Bulletin of the AAS

ArXiv in the Open Access Era: its usage and impact on physics researchers

Nagaraj Mulukunte Narayana[1], M.K. Bhandi[2]

[1]Raman Research Institute, [2]Retired University Librarian, Mangalore University

Published on: Apr 27, 2022

DOI: 10.3847/25c2cfeb.52d67090

ABSTRACT

Abstract - ArXiv is an Open Access (OA) e-prints repository of physics, mathematics, computer science, and related subjects. This study aims to determine the use and impact of arXiv on physics researchers of selected research institutes in Bengaluru, India. The study was conducted using the survey method, and a structured questionnaire was distributed to physics researchers working in five research institutions in Bengaluru. The first part of the questionnaire consists of demographic information second part consists of the usage of arXiv, and the last section includes the impact of usage and satisfaction level of users. The response was 73%, and analysis revealed that 94% use arXiv. The study showed that more than 50% contribute to arXiv and that they are highly satisfied with using this resource. Further, the study has also shown that, by contributing to arXiv, the researchers benefited from significant impacts, such as getting good feedback, collaborations, and increased citations. We believe that this study will encourage researchers to use and contribute to OA resources and also motivate libraries to promote OA repositories.

1 Introduction

1.1 arXiv

The preprint repository arXiv supports OA to scholarly communications. Physicist Paul Ginsparg established it in 1991, and as of now, it has more than 1,948,630 scholarly articles e-prints are available for free reading and download.

Ginsparg is a physicist widely known for developing arXiv.org, an e-print archive. In a talk in 2001, he said that "arXiv is an example of a service created by a group of specialists for their use. When researchers or professionals create such services, the results often differ markedly from the services provided by publishers and libraries" Further, Ginsparg [1] states that "the original objective of the e-print arXiv was to provide functionality that was not otherwise available and to provide a level playing field for researchers at different academic levels and different geographic locations." ArXiv had become a significant forum for disseminating scholarly publications in physics and mathematics. This resource has been entirely scientist-driven and is flexible enough to co-exist with the pre-existing publication system or to help it evolve to something better optimized for researcher needs.

In a slightly distinctive work, Schwarz & Kennicutt [2], studying demographic and citation trends of Astrophysical Journal papers and preprints, states that there is no obligation for any author to make papers available. Still, most physics and astronomers post their articles on the e-print arXiv.

1.2 Open Access Resource Providers

Authors contribute to green open access through self-archiving on their webpage, Blog, and ResearchGate. Publisher contributes to gold OA by publishing in Hybrid & Open access journals. Libraries also contribute through Institutional Repositories. Society, Academy & University contribute to OA through Electronic Thesis and Dissertation (ETD) repository and huge databases like NASA/ADS and arXiv.

1.3 Objectives of the Study

- To determine awareness of self-archiving and OA resources.
- To study the usage of arXiv.
- To find that there exists a relationship between the field of research and the use of the arXiv preprint repository.
- To study the usage impact of arXiv.
- To study the satisfaction level of users.

2 Methodology

2.1. Methodology

A structured questionnaire was designed using Likert's five-point scale [3] with close-ended questions. Total 480 questionnaires were distributed to physics researchers and were able to get 352 responses.

To determine physics researchers use and awareness of arXiv survey methods were used. Keeping in view of the objectives of the study, required data were collected through the distribution of a structured questionnaire to physics researchers of five research Institutions in Bangalore. The different categories of physics researchers covered under the study included faculty members, postdoc, junior and senior research scholars who may or may not be using open access resources.

3 Data Analysis

3.1 Data Analysis

The data collected from the questionnaire were analyzed through Statistical Package for Social Sciences (SPSS). The results are presented in the form of tables and figures. This analysis section depicts responses received based on questionnaire like awareness, usage, and designation of users. It also explains user's relationship with the use of arXiv and its impact on them

3.2 Population Surveyed and Response

The scope of this study was confined to five research institutions located in Bangalore, Karnataka. These Institutions are the Centre for Nano and Soft Matter Sciences (CeNS), Indian Institute of Astrophysics (IIA), Indian Institute of Science (IISc), Jawaharlal Nehru Centre for Advanced Scientific Research(JNCASR) and Raman Research Institute (RRI).

