verify the research result and its reuse. On the other hand, in modern science, the convergence of scientific processes with research data is increasingly observed because, on the one hand, data is a product of research activity. On the other hand, data is analyzed to obtain new scientific ideas and results. Thus, scientific data form an inextricable, interconnected research process, and the efforts of all possible stakeholders, guided by the principles of "open as possible, closed as necessary" (Horizon Europe, 2021), are currently focused on making data more open and available. Therefore, the conceptual foundations of "open research data" are being developed.

Regarding the definition of open research data itself, Murray-Rust combines the meaning of "open" to the definition of open-source software. Here the term is also synonymous with "free" (Murray-Rust, 2008) since the definition of "open" was originally derived from the definition of open-source code, which in turn was derived from the original Free Software Guidelines (Open Definition (a), n.d.). At the same time, according to the Open Knowledge Foundation, the interpretation of openness includes principles that define "openness" about both data and content (Open Definition (b), n. d.). Therefore, open research data fully falls under the definition of open data. According to the definition of Murray-Rust, open data in science is data freely available on the Internet, with the possibility for any user to download, copy, analyze, reprocess, transfer it to software or use it for any other purpose without financial, legal or technical obstacles (Murray-Rust et al., 2010). In turn, it is appropriate to note that the very definition of data can differ significantly. This diversity around the concept of data is because several communities and processes generate data for different purposes. Currently, a significant correlation can be observed between government (state) data and academic data (research data). Instead, it is appropriate to note that research data results from systematic research that includes observation, experimentation or hypothesis testing and consists of used and contextualized heterogeneous objects and subjects, depending on the academic discipline of origin (Koltay, 2020). While public data are more considered "assets" of a company or organization (Koltay, 2020), and include the result of the systematic performance of certain job duties by employees of the organization, and do not include research components.

The adoption of open data is mainly motivated by the expected benefits. Open data policies encourage scientific research and debate; promote innovation and potential use of new data in a complementary way; lead to the new collaboration between data users and data creators; maximize transparency and accountability; allow you to check the results of the study; encourage improvement and confirmation of research methods; reduces the costs of collecting duplicate data; increases the impact and visibility of research; promote the implementation of new research, based on already created data and their results; provides important resources for education and training (Van den Eynden, 2011). As one can see, researchers are just some of the beneficiaries. Thus, entrepreneurs can reuse research data to develop new innovative products and services or improve existing ones. At the level of public administration, open research data helps make important decisions for society and leads to economic growth. Finally, for the public, the disclosure of research data contributes to the transparency of the funding allocated to research and helps assess the benefits and benefits of such research to society.

Considering the above, any data, including research data, can be effective only if they function within the framework of an integrated ecosystem of data managers (researchers), intermediaries (research organizations and data sources), service developers (those who create new or improve existing goods and services) and users (includes third-party reuse factors of shared data) (adapted from (Kovalchuk A., 2018)). Therefore, to fully ensure the integrity of a sustainable research data ecosystem, it is important to endow data with qualitative characteristics to ensure that data can be found, accessed, interacted with, and reused by humans and machines. After all, the concept of open qualitative data includes not only openness and access to information but also legal and technical openness. At the same time, legal openness allows for the reuse and distribution of data without licensing obstacles, that is, on conditions that do not violate current legislation. Accordingly, technical openness is the ability to reuse data without technical barriers. First, technical obstacles include unstructured (unorganized) data, including those not in a machine-readable format. In 2016, members of the FORCE11 community defined key requirements for ensuring the quality of research data − FAIR principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016).

Methodology and Research Methods (for Research and Theoretical Papers)

A bibliometric analysis of the evolution of the "open science" concept was carried out to form a terminological basis and to study the dominant trends in the development of the open research data category. As an information base for the research, the authors chose one of the most authoritative multidisciplinary bibliographic databases − Scopus, which contains more than 85 million document descriptions (as of January 2022). To ensure the integrity of the research and obtain a relevant search result, we generated a sample of 1,767 documents according to the key query (TITLE ("open science") OR KEY ("open science")) in the range of years from the beginning of the first publication according to our query, which falls for 1979 to 2020 inclusive. As one can see, we focused the search only on the TITLE and KEYWORDS fields since a broader search yielded documents not related to the research topic. For further analysis and the construction of tables and graphs, the bibliometric data of the documents were exported to an Excel spreadsheet. A thorough study of the relationships between related concepts, constructing thematic categories of research, and implementing a retrospective analysis in the evolutionary time dimension were carried out using the VOSViewer tool.

