Education and Information Technologies (2024) 29:17327–17362
https://doi.org/10.1007/s10639-024-12507-3

Comparative analysis of downloads of articles by Chinese and Russian researchers using Sci-Hub

Zhenxia Gao[1] · Vladimir M. Moskovkin[2] · Nina N. Gakhova[3] · Igor V. Boichuk[4]

Received: 18 June 2023 / Accepted: 23 January 2024 / Published online: 23 February 2024

Abstract
Approximately 4.4 million articles downloaded by Chinese researchers were identified based on downloads of 28 million articles from September 1, 2015, to February 29, 2016 posted on the Internet by J. Bohannon and A. Elbakyan. They were distributed by publishers of scientific periodicals and cities in comparison with article downloads by Russian researchers. Strongly skewed distributions of these downloads were obtained towards the leading cities and publishers. It is shown that the largest downloads of articles are observed in cities concentrated in the Central, Eastern and Southern provinces of China, as well as in the capital, central and eastern regions of Russia. At the same time, downloading of articles by scientists from the first three cities of China was 1.9 times higher than such downloading for the first three cities of Russia. It is shown that for both countries, the largest number of downloads of articles by their scientists falls on journals published by Elsevier, Springer, American Chemical Society and Wiley Blackwell. At the same time, the downloading of articles by Chinese scientists from the journals of the first three publishing houses was 2.8 times higher than the downloading of articles by Russian scientists from the journals of the same publishers. In conclusion, clustering of publications based on discussions around Sci-Hub has been done Sci-Hub, obtained as a result of a search through Google Scholar and Semantic Scholar. Six clusters of publications were identified, of which the largest were clusters of publications on the legitimacy, ethics, accessibility and countermeasures of Sci-Hub activity. Methods for implementing publishing and library measures to neutralize the activities of scientific piracy are proposed. The first measure was related to the dumping price for the sale of electronic copies of articles by publishers, and the second was related to the creation of an online platform for interlibrary loan of research articles.

Keywords Sci-Hub · Elsevier · Springer · J. Bohannon · A. Elbakyan · Russia · Downloads of articles

Extended author information available on the last page of the article

1 Introduction

Since September 2011, after the pirate resource Sci-Hub was launched, all publications about it were more emotional and journalistic in nature. This topic became a subject of scientific discussion following the publication of data requested by 28 million users in Sci-Hub between September 1, 2015, and February 29, 2016, by Elbakyan & Bohannon (2016) in the public domain. This publication enabled researchers to analyze the utilization of this resource in their respective countries and specific areas of scientific research.

According to the study, Iran had more than 2.6 million download requests, while India had 3.4 million and China had 4.4 million. Among the cities included in the data for the six-month period, Tehran had the highest usage (Bohannon, 2016). The study revealed that the resource in question is utilized not only by researchers from developing countries, who face challenges in accessing expensive subscription journals, but also by researchers from developed countries. Surprisingly, a significant proportion of requests (25%) originate from OECD countries, where access to subscription journals is already easy.

The survey results obtained by Travis (2016) showed that 17% of respondents consider access to full texts through Sci-Hub much easier than through legal channels. Furthermore, the study revealed that 37% of the participants faced legal restrictions in accessing the desired articles. Notably, 23% of the respondents opted for Sci-Hub as they expressed their disagreement with the pricing strategies employed by prominent commercial scientific publishers.

Similar conclusions were drawn by Kramer (2016) relying on publicly known university IP addresses. In the city of Utrecht, it was discovered that 60% of Sci-Hub downloads were accessible through Utrecht University Library's subscriptions, while 15% of downloads were available through Gold Open Access or other free sources provided by the publishers. Additionally, 25% of downloads were not immediately accessible. This case study implies that some individuals resort to using Sci-Hub for the sake of convenience, as obtaining a PDF from a shadow library is much simpler compared to utilizing the legal access methods offered by Utrecht University. We got this information in the work (Gardner et al., 2017), since the article by Kramer (2016) is currently not in the public domain and is not available through Sci-Hub.

Singh et al. (2021), comparing the accessibility of articles in the WoS database with their accessibility through Sci-Hub and talking about the huge contrast here, concluded that legal open access models are not working.

Oxenham (2016) provides a comprehensive description of the topic in his interview with A. Elbakyan, titled 'Meet the Robin Hood of Science.' According to Oxenham, the system's efficiency is remarkably impressive and surpasses the relatively basic access methods offered to researchers at leading universities. These universities are required to allocate millions of pounds annually for such tools.

Piryani et al. (2019) demonstrated that the number of Indian research papers published in 2016 that are indexed in Web of Science and can be found in open access gold-green forms does not exceed 24%.

Parkhill (2016) uploaded TOP-100 downloads from the work (Elbakyan & Bohannon, 2016) to the PlumX tool and determined that most of the articles were from 2015, that is, Sci-Hub users prefer to download the most recent articles. Moreover, the largest number of downloads corresponded to articles from the fields of physics, engineering and life sciences.

Cabanac (2016) according to the same data found that 36% of all DOI are available on the open Library Genesis (LibGen), working in conjunction with Sci-Hub. It is noted that 68% of articles published by Elsevier, Springer and Wiley are available in LibGen. At the same time, Caffrey and Gardner (2017), according to other data, give a figure of 83% for the same publishers.

Himmelstein et al. (2017) revealed that more than 85% of scientific articles in subscription journals and more than 97% of Elsevier articles are freely available through Sci-Hub. He also showed that 68.9% of the 81.6 million scholarly articles registered with Crossref.

Björk (2017) shows that about 25% of recent research articles (after a year) can be accessed as gold OA, and another 15–20% are available as legal green copies. The overwhelming majority of the remaining 50% can be found as illegal black OA copies. Piwowar et al. (2018) also believe that initiatives such as open access have made about 50% of scientific output available without paywalls.

Two-thirds of all downloads of medical science publications through Sci-Hub occur in low-and lower-middle-income countries (Till et al., 2019), which is naturally associated with a lack of funds for journal subscriptions in these countries.

Nazarovets (2018) used data from Elbakyan & Bohannon (2016) to obtain the distribution of articles uploaded by Ukrainian researchers by publisher and region, highlighting the main areas of knowledge corresponding to these articles. He showed that chemistry, physics and astronomy accounted for 69% of articles, medical and pharmaceutical sciences - 13%, life sciences - 12%, social sciences - 6%). He also received downloads of papers from the most frequently used journals: Journal of the American Chemical Society - 6769 downloads, Organic Chemistry - 6038, Physical Rev. B - 4325, Medicinal Chemistry - 3712 downloads.

The same data was used to analyze downloads by Russian researchers in the work (Moskovkin et al., 2021), which were distributed by leading publishers and cities in comparison with downloads by Ukrainian researchers (Nazarovets, 2018). In total, over a six-month period, about 1.5 million downloaded articles by Russian researchers were identified, while the list of leading publishers in the downloads of researchers from both countries approximately coincided, and downloads of articles by Moscow researchers (about 730 thousand) exceeded downloads of articles by Kyiv researchers by 3.9 times.

For comparison, we present data for some US cities for the six-month period under consideration according to the above public data provided by Bohannon (2016). He writes that 74,000 download requests can be traced back to IP addresses in New York City where numerous universities and scientific institutions are located. 19,000 download requests came from Columbus whose population is less that 10% of New York's and another 68,000 from East Lansing, Michigan, whose population is less that 1% of New York's. In these cities are located Ohio State University and Michigan State University (MSU), respectively.

Looking specifically at Latin American countries and relying on Elbakyan and Bohannon's public data (Elbakyan & Bohannon, 2016), Machin-Mastromatteo et al. (2016) found that the top five downloaded publishers were Elsevier, Springer, Wiley Blackwell, Nature Publishing Group, and the American Chemical Society.

The same list of publishers was obtained at the global level in the work (Elbakyan & Bohannon, 2016), as well as at the regional level for article downloads by Ukrainian and Russian researchers (Nazarovets, 2018; Moskovkin et al., 2021). Obviously, when researchers from other regions of the world piracy download articles, we will see the same five publishers. Note that the first two publishers accounted for 9,296,485 and 2,630,787 downloads globally (Bohannon, 2016).

Greshake (2016) found that top five countries on a measure of downloads per 1000 inhabitants were Portugal, Iran, Tunisia, Greece, and Chile. Gross Domestic Product per capita was also positively correlated with downloads. In his calculations, he used the previously cited public data by Elbakyan & Bohannon (2016).

Timus and Babutsidze (2016), using University Association for Contemporary European Studies (UACES) access to European Studies journals, selected journals with IF (WoS) > 1 and analyzed them together with data on paper downloads (Elbakyan & Bohannon, 2016). They found that readers are mainly interested in issues related to populism, extremism and the economic crisis, and also that the leaders in European Studies are nine European countries (Spain, Italy, Germany, Ukraine, Poland, France, Russia, Belgium, Hungary) followed by Brazil, Turkey, China, and the United States.

Androćec (2017), according to the same work, studied the downloads of articles from the field of computer science, which turned out to be 5.95%. The 20 most downloaded articles are listed. The first five countries where the researchers who uploaded articles on the sciences under consideration, looked like this: India, Iran, China, USA, Indonesia. Russia in this ranked list was in seventh place with 46,659 articles.

Babutsidze (2018) studied downloads on economic topics according to the work (Elbakyan & Bohannon, 2016) for TOP-5 economic journals: American Economic Review, Quarterly Journal of Economics, Journal of Political Economy, Econometrica and Review of Economic Studies. He obtained a small number of articles from these journals, requests for which came from underdeveloped countries.

Sagemüller et al. (2021) studied downloads on development studies globally according to previous data and found that Sci-Hub is used the most by researchers from the global South, primarily from middle-income countries, whereas researchers from the poorest countries in the data set use Sci-Hub the least.

In his work, Greshake (2017) presented his own data obtained from 62 million articles that were pirated through Sci-Hub. According to his findings, there were approximately 1700 different publishers, with around 1000 of them having at least one paper downloaded. The distribution of the corpus and downloaded publications is heavily skewed towards a small number of publishers. Specifically, the top 9 publishers accounted for approximately 70% of the complete corpus and around 80% of all downloads (as of March 19, 2017).

Suh (2022) analyzed downloads from Sci-Hub by Korean scientists in 2017 and showed that the areas with most downloads were Seoul and the metropolitan area, and papers from journals in the field of natural science were downloaded the most, and about 20% of papers were in Open Access state. Both the papers published between 2010 and 2017 and IEEE's papers were the most downloaded.

A survey of 2849 employees from 30 universities in Germany, Hungary, Ireland, Italy, the Netherlands, and Sweden, conducted from June to October 2021 in (Rossello & Martinelli, 2023), showed that overall, 47% of scientists used Sci-Hub, with however, in different countries the share of scientists using Sci-Hub varying from 65% in Hungary to 32% in Sweden. In different fields, this share varied between 60% in the interdisciplinary field of Physical Sciences & Engineering and Social Sciences & Humanities to 38% in the interdisciplinary field of Social Sciences & Humanities and Life Sciences.

Monti and Unzurrunzaga (2020) obtained the following data by country for downloads from Sci-Hub for 2017: China - 24,943,832; USA – 11,991,045; Brazil – 7,004,834; Iran – 6,627,150; France – 4,210,491; Mexico - 3,272,751; Spain – 3,092,127; Colombia – 2,041,871; Chile – 1,653,808.

Significantly smaller volumes of downloads for the same year due to more thorough cleaning of the source data were obtained in the work (Geng et al., 2022): China – 11,316,634; Brazil – 5,442,464; USA – 4,082,952; Germany – 2,167,967; Mexico – 1,793,349; UK – 1,544,080; Japan – 796.035; South Africa - 195,983.

The latest data on downloads from Sci-Hub was announced in the news of the journal *Nature* by Owens (2022). They are striking, since the monthly download of articles at the beginning of 2022 for many countries was comparable to the annual download of 2017, the data for which we presented above (Monti & Unzurrunzaga, 2020; Geng et al., 2022). The leader was China (more than 25 million downloads), followed by USA (9.3), France (about 6), Brazil (2.8), India (about 2), Germany (about 2), Indonesia (1.2), Singapore (about 1), Iran (about 750,000), Mexico (750,000) and Colombia (375,000). We took the data containing "about" from the graph in the above article.

As for the revenue loss of publishers due to pirated downloads using Sci-Hub, in the work (Gardner et al., 2017) based on data from Elbakyan & Bohannon (2016) on downloads from Elsevier publishing sites (9,296,485 downloads), Royal Society of Chemistry (927,238 downloads), IOP Publishing (160,073 downloads) and JSTOR (358,786 downloads) received, respectively, the following upper bound estimate of revenue loss (6 months) –36.86 million pounds, $334.21 million, $5.28 million and $6.49 million.

In a study conducted by Parkhill (2016), the author analyzed the Elbakyan and Bohannon dataset, focusing on the 100 most downloaded articles. By estimating the cost of each item at $29.95, the study calculated a potential revenue loss of $2.7 million, which would be distributed among different publishers.