Sl. NO.	Institute	Questionnaire Distributed	Response Received	Percentage
3	IIA	110	69	62.72
5	IISc	140	98	70
1	CNES	30	23	73
2	JNCASR	80	60	75
4	RRI	120	102	85
	Total	480	352	73.33

Table 1: Distribution of questionnaire and response received

As shown in Table 1, it is clear that a maximum response of 102 (85%) was received from RRI, and a minimum response of 62% is from IIA. All the above five research institutions taken for study are major physics research institutes in Bangalore. Further, from the Table 2, we can make out that respondents are from different branches of physics.

Sr. No.	Research Field	Frequency	Percentage

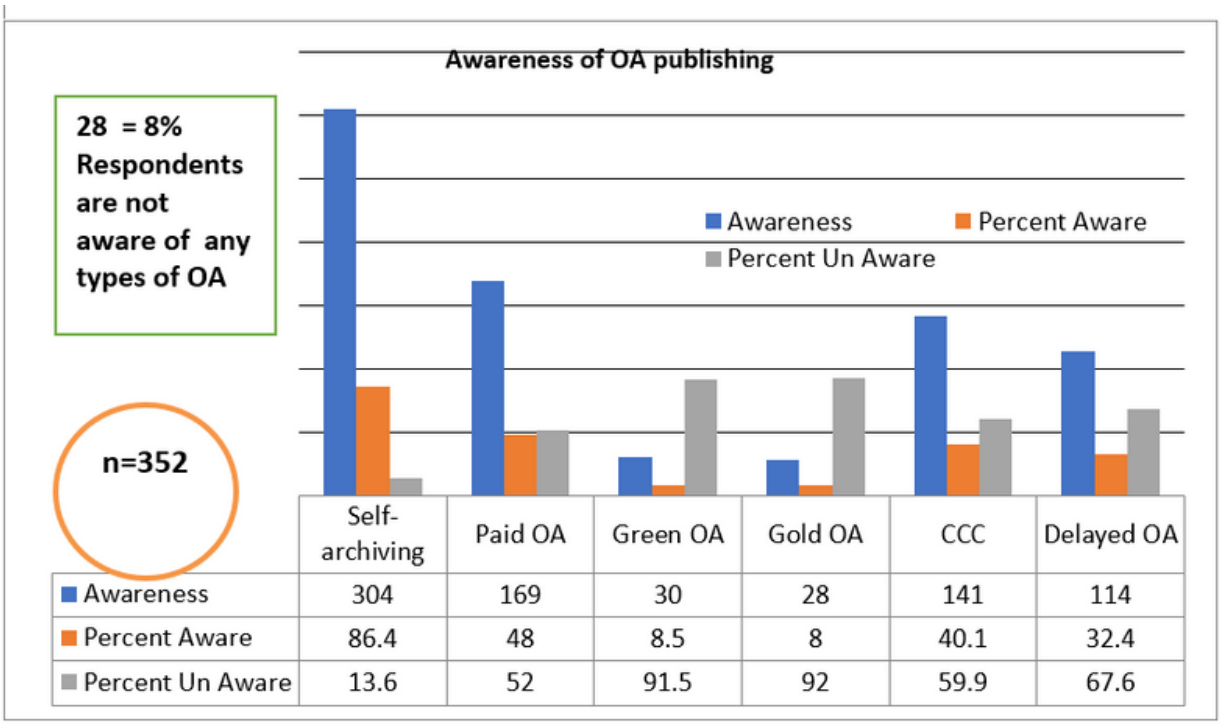
1	Optics	20	5.7
2	High Energy Physics (HEP)	22	6.3
3	Biophysics	23	6.5
4	Theoretical Physics (TP)	51	14.5
5	Astronomy & Astrophysics (A&A)	112	31.8
6	Condensed Matter Physics (CMP)	124	35.2
	Total	352	100

Table 2: Responses received as per research field of physics

Based on the responses received as shown in Table 2, six sub-branches of physics are taken for analysis as presented. It is also evident from Table 2 that the maximum response is from condensed matter physics with 124 (35.2%), followed by astronomy & astrophysics with 112 (31.8%). The least response is from optics with 20 (5.7%), and slightly better than this is high energy physics at 22 (6.3%). The percentage of condensed matter physics respondents is more because it is carried out in four research institutions.