Results

The authors of the article propose to start an in-depth analysis of the features of scientific research and the definition of the main scientific directions according to the observed topic with the analysis of the total number of studies by scientists that were carried out in the direction of "Open Science", and the trends of citations of such publications by year. Thus, according to Scopus (Scopus, n.d.), the relevance of conducting research in this direction is confirmed by their steady increase annually by an average of 42% (Fig. 1). The first mention of open science was made in the publication "Education is the key to 'open' science" by Shargool P.D. in 1979, published in the journal "Trends in Biochemical Sciences". However, until 2004, the number of publications in this direction did not exceed one per year (with the exception of 2002). The most active interest of scientists in the paradigm of open science can be noted precisely since 2004, since then there has been a steady trend towards an increase in the number of publications on the researched issues with the largest annual growth rate in 2017 and 2019 (Fig. 1).

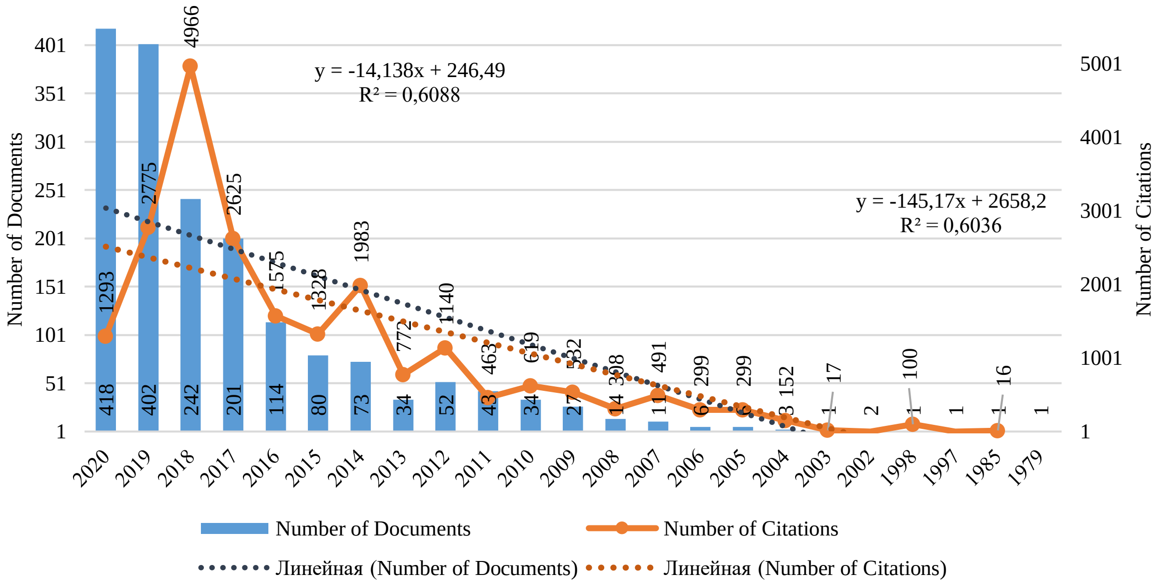

Figure 1. Dynamics of Publications and Citations for the Search Query "Open Science" in Scopus

Source: Compiled by the authors using Scopus data (Scopus, n.d.)

Thus, the number of publications in the direction of "open science" increased by 76% in 2017 compared to 2016. In 2019 it was by 66%, which is 160 units more than in 2018. The relevant trend is determined by the announcement in previous years, namely in 2016 and 2018 of the documents "Amsterdam Call for Action on Open Science" (2016), "Open innovation, open science, open to the world" (2016), "Plan S" (2018). These documents regulate the issues of full open access to all scientific publications, dissemination of research data and management of research processes, and promotion of innovative development through the discovery of scientific knowledge at the international level, and therefore directly related to the open science movement. In turn, the anticipatory growth in the number of open science research by scientists has been occurring in the last five years. It was during the period from 2016 to 2020 that 78% of all research on this issue was published. In 2016, the number of publications reached 6% of their total number. Therefore, its annual growth is further traced to 11% in 2017; 14% is typical for 2018; 23% and 24% reflect the percentage of the total number of publications in 2019 and 2020, respectively (Fig. 1).

The analysis of the number of citations is no less important in the study of the retrospective development of the open science movement and its categories. In bibliometric research, the analysis of citations is a standard method of determining the development trends of the studied issues, the degree of influence of the author, article or magazine, organization, or country on the development of certain processes taking place in the world and having their response in scientific research. Examining the data (Fig. 1), we see that the indicator of the number of citations correlates with the indicator of the number of publications. Thus, according to the data (Fig. 1), the highest rate of citations was achieved precisely in 2018. It is 23% of the total number of citations, which significantly increased the number of publications in 2019. However, the highest number of citations per publication was recorded in 1998 (100 citations per 1 publication) (Fig. 1).

