Exploring arXiv usage habits among Slovenian scientists

Zala Metelko and Jasna Maver
Department of Library and Information Science and Book Studies, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia

Received 25 July 2022
Revised 8 January 2023
Accepted 14 January 2023

Abstract

Purpose – This study investigates how important the preprint arXiv is for Slovenian scientists, whether there are differences between scientific disciplines and the reputation of arXiv among Slovenian scientists. We are also interested in what advantages and disadvantages scientists see in using arXiv.

Design/methodology/approach – A voluntary sample of active researchers from the scientific fields covered by arXiv was used. Data were collected over 21 days in September 2021 using a 40-question online survey. In addition to descriptive statistics, nonparametric statistical methods such as Pearson's chi-squared test for independence, Kruskal-Wallis' H-test and Mann-Whitney's U-test were applied to the collected data.

Findings – Among Slovenian scientists there is a wide range of different users of arXiv. The authors note differences among scientific disciplines. Physicists and astronomers are the most engaged, followed by mathematicians. Researchers in computer science, electrical engineering and systems science seem to have recognized the benefits of the archive, but are still hesitant to use it. Researchers from the other scientific fields participated in the survey to a lesser extent, suggesting that arXiv is less popular in these scientific fields. For Slovenian scientists, the main advantages of arXiv are faster access to knowledge, open access, greater impact of scientists' work and the fact that publishing in the archive is free of charge. A negative aspect of using the archive is the frustration caused by the difficulties in assessing the credibility of articles.

Research limitations/implications – A voluntary sample was used, which attracted a larger number of researchers but has a higher risk of sampling bias.

Practical implications – The results are useful for international comparisons, but also provide bases and recommendations for institutional and national policies to evaluate researchers and their performance.

Originality/value – The results provide valuable insights into arXiv usage habits and the reasons for using or not using arXiv by Slovenian scientists. There is no comparable study conducted in Slovenia.

Keywords Slovenia, Scientific literature, arXiv, Preprint servers, Scholarly publishing, International studies

Paper type Research paper

Introduction

Researchers in Slovenia, especially in the technical and natural sciences, are evaluated primarily on the basis of their publications in peer-reviewed journals (recorded in Journal Citation Reports) and the value of the journal in which the article was published (measured by the journal's impact factor), as well as the number of citations in the same group of journals. However, the peer review process has shortcomings and limitations; for example, there is concern that individuals from underrepresented minorities or less prestigious institutions may receive biased reviews (Polka *et al.*, 2018). Statistical errors can be found in peer-reviewed publications; the question is how to avoid statistical errors in submissions and optimize post-publication corrections (Allison *et al.*, 2016). There is a significant imbalance in the distribution of effort for peer review in the scientific community (Kovanis *et al.*, 2016).

Journal of Documentation
Vol. 79 No. 7, 2023
pp. 72-94
Emerald Publishing Limited
0022-0418
DOI 10.1108/JD-07-2022-0162

This work was supported by the Slovenian Research Agency (grant number P2-0214).

Balietti (2016) points out a number of problems with peer review; assessing the quality of innovative studies is a difficult task, even for trained scientists. Therefore, it is sometimes difficult for a journal editor to find competent reviewers. A review for a journal must be done in addition to normal academic duties, which often means that a scientist spends less time on it than he should. The time between submission of an article and its publication in a journal can be very long, with cycles of rejection, revision and resubmission (Powell, 2016; Vale and Hyman, 2016; Huisman and Smits, 2017). In the article by Mishkin *et al.* (2020), the authors model the article submission process as a Markov process. They note that this process can take 1 to 3 years. The authors point out the following problem: During this time, another researcher may independently develop the same or a very similar idea and publish it earlier. In such a situation, it is difficult for the authors to prove that they had the original idea months or years earlier. This is unfair and painful for the researchers, especially if the idea becomes influential and frequently cited. Smith (2006) has an extremely negative opinion and negative experience with peer review. In his view, it is poor at detecting gross errors and almost useless at detecting fraud; it is slow, expensive, wastes academic time, is highly subjective, is prone to bias and can be easily abused. The authors of the articles (Polka *et al.*, 2018; Hachani, 2015) believe that peer review would improve if journals also published peer reviews. In the paper by Spezi *et al.* (2018), the authors explore the differences between the theory and practice of peer review in open access mega-journals which follow a soundness-only peer review policy and leave the decision about the novelty or significance of the research finding to the community.

Early career researchers need to validate their research findings through publication as soon as possible. They need feedback from colleagues and peers. Preprint archives help achieve these goals (Vale and Hyman, 2016; Powell, 2016; Bourne *et al.*, 2017; Avasthi *et al.*, 2018; O'Roak, 2018; Mudrak, 2020). There are several preprint archives available. Among the best known are arXiv and bioRxiv, whose success has led to the establishment of archives in other scientific fields, for example chemRxiv in chemistry, AgriXiv in agriculture and related sciences and SocArxiv in the social sciences (McGlashan and Hadley, 2017). Most preprint archives offer free access and do not charge for article submission, making it easy for researchers with limited resources to read research articles and publish their findings (Bohannon, 2016; Till *et al.*, 2019). In the articles by Annesley *et al.* (2017)and Sarabipour *et al.* (2019), the authors enumerate many benefits of preprints, including rapid dissemination of scholarly work, open access, establishing priority or concurrence, receiving feedback and facilitating collaborations. Articles in preprint archives are copyrighted works that can be referenced because they are marked with DOI in most archives, making the work a permanent part of the scientific record. Several funders and national institutions allow researchers to cite preprints in their grant proposals (Puebla *et al.*, 2022; Crotty, 2018). Authors often upload an article to a preprint archive and then submit it to a peer-reviewed scientific journal (Abdill and Blekhman, 2019; Bastian, 2016). Very few journals reject manuscripts because they were published on a preprint server (Borgman, 2007). Some preprint archives (arXiv, 2019) allow authors to add a web link to the final version of the electronic journal article. Such web links create more confidence in the quality of articles in the archive and help authors attract new readers (Inglis and Sever, 2017; Serghiou and Ioannidis, 2018).

In this paper, we have restricted ourselves to the arXiv archive. Currently, arXiv provides free access to more than two million scientific articles in physics and astronomy, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science and economics. Article submission is free of charge. Articles do not go through a review process, but are published as submitted (arXiv, 2022).

So far, no study has been conducted in Slovenia on the use of arXiv. The aim of the study is to find out how important arXiv is for Slovenian scientists, what advantages and disadvantages they see in using arXiv for their research work and how they use it in terms of

reading, publishing and citing articles. Previous research has focused mainly on the most successful publishing countries or English-speaking countries. To get a complete picture, smaller countries such as Slovenia also need to be studied.