3.3 Awareness about Open Access Publishing

One of the objectives of the study was to understand user awareness and modes of OA publishing. It is evident from Figure 1 that the majority of respondents are unaware of green OA (91.5%) and gold (92%) routes of open access. Even though most users are using green and gold OA resources, they are unaware of the terminology. Hence below table shows less awareness of them. An exciting outcome of the analysis is that around (86.4%) are aware of self-archiving of their publications on their web pages. Delayed and paid OA modes are known by 32.4% and 48%, respectively. ArXiv preprint repository is driven by authors by self-archiving, and figure one also shows 86.4% awareness of self-archiving.



	Self-archiving	Paid OA	Green OA	Gold OA	CCC	Delayed OA
Awareness	304	169	30	28	141	114
Percent Aware	86.4	48	8.5	8	40.1	32.4
Percent Un Aware	13.6	52	91.5	92	59.9	67.6

Figure 1. Awareness of Open access publishing

3.4 Preferred Mode for Searching Information

Researchers use different search engines and databases for retrieving scholarly output. In response to a query about the most preferred mode for searching information, It is evident from Figure 2 that google is most preferred with (40.1%) followed by Google Scholar with (22.7%) and least preferred are publishers' sites with (4%) and Institutional Repository(IR) with (0.3%). Around 11.4% prefer directly access papers through arXiv.