In conclusion of the review of predominantly quantitative research, we note that we note that the ratio of citation of articles downloaded through Sci-Hub to those not downloaded through it is 2.21 to 1 (Correa et al., 2020). In conclusion of the review of predominantly quantitative studies, we note that the ratio of citations of articles downloaded through Sci-Hub to those not downloaded through it is 2.21 to 1 (Correa et al., 2020). It should be noted that quite recently Maddi and Sapinho (2023) came to similar conclusions.

This review of the above quantitative studies and our further experiments in Google Scholar and Semantic Scholar found that there are no detailed studies on downloading articles via the Sci-Hub site by Chinese researchers. Below we will fill this gap, as well as summarize the thoughts of various researchers of the Sci-Hub phenomenon, and in conclusion, measures to combat pirated scientific sites will be analyzed and the most effective such measures will be proposed. As a result, we posed the following research questions:

RQ1. Based on 28 million articles downloaded by researchers from around the world using Sci-Hub and presented by Elbakyan & Bohannon (2016), using a specially developed mathematical algorithm, identify articles downloaded by Chinese researchers?

RQ2. Construct the distribution of downloads by Chinese researchers from the Sci-Hub website for the first hundred cities and for the first hundred publishers in comparison with similar downloads of Russian researchers obtained in the work (Moskovkin et al., 2021), and prove whether these distributions satisfy the Pareto principle?

RQ3.Summarize and classify the results of discussions regarding the use of the Sci-Hub site and show what the most effective measures can be proposed to neutralize its activity?

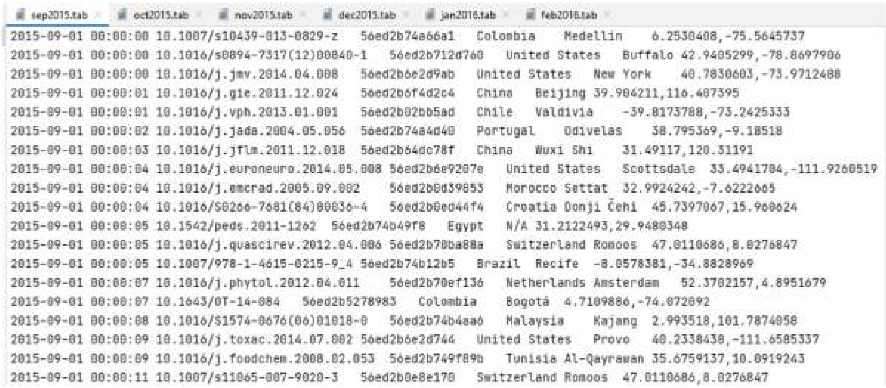
2 Materials and methods of research

To prove the absence of publications on the detailed distribution of pirated downloads using Sci-Hub by city and publisher, we conducted experiments on October 20, 2023 in the advanced search of Google Scholar and Semantic Scholar. Having proven this, in the main part of our research, we will rely on public data from the work (Elbakyan & Bohannon, 2016), which were used in the work (Moskovkin et al., 2021) to analyze downloads of articles by Russian researchers. The data on downloads of articles by Chinese and Russian scientists using Sci-Hub used in this work consisted of six files with the "*.tab" extension, each of which reflected user requests for the corresponding month. Figure 1 provides a visual representation of the data in the files.

Each file contained information about the date and time of the request, a DOI containing the publisher code and publisher article code, the user's IP address, country, city, geographic coordinates (latitude and longitude).

The difficulty of processing lies in the large amount of data, so one upload file contains several million records of downloaded publications, since the upload is carried out over six months, the total number of records is above ten million.

Since the structure of files uploaded by the Sci-Hub electronic resource is constant and unchangeable, during the research process algorithms were developed for processing primary data, cleaning and aggregating them. The block diagram of the data processor function algorithm for processing the initial data is presented in Fig. 2.



```
sep2015.tab   oct2015.tab   nov2015.tab   dec2015.tab   jan2016.tab   feb2016.tab
2015-09-01 00:00:00 10.1007/s10439-013-0829-z   56ed2b74a66a1   Colombia   Medellin   6.2530408,-75.5645737
2015-09-01 00:00:00 10.1016/s0894-7317(12)00840-1   56ed2b712d760   United States   Buffalo 42.9405299,-78.8697906
2015-09-01 00:00:00 10.1016/j.jmv.2014.04.008   56ed2b6e2d9ab   United States   New York   40.7830603,-73.9712488
2015-09-01 00:00:01 10.1016/j.gie.2011.12.024   56ed2b6f4d2c4   China   Beijing 39.904211,116.407395
2015-09-01 00:00:01 10.1016/j.vph.2013.01.001   56ed2b02bb5ad   Chile   Valdivia   -39.8173788,-73.2425333
2015-09-01 00:00:02 10.1016/j.jada.2004.05.056   56ed2b74a4d40   Portugal   Odivelas   38.795369,-9.18518
2015-09-01 00:00:03 10.1016/j.jflm.2011.12.018   56ed2b64dc78f   China   Wuxi Shi   31.49117,120.31191
2015-09-01 00:00:04 10.1016/j.euroneuro.2014.05.008 56ed2b6e9207e   United States   Scottsdale   33.4941704,-111.9260519
2015-09-01 00:00:04 10.1016/j.emcrad.2005.09.002   56ed2b0d39853   Morocco Settat   32.9924242,-7.6222665
2015-09-01 00:00:04 10.1016/S0266-7681(84)80036-4   56ed2b0ed44f4   Croatia Donji Čehi   45.7397067,15.960624
2015-09-01 00:00:05 10.1542/peds.2011-1262   56ed2b74b49f8   Egypt   N/A 31.2122493,29.9480348
2015-09-01 00:00:05 10.1016/j.quascirev.2012.04.006 56ed2b70ba88a   Switzerland Romoos   47.0110686,8.0276847
2015-09-01 00:00:05 10.1007/978-1-4615-0215-9_4 56ed2b74b12b5   Brazil   Recife   -8.0578381,-34.8828969
2015-09-01 00:00:07 10.1016/j.phytol.2012.04.011   56ed2b70ef136   Netherlands Amsterdam   52.3702157,4.8951679
2015-09-01 00:00:07 10.1643/OT-14-084   56ed2b5278983   Colombia   Bogotá   4.7109886,-74.072092
2015-09-01 00:00:08 10.1016/S1574-0676(06)01018-0   56ed2b74b4aa6   Malaysia   Kajang   2.993518,101.7874058
2015-09-01 00:00:09 10.1016/j.toxac.2014.07.002 56ed2b6e2d744   United States   Provo   40.2338438,-111.6585337
2015-09-01 00:00:09 10.1016/j.foodchem.2008.02.053   56ed2b749f89b   Tunisia Al-Qayrawan 35.6759137,10.0919243
2015-09-01 00:00:11 10.1007/s11065-007-9020-3   56ed2b0e8e170   Switzerland Romoos   47.0110686,8.0276847
```

Fig. 1 Visualization of data from the work (Elbakyan & Bohannon, 2016) in files

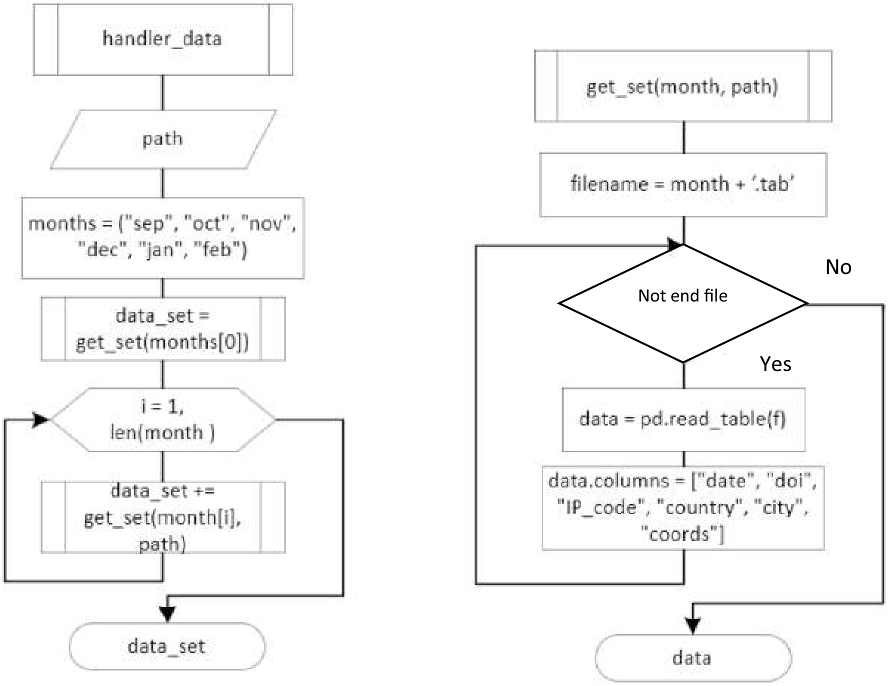


Fig. 2 Algorithm of the handler data function for processing files downloaded from Sci-Hub

The input of the called function is the directory path for storing the downloaded files. Data is uploaded for six months, starting from September 1, 2015 (Elbakyan & Bohannon, 2016), for this purpose a special variable is created that initializes the list of months. Next, the function calls the get_set subfunction, which takes two parameters: the month being processed and the path to the file. The subfunction forms the path and creates a temporary table in which the result of reading the file is placed, then column headers are formed for the table and the data is returned to the main function, where the execution result is added to the data_set variable. The developed algorithm was implemented using the Python programming language using the PyCharm integrated environment.

Poor quality data was identified among the downloaded data. As you can see in Fig. 1, some records may lack data on city names, which is marked N/A.

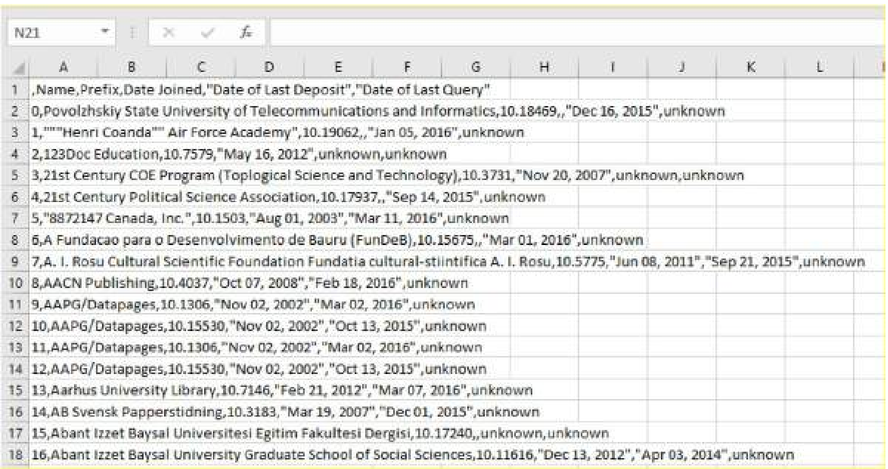
Together with these data, a DOI file was uploaded, the view of which is shown in Fig. 3, in the *.csv format, which contains: the name of the publisher and its prefix, the date of the last save and request.

Figure 3 shows that some records initially do not contain all the information about the downloaded articles, they indicate “unknown”. Data on downloading articles by Russian researchers were obtained and analyzed, which were distributed by city and publishing house. As a result, 1,492,170 downloads were made out of 521 cities in Russia (Moskovkin et al., 2021) and 1,513,292 downloads were made on articles from 583 publishers (the given data on publishers were specified in this study through more careful aggregation of downloads by branches of large publishers).

The work (Moskovkin et al., 2021) describes the results of processing source files with requests only from Russian IP addresses using the PyCharm development environment and the Python programming language. In this study, by analogy with the above work, data were selected at the request of Chinese researchers. Some of the information contained in the DOI files was also not recognized, since some records initially did not contain all the information about the downloaded articles.

An example of processing download data for Chinese cities using PyCharm development environment and the Python programming language is shown in Fig. 4.

When processing the data, it was noticed that the total number of downloads by country is not equal to the total number of downloads by city. The reason is in the source files: some data rows are missing the city name, instead there is “N/A”. An example of rows with missing city names, instead of which there is N/A, is shown in Fig. 5.


