arXiv User Survey Report

> ⚠ **The content on this page has moved**
>
> ## The content on this wiki page has been moved to **https://confluence.cornell.edu/x/pLBRF**.
>
> ## This page is no longer kept up to date.

arXiv@25: Key Findings of the User Survey

Oya Y. Rieger, arXiv Program Director, Cornell University Library, June 2016

EXECUTIVE SUMMARY

As part of its 25th anniversary vision-setting process, the arXiv team at Cornell University Library conducted a user survey in April 2016 to seek input from the global user community about the current services and future directions. We were heartened to receive 36,000 responses, representing arXiv's diverse community (See Appendix A). The prevailing message is that users are happy with the service as it currently stands. 95% of survey respondents said that they are very satisfied or satisfied with arXiv. Furthermore, 72% of respondents indicated that arXiv should focus on its main purpose, which is to quickly make available scientific papers, and this will be enough to sustain the value of arXiv in the future. This theme was pervasively reflected in the open text comments. A significant number of respondents suggested keeping to the core mission and enabling arXiv's partners and related service providers to continue to build new services and innovations on top of arXiv.

Many of the comments reflected deep satisfaction with and gratitude for arXiv. Several users referred to the significance of the service for their personal career development and expressed thanks for its continued existence; for example, a typical comment was: "Thanks for the hard work of many people over the years. My work life would be very different without your efforts." arXiv also received many plaudits for advancing the dissemination of research through the open-access system. One user referred to the service as "a beacon for scientific communication." Several commenters expressed how crucial arXiv has been for them personally in enabling them to quickly access the latest research in their field. There was an overall perception that arXiv was an important leader in the development of alternatives to traditional publishing. Independent researchers who are unaffiliated with large institutions and who might otherwise have delayed access to papers particularly emphasized the importance of arXiv for their work.

The combination of multiple choice responses (see Appendix B) and the extensive and thoughtful open text comments pinpointed areas that need to be upgraded and enhanced. Improving the search function emerged as a top priority as the users expressed a great deal of frustration with the limited search capabilities currently available, especially in author searches. Providing better support for *submitting* and *linking* research data, code, slides and other materials associated with papers emerged as another important service to expand. Regardless of their subject area, users were in agreement about the importance of continuing to implement quality control measures, such as checking for text overlap, correct classification of submissions, rejection of papers without much scientific value, and asking authors to fix format-related problems. Several users commented on the need to randomize the order of new papers in announcements and mailings. There were several useful remarks about the need to improve the endorsement system and provide more information about the moderation process and policies.

In regard to arXiv's role in scientific publishing, some users encouraged the arXiv team to think boldly and further advance open access (and new forms of publishing) by adding features such as peer review and encouraging overlay journals. On the other hand, many users strongly emphasized the importance of sticking to the main mission and not getting side-tracked into formal publishing. There was a similar divergence of opinion about encouraging an open review process by adding rating and annotation features. When it comes to adding new features to arXiv to facilitate open science, the prevailing opinion was that any such features need to be implemented very carefully and systematically, and without jeopardizing arXiv's core values.

While many respondents took the time to suggest future enhancements or the finessing of current services, several users were strident in their opposition to any changes. Throughout all of the suggestions and regardless of the topic, commenters unanimously urged vigilance when approaching any changes and cautioned against turning arXiv into a "social media" style platform. The feeling is that arXiv as it exists is working well and while there are some areas for improvement, too much change could potentially weaken the effectiveness and overall mission of arXiv.