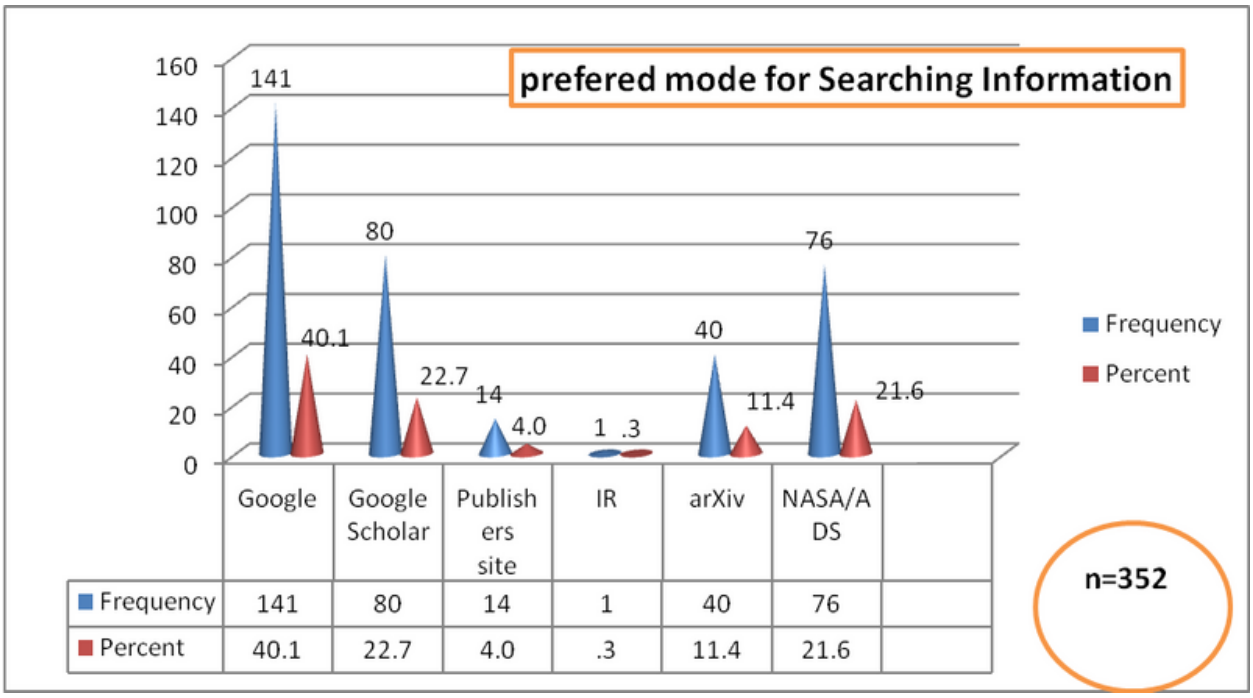


Figure 2. Preferred mode for searching information

In contrast to this study, when arXiv was 25 years, in a survey Rieger [4] has reported that to access papers directly arXiv website is used by 79% and Google 50% and Google Scholar 35%. In this study, directly accessing arXiv is 11% but accessing through Google is 40%.

3.5 Usage of arXiv

ArXiv is extensively used by all researchers in different physics research fields like astronomy and astrophysics, theoretical physics, and condensed matter physics. As shown in above Table 2 of Section 3.2, we surveyed the usage of arXiv among six specific fields of physics.

ArXiv Preprint Repository usage	Research Field of Researchers							
	Astronomy & Astrophysics	High Energy Physics	Optics	Condensed Matter Physics	Theoretical Physics	Bio-physics	Total	Percentage (%)
Aware but not used	0	0	1	2	0	3	6	1.70%

Never used	0	0	3	12	0	0	15	4.30%
Rarely	8	2	2	7	0	1	20	5.70%
Sometime	15	5	4	28	7	6	65	18.50%
Frequently	89	15	10	75	44	13	246	69.80%
	112	22	20	124	51	23	352	100%
Percentage	31.80%	6.30%	5.70%	35.20%	14.50%	6.50%	100.00%	

Table 3: Cross table showing use of arXiv preprint repository versus field of research

Table 3 shows that the majority, 246 (69.8%), of respondents use arXiv frequently, 18.5% use sometimes, and 5.7% use it rarely. Only six percent of researchers surveyed are not using arXiv. Grothkopf [5] points out that most astronomers cover almost all their need for scholarly information by using NASA ADS and arXiv.org e-print server. She addresses these two databases as gorillas of astronomy and physics. Further, both of them are free and available to everybody.

3.5.1 Chi-Square Test

To find determine if there was a statistically significant relationship between scientific field and usage of arXiv, we conducted a chi-square test to evaluate two hypotheses:

1. The Null Hypothesis (H_0): "Use of arXiv preprint repository does not depend on the field of research of physics researchers."
2. The Alternative Hypothesis (H_a): “Use of arXiv preprint repository depends on the field of research of physics researchers."

By convention, a 5.0% alpha value with 20 degrees of freedom gives us a p-value of 31.410. If the chi-square value from the data is greater than 31.41 we reject the null hypothesis, and accept the alternative hypothesis: the use of arXiv depends on a researcher’s field of study. On the other hand, if the chi-square value is less than 31.410 then we accept the null hypothesis and determine that there is no relationship between usage of arXiv and field of study.

$X^2(20)=60.102$, p=31.410

The chi-square value derived from Table 3 is 60.102. Since this is significantly greater than the p-value of 31.410, this test proves that there is significant relationship between field of physics research and use of arXiv preprint repository.

As shown in Table 3, one hundred percent usage is seen among astrophysics, theoretical physics, and high energy physics researchers. While among condensed matter physics, theoretical physics, optics, and biophysics groups usage is around 80%. This is interesting because even though arXiv comprises almost all branches of physics, those who study astrophysics, theoretical physics, and high energy physics are statistically more likely to use arXiv than other disciplines.

3.6 Citation of arXiv

Citing arXiv as a reference by users was studied to understand the usage impact of arXiv.