Literature review

More than two decades ago, an article (Brown, 1999) examining the information-seeking behaviour of scientists found that preprints are an important source of literature for research in some scientific fields, such as physics and mathematics. Further research on the citation, use and acceptance of preprints in the fields of physics and astronomy (Brown, 2001) confirmed this finding: arXiv papers accounted for a substantial and growing proportion of citations in a number of journals in physics and astronomy. The author of the article (Youngen, 1998) reached a similar conclusion. The increasing demand for current research in physics and astronomy led researchers to cite preprints rather than wait for them to appear in print. Garner *et al.* (2001) found that preprint servers provided access to needed scientific information at a reasonable price, enabled rapid dissemination of results and provided quick feedback. In 2003, ArXiv offered a separate archive section for quantitative biology. The percentage of biological publications on ArXiv reached 8% of publications submitted to arXiv in 2002 (Butler, 2003). Prakasan and Kalyane (2004) have shown that preprints are an essential part of physicists' scientific communication. Of the preprints in four different categories of high-energy physics, at least 39% from each category were cited at least once in the Science Citation Index (1991–2002). More than 60–70% of the citations occurred immediately after the preprint was published in the archive. The number of authors citing preprint archives in high-energy physics increased by a factor of about 1.25 annually, with the ratio increasing. Several other studies have examined the impact of preprints on arXiv on the number of citations (Schwarz and Kennicutt, 2004; Metcalfe, 2005, 2006). Publishing on arXiv accelerated citations because articles were available sooner (Moed, 2007). Brody *et al.* (2006) examined the correlation between the number of article downloads and the number of citations. The physics preprint archive arXiv was used for verification. They show that the short-term impact of web usage of preprints predicts a medium-term impact on citations of the final article. An analysis in Davis and Fromerth (2007) of 2,765 articles published in four mathematics journals between 1997 and 2005 found that articles deposited on arXiv received on average 35% more citations than articles not deposited, and that this difference was most pronounced for highly cited articles. Despite the citation advantage, articles published on arXiv were 23% less likely to be downloaded from the publisher's website. The data suggest that arXiv and the publisher's website may serve different functional needs of readers. In the articles by Haque and Ginsparg (2009, 2010), the authors address the interesting observation that preprints published on arXiv at the beginning and end of the day reach a larger readership and are cited more frequently in subsequent years than preprints published in the middle of the day. Shuai *et al.* (2012) analysed the online response to preprint publications on arXiv and examined the post-publication delay of article downloads and Twitter mentions. In a comprehensive analysis of data from arXiv and the Web of Science (Lariviere *et al.*, 2014), the authors find that arXiv is playing an increasingly important role as both a notification service and a preferred archive. However, the goal of most scientists is still to publish in peer-reviewed journals. In the paper by Aman (2015), the author performed an analysis of cited references on the Web of Science to preprints on arXiv. A total of 900,000 cited references to arXiv preprints were identified for the period 1991–2013. The results show that the number of citations peaks in the year following submission and then declines rapidly. The paper identifies 24 "arXiv-friendly" journals in which the majority of articles reference arXiv. In the study by Li *et al.*, 2015, the authors examined the number of citations in academic publications for four major subject repositories, arXiv, Research Papers in Economics, Social Science Research Network and PubMed Central.

The study was limited to the Scopus database from 2000 to 2013. Each subject repository was most frequently cited within its subject area, but also attracted many citations from other subject areas, with a steady increase in the number of citations. The paper by Noruzi (2016), which gives an overview of the citation rate of arXiv.org since its launch in August 1991 to 2016 based on the citation database Scopus, shows that papers published on arXiv were most cited in the following order: by physics and astronomy, mathematics, computer science and engineering. The order of countries that stand out in the number of citations were the United States, Germany, China, the United Kingdom, France and Italy. In the paper by Kim *et al.* (2020), the number of citations of preprints published on arXiv.org in three subfields of physics that were early users of arXiv is used as an indicator of the quality of individual articles that is independent of a journal's reputation or impact factor. The authors show that previous estimates of the impact of a journal's reputation on the impact of an individual article (as measured by citations) are likely inflated. They note that high quality preprints in these subfields are less likely to be published in journals today than in earlier years. In the study by Wang *et al.* (2020a), the authors analysed several factors that preprints contribute to improving scientific communication. To this end, the authors used four measures, one of which related to scholarly communication and was based on bibliometric indicators (Web of Science and Scopus citations), while the others reflected usage (Web of Science usage numbers), capture (Mendeley readers) and social media attention (Tweets). The results show that the "early-view" and "open-access" effects of preprints contribute to a measurable citation and readership advantage of preprints. Articles with preprints were more likely to be mentioned on social media and had a shorter attention delay on Altmetric. Usage and capture correlated only moderately but more strongly with citations than tweets. The results reported in Wang *et al.* (2020b) show that arXiv papers have significant citation advantage across Web of Science, Scopus and Google Scholar. The authors of the article published by Mishkin *et al.* (2020) believe that publishing research papers on arXiv has become the standard in many scientific fields covered by arXiv. They argue that arXiv offers great benefits to both the individual researcher and the field as a whole. ArXiv helps researchers at all career stages. They list the following benefits: Building a professional identity instead of an empty CV, protection from independent rediscovery or idea theft, protection from editor gatekeeping, arXiv has become the main "fodder for new results" through which people discover new work. They see the following benefits of arXiv for research and the broader community: faster and wider dissemination and better explanation, arXiv plays an important role in public access to documents for grant reports and public talks. They also note that crowdsourcing review of preprints is the way of open science and considers more aspects than traditional reviews; arXiv levels the playing field and reduces the advantages of famous labs.

Research

Methodology

The survey used a 40-question questionnaire divided into seven sections: personal information, familiarity with arXiv, advantages and disadvantages of using arXiv, reading and publishing on arXiv, reading only, publishing only and using other archives for preprints. The questionnaire contained different types of questions: single/multiple choice, liker scale, open-ended, closed-ended and matrix questions. The survey was conducted online; data were collected over 21 days in September 2021.

Sample

The study population includes 1711 active scientists conducting research in the scientific fields covered by the arXiv archive (Slovenia is a small country with a population of about

2 million). The scientists are from various institutions [1], including three public universities, one private university and one large scientific and technical research institute. The list of active scientists was compiled by visiting the websites of the above institutions and collecting the email addresses of their employees. The list also includes PhD students employed as teaching assistants or junior scientists at these institutions, as well as temporary employees. We use a voluntary sample. The scientists received an email invitation to participate in the survey, which included a link to the questionnaire. We received 218 completed questionnaires, which we included in the analysis. This represents 12.7% of the included active scientists. We present the descriptive characteristics of the data and perform various analyses using statistical methods such as Pearson's chi-square test for independence, Kruskal–Wallis' H-test and Mann–Whitney's U test.

Results

Personal information

The Personal information section contains five demographic questions about respondents' gender, age, scientific discipline, years in the profession and level of education [2]. Of the 218 respondents, 46 identified themselves as female and 172 as male. The age groups of the respondents with their number in parentheses were up to 25 (4), 25–44 (140), 45–64 (63) and 64+ (11). Thus, the largest group was respondents between the ages of 25 and 44. The highest educational level VIII/2 had 149 respondents. The groups of respondents with lower educational level, the number of which is given in parentheses, were VIII/1 (10), VII (54), VI/2 (4) and VI/1 (1). The distribution of scientists by years in the profession shows that the number of scientists decreases with the number of years in the profession: 125 with less than 15 years, 43 from 15 to 24 years, 31 from 25 to 34, 14 from 35 to 44 and five from over 44 years. It can be concluded that the majority of the respondents who participated in the study have yet to confirm themselves through intensive publication of scientific articles.

Table 1 shows the distribution of scientists by scientific discipline. The largest group comes from Computer Science, followed by Physics and Astronomy and Mathematics, while some other scientific disciplines are not so well covered, namely Biology and Economics.

Familiarity with arXiv

Respondents were asked if they were familiar with the arXiv archive. Table 2 shows that this was the case for almost four-fifths of the respondents. Large differences between scientific disciplines can be seen. The vast majority of respondents from Mathematics, Computer Science and Physics and Astronomy are familiar with the archive. In Economics and Electrical Engineering and Systems Science, the proportion drops to less than two-thirds and in Biology to less than one-third.

When asked who brought arXiv to their attention, the results were as follows: Teacher (19), colleague (55), librarian (0), I discovered the archive myself (89), others (10). The data

Table 1. Number and percentage of respondents by scientific discipline

Scientific discipline	Number	%
Physics and astronomy	60	27.5
Mathematics	39	17.9
Electrical engineering and systems science	25	11.5
Computer science	73	33.5
Biology	16	7.3
Economics	5	2.3
Total	218	100

show that librarians were not involved in introducing researchers to the archive; most respondents discovered arXiv themselves. Of the 173 researchers who were aware of the archive, only 143 used it. Table 3 shows the results by scientific discipline. A high percentage of researchers are from Mathematics (97.4%), Physics and Astronomy (85.7%) and Computer Science (81.5%). The percentage is lower in Electrical Engineering and Systems Science (69.2%). For Biology and Economics, we have only one active respondent each. Therefore, we cannot draw any conclusions from the survey questions for these two scientific disciplines. At most, we can conclude that there are individuals with certain characteristics.