N21

	A
1	,Name,Prefix,Date Joined,"Date of Last Deposit","Date of Last Query"
2	0,Povolzhskiy State University of Telecommunications and Informatics,10.18469,,"Dec 16, 2015",unknown
3	1,"""Henri Coanda"" Air Force Academy",10.19062,,"Jan 05, 2016",unknown
4	2,123Doc Education,10.7579,"May 16, 2012",unknown,unknown
5	3,21st Century COE Program (Toplogical Science and Technology),10.3731,"Nov 20, 2007",unknown,unknown
6	4,21st Century Political Science Association,10.17937,,"Sep 14, 2015",unknown
7	5,"8872147 Canada, Inc.",10.1503,"Aug 01, 2003","Mar 11, 2016",unknown
8	6,A Fundacao para o Desenvolvimento de Bauru (FunDeB),10.15675,,"Mar 01, 2016",unknown
9	7,A. I. Rosu Cultural Scientific Foundation Fundatia cultural-stiintifica A. I. Rosu,10.5775,"Jun 08, 2011","Sep 21, 2015",unknown
10	8,AACN Publishing,10.4037,"Oct 07, 2008","Feb 18, 2016",unknown
11	9,AAPG/Datapages,10.1306,"Nov 02, 2002","Mar 02, 2016",unknown
12	10,AAPG/Datapages,10.15530,"Nov 02, 2002","Oct 13, 2015",unknown
13	11,AAPG/Datapages,10.1306,"Nov 02, 2002","Mar 02, 2016",unknown
14	12,AAPG/Datapages,10.15530,"Nov 02, 2002","Oct 13, 2015",unknown
15	13,Aarhus University Library,10.7146,"Feb 21, 2012","Mar 07, 2016",unknown
16	14,AB Svensk Papperstidning,10.3183,"Mar 19, 2007","Dec 01, 2015",unknown
17	15,Abant Izzet Baysal Universitesi Egitim Fakultesi Dergisi,10.17240,,unknown,unknown
18	16,Abant Izzet Baysal University Graduate School of Social Sciences,10.11616,"Dec 13, 2012","Apr 03, 2014",unknown

Fig. 3 An example of the composition of the DOI file

Fig. 4 Processed data for Chinese cities with the number of downloads

	A	B
1	city	China
2	Anqing Shi	479
3	Anshan Shi	1708
4	Anshun Shi	138
5	Anyang Shi	1913
6	Baise Shi	167
7	Baishan Shi	7
8	Baiyin Shi	75
9	Baoding Shi	24560
10	Baoji Shi	140
11	Baotou Shi	7327
12	Bayinguoleng Men	361
13	Bazhong Shi	2
14	Beihai Shi	24
15	Beijing	746797
16	Bengbu Shi	41

While processing the original DOI file, it turned out that when selecting the names of publishers by prefix, the number of downloads by publisher does not correspond to the total number of downloads in the country. The reason for the discrepancy is duplicated rows with information on the publisher. When the number of downloads is processed and publisher names are found by prefixes, a join of two date frames is used, similar to join in SQL. Thus, duplicate rows are also taken into account and this results in an excessive number of downloads. An example of duplicate rows in a DOI file is shown in Fig. 6.

```
2015-10-06 17:55:28 10.1016/j.mib.2005.06.002    56ed2bcfeb96d   China   Hengyang Shi    26.893369,112.57
2015-10-06 17:55:28 10.1080/03089740310002498    56ed2b0bae5c3   Portugal    Lisboa  38.71573,-9.1338549
2015-10-06 17:55:28 10.1016/0921-4526(96)00008-7    56ed2b455b3aa   China   Xiamen Shi  24.479834,118.08
2015-10-06 17:55:28 10.1111/j.1432-1033.1979.tb13163.x  56ed2aff5eb72   Pakistan    Islamabad   33.72938
2015-10-06 17:55:28 10.1016/S0271-5317(00)80024-X   56ed2b38cf422   Argentina   N/A -27.4733434,-55.8913
2015-10-06 17:55:28 10.1039/C4AN00201F   56ed2b37ddcd0   Malaysia    Johor Bahru 1.492659,103.7413591
```

Fig. 5 Rows with missing city names (indicated N/A)

5237	5235,Springer Fachmedien Wiesbaden GmbH,10.1365,"Apr 03, 2003","Mar 07, 2016","Mar 10, 2016"
5238	5236,Springer Publishing Company,10.1891,"Aug 12, 2004","Feb 19, 2016","Oct 05, 2015"
5239	5237,Springer-Verlag,10.1007,"Jul 01, 2000","Mar 11, 2016","Mar 11, 2016"
5240	5238,Springer-Verlag,10.1140,"Jul 01, 2000","Mar 11, 2016","Mar 11, 2016"
5241	5239,Springer-Verlag,10.1007,"Jul 01, 2000","Mar 11, 2016","Mar 11, 2016"
5242	5240,Springer-Verlag,10.1140,"Jul 01, 2000","Mar 11, 2016","Mar 11, 2016"
5243	5241,SPVS - Sociedade Portuguesa de Vida Selvagem,10.2461,"Nov 28, 2005","Dec 30, 2015","Dec 17, 2013"

Fig. 6 Duplicate rows in the DOI file

The final part of the article analyzes the discussions around Sci-Hub in the context of the legitimacy, scientific ethics of its use and ease of accessibility, and more, the identification of academic user groups and coverage by these groups, as well as its connection with the Open Access movement. In addition, a set of measures will be developed to neutralize the activities of pirated scientific sites.

3 Results and discussions

An advanced search in Google Scholar for the term Sci-Hub together with the words China or Chinese yields about 4000 responses; viewing the first thousand of them, which the search engine allows, shows the lack of work on the detailed distribution of illegal downloads by Chinese researchers by city and publisher. It should be noted that Wan (2022), when searching for the keyword Sci-Hub.tw in this search engine, received 531 results containing this keyword. Our Semantic Scholar search for the term Sci-Hub returned 353 publications, with the first 126 containing the term in article titles (a Google Scholar search for the term in article titles returned 287 publications). In these experiments, no publications were found on the detailed distribution of downloads from Sci-Hub by Chinese researchers. The experiments were carried out on October 20. 2023. In this regard, we turned to the data from (Elbakyan & Bohannon, 2016) and the methodology for their processing described in the previous section.

As a result of data recognition for inhabited localities with Chinese IP addresses, 4,435,500 downloads of articles suitable for analysis were identified. We note that the same number of downloads was rounded off by Elbakyan & Bohannon (2016), cited in Introduction (4.4 million) with reference to (Bohannon, 2016).

After that, these records were distributed by cities, which turned out to be 253. Article downloads for the first 100 cities are presented in Table 1, where for comparison, data on the distribution of downloaded articles and Russian cities are shown (Moskovkin et al., 2021).

It should be noted that in Table 1, the prefix Shi means that an inhabited locality belongs to a city; for large cities under numbers 1, 2, 9, it is not used. Localities with two names belong to autonomous regions (only one autonomous region is marked at number 53 in the Top 100 Cities). As you can see, the largest downloads of articles are observed in cities concentrated in the Central, Eastern and Southern provinces of China, as well as in the capital, central and eastern regions

Table 1 Distribution of articles downloaded from Sci-Hub by Chinese and Russian researchers by cities in China and Russia

No	Chinese City	Downloads from Sci-Hub	No	Russian City	Downloads from Sci-Hub
1.	Shanghai	764,397	1	Moscow	731,134
2.	Beijing	746,797	2	Sankt-Peterburg	132,623
3.	Guangzhou Shi	230,957	3	Novosibirsk	57,508
4.	Wuhan Shi	227,827	4	Kazan	55,138
5.	Nanjing Shi	192,035	5	Tomsk	26,412
6.	Chengdu Shi	162,158	6	Nizhnij Novgorod	25,508
7.	Lanzhou Shi	134,120	7	Yekaterinburg	22,024
8.	Changsha Shi	131,589	8	Korolev	20,589
9.	Tianjin	123,871	9	Samara	19,401
10.	Xian Shi	109,692	10	Voronez	18,962
11.	Jinan Shi	98,137	11	Velikiy Novgorod	17,723
12.	Zhengzhou Shi	95,114	12	Irkutsk	16,752
13.	Dingxi Shi	91,478	13	Saratov	16,678
14.	Shenyang Shi	73,951	14	Rostov	15,260
15.	Haerbin Shi	72,754	15	Perm	14,740
16.	Changchun Shi	72,306	16	Krasnoyarsk	14,576
17.	Chongqing	66,651	17	Chelyabinsk	13,209
18.	Hangzhou Shi	65,674	18	Ivanovo	11,643
19.	Qingdao Shi	57,635	19	Ufa	10,905
20.	Nanchang Shi	55,132	20	Volgograd	10,798
21.	Fuzhou Shi	52,028	21	Krasnodar	10,071
22.	Kunming Shi	49,704	22	Vladivostok	9794
23.	Hefei Shi	47,627	23	Syktyvkar	9693
24.	Wenzhou Shi	45,232	24	Kemerovo	7200
25.	Suzhou Shi	43,259	25	Yaroslavl’	7172
26.	Wuxi Shi	37,389	26	Omsk	6934
27.	Dalian Shi	37,050	27	Solnechnoye	6776
28.	Dongguan Shi	35,297	28	Belgorod	6070
29.	Nanning Shi	28,715	29	Chernogolovka	6034
30.	Zhangjiakou Shi	28,239	30	Kaliningrad	5964
31.	Guiyang Shi	26,288	31	Stavropol’	4795
32.	Taiyuan Shi	25,664	32	Obninsk	4314
33.	Baoding Shi	24,560	33	Izhevsk	4205
34.	Xiamen Shi	24,098	34	Petergof	3457
35.	Ningbo Shi	22,737	35	Astrakhan’	3384
36.	Yiyang Shi	22,675	36	Pushchino	3125
37.	Wulumuqi Shi	16,732	37	Fryazino	3040
38.	Yangzhou Shi	14,242	38	Gatchina	2935
39.	Zhenjiang Shi	12,649	39	Kaluga	2892
40.	Haikou Shi	12,024	40	Berdsk	2860

Table 1 (continued)

No	Chinese City	Downloads from Sci-Hub	No	Russian City	Downloads from Sci-Hub
41.	Xianyang Shi	11,790	41	Ryazan’	2837
42.	Changzhou Shi	9480	42	Mytishchi	2732
43.	Yichang Shi	9179	43	Petrozavodsk	2695
44.	Yinchuan Shi	7811	44	Khabarovsk	2507
45.	Huhehaote Shi	7415	45	Tula	2409
46.	Baotou Shi	7327	46	Zhukovskiy	2392
47.	Mianyang Shi	6899	47	Tver’	2365
48.	Yantai Shi	6693	48	Barnaul	2351
49.	Kaifeng Shi	6576	49	Tolyatti	2293
50.	Rikaze Shi	6463	50	Arkhangel’sk	2230
51.	Zhangzhou Shi	6205	51	Kirov	2214
52.	Hengshui Shi	5956	52	Vladimir	2174
53.	Yanbian Chaoxianzuzizhizhou	5823	53	Dubna	2163
54.	Guilin Shi	5658	54	Lipeck	2108
55.	Dezhou Shi	5635	55	Tyumen’	2050
56.	Taian Shi	5609	56	Makhachkala	2012
57.	Xuzhou Shi	5503	57	Odintsovo	1994
58.	Jiaxing Shi	4820	58	Saransk	1967
59.	Zhanjiang Shi	4812	59	Podolsk	1935
60.	Jiaozuo Shi	4797	60	Chekhov	1924
61.	Xinyang Shi	4504	61	Kursk	1880
62.	Yaan Shi	4160	62	Ulyanovsk	1796
63.	Jieyang Shi	3682	63	Lyubertsy	1792
64.	Foshan Shi	3447	64	Lomonosov	1740
65.	Shaoxing Shi	3349	65	Dmitrov	1723
66.	Luoyang Shi	3297	66	Tambov	1712
67.	Ningde Shi	3237	67	Yakutsk	1658
68.	Zhuhai Shi	3017	68	Bryansk	1634
69.	Ganzhou Shi	2945	69	Taganrog	1472
70.	Qinhuangdao Shi	2904	70	Yoshkar-Ola	1453
71.	Jinzhou Shi	2853	71	Nakhodka	1354
72.	Nantong Shi	2848	72	Apatity	1347
73.	Jining Shi	2845	73	Magnitogorsk	1344
74.	Taizhou Shi	2760	74	Ivanovskoye	1270
75.	Liuzhou Shi	2574	75	Sarov	1264
76.	Xining Shi	2274	76	Oktyabrskiy	1214
77.	Xingtai Shi	2142	77	Novocheboksarsk	1169
78.	Wuhu Shi	2063	78	Stolbovaya	1125
79.	Zhuzhou Shi	1963	79	Oryol	1102
80.	Zibo Shi	1934	80	Volobuevo	1095
81.	Anyang Shi	1913	81	Volkhonshchino	1086

Table 1 (continued)

No	Chinese City	Downloads from Sci-Hub	No	Russian City	Downloads from Sci-Hub
82.	Hanzhong Shi	1829	82	Cheboksary	1069
83.	Anshan Shi	1708	83	Reutov	1054
84.	Huizhou Shi	1695	84	Chkalov	937
85.	Lishui Shi	1676	85	Kuban’	906
86.	Handan Shi	1641	86	Vidnoye	878
87.	Sanming Shi	1586	87	Penza	877
88.	Weifang Shi	1479	88	Snezhinsk	839
89.	Xinxiang Shi	1406	89	Protvino	819
90.	Jiamusi Shi	1348	90	Krasnogorsk	781
91.	Jian Shi	1316	91	Kirovsk	779
92.	Jiangmen Shi	1310	92	Sergiyev Posad	772
93.	Lianyungang Shi	1301	93	Surgut	764
94.	Weihai Shi	1284	94	Smolensk	761
95.	Maanshan Shi	1282	95	Vladikavkaz	745
96.	Hengyang Shi	1206	96	Lobnya	744
97.	Pingdingshan Shi	1199	97	Balashikha	717
98.	Maoming Shi	1145	98	Dzerzhinskiy	714
99.	Quanzhou Shi	1068	99	Domodedovo	706
100.	Huaibei Shi	1054	100	Lytkarino	681

of Russia. At the same time, downloading of articles by scientists from the first three cities of China is 1.9 times higher than such downloading for the first three cities of Russia, and downloading of articles by scientists from the 100th city of China by the ordinal number is approximately 1.6 times higher than such downloading for the 100th by the ordinal city of Russia. If Shanghai accounted for 17.2% of the total number of downloads, then Moscow accounted for 49%, which is very close to the data given by Bohannon (2016) for this city (48%).