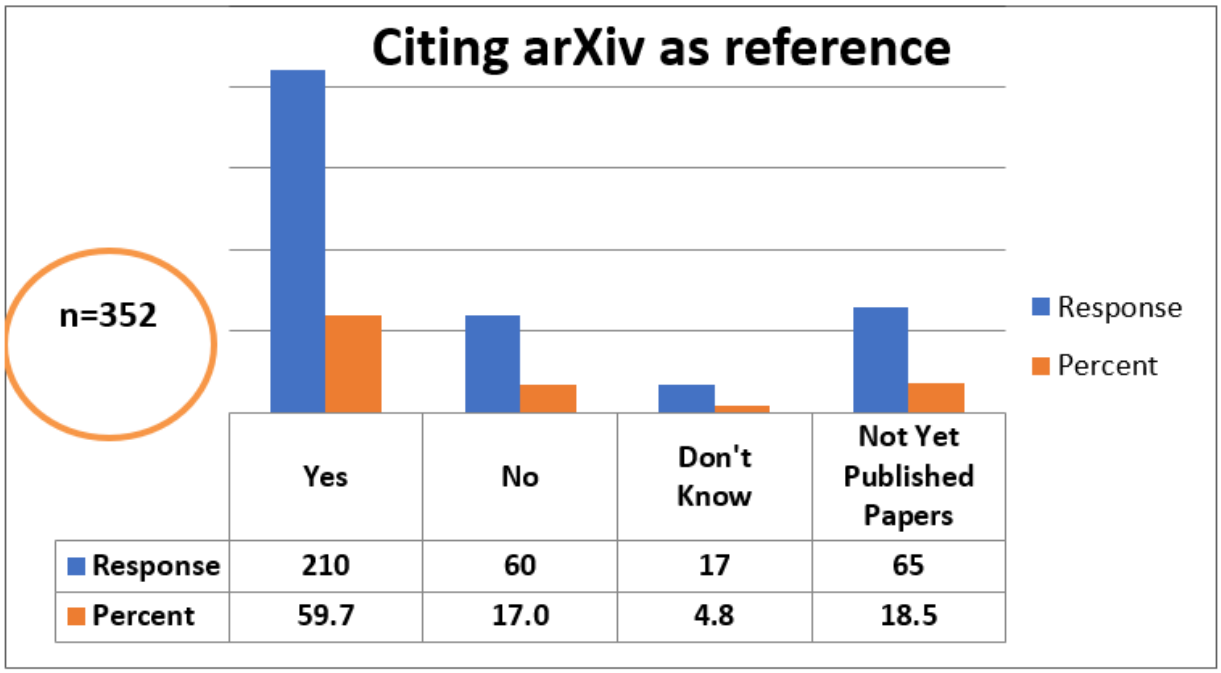
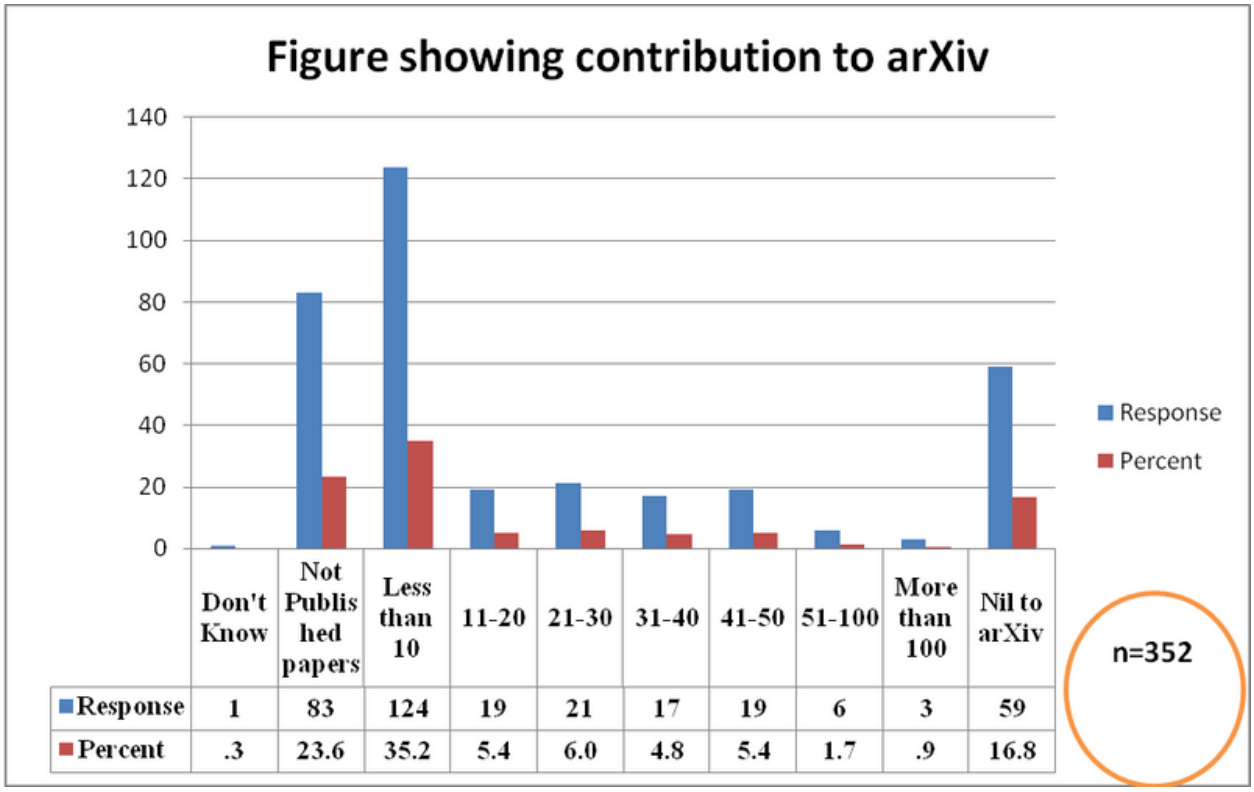