We were interested in the reasons for not using the archive. Respondents had the option of either selecting the answer "Use of the archive is not widespread in my research area" or entering their own response. Of the 30 respondents, 16 chose the second option. Some of them felt that the quality of articles in arXiv is insufficient. Someone thought that articles that are not published in peer-reviewed journals end up there. Someone noted that the rules for what is and is not appropriate for publication are too loose, while several others saw no reason to publish in the archive and said that peer-reviewed journals are a better option. Some respondents pointed out that publishing on arXiv does not earn points toward promotion to a higher scientific or academic title. One of the reasons for not using the archive was the possibility of a dispute with scientific journals that do not allow pre-publication in the archive. Some use the archive indirectly through Google scholar. One researcher had just started his research and believed he would use the archive in the near future.

Advantages and disadvantages of using arXiv

Respondents rated eight positive benefits of using arXiv. Table 4 shows the benefits offered in the survey and the results. The most important benefits for respondents were "Faster access to knowledge" and "Access to articles is not dependent on financial resources." Somewhat surprisingly, the benefit "More accurate recognition of contributions" received the most "Not important" responses. More than one in four respondents answered this way, 28% to be exact. For all benefits offered, the "Moderately important" or "Very important" responses combined accounted for at least 72%.

	No	Yes	Yes [%]
Physics and astronomy	11	49	81.7
Mathematics	1	38	97.4
Electrical engineering and systems science	12	13	52.0
Computer science	8	65	89.0
Biology	11	5	31.3
Economics	2	3	60.0
Total	45	173	79.4

Table 2. Number and percentage of respondents familiar with arXiv by scientific discipline

	No	Yes	Yes [%]
Physics and astronomy	7	42	85.7
Mathematics	1	37	97.4
Electrical engineering and systems science	4	9	69.2
Computer science	12	53	81.5
Biology	4	1	20.0
Economics	2	1	33.3
Total	30	143	82.7

Table 3. Number and percentage of respondents using arXiv by scientific discipline

The graphs in Figure 1 show the results by scientific discipline. The graphs show that the percentage of "Very important" responses for "Faster access to knowledge" is lower in Electrical Engineering and Systems Science and Computer Science than in Physics and Astronomy and Mathematics. The same is true for the benefits "Greater impact of the work of the scientist" and "Peer and colleague response to published articles." For each of the specified benefits, we performed the Kruskal–Wallis H-test (see Table 5) to test the null hypothesis that none of the scientific disciplines is stochastically dominant over the others in terms of the benefit tested. Because the test requires that the size of all groups be at least 5, we did not include a biologist and an economist in the test. The H-statistics were corrected for ties. If the p-value is less than 0.05, the null hypothesis is rejected and the differences between the scientific disciplines are confirmed. The results of the Kruskal–Wallis H-test are consistent with the above observations.

Other benefits of using arXiv mentioned by respondents included daily review of new articles in their research area in one place, ease of review, quick and easy access to articles, ease of searching for specific information and not having to search many scholarly journals. Then most of the articles in one place, articles are available immediately and there is no waiting, publications of the latest results, easy publication of new ideas, faster progress in the field and a very good publication ensures that no one steals your idea. For one respondent, two features were crucial: fast notification of new results and "registration" (timestamp) of the result.

The quality of articles in the archive is an important issue. In the survey, we asked first about the characteristics of a poor quality article and then about the percentage of such articles in arXiv. Tables 6 and 7 show the results.

For most respondents, a poor quality article violates the principles of scientific research (generality, objectivity, verifiability, validity, reliability, accuracy and systematicity), does not take into account existing research and does not contribute to it. Poor qualities that respondents can more easily tolerate include spelling and grammatical errors and poor style of expression. In addition, respondents described poor articles as boring articles with no structure, articles with errors in content, articles that lie – they talk about things that are not true, articles with a lot of form and little content and plagiarism. Two of the respondents indicated that bad articles are primarily those that have not been peer reviewed.

As shown in Table 7, which shows the results for the percentage of poor quality articles in arXiv reported by respondents, none of the respondents chose the "no poor articles" option. Half of the respondents (50.3%) indicated 10–30%. The percentages of respondents who indicated that the percentage of poor quality articles was smaller or larger were similar at 24.5% and 25.2%, respectively. However, there are differences between scientific disciplines. In Computer Science and Electrical Engineering and Systems Science, the percentage of

Table 4.
Number of responses in each category for each of the 8 benefits of using arXiv

	Not important	Moderately important	Very important
Faster access to knowledge	3	24	116
Greater impact of the work of the scientist	25	57	61
More accurate recognition of contributions	40	57	46
Peer and colleague response to published articles	30	70	33
Articles do not have to be changed according to publisher standards before publication	39	71	33
Publication of an article is not dependent on financial resources	26	55	62
Access to articles is not dependent on financial resources	13	21	109
Publication of articles helps with career development	33	68	42

Figure 1.
Number of responses in each category for each of the 8 benefits of using arXiv, shown by scientific discipline

respondents who indicated that the percentage of bad articles is greater than 10–30% is higher (36% and 33%, respectively), while this percentage is lower in Physics and Astronomy and Mathematics (18% and 19%, respectively).The Kruskal–Wallis H-test (see Table 7) confirms significant differences between scientific disciplines ($p < 0.05$).

The questioner also asked about the disadvantages of using arXiv. Table 8 shows the bad aspects of arXiv use offered and the results. As many as 27% of the respondents believed that publishing a bad article on arXiv would have negative consequences for their career. If one adds the answers "partially applies" here, one even comes to 64%. The aspect "I find it difficult to judge whether the articles are credible" also stands out. Almost 52% of respondents selected "applies" or "partially applies". Figure 2 shows the distribution of responses by scientific discipline. For the responses "I spend too much time reading articles

Table 5. Results of Kruskal–Wallis H-tests for each of the 8 benefits of using arXiv

	$H \approx \chi^2(k-1)$	$H_{corrected}$	k	p-value
Faster access to knowledge	5.901	12.649	4	*0.0055*
Greater impact of the work of the scientist	12.655	14.853	4	*0.0019*
More accurate recognition of contributions	3.041	3.444	4	0.3281
Peer and colleague response to published articles	18.067	21.361	4	*8.9E−05*
Articles do not have to be changed according to publisher standards before publication	1.479	1.75	4	0.6258
Publication of an article is not dependent on financial resources	0.559	0.651	4	0.8846
Access to articles is not dependent on financial resources	1.237	2.213	4	0.5294
Publication of articles helps with career development	6.721	7.858	4	*0.0490*

Table 6. Number of responses for each of the 7 characteristics of a poor quality article and percentage of respondents who selected the characteristic

	Number	%
The article does not contribute to science	118	83
Failure to take into account existing research	118	83
Failure to adhere to the principles of scientific research (generality, objectivity, verifiability, validity, reliability, accuracy, systematicity)	125	87
The topic, abstract and content of the article are inconsistent	73	51
Poor style of expression	66	46
Spelling and grammatical errors	49	34
Incomplete references	77	54

Table 7. Results for percentage of poor quality articles in arXiv by scientific discipline and results of Kruskal–Wallis H-test

	No poor articles	<10%	10%–30%	30%–50%	>50%
Physics and astronomy	0	15	20	7	0
Mathematics	0	12	18	6	1
Electrical engineering and systems science	0	2	4	1	2
Computer science	0	5	29	15	4
Biology	0	1	0	0	0
Economics	0	0	1	0	0
Total	0 (0%)	35 (24.5%)	72 (50.3%)	29 (20.3%)	7 (4.9%)

Kruskal–Wallis H-test			
$H \approx \chi^2(k-1)$	$H_{corrected}$	k	p-value
10.72	12.62	4	*0.0055*

that turn out to be unreliable" and "Publishing an article on arXiv would prevent the article from being published in a scientific journal," the responses "partially applies" and "applies" are more likely among scientists in Computer Science and Electrical Engineering and Systems Science than in Physics and Astronomy and Mathematics. For the aspect "I find it difficult to judge whether the articles are credible," the answers "partially applies" and "applies" together predominate in all science fields except Physics and Astronomy. Table 9 contains the results of the Kruskal–Wallis H-tests, with H-statistics corrected for ties, and