According to the number of citations, we divided the publications according to seven indicators (< 5, ≥ 5, ≥ 50, ≥ 100, ≥ 200, ≥ 400, ≥ 1000). So, we have a relatively small number of publications that have a high percentage of citations: only one publication (0.1%) has more than 1000 citations; 3 papers (0.2%) have 400 or more citations; 40 works (2.3%) have more than 50 citations; 31%, namely 540 works, are publications with the number of citations from 5 to 50. One should note that more than 65% of publications do not exceed four citations or have no citations. This trend is due to two factors: the publication data are not relevant enough for the scientific community, or they were published recently and did not have time to collect the required citations.

Scientific publications describing open-source software products and the benefits of sharing primary data, such as open-source software in open access, are the most cited (Price-Whelan et al., 2018; Peirce et al., 2019; Mueller & Piper, 2014) (Table 2). No less cited (479 citations) is the scientific work of scientists Nosek, B.A. and others (Nosek et al., 2018), which addresses the issue of pre-registration of the analysis plan for the execution of the analytical steps of the study and the importance of publishing such a plan on relevant open platforms (Table 2). The work of scientist Piwowar H. which highlights the benefits of open access on the impact of the number of citations is highly cited (341 citations) (Piwowar et al., 2018). Highly rated (335 citations) is also a publication that characterizes the relationship between open science and open scientific networks (open science grid) (Pordes et al., 2007) (Table 2).

Table 2. The Most Cited Works by the Search Query "Open Science" in Scopus

Year	Author	Title	Source	Number of citations	Citations per year
2018	Price-Whelan, A.M. (et al)	The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package	Astronomical Journal	1656	331
2019	Peirce, J. (et al)	PsychoPy2: Experiments in behavior made easy	Behavior Research Methods	586	146,5
2018	Nosek, B.A. (et al)	The preregistration revolution	Proceedings of the National Academy of Sciences of the United States of America	479	95,8
2014	Mueller, S.T. (et al)	The Psychology Experiment Building Language (PEBL) and PEBL Test Battery	Journal of Neuroscience Methods	427	47,4
2018	Piwowar, H. (et al)	The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles	PeerJ	341	68,2
2007	Pordes, R. (et al)	The open science grid	Journal of Physics: Conference Series	335	22,3

Source: Compiled by the authors using Scopus data (Scopus, n.d.)

Thus, the six most cited works have more than 300 citations. Separately, it is worth noting that the scientific publications with the highest number of citations per year are the scientific works of scientists of the last five years (Table 2) and correlate with such directions in open science as open-source code; data/research reproducibility, research data management; open access to publications (open access). Such a tendency for citations shows that these publications are highly appreciated by the scientific community and emphasize the relevance of the research areas. Analyzing the affiliation of scientists to the country, the movement of open science is global, and the implementation of research by scientists in this direction is becoming relevant in many countries of the world. 1,767 works included in our sample are covered in the publications of scientists from 87 countries, which means that at least one work on the studied issue was published in each of these countries.

Our analysis of indicators of the number of publications by country demonstrates at least 10 countries that occupy leading positions within the framework of the studied issues (Fig. 2). Together, the publications of the ten most productive countries make up 67% (1,873 articles) of the total number of all publications by country (2,793 publications). According to the data, the undisputed leader is the United States of America (649 articles), followed by Great Britain (246 articles), Germany (237 articles), Italy (136 articles), and the Netherlands (133 articles) (Figure 2).