Figure 3. showing citing arXiv as reference

In response to this question, as shown in figure 3, out of 352 respondents, 210 (59.7%) replayed that they are citing, 17% are not citing, 4.8% don't know, and 18.5% do not yet publish papers. It is clear from Figure 3 that around 60% are citing arXiv, which is a positive impact on arXiv.

3.7 Contribution to arXiv

In response to the question to know contribution made to the arXiv preprint repository, it is evident from figure 4 that a maximum percentage of 35.2% has contributed less than ten articles.



	Don't Know	Not Published papers	Less than 10	11-20	21-30	31-40	41-50	51-100	More than 100	Nil to arXiv
Response	1	83	124	19	21	17	19	6	3	59
Percent	.3	23.6	35.2	5.4	6.0	4.8	5.4	1.7	.9	16.8

Figure 4. showing citing arXiv as reference

As per Figure 4, out of 352 respondents, 209 (59.4%) of respondents have contributed to arXiv, and 83 (23.6%) have indicated that they are not published any papers, and 59 (16.8%) have not contributed to arXiv. Further, it is evident from figure four that 5.4% of respondents have contributed 41-50 articles, another 5.4% have contributed 11-20 articles, and 1.7% have contributed 51-100 articles. The maximum number of contributions, 35.2%, is from less than ten publications categories, and the minimum is 0.9% from more than 100 publications categories. However, more than 50% of respondents contributing to the arXiv preprint repository showed significant impact by increase in number of citations, resulted in collaboration and helped to improve research results. The same has been explained in section 3.8.

3.8 Impact of Contribution to arXiv

As seen from Figure 4, there is more contribution to arXiv, which has motivated to study the impact of the contribution. Hence, in this study, an attempt has been made to study the impact of the contribution.