	Applies	Partially applies	Not applies
I find it difficult to judge whether the articles are credible	11	63	69
I spend too much time reading articles that turn out to be unreliable	5	33	105
Publishing a bad article would have negative consequences for my career	39	52	52
Publishing an article exposes me to plagiarists	9	41	93
Publishing an article on arXiv would prevent the article from being published in a scientific journal	6	44	93

Table 8.
Number of responses in each category for each of the 5 bad aspects of arXiv use

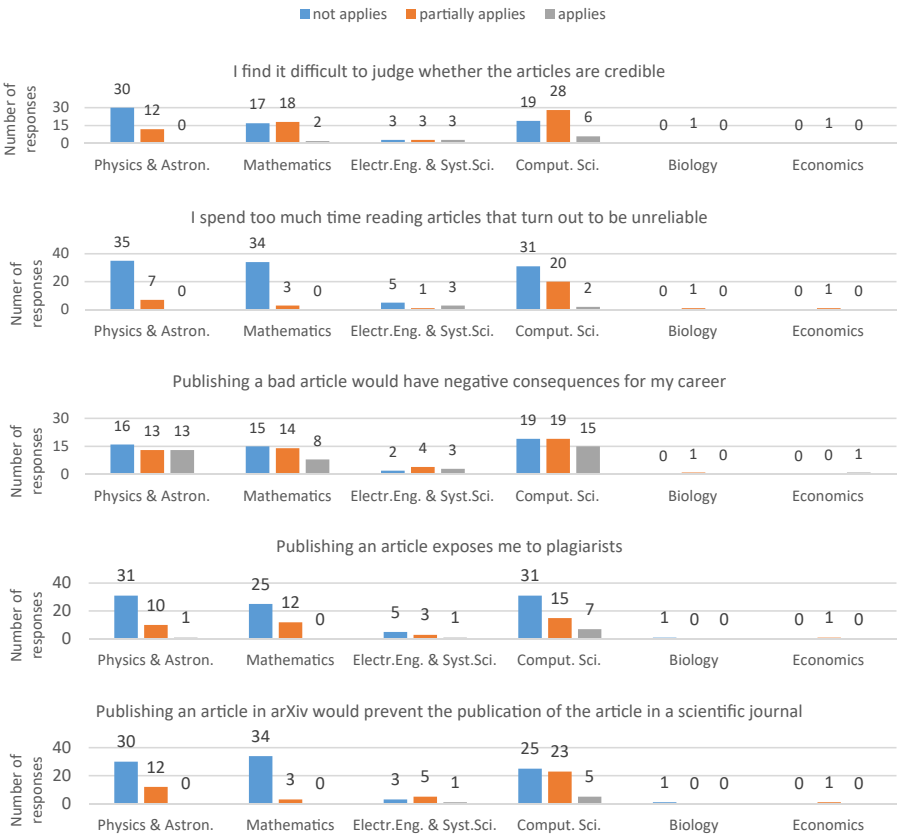

Figure 2.
Number of responses in each category for each of the 5 bad aspects of using arXiv, shown by scientific discipline

groups with a count of less than 5 not included in the tests. Respondents pointed out several other disadvantages of archive use. Here are their comments. The article published on arXiv may receive more citations than the same article in a scientific journal. These citations do not contribute to academic promotion. It is not clear how protected the published content is. Bad articles published in the archive quickly reach readers and can cause harm. In arXiv, scientific subfields are not evenly represented. Thus, it is almost obligatory to follow new arXiv articles in one subfield but not in another because there is no separate "category" in arXiv; some fields are too broad. Some researchers do not publish their articles on arXiv. Some publications are too fast and are insufficiently verified. An article published on arXiv is less valid because it has not gone through a review process.

One characteristic that indicates the popularity of an archive is the frequency of its use. Most respondents, 22%, use arXiv once or twice a week (see results in Table 10). However, the data show that frequency varies by scientific discipline. Most respondents from Physics and Astronomy use the archive daily, from Mathematics once or twice a week, from Computer Science every 14 days and from Electrical Engineering and Systems Science less than once a month. The Kruskal–Wallis H-test confirms significant differences between scientific disciplines in the frequency of arXiv use ($p < 0.05$).

Three different groups of arXiv users

The last question in the "Advantages and disadvantages of using arXiv" section divided respondents who use arXiv into three groups: those who read and publish articles in the archive, those who only read and those who only publish. Figure 3 shows the number of

Table 9.
Results of Kruskal–Wallis H-tests for each of the 5 bad aspects of using arXiv

	$H \approx \chi^2(k-1)$	$H_{corrected}$	k	p-value
I find it difficult to judge whether the articles are credible	12.723	15.875	4	*0.0012*
I spend too much time reading articles that turn out to be unreliable	10.145	17.601	4	*5.3E−04*
Publishing a bad article would have negative consequences for my career	1.038	1.174	4	0.7592
Publishing an article exposes me to plagiarists	2.821	4.036	4	0.2576
Publishing an article in arXiv would prevent the publication of the article in a scientific journal	17.386	25.059	4	*1.5E−05*

Table 10.
Number of responses for each category of frequency of arXiv use, by scientific discipline and results of the Kruskal–Wallis H-test

	Daily	Every other day	Once or twice a week	Every 14 days	Once a month	Less than monthly
Physics and astronomy	18	7	8	4	2	3
Mathematics	6	0	12	7	8	4
Electrical engineering and systems science	1	1	2	2	0	3
Computer science	2	7	10	13	12	9
Biology	0	0	0	0	1	0
Economics	1	0	0	0	0	0
Total	28	15	32	26	23	19

Kruskal–Wallis H-test $H \approx \chi^2(k-1)$	$H_{corrected}$	k	p-value
23.412	24.196	4	2.27E−05

respondents for each of the three group categories by scientific discipline. There were 95 respondents in the "read and publish" group, 45 in the "read only" group and only three in the "publish only" group. The majority of respondents from Physics and Astronomy and Mathematics belong to the "read and publish" group. In Computer Science, the "read and publish" and "read only" groups are similar in size. In Electrical Engineering and Systems Science, two-thirds of respondents are in the "read only" group and one-third are in the "read and publish" group. Three respondents (2% of arXiv users) would publish only. One is from Mathematics, two from Computer Science.

Comparison of the "read and publish" and "read only" groups

Forty-five respondents use arXiv only to read articles. When asked why they do not publish on arXiv, nineteen researchers chose the response "Publishing on arXiv is not encouraged in the work environment." Five researchers do not publish on arXiv because it would prevent publication in a scientific journal. Four researchers selected the response "I have not received an endorsement from an established author," and eleven respondents selected the option "Other," most of whom indicated that they had only been in the profession a short time and had not yet written an article. A few researchers had so far written only one article, which they preferred to publish in a scientific journal. One researcher published articles on bioRxiv.

As shown in Table 11, the majority of respondents in both groups "read and publish" and "read only" have already cited an article published in the archive. However, there is an important difference between the groups; the test for independence confirms connection with the groups with statistical significance ($p < 0.05$).

Table 12 shows the same data broken down by scientific discipline. If we exclude the groups with only one member, we find that the percentage of researchers who have already cited an article in arXiv is highest in Mathematics, followed by Physics and Astronomy, Computer Science and Electrical Engineering and Systems Science. The same order according to scientific disciplines was also found in the article (Li *et al.*, 2015) and with reverse order for the first two in the work of Noruzi (2016).

Table 13 shows how many arXiv articles respondents typically read per month. Most respondents in both groups read 1 to 4 articles per month. A difference between the groups, about 10% or more, can be seen for the choices "<1" and "≥15". The results suggest that researchers who publish articles on arXiv also read more of the articles published there.