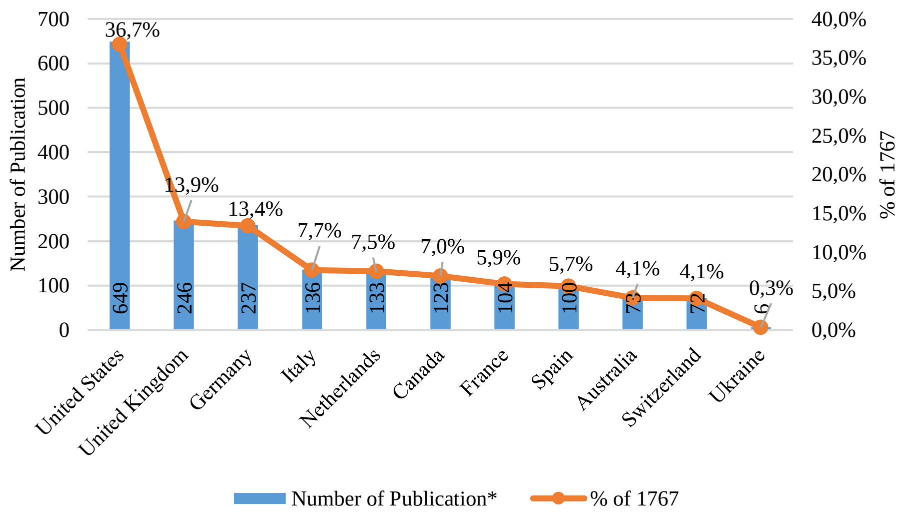

Figure 2. Countries with the Largest Number of Publications for the Search Query "Open Science" in Scopus

**Note: The total number of publications is greater than 1767, since some studies were carried out in collaboration with scientists from other countries*

Source: Compiled by the authors using Scopus data (Scopus, n.d.)

However, despite the significant geographical distribution, this direction of research does not cover even half of all the countries in the world. Separately, it is appropriate to note that the indicator of the number of publications by research geography correlates with internal national initiatives for the regulation of issues of open science and research data at the level of state administration. Therefore, this direction of research needs the development and expansion of the geographical component, however, to achieve a synergistic effect, there is a need to regulate the issues of open science at the state level.

In the context of the research, it is appropriate to analyze the branch structure of open science research. Thus, the largest number of scientific works (638 publications) is concentrated in the field of "Computer Sciences", which is equivalent to 21% of all publications; "Social sciences" make up 17% (528 publications); "Medicine" covers 7%, namely 204 publications of the researched field (Fig. 3). Open science and open data contribute to many economic processes, so the focus of 5% of publications on economic fields of knowledge cannot be overlooked. Among the publications of the economic direction, studies in the direction of "Business and Management" prevail, which make up 4% of all thematic groups; "Economics, econometrics and finance" cover 1%. As you can see, the share of economic publications is relatively low, but the rapid increase in the number of such scientific works has been going on for three years in a row (2018-2020), which is confirmed by their 114% annual growth.

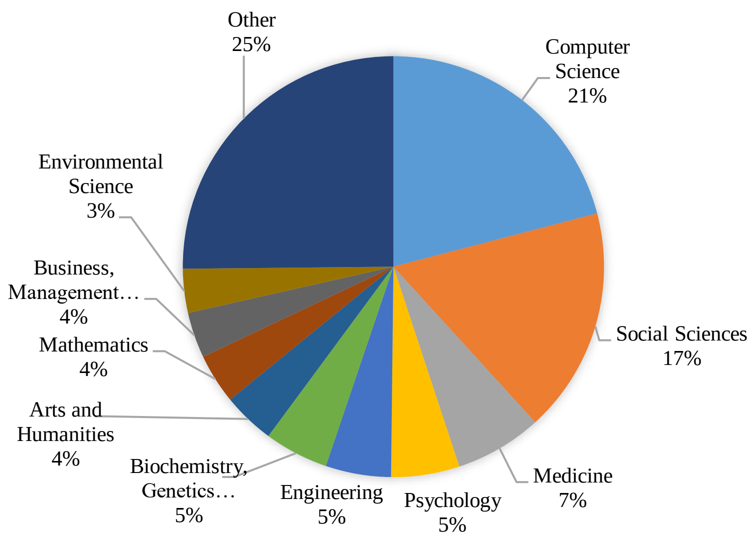

Figure 3. Structure of Publications by Fields of Knowledge of the "Open Science" Research Field in Scopus

Source: Compiled by the authors using Scopus data (Scopus, n.d.)

A thorough study of the trends in the development of open science and its concepts, the construction of relationships between key thematic categories, as well as the determination of relevant directions for further research, led to the use of the visualization and analysis tool of bibliographic data VOSViewer. As an information base for the analysis, we used the author's keywords in the VOSViewer tool. From 1767 publications, we received 3141 keywords. 79% of them, namely 2483 keywords, were used (occurred) in publications only once. 134 words appeared more than five times, equivalent to 4%, 132 words appeared more than 20 times, and only 0.2%, namely 5 keywords, were mentioned more than 50 times in our sample. On the other hand, in the analysis of keywords, not only the frequency of their occurrences (occurrences) in publications is important, but also the total link strength (total link strength), which indicates the overall strength of the links of an element with other elements. Therefore, a higher value of this indicator demonstrates that the element (keyword) was repeatedly associated with other elements, and therefore gives weight to the element (keyword) under consideration.