Features of the distribution curves of downloaded articles for Top - 50 cities are shown in Fig. 7.

On the interval Top - 10 cities, the curves are broken lines with sharp breaks. There is a very sharp decline in the curves, which suggests that a small number of cities account for the lion’s share of article downloads. We also see that the curve for downloading articles by Chinese researchers is higher than the similar curve for downloading articles by Russian researchers.

Data processing was also carried out on the distribution of downloaded articles by Chinese researchers by publishers. It should be noted that in the obtained data on publishers, some names of publishers were not indicated or not recognized. In addition, the list of publishers sometimes included the names of journals (Fig. 8). Such incorrect data were not taken into account in further analysis.

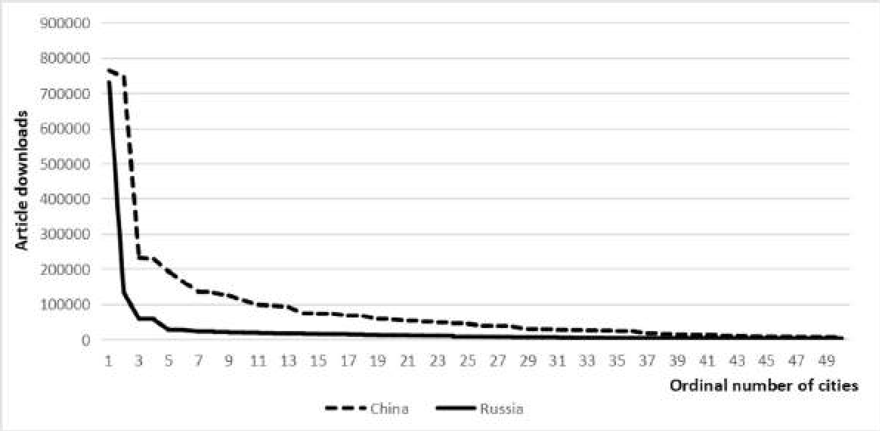


Fig. 7 Downloads of articles by Chinese and Russian researchers by Top - 50 cities

It should be noted that in Fig. 8, the list of publishers includes journals that are listed under numbers 31 and 33. Of the prestigious journals with large downloads, we note the New England Journal of Medicine and BMJ.

The number of incorrect downloads of articles with missing or duplicate publisher names and journal names was 40,890 out of a total of 4,456,076 downloads. Thus, 4,415,186 (4,456,076 - 40,890) downloads of articles turned out to be suitable for processing, which were distributed among 627 publishers. With this distribution, affiliated publishing houses were merged with their main offices. This mainly concerned publishers Elsevier, Springer, Wiley, Informa UK, Bio One, Ovid Technologies. Thus, 4,415,186 downloads were analyzed, which are grouped by 627 publishers.

Table 2 presents data on the first 100 publishers downloaded from Sci-Hub articles by Chinese researchers, sorted by the number of downloaded articles. For comparison, this table contains data on downloading articles by Russian researchers. These were the data obtained in the work (Moskovkin et al., 2021), but corrected in this study (merging subsidiary publishers with the head office of the publisher, removing downloads with the titles of journal articles that were erroneously included in the sample for publishers). As a result, more careful aggregation of article downloads by Russian researchers by publisher in this study allowed us to obtain more accurate data compared to the underestimated data of the work (Moskovkin et al., 2021). For example, data on downloads from journals published by Elsevier, Springer, Wiley Blackwell and Informa UK were 2.3%, 8.75%, 43.2% and 33.2% higher, respectively.

As noted in the previous section of the article, the total adjusted data was: 1,513,292 downloads, 583 publishers.

As we can see for both countries, the largest number of downloads of articles by their scientists comes from journals published by Elsevier, Springer, American Chemical Society and Wiley Blackwell. At the same time, the downloading of articles by Chinese scientists from the journals of the first three publishing houses is 2.8

24	Name	downloads
25	Chemical Science Transactions	1
26	Wildlife Disease Association	85
27	World Literature Today	1
28	Scrivener Publishing	13
29	Japan Society of Applied Physics	1299
30	Physical Society of Japan	446
31	Journal of Biomedical Research	8
32	eLife Sciences Organisation, Ltd.	14
33	Revista Mexicana de Biodiversidad	1
34	American Podiatric Medical Association	76
35		14
36		3
37		1
38	James Nicholas Publishers	2
39	Astro, Ltd.	11
40	American College of Physicians	85
41	Society of CAD/CAM Engineers	1
42	Asian Pacific Organization for Cancer Prevention	1
43	Columbia University Press	1
44	National Science Museum of Korea	2
45	Manchester University Press	27
46	AMSUS	2

Fig. 8 An example of incorrect data with missing names of publishers and the presence of journal names

times higher than the downloading of articles by Russian scientists from the journals of the same publishers, and the downloading of articles by Chinese scientists from the journals of the 100th publishing house by serial number is approximately 3.4 times higher than such downloading of articles by Russian scientists for the corresponding 100th by publisher number. Features of the distribution curves of downloaded articles for the Top - 100 publishers are shown in Fig. 9.

The resulting curves are broken lines with sharp breaks in the interval Top - 20 publishers. If we draw these curves on a normal scale, we would see a very sharp decline in the curves similar to the curves in Fig. 7, which suggests that a small number of publishers account for the lion's share of article downloads. Just as in Fig. 7, the curve for article downloads by Chinese researchers is higher than the similar curve for article downloads by Russian researchers.

Table 2 Comparative distribution of downloaded articles by Chinese and Russian researchers by publisher, sorted by the number of downloads (top 100 places)

	China			Russia	
№	Name	Downloads		Name	Downloads
1	Elsevier	1,191,461	1	Elsevier	463,869
2	American Chemical Society	508,097	2	Springer	160,712
3	Wiley Blackwell	445,995	3	American Chemical Society	141,852
4	Springer	370,032	4	Wiley Blackwell	124,612
5	Nature Publishing Group	322,017	5	Institute of Electrical and Electronics Engineers	83,952
6	The Royal Society of Chemistry	314,080	6	Nature Publishing Group	58,734
7	Informa UK	170,698	7	The Royal Society of Chemistry	48,529
8	Institute of Electrical and Electronics Engineers	79,211	8	Informa UK	47,042
9	Oxford University Press	68,233	9	American Physical Society	41,296
10	American Association for the Advancement of Science (AAAS)	64,549	10	American Institute of Physics	29,234
11	American Physical Society	62,795	11	The Optical Society	25,691
12	American Institute of Physics	55,561	12	JSTOR	19,636
13	SAGE Publications	41,490	13	IOP Publishing	19,624
14	Trans Tech Publications	33,874	14	Pleiades Publishing	18,742
15	IOP Publishing	33,373	15	SPIE - International Society for Optical Engineering	15,306
16	Ovid Technologies (Wolters Kluwer)	24,839	16	Oxford University Press	14,952
17	Proceedings of the National Academy of Sciences	24,743	17	American Association for the Advancement of Science (AAAS)	12,045
18	American Society for Microbiology	22,487	18	SAGE Publications	10,890
19	Annual Reviews	21,877	19	Walter de Gruyter GmbH	7210
20	JSTOR	21,636	20	Trans Tech Publications	6256
21	American Association for Cancer Research	21,338	21	American Institute of Aeronautics and Astronautics	6190
22	The Electrochemical Society	20,838	22	Annual Reviews	5373
23	American Society for Biochemistry and Molecular Biology	18,453	23	Ovid Technologies (Wolters Kluwer)	5226
24	Walter de Gruyter GmbH	17,049	24	Proceedings of the National Academy of Sciences	5219

Table 2 (continued)

	China			Russia	
№	Name	Downloads		Name	Downloads
25	SPIE - International Society for Optical Engineering	16,792	25	Cambridge University Press	5211
26	Mary Ann Liebert	15,506	26	Mary Ann Liebert	4914
27	Thieme Publishing Group	15,480	27	World Scientific	4475
28	World Scientific	13,257	28	The Electrochemical Society	4342
29	Pleiades Publishing	12,818	29	Canadian Science Publishing	4317
30	The Optical Society	12,358	30	Japan Society of Applied Physics	4233
31	American Geophysical Union	11,657	31	Thieme Publishing Group	4161
32	American Medical Association	11,295	32	International Union of Crystallography	4131
33	Cambridge University Press	10,587	33	Institution of Electrical Engineers	3641
34	American Institute of Aeronautics and Astronautics	10,019	34	American Society for Microbiology	3566
35	American Society of Plant Biologists	9624	35	Association for Computing Machinery	3530
36	Canadian Science Publishing	8491	36	American Medical Association	3237
37	S. Karger AG	8228	37	American Geophysical Union	3146
38	Cold Spring Harbor Laboratory Press	7367	38	S. Karger AG	2847
39	The Endocrine Society	7295	39	Turpion	2793
40	American Physiological Society	6990	40	American Society for Biochemistry and Molecular Biology	2513
41	Japan Society of Applied Physics	6955	41	The Royal Society	2496
42	American Society of Civil Engineers	6954	42	BioOne	2445
43	Society for Industrial and Applied Mathematics	6693	43	Brill Academic Publishers	2159
44	American Diabetes Association	6663	44	American Vacuum Society	2026
45	American Scientific Publishers	6374	45	Society for Industrial and Applied Mathematics	2010
46	American Vacuum Society	6308	46	Cold Spring Harbor Laboratory Press	1700
47	BioOne	6114	47	ASME International	1674

Table 2 (continued)

№	China Name	China Downloads	№	Russia Name	Russia Downloads
48	The American Association of Immunologists	6041	48	Future Medicine	1664
49	ASME International	5892	49	Bentham Science	1635
50	Association for Computing Machinery	5796	50	Maney Publishing	1631
51	Landes Bioscience	5758	51	Allerton Press	1630
52	Future Medicine	5616	52	American Association for Cancer Research	1451
53	International Union of Crystallography	5564	53	The Endocrine Society	1392
54	American Society of Clinical Oncology	5536	54	Nature Publishing Group - Macmillan Publishers	1308
55	The Rockefeller University Press	5173	55	Emerald (MCB UP)	1202
56	Bentham Science	5137	56	American Physiological Society	1199
57	The Royal Society	5031	57	The University of Chicago Press	1180
58	Society for Neuroscience	5017	58	American Association of Physics Teachers	1173
59	Geological Society of America	4793	59	Acoustical Society of America	1150
60	The Company of Biologists	4763	60	American Scientific Publishers	1130
61	Maney Publishing	4747	61	Oldenbourg Wissenschaftsverlag	1125
62	American Thoracic Society	3970	62	CSIRO Publishing	1121
63	Institution of Electrical Engineers	3939	63	American Society of Civil Engineers	1104
64	Microbiology Society	3881	64	Woodhead Publishing	1088
65	Bioscientifica	3752	65	Geological Society of London	1048
66	Future Science, LTD	3752	66	Society for Neuroscience	1022
67	American Society for Pharmacology & Experimental Therapeutics	3720	67	Muse - Johns Hopkins University Press	916
68	Institutional Investor Journals	3591	68	Geological Society of America	908
69	The University of Chicago Press	3271	69	The Rockefeller University Press	879

Table 2 (continued)