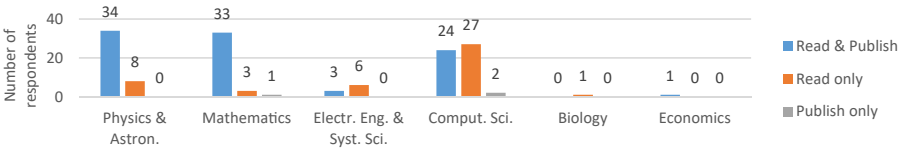

Figure 3.
Number of respondents for each of the three groups of arXiv usage types, presented by scientific discipline

	Yes	No
Read and publish group	87 (91.6%)	8 (8.4%)
Read only group	32 (71.1%)	13 (28.9%)
Pearson's chi-squared test for independence		
χ^2	df	p-value
10.03	1	0.0015

Table 11.
Number and percentage of respondents who have already cited an article on arXiv, broken down by group type and results of Pearson's chi-squared test for independence

Table 13 also shows the result of the Mann–Whitney U test. The test approximated the U statistic by the normal distribution, with variance corrected for ties. The calculation includes a correction for continuity. The test confirms the differences between the groups ($p < 0.05$).

Comparison of the "read and publish" and "publish only" groups

Three of the respondents use arXiv only to publish articles, two are researchers in computer science and one is a mathematician. When asked why they do not read articles published on arXiv, they all chose the answer "I only read professionally peer-reviewed articles." Two other response options were available: "Many articles on arXiv are of poor quality" and "Other".

During the submission process to arXiv, authors submitting a paper in a subject category for the first time may be required to obtain an endorsement from an established arXiv author. When asked if they needed an endorsement, 14 respondents answered yes and 84 answered no. Several respondents, 21, indicated that they had already endorsed a new author. Table 14 shows the responses by scientific discipline.

Table 12. Number and percentage of respondents who have already cited an article on arXiv by scientific discipline

	Yes	No	Yes [%]
Physics and astronomy	35	7	83.3
Mathematics	36	0	100
Electrical engineering and systems science	6	3	66.7
Computer science	40	11	78.4
Biology	1	0	100
Economics	1	0	100
Total	119	21	85.0

Table 13. Number of articles read by respondents per month and results of Mann–Whitney U test

Number of articles read per month	<1	1–4	5–9	10–14	≥15
Read and publish group	10 (10.5%)	41 (43.2%)	24 (25.3%)	7 (7.4%)	13 (13.7%)
Read only group	9 (20%)	22 (48.9%)	12 (26.7%)	2 (4.4%)	0 (0%)
Total	19 (13.6%)	63(45%)	36 (25.7%)	9 (6.4%)	13 (9.3%)

Mann–Whitney U test					
U_R	$U_{R\&P}$	M	σ_{corr}	z_U	p-value (one-sided)
1629	2,646	2137.5	211.2	2.49	0.0064

Note(s): Numbers and percentages are given for each category, broken down by group type

Table 14. Results for new author endorsement by scientific discipline

	Did you need to obtain an endorsement?		Have you ever endorsed a new author?	
	Yes	No	Yes	No
Physics and astronomy	8	26	9	25
Mathematics	4	30	8	26
Electrical engineering and systems science	0	3	1	2
Computer science	2	24	3	23
Economics	0	1	0	1
Total	14	84	21	77

Figure 4 shows the number of articles published by respondents on arXiv. The average number was 30 and the median was 10. The averages for the scientific disciplines whose numbers are in parentheses were Physics and Astronomy (54), Mathematics (20), Computer Science (6) and Electrical Engineering and Systems science (3). Nine of the respondents published 100 articles or more each. The largest number of articles published on arXiv by a respondent was 250; the researcher is a mathematician.

Table 15 shows data on the proportion of articles that respondents publish on arXiv. The response "less than half" predominates in Mathematics, Computer Science and Electrical Engineering and Systems Science. The answer "all" is predominant in Physics and Astronomy (47%). It is 17.6% in Mathematics and 0% in the other scientific disciplines. The differences between the groups are confirmed.

Table 16 shows the results to the question "How many of the articles you published on arXiv were published in a scientific journal?" More than half of the respondents (60%) chose the answer "all" and almost one in four respondents chose the answer "more than half". The Kruskal–Wallis H-test confirms the differences between scientific disciplines.

It is not common practice to publish articles first in a scientific journal and only then on arXiv (see Table 17). Nevertheless, quite a number of researchers use this practice. We also calculated the Kruskal–Wallis H-test. The calculated p-value is just above 0.05.

The time between publication of an article on arXiv and publication in a scientific journal is typically up to six months or up to a year, and less frequently up to two months or more than a year, although the time span can vary widely (see Figure 5). There are differences between scientific disciplines. In Physics and Astronomy, the most frequently selected value is two to six months; in Mathematics, more than one year and in Computer Science and Electrical Engineering and Systems Science six months to one year. The results for

Figure 4. Results for the number of articles published on arXiv

	Less than half	Half	More than half	All
Physics and astronomy	8	0	10	16
Mathematics	13	6	9	6
Electrical engineering and systems science	2	0	1	0
Computer science	20	4	2	0
Economics	1	0	0	0
Total	44	10	22	22

Kruskal–Wallis H-test $H \approx \chi^2(k-1)$	$H_{corrected}$	k	p-value
23.421	26.258	3	*1.99E−06*

Table 15. Results for the proportion of articles that respondents publish on arXiv, by scientific discipline and results of the Kruskal–Wallis H-test

mathematics are consistent with the empirical analysis for mathematics published in the article by Wang *et al.* (2020a), in which the delay is estimated to be a median of 13 months and an average of 16.13 months. The results are also consistent with the figures in (Lariviere *et al.*, 2013), where the delay in physics and astronomy is less than half a year on average, while it is longer in mathematics (>1 year).

Two researchers from physics and astronomy, two from mathematics and two from computer science, that is 6.1% of respondents from the "publish only" and "read and publish" groups combined, have already had the experience of having a publication in a scientific journal rejected because they had previously published the article on arXiv. This is still a small percentage. In most cases, scientific journals accept articles with previous publication on arXiv.

Table 16.
Results for the proportion of the researcher's articles deposited in arXiv and published in a scientific journal, by scientific discipline and results of the Kruskal–Wallis H-test

	None	Less than half	Half	More than half	All
Physics and astronomy	1	0	1	6	26
Mathematics	0	0	2	10	22
Electrical engineering and systems science	0	1	1	0	1
Computer science	4	3	2	7	10
Economics	0	0	1	0	0
Total	5	4	7	23	59

Kruskal–Wallis H-test $H \approx \chi^2(k-1)$	$H_{corrected}$	k	p-value
9.221	9.361	3	*0.0093*

Table 17.
Articles already accepted in scientific journals at the time of submission to arXiv, by scientific discipline and results of the Kruskal–Wallis H-test

	None	Less than half	Half	More than half	All
Physics and astronomy	0	22	2	6	4
Mathematics	0	30	3	1	0
Computer science	0	20	0	2	4
Economics	0	1	0	0	0
Electrical engineering and systems science	0	0	2	1	0
Total	0	73	7	10	8

Kruskal–Wallis H-test $H \approx \chi^2(k-1)$	$H_{corrected}$	k	p-value
3.273	5.955	3	0.051

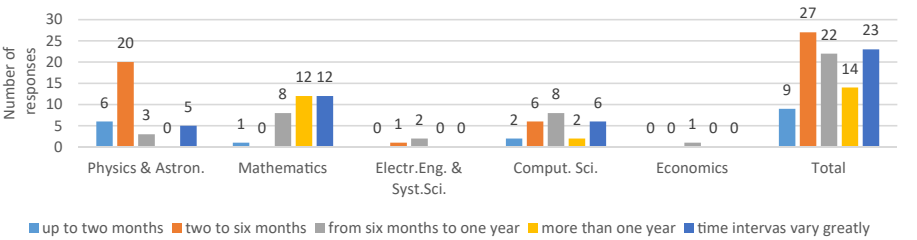

Figure 5.
Time between publication in arXiv and in a scientific journal overall and by scientific discipline

Use of other archives for preprints

In the last section of the survey, we asked respondents about other archives for preprints that they use. They indicated a number of archives. Responses included bioRxiv, ChemRxiv, Preprints, Sci-Hub, SSRN, PeerJ, Inspire, CERN Archives, Zenodo, EuroFusion, ResearchGate, JSTOR and ECC Research Papers. The reader can find answers to some less important questions in the supplementary materials.