Thus, according to the data (Table 3), the keywords open access (open access), open data (open data), reproducibility, and data sharing. They are followed by the keywords open source, research data management, and scholarly communication (Table 3). As we can see (Table 3), according to the indicator of prevalence/use (occurrences) in publications, such keywords as "transparency" or "scholarly communication" have a higher frequency of appearance than, for example, "research data management", but the overall power of the total link strength of the keyword "research data management" is higher, and therefore indicates its importance and relevance in scientific research.

Table 3. Trend of Keywords in Publications for the Search Query "Open Science" in Scopus

Keyword	Occurrences	Total link strength
open science	869	4296
open access	153	832
open data	109	611
reproducibility	81	412
data sharing	66	344
open source	30	211
research data management	35	208
scholarly communication	41	197
replication	37	197
transparency	39	196
research data	29	191
big data	30	175
open innovation	19	138

Source: Compiled by the authors using VOSViewer (VOSViewer, n.d)

Further construction of the terminological map of relationships was carried out with the restrictions of the minimum number of occurrences of keywords of five repetitions. Thus, we obtained 134 keywords, which formed six main clusters among themselves (Fig. 4). Thus, the basis of the formation of the first most powerful cluster (red color), which reflects the conceptual foundations of the development of open science, is the concept of open science to such keywords as open access, scientific communication), open peer review, predatory journals, science 2.0, scholarly publishing, copyright (Fig. 4).

The second cluster (green color) highlights open science in the paradigm of digital transformations and covers such keywords as open science grid, electronic infrastructures (e-infrastructure), electronic science (e-science), cyberinfrastructure (cyberinfrastructure), cloud computing, etc., and is formed around the concept of big data (Fig. 4). The formation of the third cluster (blue color) is based on the concept of reproducibility, which, in conjunction with keywords such as transparency, database, replication, and others, highlights the key practices of ensuring reproducibility data usage (Fig. 4).

The fourth cluster (yellow color) is formed around the concept of data sharing and to the keywords interoperability, machine learning, data science, and open innovation reflects the development of such interconnected concepts as "data-driven science" and "data-driven innovation" (Fig. 4). The formation of the fifth cluster (purple color) provides the concept of research data and illuminates open science in terms of a "quality researcher" due to the presence of such keywords as data quality, research data management, FAIR principles, FAIR data, data management plan (Fig. 4). The last, sixth cluster (orange color), which highlights open science in terms of alternative sources of dissemination of scientific knowledge, includes such keywords as open repositories (open repositories), institutional repositories (institutional repositories), gray literature (grey literature), − is formed around the key concept − universities (Fig. 4).

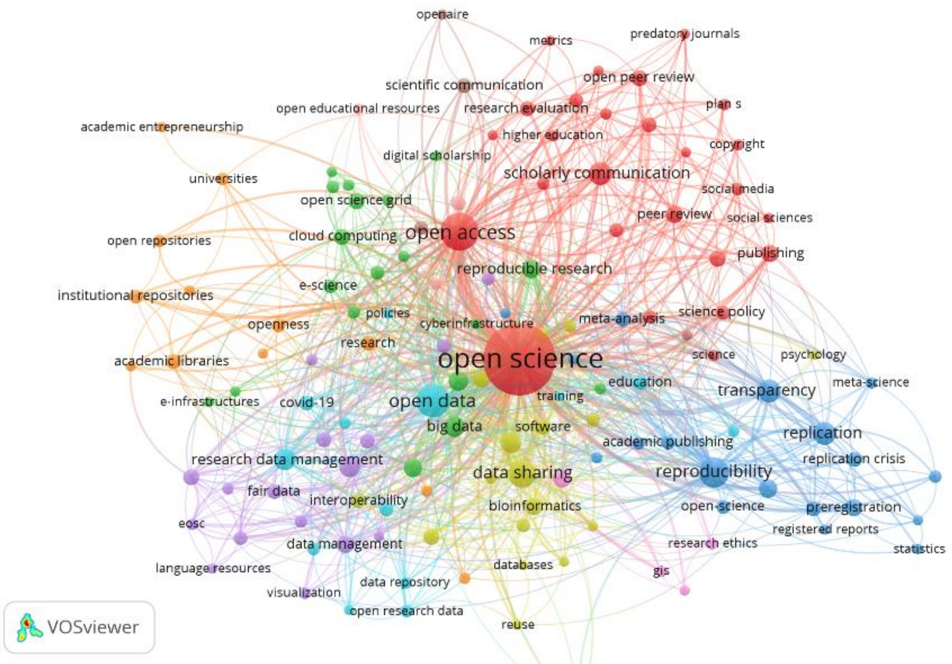

Figure 4. The Result of the Clustering of Categories in Publications Based on the Search Query "Open Science" in Scopus