KEY FINDINGS

Improving the Current arXiv Services

- When asked about the importance of improving a specific range of services, more than 70% of respondents said that improving search functions to allow more refined results was *very important/important* across all groups by years of use, age groups, number of articles published, country groups, and subject areas. Many commenters requested enhanced functions such as author search, date-limited searching, and searching non-English languages. Search was equally problematic regardless of whether the user searched for a known paper, was browsing a subject category, or looking for specific authors.
- A series of questions asked users about improving the submission process specifically with (1) support for submitting research data, code, slides and other materials; (2) improving support for linking research data, code, slides, etc., with a paper; and (3) updating the TeX engine and various other enhancements. Support About 40% of respondents rated each one as *very important/important.* The open text responses also displayed considerable interest in better support for supplemental materials, although respondents disagreed as to whether they should be hosted by arXiv or another party. Many respondents are supportive of integrating or linking to other services (especially GitHub), while a significant number of respondents also indicated doubts about long-term availability and link rot for content not hosted within arXiv. Some expressed concerns regarding the resources required for arXiv to improve this. There was some interest in including the data underlying figures in arXiv papers.
- Among other services and improvements recommended by respondents were:
    - Consistent inclusion of information and links about the published versions of the papers.
    - More refined options for alerting, both email and RSS. Several respondents specifically requested email alerts for works by a particular author, and there was some interest in HTML-formatted email with live links.
    - Updating and keeping current arXiv's TeX engine and provide TeX templates or style files to make submission easier.
    - Linking papers to each other via citations and actionable links in bibliographies.
    - Ability to submit a PDF, an increase in the file size limit (often with specific request to link to figures), and the ability to upload multiple files at once.
    - Allowing submission directly from authoring platforms (such as Overleaf or Authorea).
    - Providing use statistics such as paper downloads and views
    - A much larger percentage of recent arXiv users (five years or less) selected the "no opinion" option about current service upgrades. For all the questions in this category, the same trend is visible: a higher percentage of recent users expressed that they had no opinion and this percentage of respondents decreased with each level of increase in years of use. *Interestingly, this same trend is not visible by age group; i.e., our data do not show that a higher percent of younger users have no opinion.*

Importance of Quality Control Measures

- arXiv's users were asked a series of questions regarding quality-control measures. Based on the 26,430 responses to specific controls, the most important of these (ranked *very important/important*) were:
    - Check papers for text overlap, i.e., plagiarism   77%
    - Make sure submissions are correctly classified   64%
    - Reject papers with no scientific value   60%
    - Reject papers with self-plagiarism   58%

- A large percentage of all demographic groups found checking for plagiarism to be important and a slightly smaller group found checking for self-plagiarism as important. There was no discernible difference across demographic groups for the other measures. Similarly, self-plagiarism was also mentioned as another area for improvement. Some noted that context is the key; for example, conference papers are a common and typical area where self-plagiarism could occur in an otherwise scientifically sound submission.
- Several respondents said they were unaware of precisely what quality-control measures were already in place, and felt that the process is too opaque. Others acknowledged the difficult balance between rejecting papers that are clearly unworthy—"crackpot"—and rejecting papers for other, perhaps less obvious, and anonymized reasons. However, even in the face of such criticisms there was a strong thread of satisfaction with arXiv's current quality-control process and users cautioned against going too far in the other direction.
- Some users would prefer that arXiv embrace a more open peer review and/or moderation process, while others were adamant that current controls allow arXiv the freedom and speed of access that is otherwise unobtainable through traditional publishing.
- Overall, the feeling was that quality control matters but user comments varied greatly in relation to how arXiv could practically achieve these goals. As one respondent wrote, "Judgment about quality control is a very relative issue."

Adding New Subject Categories:

- 73% of the respondents were not interested in seeing new subject categories added to arXiv. 26% of respondents would like to see new subject categories added and suggested chemistry (881), engineering (483), biology (429), economics (248), philosophy (220), and social sciences (106). There were also several smaller categories such as Machine Learning (82 responses) and Artificial Intelligence (27 responses).
- A frequently repeated theme was that arXiv does not need to focus particularly on additional subjects but instead should focus on the refinement and addition of subfields and subcategories, especially in High Energy Physics Theory as well as Mathematics.

Developing New Services

- Users were asked to rate a range of proposed new services for arXiv. In the ranked responses, more than 63% of users rated adding direct links to papers in the references (reference extraction) *as very important/important.* Citation export in formats such as BibTex, RIS was rated as *very important/important* by more than 57% of users, and extraction for the BibTeX entry for the arXiv citation was similarly rated by more than 55% of respondents. Citation analysis tools in general were ranked as *very important/important* by almost 53% of respondents.
- In the open text comments, opinions were divided on the need for enhanced citation-analysis capabilities. While users were generally in favor of citation tools many of the same users noted that other systems are already doing this, and that this was sufficient for their needs.
- In the multiple choice survey responses the option to "offer a rating system so readers can recommend arXiv papers that they find valuable" was closely split between *very important/important* (36%) and *not important/should not* be doing this (36%). This matches the way the comments were closely split between those in favor and those less certain. Also, it was found valuable by 50% of recent users as compared with 28% of seasoned users. In addition, a larger percentage of younger users find it important (42% of those under 30 years), as compared to 28% of those 60 and above. Opinions were divided in the open text comments but overall the respondents were hesitant about the idea. Some users liked the rating feature "in an ideal world" setting, but did not think it was appropriate for arXiv; others expressed concern that it would dilute the mission of arXiv, or simply appears unfeasible in arXiv's current incarnation. However, even users directly in favor of a rating system raised issues about whether it would be open to the public, rated by peers, anonymous, etc. Several respondents stressed that such a feature would need to be implemented very carefully.