Discussion

The survey was sent to 1711 active scientists doing research in the scientific fields covered by arXiv, and 218 volunteers responded. Thus, this is a voluntary sample, which is characterized by a higher response rate, but also carries a higher risk of sampling bias. According to the questionnaire survey, 79.4% of the survey participants knew arXiv and 82.7% of them used it. We found differences between scientific disciplines. The most engaged users of the archive, who can get the most benefit from it, are physicists and astronomers, followed by mathematicians. In these scientific disciplines, we find researchers who deposit all or most of their articles in the archive. These researchers read new articles in the archives every day to learn about new discoveries in their research area. They believe that this way of working enables faster development of science. Some of them are endorsers for their research field. Researchers in Computer Science and Electrical Engineering and Systems Science seem to have recognized the benefits of the archive, but encounter a higher percentage of poor quality articles when using it, which discourages them from using it. In this group, there is more uncertainty about which article to read. This group also includes researchers who publish in the archive, but they are generally not researchers who track newly published articles on a daily basis. The reaction of colleagues and peers to an article published on arXiv is less important to this user group. In this group, there is more of a fear that an article already published in the archive will not be accepted by a scientific journal. Researchers from the other scientific fields participated in the survey to a lesser extent, suggesting that arXiv is less popular in these scientific fields. The survey also found that some researchers refuse to use arXiv.They believe that the quality of articles in the archive is insufficient. Other reasons for non-use are that arXiv does not cover a particular scientific subfield and that publications in the archive do not contribute as much to academic promotion as publications in peer-reviewed scientific journals.

Several results of the survey are consistent with published results in research papers. The intensity of arXiv use is greater in Physics and Astronomy and Mathematics than in the other sciences. This is also noted in Noruzi (2016). Similarly, in Lariviere *et al.* (2014), we find that arXiv plays an important role both as a notification service and as a preferred archive, and that the goal of most scientists is still to publish in peer-reviewed journals. Most Slovenian scientists in Physics and Astronomy and Mathematics, as well as some in Computer Science, upload an article to arXiv and then submit it to a peer-reviewed scientific journal, while some upload it to arXiv only after it has already been published in a scientific journal. The delay reported by Slovenian scientists between publication in the archive and publication in a scientific journal is similar to delays reported in articles by Lariviere *et al.* (2013) and Wang *et al.* (2020a). For Physics and Astronomy, it is two to six months; for Mathematics, it is longer, usually more than a year. The order of science fields by citation frequency of articles published on arXiv based on the survey is the same as in the article by Li *et al.* (2015). It is highest in Mathematics, followed by Physics and Astronomy, Computer Science and Electrical Engineering and Systems Science. Very few researchers have had the experience of having their manuscripts rejected because they were published on a preprint server, as more and more academic journals now accept submissions of preprinted manuscripts (Wikipedia, 2022).

The main advantages of using arXiv for Slovenian scientists are quick access to knowledge, open access and greater impact of scientists' work. It is also important for them that publishing in the archive is free of charge. Some of the challenges of preprints listed in the paper by Chiarelli *et al.* (2019) are also reflected in the responses to the question about the disadvantages of using arXiv. These include the lack of quality assurance, the risk of media reporting inaccurate research results and causing harm through rapid dissemination and journals rejecting submissions if they have been published as preprints.

Our study also showed that librarians did not inform young researchers about the possibility of arXiv. We agree with Frederick (2021) that librarians, especially in academic libraries, need to understand the value of preprint archives for scholarly work and introduce students to preprint archives. Reading articles and commentaries in preprint archives can help students discover areas of research that are of greatest interest to them.

Conclusions

Only 66% of the scientists who responded to the survey use arXiv, and only 45% publish their research articles there. The vast majority of physicists, astronomers and mathematicians who use arXiv also publish there. In computer science, it is only half, and in electrical engineering and systems science, only one-third. For these scientists, arXiv plays an important role both as a notification service and as a preferred archive, but the goal of most scientists is still to publish in peer-reviewed journals. A similar conclusion is found in the paper by Lariviere *et al.* (2014). For a large part of Slovenian scientists, publishing research articles on arXiv has not yet become a standard practice in terms of belief in the work of Mishkin *et al.* (2020). There are several reasons for this: belief in the inadequate quality of articles in some scientific fields; articles deposited in the archive and the citation of these articles do not contribute to academic promotion, as is the case with articles in peer-reviewed scientific journals; rejection by some journals; sometimes the fear that the idea will be stolen. Scientists who publish in the archive are aware of the advantages that the archive offers: rapid publication of articles, minimal cost, open access and wider readership.

Notes

1. Institution included in the survey were:
   - University of Ljubljana (the Faculty of Mathematics and Physics, the Faculty of Computer and Information Science, the Faculty of Electrical Engineering, the Biotechnical Faculty and the School of Economics and Business).
   - University of Maribor (the Faculty of Science and Mathematics and the Faculty of Electrical Engineering, Computer Science and Informatics).
   - University of Primorska (the Faculty of Mathematics, Natural Sciences and Information Technologies).
   - University of Nova Gorica (the Faculty of Natural Sciences).
   - Relevant departments of the Jožef Stefan Institute
2. Educational levels: VIII/2 (3rd Bologna cycle, doctoral degree), VIII/1 (specialization by university program, master of science), VII (2nd Bologna cycle, a master's degree), VI/2 (1st Bologna cycle, a bachelor's degree) and VI/1 (post-secondary professional program).

References

Abdill, R.J. and Blekhman, R. (2019), "Meta-Research: tracking the popularity and outcomes of all bioRxiv preprints", *eLife*, Vol. 8, e45133, doi: 10.7554/eLife.45133.

Allison, D.B., Brown, A.W., George, B.J. and Kaiser, K.A. (2016), "Reproducibility: a tragedy of errors", *Nature*, Vol. 530, pp. 27-29, doi: 10.1038/530027a.

Aman, V. (2015), *Citing E-Prints on arXiv A Study of Cited References in WoS-Indexed Journals from 1991-2013*, ISSI, available at: https://www.issi-society.org/proceedings/issi_2015/1107.pdf

Annesley, T., Scott, M., Bastian, H., Fonseca, V., Ioannidis, J.P.A., Keller, M.A. and Polka, J. (2017), "Biomedical journals and preprint services: friends or foes?", *Clinical Chemistry*, Vol. 63 No. 2, pp. 453-458, doi: 10.1373/clinchem.2016.268227.

arXiv (2019), "To add a journal reference, DOI or report number to an article", available at: https://arxiv.org/help/jref

arXiv (2022), "About arXiv", available at: https://arxiv.org/about

Avasthi, P., Soragni, A. and Bembenek, J.N. (2018), "Journal clubs in the time of preprints", *Elife*, Vol. 11 No. 7, e38532, doi: 10.7554/eLife.38532.

Balietti, S. (2016), "Science is suffering because of peer review's big problems", *The New Republic*, available at: https://newrepublic.com/article/135921/science-suffering-peer-reviews-big-problems

Bastian, H. (2016), "Breaking down pros and cons of preprints in biomedicine", *PLOS Blogs*, available at: https://blogs.plos.org/absolutely-maybe/2016/05/01/breaking-down-pros-and-cons-of-preprints-in-biomedicine/

Bohannon, J. (2016), "Who's downloading pirated papers? Everyone", *Science*, Vol. 352 No. 6285, pp. 508-512, doi: 10.1126/science.352.6285.508.

Borgman, C.L. (2007), *Scholarship in the Digital Age: Information, Infrastructure and the Internet*, MIT Press, doi: 10.7551/mitpress/7434.001.0001.