Source: Compiled by the authors using VOSViewer (VOSViewer, n.d)

Thus, our bibliometric analysis of the clustering of categories of open science proved the presence of stable relationships between the concept of "open research data" and all categories. After all, the formation of the most powerful three clusters takes place around data-related concepts. Figure 5 shows the three most important stages of development and formation in scientific research of open research data and its components in the open science paradigm. From the point of view of the time dimension, the first stage falls in 2016-2017. The rapid growth of big data characterizes this period. The attention of researchers focuses on high-performance distributed computing, and computing networks. Therefore, the infrastructure of smart measurement finds its response in scientific publications, which form the first stage of the development of the concept of "open research data".

The second stage falls in the period from 2017 to 2018. It is characterized by the transition from smart measurement infrastructure to the formation of the concept of "open research data" as such. Therefore, in this period, the importance of unimpeded open access to research data becomes a fundamental issue for the scientific community and society. At the same time, national initiatives on open science and initiatives of funding organizations become drivers of the formation of this period. The last the third stage falls in 2018-2019. The data-driven open science movement is undeniably helping to increase open data research. On the other hand, it is during this period that researchers need qualitative data rather than increasing their quantitative indicators becomes obvious. Therefore, the focus of society and the scientific community is focused on ways to provide data with qualitative characteristics.

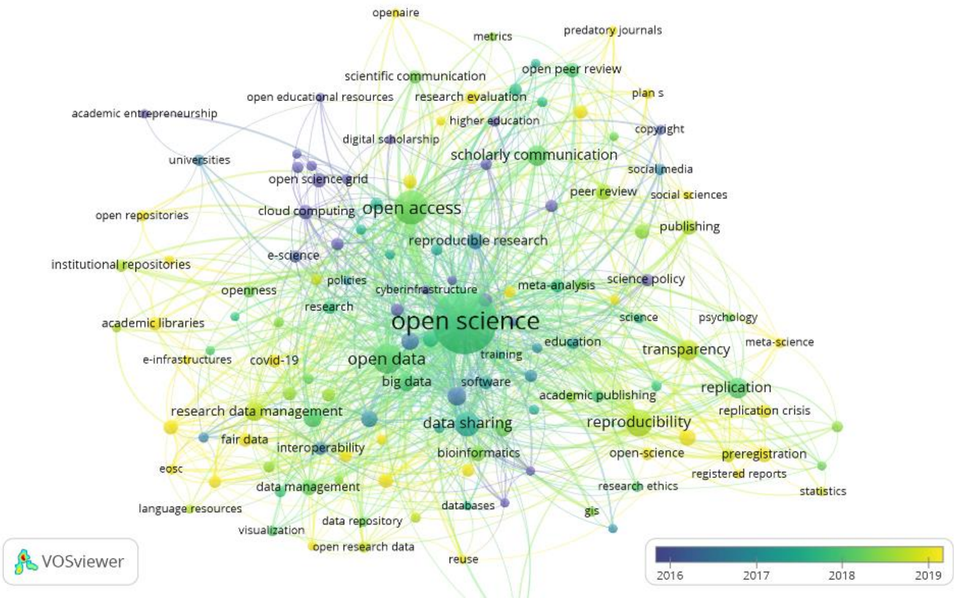

Figure 5. Visualization of the Evolution of Relationships Between Categories

Source: Compiled by the authors using VOSViewer (VOSViewer, n.d)

Conclusions

Research on open data research in the development paradigm of open science is growing and developing. Despite the rapid development of all components of open science all over the world, in Ukraine, especially in terms of disseminating scientific research data, this is still an issue that needs to be resolved at the micro and macro levels. Undoubtedly, most of the problems associated with the transition to open science regarding disseminating research data cannot be fully mitigated. Nevertheless, one effective way to build an effective data management ecosystem at the institutional level is to combine both institutional mechanisms, which, among other things, include the implementation of data disclosure policies and training programs on research data management in higher education institutions and the involvement of external stakeholders, as well as infrastructural mechanisms, which mostly cover a combination of tools related to technical issues, such as the development of open environments for data sharing and reuse, such as repositories, archives, portals and platforms; data management tools, their audit, analysis and visualization, etc.