№	China			Russia	
	Name	Downloads		Name	Downloads
70	Emerald (MCB UP)	3198	70	Radiological Society of North America	867
71	Institute for Operations Research and the Management Sciences	3196	71	American Society of Clinical Oncology	865
72	De Gruyter Open Sp. z o.o.	3140	72	De Gruyter Open Sp. z o.o.	828
73	Geological Society of London	3111	73	Mineralogical Society of America	766
74	American Dairy Science Association	2891	74	Microbiology Society	745
75	Ecological Society of America	2844	75	Society of Petroleum Engineers	743
76	American Meteorological Society	2810	76	American Diabetes Association	726
77	Radiological Society of North America	2424	77	Landes Bioscience	715
78	Oldenbourg Wissenschaftsverlag	2342	78	Society of Exploration Geophysicists	695
79	Brill Academic Publishers	2315	79	The American Association of Immunologists	682
80	American Psychological Association	2302	80	GeoScienceWorld	655
81	CSIRO Publishing	2272	81	American Meteorological Society	654
82	Acoustical Society of America	2264	82	American Psychological Association	653
83	Rubber Division, ACS	2261	83	The Company of Biologists	646
84	Federation of American Society for Experimental Biology	2217	84	MIT Press	620
85	American Society of Nephrology	2201	85	Future Science, LTD	574
86	Research Publishing Services - Professional Engineering Publishing	2160	86	American Society for Pharmacology & Experimental Therapeutics	568
87	Spandidos Publications	2093	87	American Society of Plant Biologists	551
88	Society of Exploration Geophysicists	2051	88	Schweizerbart	540
89	MIT Press	1951	89	Brepols Publishers, NV	465
90	American Association for Clinical Chemistry	1878	90	Institute of Organic Chemistry & Biochemistry, Academy of Sciences of the Czech Republic	456

Table 2 (continued)

	China			Russia	
№	Name	Downloads		Name	Downloads
91	NACE International	1840	91	Hindawi Publishing Corporation	454
92	Hindawi Publishing Corporation	1589	92	Institute for Operations Research and the Management Sciences	452
93	American Society for Nutrition	1515	93	Duke University Press	445
94	Mineralogical Society of America	1428	94	MDPI AG	437
95	Society of Petroleum Engineers	1397	95	Japan Institute of Heterocyclic Chemistry	429
96	American Roentgen Ray Society	1363	96	American Psychiatric Publishing	410
97	Society for the Study of Reproduction	1348	97	American Thoracic Society	410
98	Alphamed Press	1341	98	Bioscientifica	405
99	American Association of Physicists in Medicine	1330	99	American Roentgen Ray Society	379
100	Carl Hanser Verlag GMBH & Co KG	1259	100	American Academy of Pediatrics	366

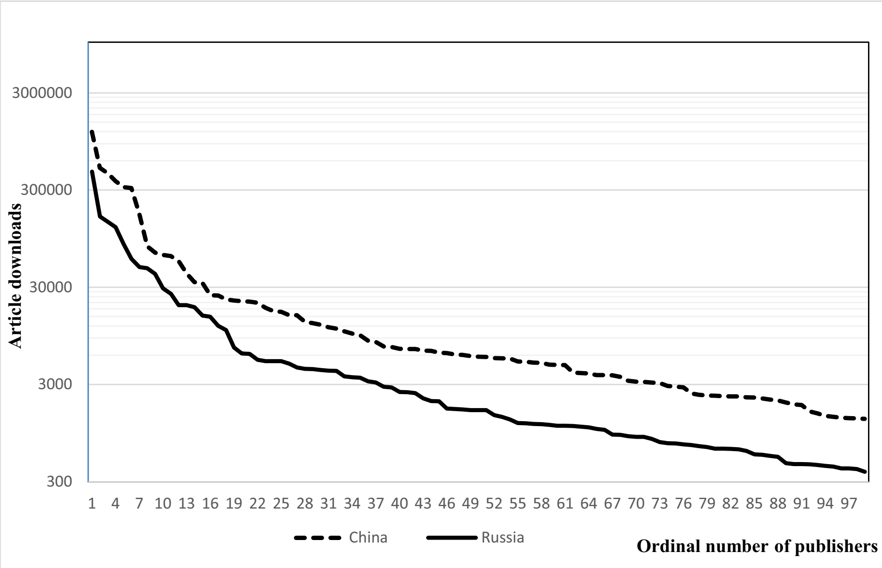


Fig. 9 Downloads of articles by Top-100 publishing houses by researchers from China and Russia

An analysis was performed on the correspondence of the calculations to the Pareto distribution. It is shown that 20% of Russian and Chinese cities account for about 96–98% of the total downloads of articles, and 20% of publishers in both cases account for about 99% of the total downloads of articles. Consequently, these distributions are far from the Pareto distribution, which is due to the bias towards large volumes of downloads of articles falling on a small number of cities and publishers, and the presence of long tails of these distributions with small numbers of downloads.

The same heavily skewed data distribution was obtained by Greshake (2017) on the sample of 62 million downloads, as we noted in the Introduction. According to him, the first 9 publishers out of 1000 publishers accounted for 80% of downloads. According to our data (Table 2), for downloads by Chinese researchers, the top 9 publishers account for (3,469,824/4,415,186) × 100% = 78.6%, and for downloads by Russian researchers - (1,170,598/1,513,212) × 100% = 77.4%, which is very close to the calculations of Greshake (2017).

Quite recently, Sci-Hub article downloads across 29 countries in all of 2017 were compared to legal article downloads from the Springer publishing site for 9 days in April 2012 (Geng et al., 2022; Wang et al., 2012). It should be noted that due to the incompatibility of the time intervals, the country rankings of downloads were compared. Let us compare the obtained data with our data for China and Russia.

According to the above work, 11,316,634 and 2,778,061 articles were uploaded by Chinese and Russian researchers, respectively. If our semi-annual data is approximated to a yearly period (from September 1, 2015 to August 30, 2016), then we get, approximately, 9 and 3 million downloads. Consequently, 3.3 million more downloads were downloaded by Chinese researchers in 2017 than in the previous year period, consisting of 4 months of 2015 and 6 months of 2016. At the same time, downloads of articles by Russian researchers for the periods under consideration were approximately the same.

Regarding the comparison of data on legal and illegal downloads of articles from the Springer platform, which, as we have shown, is the second most popular publishing platform, both for the two countries in question, and in general, our joint analysis of the relative characteristics of our data and work data (Geng et al., 2022) showed that Russian researchers use illegal downloads from the Springer platform to a much greater extent than legal downloads.

The total volume of illegal downloads in China for the entire 2017 was also given in (Monti & Unzurrunzaga, 2020), which was more than twice as large as in (Geng et al., 2022) and amounted to 24,943,832. This is apparently due to the careful data cleaning of the work of the Chinese authors (with a total of 140 million downloads to 74,742,113 downloads), in contrast to the work of the Argentine authors. The latest data on downloads from Chinese researchers is presented in the article (Owens, 2022) with a link to the Sci-Hub resource. It noted that China ranks first in the world with more than 25 million downloads over the past month. From the data provided in the Introduction, we can clearly see that the monthly download traffic from Sci-Hub by Chinese researchers at the beginning of 2022 was more than double the annual download traffic in 2017. We attribute this not only to the regular positive trend in the growing popularity of Sci-Hub, but also, especially, to the COVID-19 pandemic, when researchers worked exclusively from home and were no longer constrained by the institutional environment of their organizations.

The above experiments in Google Scholar and Semantic Scholar, as well as content analysis of their results, made it possible to identify the following 6 partially overlapping clusters of publications:

1. Debate over the legitimacy of Sci-Hub and its relationship to OA (Ajani et al., 2023; Björk, 2017; Brembs, 2016); Buehling et al., 2022; Boudry et al., 2019; Cochran, 2016; Cross, 2017; Deshpande, 2019; Esposito, 2016, 2017; Ferreira, 2023; González-Solar & Fernández-Marcial, 2019; Green, 2017; Lawson, 2017; Machin-Mastromatteo et al., 2016; Maddi & Sapinho, 2023; Moskovkin et al., 2021; Nicholas et al., 2019; Pastor-Ramon, 2023; Pecastre & Correa, 2016; Penn, 2018; Piwowar et al., 2018; Singh et al., 2020; Swartz, 2008). There are 23 publications in total.
2. Ethical Issues in Using Sci-Hub: Pros and Cons (Ajani et al., 2023; Bendezú-Quispe et al., 2016; Boudry et al., 2019; Brembs, 2016; Cross, 2017; Deshpande, 2019; González-Solar & Fernández-Marcial, 2019; Kipnis, 2023; Lawson, 2017; Łuczaj & Holy-Łuczaj, 2020; Nicholas et al., 2019; Pastor-Ramon, 2023; Plutchak, 2019; Rossello & Martinelli, 2023; Saleem et al., 2017; Till et al., 2019; Travis, 2016; Swartz, 2008). There are 19 publications in total.
3. Speed and ease of access to scientific information is the main advantage of Sci-Hub (Ajani et al., 2023; Cross, 2017; Esposito, 2017; Geng et al., 2022; Greco, 2017; Faust, 2016; Felts et al., 2020; Gardner et al., 2017; González-Solar & Fernández-Marcial, 2019; Heathers, 2016; Hurst & Schira, 2019; Kipnis, 2023; Kramer, 2016; Mellins-Cohen, 2017; Moore, 2020; Oakley, 2016; Oxenham,

2016; Pastor-Ramon, 2023; Plutchak, 2019; Sagemüller et al., 2021; Singh et al., 2021; Steel, 2016; Travis, 2016; Tury et al., 2015; Wilcock, 2018). There are 25 publications in total.
4. Concerns about espionage and cybersecurity associated with the use of Sci-Hub (Kipnis, 2023; Martin, 2021; Mellins-Cohen, 2017; Plutchak, 2019). There are 4 publications in total.
5. Digital socialism and academic capitalism in the Sci-Hub debate (Brembs, 2016; Couldry & Mejias, 2019; Deshpande, 2019; Drahos & Braithwaite, 2002; Kelly, 2009; Liang, 2018; Łuczaj & Holy-Łuczaj, 2020; Moore, 2020; Moskovkin & Serkina, 2016; Peters, 2020; Swartz, 2008). There are 12 publications in total.
6. Anti-Sci-Hub measures, including library training measures (Ajani et al., 2023; Björk, 2017; Cook et al., 2017; Esposito, 2016, 2017; Ferreira, 2023; Geng et al., 2022; Hoy, 2017; Kipnis, 2023; Mellins-Cohen, 2017; Moskovkin et al., 2021; Pastor-Ramon, 2023; Plutchak, 2019; Rossello & Martinelli, 2023; Steel, 2016; Saarti & Tuominen, 2021; Strielkowski, 2017). There are 17 publications in total.

As we can see, the authors most often discuss issues of legitimacy, ethics, accessibility and countermeasures for Sci-Hub activity. Let us consider the most important, in our opinion, articles from these six publication clusters, and conclude with our suggestions for the most effective measures to combat Sci-Hub.

3.1 Debate over the legitimacy of Sci-hub and its relationship to OA

After the launch of Sci-Hub, the debate about its legitimacy continues to this day, and scientists have not yet come to a consensus about this phenomenon. According to Pecastre and Correa (2016), there is a surprising lack of scholars and librarians who argue that Sci-Hub is a 'mass-piracy criminal enterprise' or that it aims to become the 'WikiLeaks of scientific information' or the 'Pirate Bay of research', despite these claims being central to the ongoing debate about the website.

According to Machin-Mastromatteo et al. (2016), a significant portion of the open access (OA) community disputes the notion that Sci-Hub qualifies as OA. This is primarily due to the fact that the articles on Sci-Hub are not accessible under an open licensing scheme like Creative Commons.

As Buehling et al. (2022) write, Sci-Hub, along with other shadow libraries, is often not considered a legitimate form of Open Access (OA) due to their modes of conduct being incompatible with widely accepted definitions of OA, (such as the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (Max Planck Society, 2003)).

Boudry et al. (2019) also believe that Sci-Hub does not fall under common definitions of Open Access, but it does provide access to otherwise costly literature.

Although many researchers do not classify the Sci-Hub as Open Access, since it ignores copyright and ethical standards (Piwowar et al., 2018; Saarti & Tuominen, 2021; Ferreira, 2023; Maddi & Sapinho, 2023), Sci-Hub actually provides 100% access to current and in many cases retrospective journal publications. Therefore, the Sci-Hub phenomenon is often referred to as Black

Open Access (Björk, 2017; Green, 2017; Penn, 2018; Singh et al., 2020), with Björk (2017) and Penn (2018) regarding Black Open Access as one of the greatest challenges to the traditional model of academic publication. We also note, referring to the work of (Ferreira, 2023), that in some cases Black Open Access is called a subtype of Open Access, following the ideas put forward by the “Guerilla Open Access Manifesto” (Swartz, 2008).

Maddi and Sapinho (2023) note that many publishers and scientist-editors condemn shadow libraries for copyright infringement and loss of publication usage information, while some scientists and institutions support them, sometimes in a roundabout way, for their role in reducing inequalities in access to knowledge, especially in low-income countries. These same authors showed that while Sci-Hub facilitates access to scientific knowledge, it has a negative impact on the open access movement as a whole, reducing the comparative advantage of open access publications in terms of visibility for researchers (Maddi & Sapinho, 2023).

At the same time, Deshpande (2019) believes that Sci-Hub is a pioneering search engine for free and massive open access to research articles and therefore requires support, and Esposito (2017) argues that no publisher will be able to act in the foreseeable future without taking into account the Sci-Hub factor, which has now entered into the decision-making process.

Ajani et al. (2023), while noting that Sci-Hub is problematic from a legal perspective, consider it a valuable resource that has provided “blessing in disguise” to library users who do not have access to academic resources.