Bourne, P.E., Polka, J.K., Vale, R.D. and Kiley, R. (2017), "Ten simple rules to consider regarding preprint submission", *PLoS Computational Biology*, Vol. 13 No. 5, p. e1005473, doi: 10.1371/journal.pcbi.1005473.

Brody, T., Harnad, S. and Carr, L. (2006), "Earlier web usage statistics as predictors of later citation impact", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 8, pp. 1060-1072, doi: 10.1002/asi.20373.

Brown, C. (1999), "Information seeking behavior of scientists in the electronic information age: astronomers, chemists, mathematicians, and physicists", *Journal of the American Society for Information Science*, Vol. 50 No. 10, pp. 929-943, doi: 10.1108/00012530810908184.

Brown, C. (2001), "The E-volution of preprints in the scholarly communication of physicists and astronomers", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 3, pp. 187-200, doi: 10.1002/1097-4571(2000)9999:99993.0.CO;2-D.

Butler, D. (2003), "Biologists join physics preprint club", *Nature*, Vol. 425 No. 6958, p. 548, doi: 10.1038/425548b.

Chiarelli, A., Johnson, R., Pinfield, S. and Richens, E. (2019), "Preprints and scholarly communication: an exploratory qualitative study of adoption, practices, drivers and barriers", [version 2; peer review: 3 approved, 1 approved with reservations], *F1000Research*, Vol. 8, p. 971, doi: 10.12688/f1000research.19619.2.

Crotty, D. (2018), "Preprints and citations: should non-peer reviewed material Be included in article references?", *Scholarly Kitchen*, available at: http://bit.ly/2FSR0un

Davis, P.M. and Fromerth, M.J. (2007), "Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?", *Scientometrics*, Vol. 71 No. 2, pp. 203-215, doi: 10.1007/s11192-007-1661-8.

Frederick, D.E. (2021), "Will preprint servers disrupt scientific publishing, reference work and information science?", *Library Hi Tech News*, Vol. 38 No. 6, pp. 7-14, doi: 10.1108/LHTN-10-2021-0071.

Garner, J., Horwood, L. and Sullivan, S. (2001), "The place of eprints in scholarly information delivery", *Online Information Review*, Vol. 25 No. 4, pp. 250-253.

Hachani, S. (2015), "Open peer review: fast forward for a new science", in *Current Issues in Libraries, Information Science and Related Fields (Advances in Librarianship)*, Vol. 39, pp. 115-141, doi: 10.1108/S0065-283020150000039012.

Haque, A. and Ginsparg, P. (2009), "Positional effects on citation and readership in arXiv", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 11, pp. 2203-2218, doi: 10.1002/asi.21166.

Haque, A. and Ginsparg, P. (2010), "Last but not least: additional positional effects on citation and readership in arXiv", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 12, pp. 2381-2388, doi: 10.1002/asi.21428.

Huisman, J. and Smits, J. (2017), "Duration and quality of the peer review process: the author's perspective", *Scientometrics*, Vol. 113 No. 1, pp. 633-650, doi: 10.1007/s11192-017-2310-5.

Inglis, J.R. and Sever, R. (2017), "bioRxiv: a progress report", available at: https://asapbio.org/biorxiv

Kim, L., Portenoy, J.H., West, J.D. and Stovel, K.W. (2020), "Scientific journals still matter in the era of academic search engines and preprint archives", *Journal of the Association for Information Science and Technology*, Vol. 71 No. 10, pp. 1218-1226, doi: 10.1002/asi.24326.

Kovanis, M., Porcher, R., Ravaud, P. and Trinquart, L. (2016), "The global burden of journal peer review in the biomedical literature: strong imbalance in the collective enterprise", *PLoS ONE*, Vol. 11 No. 11, p. e0166387, doi: 10.1371/journal.pone.0166387.

Lariviere, V., Sugimoto, C.R., Macaluso, B., Milojević, S., Cronin, B. and Thelwall, M. (2013), "arXiv E-prints and the journal of record: an analysis of roles and relationships", *arXiv*, available at: https://arxiv.org/abs/1306.3261

Lariviere, V., Sugimoto, C.R., Macaluso, B., Milojević, S., Cronin, B. and Thelwall, M. (2014), "arXiv E-prints and the journal of record: an analysis of roles and relationships", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 6, pp. 1157-1169, doi: 10.1002/asi.23044.

Li, X., Thelwall, M. and Kousha, K. (2015), "The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication", *Aslib Journal of Information Management*, Vol. 67 No. 6, pp. 614-635, doi: 10.1108/AJIM-03-2015-0049.

McGlashan, D. and Hadley, C. (2017), "The rising tide of preprint servers", *INLEXIO*, available at: https://www.inlexio.com/rising-tide-preprint-servers/

Metcalfe, T.S. (2005), "The rise and citation impact of astro-ph in major journals", *Bulletin of the American Astronomical Society*, Vol. 37, pp. 555-557.

Metcalfe, T.S. (2006), "The citation impact of digital preprint archives for solar physics papers", *Solar Physics*, Vol. 239, pp. 549-553, doi: 10.1007/s11207-006-0262-7.

Mishkin, D., Tabb, A. and Matas, J. (2020), "ArXiving before submission helps everyone", doi: 10.48550/arXiv.2010.05365.

Moed, H.F. (2007), "The effect of 'Open Access' on citation impact: an analysis of arXiv's condensed matter section", *Journal of the American Society for Information Science andTechnology*, Vol. 58 No. 13, pp. 2047-2054, doi: 10.1002/asi.20663.

Mudrak, B. (2020), "What are preprints, and how do they benefit authors?", *AJE Scholar*, available at: https://www.aje.com/arc/benefits-of-preprints-for-researchers/

Noruzi, A. (2016), "ArXiv popularity from a citation analysis point of view", *Webology*, Vol. 13 No. 2, pp. 1-7, available at: http://www.webology.org/2016/v13n2/editorial22.pdf

O'Roak, B. (2018), "How I learned to stop worrying and love preprints", *Spectrum*, available at: https://www.spectrumnews.org/opinion/learned-stop-worrying-love-preprints/

Polka, J.K., Kiley, R., Konforti, B., Stern, B. and Vale, R.D. (2018), "Publish peer reviews", *Nature*, Vol. 560 No. 7720, pp. 545-547, doi: 10.1038/d41586-018-06032-w.

Powell, K. (2016), "Does it take too long to publish research?", *Nature*, Vol. 530 No. 7589, pp. 148-151, doi: 10.1038/530148a.

Prakasan, E.R. and Kalyane, V.L. (2004), "Citation analysis of LANL high-energy physics E-prints through science citation Index (1991-2002)", *eLis*, available at: http://eprints.rclis.org/5406/

Puebla, I., Polka, J. and Rieger, O. (2022), "Preprints: their evolving role in science communication", *MetaArXiv*. doi: 10.3998/mpub.12412508.

Sarabipour, S., Debat, H.J., Emmott, E., Burgess, S.J., Schwessinger, B. and Hensel, Z. (2019), "On the value of preprints: an early career researcher perspective", *PLoS Biology*, Vol. 17 No. 2, e3000151, doi: 10.1371/journal.pbio.3000151.

Schwarz, G.J. and Kennicutt, R.C.J. (2004), "Demographic and citation trends in astrophysical journal papers and preprints", *Bulletin of the American Astronomical Society*, Vol. 36, pp. 1654-1663.

Serghiou, S. and Ioannidis, J.P.A. (2018), "Altmetric scores, citations, and publication of studies posted as preprints", *JAMA*, Vol. 319 No. 4, pp. 402-404, doi: 10.1001/jama.2017.21168.

Shuai, X., Pepe, A. and Bollen, J. (2012), "How the scientific community reacts to newly submitted preprints: article downloads, twitter mentions, and citations", *PLoS ONE*, Vol. 7 No. 11, e47523, doi: 10.1371/journal.pone.0047523.

Smith, R. (2006), "Peer review: a flawed process at the heart of science and journals", *Journal of the Royal Society of Medicine*, Vol. 99 No. 4, pp. 178-182, doi: 10.1177/014107680609900414.