Author Contributions: Conceptualization: Anton Boiko and Olha Kramarenko; methodology: Anton Boiko and Olha Kramarenko; software: Anton Boiko and Olha Kramarenko; validation: Anton Boiko, Olha Kramarenko and Tetyana Mayboroda; formal analysis Olha Kramarenko; investigation: Anton Boiko and Olha Kramarenko; resources: Olha Kramarenko and Tetyana Mayboroda; data curation: Anton Boiko and Olha Kramarenko; writing-original draft preparation: Anton Boiko and Olha Kramarenko; writing-review and editing: Anton Boiko, Olha Kramarenko and Tetyana Mayboroda; visualization: Olha Kramarenko; supervision: Anton Boiko; project administration: Olha Kramarenko; funding acquisition: Tetyana Mayboroda.

Conflicts of Interest: Authors declare no conflict of interest.

Data Availability Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

References

1. Amsterdam Call for Action on Open Science (2016). Information from the Government of The Netherlands. Available at: [Link]
2. Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 205395171559463. [Google Scholar] [CrossRef]
3. Bartling S., Friesike S. (2014). Towards Another Scientific Revolution. Opening Science. Springer, Cham. [Google Scholar] [CrossRef]

4. Björk, B.-C. (2017). Scholarly journal publishing in transition- from restricted to open access. *Electronic Markets*, *27*(2), 101-109. [Google Scholar] [CrossRef]
5. Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*, *2*. [Google Scholar] [CrossRef]
6. Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*, 2. [Google Scholar] [CrossRef]
7. Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, *66*(9), 1747-1762. [Google Scholar] [CrossRef]
8. CERN (n.d.). The birth of the Web. Available at: [Link]
9. Class, B., de Bruyne, M., Wuillemin, C., Donzé, D., & Claivaz, J.-B. (2021). Towards Open Science for the Qualitative Researcher: From a Positivist to an Open Interpretation. *International Journal of Qualitative Methods*, *20*, 160940692110346. [Google Scholar] [CrossRef]
10. De Smedt, K., Koureas, D., & Wittenburg, P. (2020). FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. *Publications*, *8*(2), 21. [Google Scholar] [CrossRef]
11. Easterbrook, S.M. (2014). Open code for open science? *Nature Geoscience*, *7*(11), 779-781. [Google Scholar] [CrossRef]
12. Fecher B., Friesike S. (2014) Open Science: One Term, Five Schools of Thought. In: Bartling S., Friesike S. (eds) Opening Science. Springer, Cham. [Google Scholar] [CrossRef]
13. Friesike, S., Widenmayer, B., Gassmann, O., & Schildhauer, T. (2014). Opening science: towards an agenda of open science in academia and industry. *The Journal of Technology Transfer*, *40*(4), 581-601. [Google Scholar] [CrossRef]
14. Hajduk, G.K., Jamieson, N.E., Baker, B.L., Olesen, O.F., & Lang, T. (2019). It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too! *BMJ Global Health*, *4*(4), Article e001550. [Google Scholar] [CrossRef]
15. Hauschke, C., Nazarovets, S., Altemeier, F., & Kaliuzhna, N. (2021). Roadmap to FAIR Research Information in Open Infrastructures. *Journal of Library Metadata*, *21*(1-2), 45-61. [Google Scholar] [CrossRef]
16. Hey, T. (2005). Cyberinfrastructure for e-Science. *Science*, *308*(5723), 817-821. [Google Scholar] [CrossRef]
17. Hey, T., & Payne, M.C. (2015). Open science decoded. *Nature Physics*, *11*(5), 367-369. [Google Scholar] [CrossRef]
18. Horizon Europe (HORIZON): Programme Guide. (2021). European Commission. Available at: [Link]
19. Jamali, D., Barkemeyer, R., Leigh, J., & Samara, G. (2020). Open Access, Open Science, and Coronavirus: Mega trends with historical proportions. *Business Ethics: A European Review*, *29*(3), 419-421. [Google Scholar] [CrossRef]
20. Kattge, J., Díaz, S., & Wirth, C. (2014). Of carrots and sticks. *Nature Geoscience*, *7*(11), 778-779. [CrossRef]
21. Koltay, T. (2020). Quality of Open Research Data: Values, Convergences and Governance. *Information*, *11*(4), 175. [Google Scholar] [CrossRef]
22. Kovalchuk, A., Khanzhyn, V., & Kudlatskyi, Ya. (2018). Ekonomichnyi potentsial vidkrytykh danykh v Ukraini. [Economic potential of the open data in Ukraine] TAPAS. Available at: [Link]
23. Matyushenko, I.Y., Khaustova, V.Y., & Kniaziev, S.I. (2017). Institutional Support of Innovative R&D in the Formation of Single Research Area in the EU and Ukraine. *Science and innovation*, *13*(2), 5-23. [Google Scholar] [CrossRef]
24. McKiernan, E.C., Bourne, P.E., Brown, C.T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B.A., Ram, K., Soderberg, C.K., Spies, J.R., Thaney, K., Updegrove, A., Woo, K.H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, *5*. [Google Scholar] [CrossRef]
25. Ministry of economic development and trade of Ukraine. (2017). Sustainable Development Goals: Ukraine. Available at: [Link]
26. Mueller, S.T., & Piper, B.J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250-259. [Google Scholar] [CrossRef]
27. Murray-Rust P., Neylon Cameron, Pollock Rufus, Wilbanks John (2010). Panton Principles, Principles for open data in science. Available at: [Link]
28. Murray-Rust, P. (2008). Open Data in Science. Nature Precedings. [CrossRef]