The main argument in support of Sci-Hub is that the scientific publishing system is unequal and that the fundamental goal of changing it to make it fairer requires any means necessary. Sci-Hub is a symptom, a reaction to a problem that must be solved at all costs (González-Solar & Fernández-Marcial, 2019). Lawson (2017) writes about the same thing, noting that Sci-Hub may not be the solution, but it is an alarming signal about the need to create a common scientific and academic heritage that goes beyond the restrictions imposed by intellectual property. Ferreira (2023) also writes about such an alarming signal, arguing that while the case for using Sci-Hub may be compelling in some cases, the disruptive options are not sustainable and it cannot be the solution to a viable publishing model.

However, there is clear support for Sci-Hub within the open access ranks (Cochran, 2016), and tweets often mention their new domains. Heather Piwowar plays an important role in scaring and pushing editors to “do the right thing” and go for open access. (González-Solar and Fernández-Marcial, 2019; Piwowar et al., 2018).

In this regard, Steel (2016) writes: “I suspect many academic librarians and open access advocates support Sci-Hub’s ends if not its means”, and González-Solar and Fernández-Marcial (2019) suggest that there is some latent support for Sci-Hub among librarians. As Esposito (2017) writes for librarians, “Sci-Hub is an unacknowledged reserve army prepared to enter the battle with publishers,” and for publishers, “Sci-Hub is an enemy, clear and simple.”

3.2 Ethical issues in using Sci-hub: Pros and cons

The attitude to this phenomenon in the scientific community is ambiguous. So, Saleem et al. (2017) are of the opinion that for the sake of scientific ethics illegal practices should be prohibited, even if they are intended for progress. Nicholas et al. (2019) believe that Sci – Hub is a pure and unashamed ‘pirate’. In contrast, according to a survey of 11,000 scientists, published by Travis (2016), 88% of respondents do not criticize downloading articles from illegal sources. This survey shows a purely consumer attitude towards the phenomenon under consideration, people are simply comfortable using this tool, even if they have licensed access to scientific publications in their institutes.

González-Solar and Fernández-Marcial (2019) suggest that one of the problems raised by Sci-Hub is that users or part of them give priority to accessing scientific content regardless of the legal or ethical connotations that it may imply.

Rossello and Martinelli (2023) proved through their empirical research that illegal access to scientific literature through Sci-Hub can be considered a mild deviant behavior, despite the fact that its widespread prevalence is documented, since almost half scientists at leading European universities are illegally uploading papers using Sci-Hub.

There is large group of researchers who care about ethical standards in science, but who also believe that Sci-Hub is critical in many cases. Most of these researchers are among those involved in biomedical research, and especially those who are medical practitioners (Bendezú-Quispe et al., 2016; Boudry et al., 2019; Cross, 2017; Faust, 2016; Ferreira, 2023; González-Solar & Fernández-Marcial, 2019; Machin-Mastromatteo et al., 2016; Till et al., 2019). They wonder if they have a moral duty to use pirated information to improve health care or save lives, even if they do so in violation of the law or ease of access?

This problem is especially acute in developing countries, where universities, independent biomedical laboratories and hospitals do not have enough funds to subscribe to very expensive biological and medical journals. As we noted in the Introduction, two-thirds of all downloads of medical science literature via Sci-Hub happen in low-and lower-middle-income countries (Till et al., 2019).

As noted by Bendezú-Quispe et al. (2016) physicians are faced with a dilemma about whether to access through Sci-Hub the information that they need to provide their patients with optimum care.

Boudry et al. (2019) compare the accessibility of recent full text articles in the field of ophthalmology in 27 established institutions located worldwide and conclude that the paucity of full-text ‘paiwalled articles’ that can be found in most of the institutions studied stimulates ophthalmology researchers to use Sci-Hub to seek for scientific information. Both scientific community and decision-makers must join forces and intensify their efforts to find ways to improve access to scientific literature worldwide and obviate the collapse of the scientific publishing model.

This ethical dilemma was especially acute during the COVID-19 pandemic, when the lack of important medical information cost the lives of patients. Thus, this urgent need for access to the results of medical and pharmacological research was confirmed in the search for a vaccine against COVID-19 (Ferreira, 2023).

3.3 Speed and ease of access to scientific information is the main advantage of Sci-hub

When it comes to comparing preferences between using Sci-Hub and using library services, there is a tendency to choose Sci-Hub due to its convenience, ease of use and speed of access to information, as well as recognition and popularity among the academic community (Björk, 2017; Felts et al., 2020; Geng et al., 2022; González-Solar & Fernández-Marcial, 2019; Hurst & Schira, 2019; Kipnis, 2023; Lawson, 2017; Mellins-Cohen, 2017; Pastor-Ramon, 2023; Plutchak, 2019; Sagemüller et al., 2021; Tury et al., 2015; Wilcock, 2018).

Let us present the most interesting excerpts from a number of the listed works.

1. Plutchak (2019) writes that legal arguments are not enough and regardless of whether one views the use of Sci-Hub as a noble act of civil disobedience and a strike against an unfair system of access to scientific knowledge, people will continue to use Sci-Hub as long as it is so useful and easier than anything legal that librarians and publishers have come up with.
2. The advent of Sci-Hub marked a turning point for academic and research libraries, presenting a challenge to the information retrieval systems offered through them. Sci-Hub's interface is simple and friendly, similar to Google. Users access document content using only the DOI or text header (González-Solar & Fernández-Marcial, 2019).
3. Pastor-Ramon (2023) based on a survey of Spanish-speaking science and social science researchers on the use of library resources and Sci-Hub showed that researchers know that they use pirated access even when they have access to a virtual library, but they do not mind because they believe that efficiency is more important than copyright issues, ignoring the stages of access through the library, and in some cases the library does not have remote access.
4. Tury et al. (2015) studied remote information access behavior by surveying 649 students from 81 countries. They found that the most important elements were accessibility and speed of access to information, as well as familiarity with the source, and concluded that these characteristics were more relevant than issues of quality, reliability or completeness.
5. Researchers may prefer sites like Sci-Hub and PubMed due to the lack of a better alternative provided by publishers. If the publisher's site does not offer significant qualitative differences compared to Sci-Hub, readers are more inclined to choose Sci-Hub. In order to compete with Sci-Hub, publishers must not only offer a more enriching user experience but also match or surpass Sci-Hub's seamless user interface (Wilcock, 2018).
6. Albeit there are manifold obstacles in the way of scientific research, paywalls implemented by publishers, difficult authentication methods, and other factors undoubtedly use up and waste a considerable amount of scientists' time. This may complicate their task and extend their working hours. No one likes to be bothered (Geng et al., 2022).

7. Present barriers to access comprise compelling the user to click through a lot of pages to get to the content protected by a paywall. This is typically complicated by the fact that users' credentials are distributed among numerous platforms the management of which becomes increasingly exacting… Since libraries offer such burdensome procedures to go beyond a paywall, fully entitled end users seeking easily accessible content, could be unintentionally encouraged thereby to benefit by such alternative resources as Sci-Hub or ResearchGate (Felts et al., 2020).
8. As professors noted, students anticipate unhindered round-the-clock access to library resources and electronic learning tools. Additionally, professors report that sometimes it is impossible for students to postpone their other occupations and pursuits. Students are dependent on the availability of resources whenever needed. If this is not the case, it increases their stress (Hurst & Schira, 2019).

Due to widespread use of Sci-Hub by researchers around the world, libraries have realized that legal access to subscription resources is difficult and requires many more steps than those required through Sci-Hub (Faust, 2016; Greco, 2017; Heathers, 2016; Oakley, 2016; Plutchak, 2019). For example, Heathers (2016) notes 4 steps to access an article through Sci-Hub and 10 steps to such access through Institutional Access, and Oakley (2016) compares a paper search in the Georgetown University Library and Sci-Hub: "Article in journal to which we subscribe:–GU Library: 6 clicks, 24 seconds–Sci -Hub: 2 clicks, 5 seconds and Article in a journal to which we do not subscribe:–GU Library: 1 minute 45 seconds + ILL delivery–Sci-Hub: 2 clicks, 10 second". Faust (2016) talks about a similar experience. Faust (2016), emphasizing not only the ease of searching for articles, but also its reliability: "Sci-Hub's appeal does not rest on speed alone but rather its reliability… In contrast, when Sci-Hub finds an article, you're always 1 click away from the pdf file".

The analysis of Sci-Hub users in the previous three paragraphs of the article allowed us to identify four groups of representatives of the academic community in terms of their attitude towards the Sci-Hub phenomenon:

1) ardent opponents of Sci-Hub who want to ban it.
2) ardent supporters of Sci-Hub, proving its validity;
3) users of Sci-Hub who have licensed access to scientific information, but use it only for the sake of convenience.
4) Sci-Hub users who do not have licensed access to scientific information and are forced to use Sci-Hub for the sake of conducting scientific research or saving the lives of patients using previously obtained results of biomedical research.

The first two groups of representatives of the academic community are small compared to the last two groups. The third group of researchers work mainly in developed countries, and the fourth - in developing ones.

3.4 Concerns about espionage and cybersecurity related to the use of Sci-Hub

We found only four publications on this issue of discussions around Sci-Hub, which indicates its low relevance. Here, cybersecurity should be considered from two sides. From the security side of Sci-Hub users from their personal or university computers, as well as from publishers, who are subject to continuous attacks from the above-mentioned users.

Thus, a survey conducted by Kipnis (2023), showed concerns among academic librarians and professors about espionage, viruses and malware when using Sci-Hub. Martin (2021) cites a statement from the Scholarly Networks Security Initiative (SNSI), which welcomes the City of London Police's warning and added: "Pirate sites like Sci-Hub threaten the integrity of the scientific record, and the safety of university and personal data which could result in wider institutional data access and potential misuse." Plutchak (2019) notes that while there is no evidence that simply using Sci-Hub on an individual basis poses a security risk, compromised credentials, however obtained, certainly pose a threat.

As for the cybersecurity of publishers, we saw a detailed discussion of this issue in (Mellins-Cohen, 2017). Here are the most important excerpts from his work: "First, publishers can undertake a simple audit of IP ranges to clean up institutional subscription IPs and prevent content being opened up to the wrong customers without payment…. Second, monitoring usage activity and setting alerts for unusual spikes or surges in activity can help to spot piracy."

In theory, Sci-Hub could embed viruses into the site itself or into PDF files, thereby attacking users' computers. In practice, if this were possible, then someone would have already discovered these "bookmarks" and written a detailed expose. And it's hard to imagine that Alexandra Elbakyan, given her volunteer and altruistic Sci-Hub project, would do this.

3.5 Digital socialism and academic capitalism in the Sci-Hub debate

The Sci-Hub phenomenon is based on an ideology that is opposite to the concept of "academic capitalism", and the concept of "digital socialism" is close to it (Kelly, 2009; Peters, 2020). Supporters of the Sci-Hub phenomenon, as well as supporters of the Open Access movement, believe that scientific knowledge created at the expense of taxpayers should be freely available without any restrictions to the creators of this knowledge and to society as a whole. Academic capitalism (or knowledge feudalism), which exists within the neo-liberal agenda, cannot lead to sustainable development of scientific systems (Moskovkin & Serkina, 2016), which naturally caused the above phenomena and processes.

Based on the concept of "Digital capitalism" and the concept of "Data colonialism" included in it (Couldry & Mejias, 2019), Moore (2020) discusses the ability of publishers to develop systems that promote data colonialism, which is largely due to the fact that universities have transferred control of scholarly communication infrastructures to private for-profit companies through processes of marketization and monetization. He believes that this is the direction the publishing industry is heading, with all future scientific research being colonized by the publishing oligopoly, despite the Open Access movement. He reasonably notes that "in practice open access has allowed commercial publishers to increase their stranglehold through article-processing charges and has even provided the conditions for data extraction." It should be noted that a similar concept to "Digital capitalism" was proposed by Drahos and Braithwaite (2002) under the name "Information feudalism", which was the basis of the above-mentioned work by Moskovkin and Serkina (2016).

As noted by Łuczaj and Holy-Łuczaj (2020), Open access policy corresponds to the "digital socialism" model, but open access journals only partially satisfy the demand. The gap between needs and resources leads scientists to breach copyright using such sites as ResearchGate, Academia.edu, or Sci-Hub.

However, the Sci-Hub users interviewed for this article did not believe that using the service was a significant ethical problem. It should be noted that the authors conducted 100 in-depth interviews (2018 and 2019 years) among scientists who worked in Poland. The results of these interviews showed that researchers from peripheral countries usually perceive global commercial publishers as actors who deprive them of profits, acting as capitalists in a global knowledge economy where scientists are workers (Łuczaj & Holy-Łuczaj, 2020).