Spezi, V., Wakeling, S., Pinfield, S., Fry, J., Creaser, C. and Willett, P. (2018), "'Let the community decide'? The vision and reality of soundness-only peer review in open-access mega-journals", *Journal of Documentation*, Vol. 74 No. 1, pp. 137-161, doi: 10.1108/JD-06-2017-0092.

Till, B.M., Rudolfson, N., Saluja, S., Gnanaraj, J., Samad, L., Ljungman, D. and Shrime, M. (2019), "Who is pirating medical literature? A bibliometric review of 28 million Sci-Hub downloads", *The Lancet*, Vol. 7 No. 1, pp. 30-31, doi: 10.1016/S2214-109X(18)30388-7.

Vale, R.D. and Hyman, A.A. (2016), "Point or view: priority of discovery in the life sciences", *Elife*, Vol. 5, e16931, doi: 10.7554/fLife.16931.

Wang, Z., Chen, Y. and Glänzel, W. (2020a), "Preprints as accelerator of scholarly communication: an empirical analysis in Mathematics", *Journal of Informetrics*, Vol. 14 No. 4, 101097, doi: 10.1016/j.joi.2020.101097.

Wang, Z., Glänzel, W. and Chen, Y. (2020b), "The impact of preprints in Library and Information Science: an analysis of citations, usage and social attention indicators", *Scientometrics*, Vol. 125 No. 2, pp. 1403-1423, doi: 10.1007/s11192-020-03612-4.

Wikipedia (2022), "List of academic journals by preprint policy", available at: https://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy

Youngen, G.K. (1998), "Citation patterns to traditional and electronic preprints in the published literature", *College and Research Libraries*, Vol. 59 No. 5, pp. 448-456.

Supplementary material

Supplementary material contains results on questions that we consider less important.

The archive allows researchers to comment on the published articles. Table S1 shows the results for this type of activity among Slovenian researchers. The questioner offered two additional options not included in Table S1: "I comment on more than half of the articles I read" and "I comment on all articles I read," but neither respondent selected them. The data suggest that respondents from the "read and publish" group are more likely to comment on articles they read on arXiv. The Mann–Whitney U test

(see Table S1) confirms the difference between the groups. The detailed presentation of the data in Table S1 shows that the majority of scientists responding to articles are physicists, astronomers and mathematicians.

	I do not comment on articles I read		I comment on less than half of the articles I read		I comment on half of the articles I read	
	Read&Publish	Read only	Read&Publish	Read only	Read&Publish	Read only
Physics and astronomy	23	7	10	1	1	0
Mathematics	23	3	10	0	0	0
Electrical engineering and systems science	2	6	1	0	0	0
Computer science	22	24	2	3	0	0
Biology	0	1	0	0	0	0
Economics	0	0	0	0	1	0
Total	70	41	23	4	2	0

Mann–Whitney U test					
U_R	$U_{R\&P}$	μ	σ_{corr}	z_U	p-value (one-sided)
2,514	1761	2137.5	157.6	2.39	0.0085

Note(s): Numbers are given for each category, broken down by group type and scientific discipline

Table S1. Frequency of comments on articles read and results of the Mann–Whitney U test

Most of the changes to the article published in arXiv and subsequently in the journal are minor or moderate (see Figure S1). The most common change is to the graphical representation of the article. Sometimes the content of the article, the structure and the style also change.

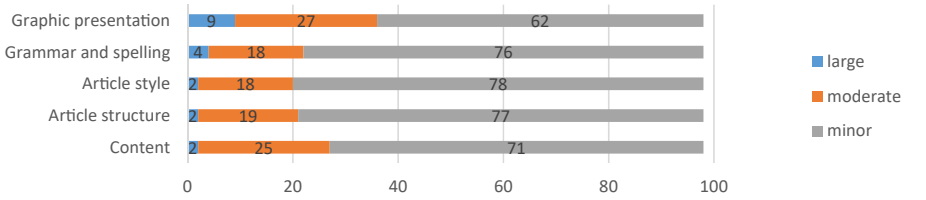

Figure S1. Changes that an arXiv article undergoes when published in a scientific journal

Only four of the respondents had the experience of having an article rejected on arXiv. For three of them this was the case only once, for one three times.

Seven out of 98 respondents, or 7.1%, felt pressured by colleagues to publish as many articles as possible on arXiv. Figure S2 shows the results by scientific discipline.

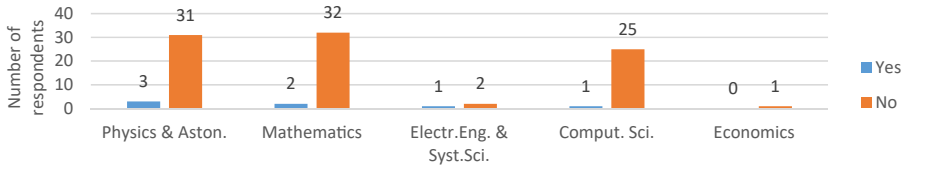

Figure S2. Pressure by colleagues to publish as many articles as possible on arXiv

Table S2 shows the data and the results of the Kruskal–Wallis H test for the frequency of receiving responses to articles published on arXiv. The calculated p-value is just above 0.05. The response "to less than half of the articles" predominates. However, three respondents received responses to all published articles. A closer look at the data shows that two published only one article and one published 10 articles.

	To no article	To less than half of the articles	To half articles	To more than half	To all articles
Physics and astronomy	8	12	10	3	1
Electrical engineering and systems science	1	2	0	0	0
Mathematics	7	22	3	1	1
Computer science	11	12	1	1	1
Economics	0	1	0	0	0
Total	27	49	14	5	3

Kruskal–Wallis H test $H \approx \chi^2(k-1)$	$H_{corrected}$	K	p-value
5.08	5.92	3	0.052

Table S2. Feedback from colleagues on an article published on arXiv

Figure S3 shows employers' attitudes toward publishing articles on arXiv. The majority of employers, 63%, leave publication to the personal discretion of the researcher, 35% of employers support publication on arXiv and 2% do not support it.

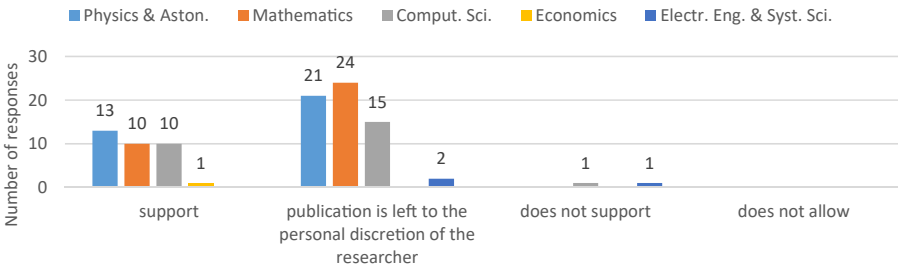

Figure S3. Employers' attitudes toward publishing articles on arXiv

Use of other archives for preprints

In the last section of the survey, we asked respondents about other archives for preprints that they use. Table S3 shows the results. The answer "yes" is more common in the group of respondents who do not use arXiv. The test for independence confirms the relationship to the groups with statistical significance.

	Do you use any other archive for preprints of scientific articles?	
	Yes	No
not using arXiv	13 (43.3%)	17 (56.7%)
using arXiv	29 (20.3%)	114 (79.7%)

Pearson's chi-squared test for independence		
χ^2	*df*	*p*-value
7.17	1	*0.0074*

Table S3.
Use of other archives for preprints of scientific articles

Respondents were asked to name the preprint archives they used. They indicated a number of archives. However, preprint archives were not the only ones among the responses. Responses included bioRxiv, ChemRxiv, Preprints, Sci-Hub, SSRN, PeerJ, Inspire, CERN Archives, Zenodo, EuroFusion, ResearchGate, JSTOR and ECC Research Papers.

Corresponding author
Jasna Maver can be contacted at: jasna.maver@ff.uni-lj.si