29. Nosek, B.A., Ebersole, C.R., DeHaven, A.C., & Mellor, D.T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. [Google Scholar] [CrossRef]
30. OECD (2015). Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. [CrossRef]
31. Open Definition (a). Defining open in open data, open content and open knowledge. Open Definition 2.1. (n.d.). The Open Knowledge Foundation. Available at: [Link]
32. Open Definition (b). Defining Open in Open Data, Open Content and Open Knowledge. (n. d.). Open Knowledge Foundation. Available at: [Link]
33. Open innovation, open science, open to the world: a vision for Europe (2016). European Commission, Directorate-General for Research and Innovation. [CrossRef]
34. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. [Google Scholar] [CrossRef]
35. Piwowar, H.A., & Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. [Google Scholar] [CrossRef]
36. Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, Article e4375. [Google Scholar] [CrossRef]
37. Plan S. (2018). Part III: Technical Guidance and Requirements. Available at: [Link]
38. Pordes, R., Petravick, D., Kramer, B., Olson, D., Livny, M., Roy, A., Avery, P., Blackburn, K., Wenaus, T., Würthwein, F., Foster, I., Gardner, R., Wilde, M., Blatecky, A., McGee, J., & Quick, R. (2007). The open science grid. *Journal of Physics: Conference Series*, *78*, 012057. [CrossRef]
39. Price-Whelan, A.M., Sipőcz, B.M., Günther, H.M., Lim, P.L., Crawford, S.M., Conseil, S., Shupe, D. L., Craig, M.W., Dencheva, N., Ginsburg, A., VanderPlas, J.T., Bradley, L.D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T.L., Cruz, K.L., Robitaille, T.P., Tollerud, E.J., Ardelean, C., Zabalza, V. (2018). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *The Astronomical Journal*, *156*(3), 123. [Google Scholar] [CrossRef]
40. Pro naukovu i naukovo-tekhnichnu diialnist, Zakon Ukrainy № 848-VIII (2021). (Ukraine). Available at: [Link]
41. Pro vykonannia Uhody pro asotsiatsiiu mizh Ukrainoiu, z odniiei storony, ta Yevropeiskym Soiuzom, Yevropeiskym spivtovarystvom z atomnoi enerhii i yikhnimy derzhavamy-chlenamy, z inshoi storony, Postanova Kabinetu Ministriv Ukrainy № 1106 (2020) (Ukraina). Available at: [Link]
42. Pro zatverdzhennia dorozhnoi karty z intehratsii naukovo-innovatsiinoi systemy Ukrainy do yevropeiskoho doslidnytskoho prostoru. (2021). Ministerstvo osvity i nauky Ukrainy. Available at: [Link]
43. Ramjoué, C. (2015). Towards Open Science: The vision of the European Commission. *Information Services & Use*, *35*(3), 167-170. [Google Scholar] [CrossRef]
44. Scopus (n.d.). Official site. Available at: [Link]
45. Suber, P. (2012). What is open access? In Open access. The MIT Press. Available at: [Link]
46. Svystunov, S.Ya., Shevchenko, A.Yu. (2014). Stan ta perspektyvy rozvytku Ukrainskoho natsionalnoho hrid. Analiz ta lohika mozhlyvoho rozvytku. [State and prospects of the development of the Ukrainian National Grid. Analysis and logic of possible development]. *Systemni doslidzhennia ta informatsiini tekhnolohii*, *2*, 40-52. Available at: [Link]
47. UNESDOC Digital Library (2021). UNESCO Recommendation on Open Science. Available at: [Link]
48. Van den Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). Managing and sharing data. University of Essex. Available at: [Link]
49. Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428-436. [Google Scholar] [CrossRef]
50. VOSViewer (n.d). Official site. Available at: [Link]
51. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1). [Google Scholar] [CrossRef]