3.6 Countermeasures against Sci-Hub, including library training measures

Publishers, as Pastor-Ramon (2023) notes, have two positions. On the one hand, some are keen to protect content and want to strengthen security. In this sense, some publishers are considering stopping offering articles in PDF format to prevent illegal downloading (Cochran, 2016; Esposito, 2016; Cook et al., 2017, Steel, 2016). On the other hand, some publishers are betting on making access more convenient and user-friendly (Pastor-Ramon, 2023).

The first position was well voiced by Steel (2016). In a web-centric world, PDFs, which are most heavily used for scholarly communication, must take a backseat as a means of sharing knowledge, he says, so it's high time publishers and librarians work together to move beyond PDFs. But unfortunately, as Plutchak (2019) notes, much of the attention of publishers and network security specialists has been aimed at making it more difficult to penetrate university systems by adding more obstacles for authorized users.

But regarding PDF files, we want to note that projects such as archive.md or archive.org allow you to download html versions of any pages on the Internet to make an archive copy. Therefore, theoretically, this can be implemented on Sci-Hub, if publishers switch exclusively to html versions of articles.

At the same time, in contrast to the above opinion of publishers and network security specialists, many researchers believe that academic publishers, academic libraries and institutional IT – services could collaborate to create a better and easier environment for researchers to access academic publications, removing barriers to their legitimate use, for example by providing seamless, secure and one-click access to content subscription so that scientists can legally access articles anytime, anywhere (Felts et al., 2020; Moore, 2020; Geng et al., 2022).

Pastor-Ramon (2023) argues that libraries should learn from pirated resources how to improve access to their scholarly resources and offer editors a different way of doing business, more akin to Spotify or Netflix than huge journal packages. Mellins-Cohen (2017) wrote vividly about this much earlier: “After all, if you can stream a film in HD on Netflix for a few pennies, why bother scrounging around the internet for a low-resolution pirate copy?”. Cross (2017) and González-Solar and Fernández-Marcial (2019) wrote the same about the Napster file sharing service. And finally, as Plutchak (2019) writes: “The death knell for music piracy (not that music piracy has been completely eradicated) was iTunes introducing $.99 songs with an easy-to-use interface. … Sci-Hub will thrive until we can come up with something comparable”.

This experience in combatting music piracy and the insistence of Open Access supporters to use it to combat scientific piracy will allow us to make more specific proposals on this issue later.

Another very important issue contributing to the spread of scientific piracy is the outdated model of interlibrary loan operation, which does not take into account new methods of peer-to-peer communication and the advantages of digital technologies, making it almost irrelevant in the current research scenario (Björk, 2017; Ferreira, 2023; Hoy, 2017; Saarti & Tuominen, 2021). Below we will propose a technology for transforming this model for the digital environment.

Regarding library training efforts, Pastor-Ramon (2023) notes that there are tools to help libraries avoid or minimize users’ use of Sci-Hub instead of library services and resources, such as EndNote Click or Unpaywall; both can be configured to be accessed directly from search platforms such as PubMed or directly from the web page of the journal to which the library subscribes, or other legal resources such as Institutional OA repositories. He writes: “Teach to the users how to install and use tools such as EndNote Click or Unpaywall, is mandatory to fight against piracy”.

Above we noted two measures that can help neutralize scientific piracy. Let us outline the methods for their implementation.

Publishing measure Publishers are organizing a survey of Sci-Hub users to determine at what cost of electronic copies of articles they are willing to stop using Sci-Hub. The estimated number of such users and the income for publishers when selling these copies at the price obtained as a result of the survey are estimated. At the same time, publishers are calculating the losses they will sustain from refusing to sell electronic copies of articles at the previously established price of 30–40 dollars per article. The ratio of income to losses will indicate the profitability of this measure for publishers. This calculation should also be made based on the predetermined cost of selling such copies of articles at a price of $1.

But there is a very serious limitation in implementing this measure, which was discussed when describing the third cluster of publications in discussions on Sci-Hub. In addition to being cheap, the payment system should be convenient. People don’t want to spend 10 minutes registering their account on the publisher’s website, indicating their real full name and postal address with a zip code (which is checked for correctness), then confirming registration by email, then entering their card number, receiving an SMS with confirmation, and so on. Therefore, even after a serious price reduction, traffic to Sci-Hub may not decrease. So, this measure will be viable if only the loading time of the article on the publisher's site is no longer than its loading time on the Sci-Hub site.

Library measure Previously, we noted the inefficiency of interlibrary loan within the traditional subscription model. It allows you to exchange books and magazines in paper form. But with the development of the Internet, when publishers sell electronic copies of articles, the idea of exchanging such copies arises. Consequently, an Online platform of interlibrary loan for research articles can be created. It, in turn, can be created by the International library consortium, whose libraries collectively have subscriptions to almost all journals of commercial publishers, indexed in the Scopus&WoS databases. But this measure, like the previous one, will be effective if the speed and ease of access to articles is comparable to the speed and ease of access through Sci-Hub.

4 Conclusion

According to our first research question the article provides a detailed analysis of downloads of articles by Chinese researchers by city and publisher in comparison with downloads of articles by Russian researchers over the same six-month period based on public data from Elbakyan & Bohannon (2016). We obtained highly skewed distributions of these downloads towards leading cities and publishers, consistent with the results of Greshake (2017) obtained on a much larger sample of downloads. The correct use of methods of mathematical statistics and data mining (development of algorithms for processing files downloaded from Sci-Hub, data cleaning and aggregation, their implementation using the Python programming language in the PyCharm environment) speaks in favor of the reliability of our research.

Some generalized data on downloads corresponded to the data of the work (Elbakyan & Bohannon, 2016) noted by Bohannon (2016). This concerned the total volume of downloads by Chinese researchers and the share of downloads attributable to researchers from Moscow in the total volume of downloads by Russian researchers. We also showed the consistency of our data with the data of Geng et al. (2022). Indirect evidence of the reliability of our research is also the compliance of the obtained distributions of downloads from the Sci-Hub website by city and publishing house of the two countries under consideration with the Pareto principle.

As part of the second research question, we clustered publications on discussions around Sci-Hub, obtained as a result of a search through Google Scholar and Semantic Scholar. Six clusters of publications were identified, of which the largest were clusters of publications on the legitimacy, ethics, accessibility and countermeasures of Sci-Hub activity. At the conclusion of the last cluster of publications, methods for implementing publishing and library measures to neutralize the activities of scientific piracy were proposed. The first measure was related to the dumping price for the sale of electronic copies of articles by publishers, and the second was related to the creation of an Online platform of interlibrary loan for research articles.

In conclusion, we should note that under the pressure of the legal and pirate Open Access movements, commercial publishers are forced to transform their activities and move to the Open Access model. When this happens, then the Sci-Hub pirate project will die out by itself, as Alexandra Elbakyan always said in her interviews. A similar thought was expressed by Björk (2017) speaking about 100% Gold Open Access. This is precisely what Plan S that was launched in September 2018 aims to develop as the goals and principles of the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2013).

Data availability All data generated or analysed during this study are included in this published article.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

References

Ajani, Y.A., Tella, A., & Okere, S. (2023). Access to full-text documents in libraries via Sci-Hub: A blessing in disguise to library users. Library Hi Tech News. Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/LHTN-03-2023-0053

Androćec, D. (2017). Analysis of Sci-Hub downloads computer science papers. *Acta Universitatis Sapientiae, Informatica, 9*(1), 83–96. https://sciendo.com/article/10.1515/ausi-2017-0006

Babutsidze, Z. (2018). Pirated economics. *South – Eastern Europe Journal of Economics, 16*(2), 209–219. https://ojs.lib.uom.gr/index.php/seeje/article/view/9619

Bendezú-Quispe, G., Nieto-Gutiérrez, W., Pacheco-Mendoza, J., & Taype-Rondan, A. (2016). Sci-Hub and medical practice: An ethical dilemma in Peru. *The Lancet Global Health, 4*(9), e608. https://doi.org/10.1016/S2214-109X(16)30188-7

Björk, B.-C. (2017). Gold, green, and black open access. *Learned Publishing, 30*, 173–175. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/leap.1096

Bohannon, J. (2016). Who's downloading pirated papers? *Everyone. Science, 352*(6285), 508–512. https://www.science.org/doi/10.1126/science.352.6285.508

Boudry, C., Alvarez-Muñoz, P., Arencibia-Jorge, R., Ayena, D., Brouwer, N. J., Chaudhuri, Z., Chawner, B., Epee, E., Erraïs, K., Fotouhi, A., Gharaibeh, A. M., Hassanein, D. H., Herwig-Carl, M. C., Howard, K., Kaimbo Wa Kaimbo, D., Laughrea, P.-A., Lopez, F. A., Machin-Mastromatteo, J. D., Malerbi, F. K., … Mouriaux, F. (2019). Worldwide inequality in access to full text scientific articles: The example of ophthalmology. *PeerJ, 7*, e7850. https://doi.org/10.7717/peerj.7850

Brembs, B (2016). Sci-Hub as necessary, effective civil disobedience. In: bjoern.brembs.blog. Available at: http://bjoern.brembs.net/2016/02/sci-hub-as-necessary-effective-civil-disobedience/

Buehling, K., Geissler, M., & Strecker, D. (2022). Free access to scientific literature and its influence on the publishing activity in developing countries: The effect of Sci-Hub in the field of mathematics. *Journal of the Association for Information Science and Technology, 73*(9), 1336–1355. https://doi.org/10.1002/asi.24636

Cabanac, G. (2016). Bibliogifts in LibGen? Study of a text sharing platform driven by biblioleaks and crowdsourcing. *Journal of the Association for Information Science and Technology*, *67*(4), 874–884. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23445

Caffrey, C., & Gardner, G.J. (2017). Fast and furious (at publishers): The motivations behind crowdsourced research sharing. *College & Research Library, 78*(2), 131–149. https://crl.acrl.org/index.php/crl/article/view/16578

Cochran, A. (2016). A funny thing happened on the way to OA. The scholarly kitchen, 25 Feb. https://scholarlykitchen.sspnet.org/2016/02/25/a-funny-thing-happened-on-the-way-to-oa

Cook, E. I., Levine-Clark, M., & Katz, P. (2017). Sci-Hub: What is it, and why does it matter to academic libraries? An interview with Michael Levine-Clark and Peter Katz. *The Serials Librarian, 73*(2), 79–83. https://doi.org/10.1080/0361526X.2017.1361886

Correa, J.C., Laverde-Rojas, H., Marmolejo-Ramos, F., Tejada, J., & Bahnik, Š. (29 Jan 2020). The Sci-Hub Effect: Sci-Hub downloads lead to more article citations. arXiv:2006.14979v2[cs.Dl]. http://arxiv.org/abs/2006.14979

Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism.- Stanford: Stanford University Press. 352 p. http://www.sup.org/books/title/?id=28816

Cross, W. (2017). Legal issues in information sharing in the era of Sci-Hub. *Against the Grain, 29*(4), Article 42. https://doi.org/10.7771/2380-176X.7816

Deshpande P.R. (2019). Why should Sci-Hub be supported? *International Journal of Health Allied Sciences, 8*(3), 210–212. https://www.ijhas.in/article.asp?issn=2278-344X;year=2019;volume=8;issue=3;spage=210;epage=212;aulast=Deshpande

Drahos, P, & Braithwaite, J (2002). Information feudalism: Who owns the knowledge economy? Routledge, 272 p. https://www.routledge.com/Information-Feudalism-Who-Owns-the-Knowledge-Economy/Drahos-Braithwaite/p/book/9781853839177

Elbakyan, A., & Bohannon, J. (2016). Data from: Who's downloading pirated papers? Everyone. Dryad Digital Repository. Retrieved from https://datadryad.org/stash/dataset/doi:10.5061/dryad.q447c

Esposito J. (2016). Sci-Hub and the Four Horsemen of the Internet. The Scholarly Kitchen, 2 March https://scholarlykitchen.sspnet.org/2016/03/02/sci-hub-and-the-four-horsemen-of-the-internet

Esposito, J. (2017). Sci-Hub moves to the center of the ecosystem. The Scholarly Kitchen, 5 September. https://scholarlykitchen.sspnet.org/2017/09/05/sci-hub-moves-center-ecosystem

Faust, J. S. (2016). Sci-Hub. A solution to the problem of paywalls, or merely a diagnosis of a broken system? *Annals of Emergency Medicine, 68*(1), A15–A17. https://doi.org/10.1016/j.annemergmed.2016.05.010

Felts, J., Lloyd, T., & Singley, E. (2020). Introducing seamless Access.org: Delivering a simpler, privacy-preserving access experience (pp. 386–390). Purdue University Purdue University Purdue e-PubsPurdue e-Pubs Charleston Library Conference https://doi.org/10.5703/1288284317205

Ferreira, B. B. (2023, May 16). Illegal access to scholarly information: Considerations regarding the use of Sci-Hub and its constraints to academic libraries. OSF Preprints. https://doi.org/10.31219/osf.io/mer5n

Gardner, G. J., McLaughlin, S. R., & Asher, A. D. (2017). Shadow libraries and you: Sci-Hub usage and the future of ILL. In: ACRL 2017. At the Helm: Leading Transformation. Baltmore, Maryland. p. 568–587. Retrieved from https://alair.ala.org/bitstream/handle/11213/17772/ShadowLibrariesandYou.pdf?sequence=1