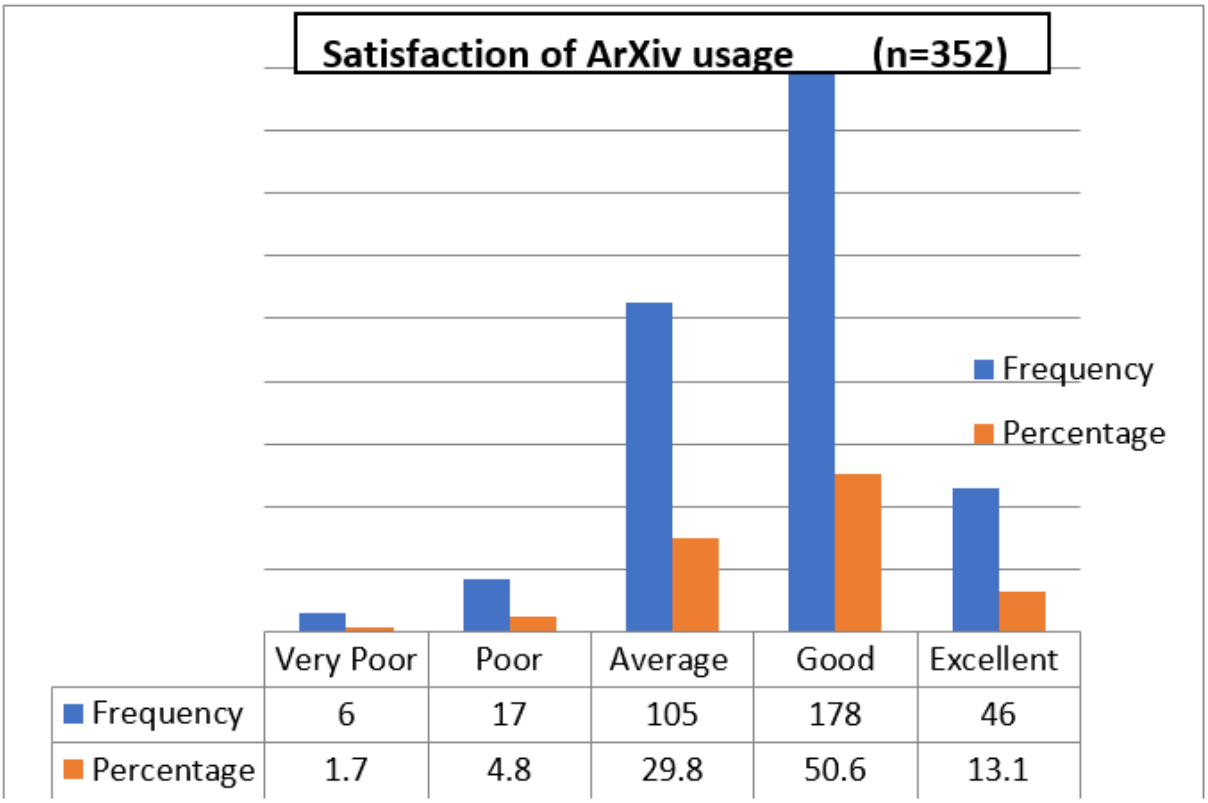
The correlation between physics researchers' experience and impact due to contribution to arXiv preprint open access repository shows significance with 36 degrees of freedom, an alpha of .05, and a p-value of 50.998 implying that respondents have significant impact. Table 4 shows users have impact due to contribution in terms of feedback, resulting in collaboration, improved results and increased citations. The reason for collaboration and increases in citations among experienced researchers may be because faculty contribute more articles. In contrast to this junior and senior researchers will seldom have publications.

SL NO	Physics Researchers	Impact due to contribution to arXiv	Chi Square Test	Relationship due to contribution
1	Research Experience	Feedback received for contributing articles	$X^2(36) = 89.674$ $p=50.998$ Highly significant	Yes
2	Research Experience	Resulted in collaboration	$X^2(36) = 94.267$ $p=50.998$ Highly significant	Yes
3	Research Experience	Helped to improve results	$X^2(36) = 107.963$ $p=50.998$ Highly significant	Yes
4	Research Experience	Increased number of citations	$X^2(36) = 98.551$ $p=50.998$ Highly significant	Yes

Table 4: Relationship between research experience and impact due to contribution to arXiv

3.9 Usage Satisfaction of arXiv

It was most important to understand how users are satisfied with usage. Hence, in response to this, as shown in figure 5, most researchers (93.5%) reported that they are satisfied with the use of arXiv. Out of 352, 13.1% feel excellent, 50.6% feel good, and 29.8% are averagely satisfied. Totally 6.5% are not satisfied with arXiv grading it as poor and very poor.



	Very Poor	Poor	Average	Good	Excellent
Frequency	6	17	105	178	46
Percentage	1.7	4.8	29.8	50.6	13.1

Figure 5. Satisfaction level of Users

Working on the same lines, Larivière, Vincent et al. [6], in their study on arXiv points that unpublished e-prints are being cited, raises the question of peer-review in those fields. According to them, citing authors either evaluate papers themselves or trust the results presented, which might result from huge collaborations.

4 Findings

- ArXiv preprint repository is driven by authors through self-archiving. This study shows that 86.4% are aware of self-archiving their publications, and the remaining 13.6% are not aware of self-archiving.

- Out of 352 respondents, 313 (94%) are using the arXiv preprint repository, and 1.7% are aware but not using it 4.3% are neither aware nor using it.
- This study shows that around 60% cite arXiv, which positively impacts the usage of OA resources.
- This study also shows that 63.7% are satisfied with ariXiv and grade it as excellent and good. It is evident from this study around (29.8%) are partially satisfied with arXiv respondents have graded as average, and 6.5% are not at all satisfied grading as poor and very poor.

5 Conclusion

So far, many studies have been conducted on the use and awareness of OA journals. This paper may be a unique attempt to study the use and awareness of OA resources like the arXiv preprint repository among physics researchers consisting of faculty and junior and senior research fellows. The result may benefit publishers and information specialists to improve the use and awareness of OA resources. It is also found that physics researchers have a high level of awareness and use of arXiv. The study found no association between uses of arXiv among physics researchers working in different research institutions in Bangalore. It is observed from the study arXiv is not only extensively used, but it is also cited, thus indicating that physics researchers have a positive perception about arXiv. As this database consists of many author manuscripts, further study can be carried out on the evaluation of this database. Further, similar studies can be carried out in other fields like medicine and biology.

6 Acknowledgments

The authors would like to thank all the physics researchers who participated in this survey. Furthermore, authors also thank all the five research institution who supported this project and made this successful.

Citations

1. Ginsparg, Paul. "Creating a global knowledge network." In Second Joint ICSU Press-UNESCO Expert Conference on Electronic Publishing in Science, pp. 19-23. 2001.: http://arxiv.org/blurb/pg01unesco.html (accessed on 17-08-2021). ↩
2. Schwarz GJ, KennicuttR CJ. "Demographic and citation trends in astrophysical journal papers and preprints." Bulletin of the American Astronomical society 36, no.

5 (2004): 1654-1663. (http://arxiv.org/abs/astro-ph/0411275 - Accessed on 20/08/2021) ↩

3. Wikipedia. 2021. “Likert scale.” Last modified May 17 2021. https://en.wikipedia.org/wiki/Likert_scale ↩

4. Rieger, Oya Y., Gail Steinhart, and Deborah Cooper. "arXiv@ 25: Key findings of a user survey." arXiv preprint arXiv:1607.08212 (2016). ↩

5. Grothkopf, Uta. A guide to effective publishing in astronomy, Astronomy libraries–Your gateway to information. (EDP Sciences, France, 2012) pp.99-128 ↩

6. Larivière, Vincent, Cassidy R. Sugimoto, Benoit Macaluso, Staša Milojević, Blaise Cronin, and Mike Thelwall. "arXiv E-prints and the journal of record: An analysis of roles and relationships." Journal of the Association for Information Science and Technology 65, no. 6 (2014): 1157-1169. ↩