Geng, Y., Cao, R., Han, X., Tian, W., Zhang, G., & Wang, X. (2022). Scientists are working overtime: When do scientists download scientifc papers? *Scientometrics.* https://doi.org/10.1007/s11192-022-04524-1

González-Solar, L., & Fernández-Marcial, V. (2019). Sci-Hub, a challenge for academic and research libraries. *El professional de la información, 28*(1), e280112. https://doi.org/10.3145/epi.2019.ene.12

Greco, A. N. (2017). The Kirtsaeng and SCI-HUB cases: The major U.S. copyright cases in the twenty-first century. *Publishing Research Quarterly, 33*, 238–253. https://doi.org/10.1007/s12109-017-9522-7

Green, T. (2017). We've failed: Pirate black open access is trumping green and gold and we musn change our approach. *Learned Publishing, 30*, 325–329. https://doi.org/10.1002/leap.1116

Greshake, B. (2016). Correlating the Sci-Hub data with World Bank indicators and identifying academic use. *The Winnower, 3*, e146485–e157797. https://doi.org/10.15200/winn.146485.57797

Greshake, B. (2017). Looking into Pandora's box: The content of Sci-Hub and its usage. PMC. Retrieved from https://doi.org/10.12688/f1000research.11366.1

Heathers J. (May 2, 2016). Why Sci-Hub will win. Medium, https://jamesheathers.medium.com/why-sci-hub-will-win-595b53aae9fa

Himmelstein, D. S., Romero, A. R., McLaughlin, S. R., Greshake, B., & Greene, C. S. (2017). Sci-Hub provides access to nearly all scholarly literature. *Peer J. Preprints, 5*, e3100v2. https://doi.org/10.7287/peerj.preprints.3100v1

Hoy, M. B. (2017). Sci-Hub: What librarians should know and do about article piracy. *Medical Reference Services Quarterly, 36*(1), 73–78. https://doi.org/10.1080/02763869.2017.1259918

Hurst, C., & Schira, H. R. (2019). Going dark: What are the consequences of losing of-campus access to library resources? *Partnership: The Canadian Journal of Library and Information Practice and Research*, *14*(1). https://pdfs.semanticscholar.org/31b1/7a1c7bad191b3579ab352c34beeffb21dba8.pdf

Kelly, K. (May 22, 2009). The new socialism: Global collectivist society is coming online. Wired. Retrieved from https://www.wired.com/2009/05/nep-newsocialism/

Kipnis, D. G. (2023). Survey results from academic librarians and professors on teaching and using pirate websites. *Internet Reference Services Quarterly, 27*(1), 1–23. https://doi.org/10.1080/10875301.2022.2095474

Kramer, B. (June 20, 2016). Sci-Hub: Access or Convenience? A Utrecht Case Study (Part 2). I&M / I&O 2.0. Retrieved from https://im-2punt0.wordpress.com/2016/06/20/sci-hub-access-or-convenience-a-utrecht-case-study-part-2/

Lawson, S. (2017). Access, ethics and piracy. *Insights UKSG J, 30*, 25–30.

Liang, L. (2018). India: The knowledge thief. In J. Karaganis (Ed.), *Shadow libraries: Access to knowledge in global higher education* (pp. 183–222). MIT Press.

Łuczaj, K., & Holy-Łuczaj, M. (2020). How academics outside global research centers challenge limitationsin access to scholarly literature. *Portal: Libraries and the Academy, 20*(4), 715–731. https://preprint.press.jhu.edu/portal/sites/default/files/20.4luczaj-pre.pdf

Machin-Mastromatteo, J. D., Uribe-Tirado, A., & Romero-Ortiz, M. E. (2016). Piracy of scientific papers in Latin America: An analysis of Sci-Hub usage data. *Information Development, 32*(5), 1806–1814. https://doi.org/10.1177/0266666916671080

Maddi, A., & Sapinho, D. (2023). On the culture of open access: The Sci-hub paradox. *Scientometrics, 128*, 5647–5658. https://doi.org/10.1007/s11192-023-04792-5

Martin, A. (19 March 2021). Sci-Hub: Police warn and universities against using the Pirate Bay of scince. Sky News. https://news.sky.com/story/police-warn-students-and-universities-against-using-the-pirate-bay-of-science-12250407

Max Planck Society, (2003). https://openaccess.mpg.de/Berlin-Declaration

Mellins-Cohen, T. (2017). Easy access to the version of record (VoR) could help combat piracy: Views from a publishing technologist. *Insights the UKSG journal, 30*(2), 47–54. https://doi.org/10.1629/uksg.360

Monti, C., & Unzurrunzaga, C. (2020). Acceso a la literatura científica desde Sci-Hub: Análisis y reflexión de las descargas en Argentina. *Revista Hipertextos, 8*(14), 111–136. https://doi.org/10.24215/23143924e022

Moore, S. A. (2020). Individuation through infrastructure: Get full text research, data extraction and the academic publishing oligopoly. *Journal of Documentation, 77*(1), 129–141. https://doi.org/10.1108/JD-06-2020-0090

Moskovkin, V. M., & Serkina, O. V. (2016). Is sustainable development of scientific systems possible in the neo-liberal agenda? *Ethics Sci Environ Polit., 16*, 1–9. https://doi.org/10.3354/esep00165

Moskovkin, V. M., Gakhova, N. N., & Yu, N. A. (2021). Downloading articles by russian researchers using the SCI-Hub resource. *Scientific and Technical Information Processing, 48*(3), 159–167. https://doi.org/10.3103/S0147688221030059

Nazarovets, S.A. (23, Apr. 2018). Black open access in Ukraine: Analysis of downloading Sci-Hub publications by Ukrainian Internet Users. arXiv 1804.08479v1. https://arxiv.org/ftp/arxiv/papers/1804/1804.08479.pdf

Nicholas, D., Boukacem-Zeghmouri, C., Xu, J., Herman, E., Clark, D., Abrizah, A., Rodríguez-Bravo, B., & Świgon, M. (2019). Sci-Hub: The new and ultimate disruptor? View from the front. *Learned Publishing, 32*(2), 147–153. https://doi.org/10.1002/leap.1206

Oakley, M. (2016). *Online piracy: Why Sci-Hub is disrupting scholarly publishing*. Georgetown University Library. https://www.library.georgetown.edu/sites/default/files/sci-hub-intro.pdf

Owens, B. (Feb. 2022). Sci-Hub downloads show countries where pirate paper site is most used. *Nature*. https://doi.org/10.1038/d41586-022-00556-y

Oxenham, S. (9 February 2016). Meet the robin good of science. Big Think. Retrieved from https://bigthink.com/culture-religion/a-pirate-bay-for-science/

Parkhill, M. (May 16, 2016). Sci-Hub: The Academic Cat Is Out of the Bag. Blog-Plum Analytics. Retrieved from http://plumanalytics.com/sci-hub-academic-cat-bag-post/

Pastor-Ramon, E. (2023). Sci-Hub use among Spanish researchers: Enemy or a learning opportunity for libraries? *Journal of Information Science*. https://doi.org/10.1177/01655515221142432

Pecastre, I. F., & Correa, F. G. (2016, September 26). The current system of knowledge dissemination isn't working and Sci-Hub is merely a symptom of the problem. Impact of Social Sciences. London School of Economics and Political Science. Retrieved Jan 30, 2023, from https://blogs.lse.ac.uk/impactofsocialsciences/2016/09/26/the-current-system-of-nowledge-dissemination-isnt-working-and-sci-hub-is-merely-a-symptom-of-the-problem/

Penn, L. (2018). Alternative ways of obtaining scholarly articles and the impact on traditional publishing models from a UK/European perspective. *Serials Review, 44*(1), 40–50. https://doi.org/10.1080/00987913.2018.1433906

Peters, M. A. (2020). Digital socialism or knowledge capitalism? *Educational Philosophy and Theory, 52*(1), 1–10. https://doi.org/10.1080/00131857.2019.1593033

Piryani, R., Dua, J., & Singh, V. K. (2019). Open access levels and patterns in scholarly articles from India. *Current Science, 117*(9), 1435–1440. https://doi.org/10.18520/cs/v117/i9/1435-1440

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*, e4375. https://doi.org/10.7717/peerj.4375

Plutchak, T.S. (2019). Epistemology: Three ways of talking about Sci-Hub. Against the Grain. Manuscript 8389, 60–62. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=8389&context=atg

Rossello, G., & Martinelli, A. (2023). *Breach of academic values and digital deviant behaviour: The case of Sci-Hub.* UNU-MERIT. UNU-MERIT Working Papers No. 009 https://www.merit.unu.edu/publications/wppdf/2023/wp2023-009.pdf

Saarti, J., & Tuominen, K. (2021). Openness, resource sharing and digitalization – An examination of the current trends in Finland. *Information Discovery and Delivery, 49*(2), 97–104. https://doi.org/10.1108/IDD-01-2020-0006

Sagemüller, F., Meißner, L., & Mußhoff, O. (2021). Where can the crow make friends? Sci-Hub's activities in the library of development studies and its implications for the field. *Development and Change, 52*(3), dech.12638. https://doi.org/10.1111/dech.12638

Saleem, F., Hasaali, M. A., & ul Haq, N. (2017). Sci-Hub & ethical issues. *Research in Social and Administrative Pharmacy, 13*(1), 253. https://doi.org/10.1016/j.sapharm.2016.09.001

Singh, V. K., Piryani, R., & Srichandan, S. S. (2020). The case of significant variations in gold–green and black open access: Evidence from Indian research output. *Scientometrics, 124*, 515–531. https://doi.org/10.1007/s11192-020-03472-y

Singh, V. K., Srichandan, S. S., & Bhattacharya, S. (2021). Is Sci-Hub increasing visibility of Indian research papers? An analytical evaluation. *Journal of Scientometric Res., 10*(1), 130–134. https://doi.org/10.5530/jscires.10.1.16

Steel, G. (2016). Wallets with a serious case of Stockholm syndrome: Sci-Hub and the future of scholarly communication. *The Winnower, 5*, e145677–e173287. https://doi.org/10.15200/winn.145677.73287

Strielkowski, W. (2017). Will the rise of Sci-Hub pave the road for the subscription-based access to publishing databases? *Information Development, 33*(5), 540–542. https://doi.org/10.1177/0266666917728674

Suh, J. (2022). An Analysis of Usage of Sci-Hub in Korea. *Journal of the Korean Society for Library and Information Science, 56*(4), 473–490. https://doi.org/10.4275/KSLIS.2022.56.4.473

Swartz, A. (2008). Guerilla open access manifesto. https://archive.org/details/GuerillaOpenAccessManifesto

Till, B. M., Rudolfson, N., Saluja, S., Gnanaraj, J., Samad, L., Ljungman, D., & Shrime, M. (2019). Who is pirating medical literature? A bibliometric review of 28 million Sci-Hub downloads. *Lancet Global Health, 7*(1), e30–e31. https://doi.org/10.1016/S2214-109X(18)30388-7

Timus, N., & Babutsidze, Z. (2016). Pirating European studies. *Journal of Contemporary European Research, 12*(3), 783–791. https://doi.org/10.30950/jcer.v12i3.760

Travis, J. (2016). In survey, Most give thumbs – Up to pirated papers. *Science News.* https://doi.org/10.1126/science.aaf5704

Tury, S., Robinson, L., & Bawden, D. (2015). The information seeking behaviour of distance learners: A case study of the University of London International Programmes. *The Journal of Academic Librarianship, 41*(3), 312–321. https://doi.org/10.1016/j.acalib.2015.03.008

Wan, S. (2022). Which nationals use Sci-Hub mostly? *The Serials Librarian, 83*(3–4), 228–232. https://doi.org/10.1080/0361526X.2023.2193613

Wang, X. W., Xu, S. M., Peng, L., Wang, Z., Wang, C. L., Zhang, C. B., & Wang, X. B. (2012). Exploring scientists' working timetable: Do scientists often work overtime? *Journal of Informetrics, 6*(4), 655–660. https://doi.org/10.1016/j.joi.2012.07.003

Wilcock, J. (2018). Putting an end to download-and-go: The website's role in a content marketing ecosystem. *Learned Publishing, 31*, 173–179. https://doi.org/10.1002/leap.1131

Authors and Affiliations

Zhenxia Gao[1] · **Vladimir M. Moskovkin**[2] · **Nina N. Gakhova**[3] · **Igor V. Boichuk**[4]

✉ Zhenxia Gao
gaozhenxia@lcu.edu.cn

1 Department of Film and TV Technology, School of Media and Technology, Liaocheng University, Liaocheng, China

2 Independent Researcher, Vimperk, Czechia

3 Department of Applied Informatics and Information Technologies, Belgorod National Research University, Belgorod, Russia

4 Department of Foreign Languages and Professional Communication, Belgorod National Research University, Belgorod, Russia