- Like the question about offering a rating system, the idea of adding an annotation feature to allow readers to comment on papers was almost evenly split, with 34.89% of users ranking it as *very important/important* and 34.08% as *not important/should not be doing this.* In the open text responses, the trend opposed the idea and some of the responses reflected strongly negative feelings. Those in favor or open to the idea of a commenting system often added a caveat and in general there was a sense of caution even for those responding positively. A common theme of concern was that a moderated system and verifiable accounts would be necessary to prevent a free-for-all. Unlike the question about offering a rating system, there were no discernible differences in opinion based on different demographic characteristics.

Finding arXiv Papers:

- The vast majority of arXiv's users access the papers directly from the homepage (79%), followed by using Google to search (50%) and Google Scholar (35%).
- Once on the homepage, reactions were mixed regarding the ease of use and navigation. 32% rated this as *easy,* but only 25% find it *somewhat easy* and 21.6% rated it *somewhat difficult* to use.
- To discover content, 63% of users *go to the link for new or recent under a particular category* and equally 63% of users *use arXiv's search engine and enter a specific arXiv ID, author name or search term.* A small number of users, 14%, rely on the daily mailing list and then look for a particular article in the search field.
- In the open text comments, opinion was divided about the user interface. The majority of respondents disliked the outdated style, but a definite subgroup appreciated the interface's simplicity, which these users feel helps arXiv efficiently carry out its mission. The main issues mentioned aside from the homepage's look were the number of links, layout and finding submission information. The lack of hierarchy in organization was found challenging to understanding arXiv's navigation.
- Requests for enhancements related to UX included greater personalization of arXiv for readers; for example, the ability to "favorite" papers, curate a personal library, and see recommendations when users visit the site. Other users mentioned the development of APIs to further facilitate the development of overlay journals. Some users also suggested the development of a mobile-friendly version.
- Many commenters either described how they rely on other services to interact with arXiv content (site-specific searches, ADS, INSPIRE) or recommended features based on their experience with other information systems. Among those frequently praised were ADS, INSPIRE, Google Scholar, gitxiv.com and arxiv-sanity.com.

Also see:

About arXiv: arXiv, an open-access scientific digital archive, is funded by the Simons Foundation, Cornell University Library, and about 190 member libraries from all around the world. The site is collaboratively governed and supported by the research communities and institutions that benefit from it most directly, ensuring a transparent and sustainable resource. It is a moderated scholarly communication forum informed and guided by scientists and the scientific cultures it serves. As of June 2016, arXiv contains more than 1,110,000 e-prints. In 2015, the repository saw 105,000 new submissions and close to 139 million downloads from all over the world.

APPENDIX A: DEMOGRAPHICS OF RESPONDENTS

I use arXiv in the following ways: (Please choose all that apply)

Answer	%	Count
I am an arXiv reader	**93%**	**31862**
I am an arXiv author	**53%**	**18270**
I am an arXiv submitter	**50%**	**17189**
I am an arXiv (other type of user): Please describe	**2%**	**845**

The number of articles I have published/submitted on arXiv is:

Answer	%	Count
1 article	**11.99%**	**2570**
2 articles	**8.96%**	**1920**
3 - 4 articles	**15.19%**	**3254**
5-10 articles	**23.06%**	**4941**
More than 10 articles	**40.80%**	**8743**
Total	**100%**	**21428**

My current occupation is: (Please choose ALL that apply)

Answer	%	Count
I am an academic faculty member (professor) at a college or university	27%	8868
I am an academic staff member (researcher or postdoc) at a college or university	22%	7207
I am a researcher at a non-profit or governmental agency	8%	2707
I am a Masters/Ph.D. student	30%	9890
I am an undergraduate student	5%	1514
I am (please describe)	13%	4353

13% of respondents (4353) indicated a different occupation category. The top ones included researchers at a company or industry (900), engineer (515), and retired individuals (478). There were also respondents who described themselves as science writers, editors, or freelance editors. Other response types included data scientist, self-described amateur researchers, self-described laypeople, unemployed, teachers, and the generally curious (e.g., "a man doing research as hobby").

As a user, my main subject area of interest in arXiv is: (please choose all that apply)
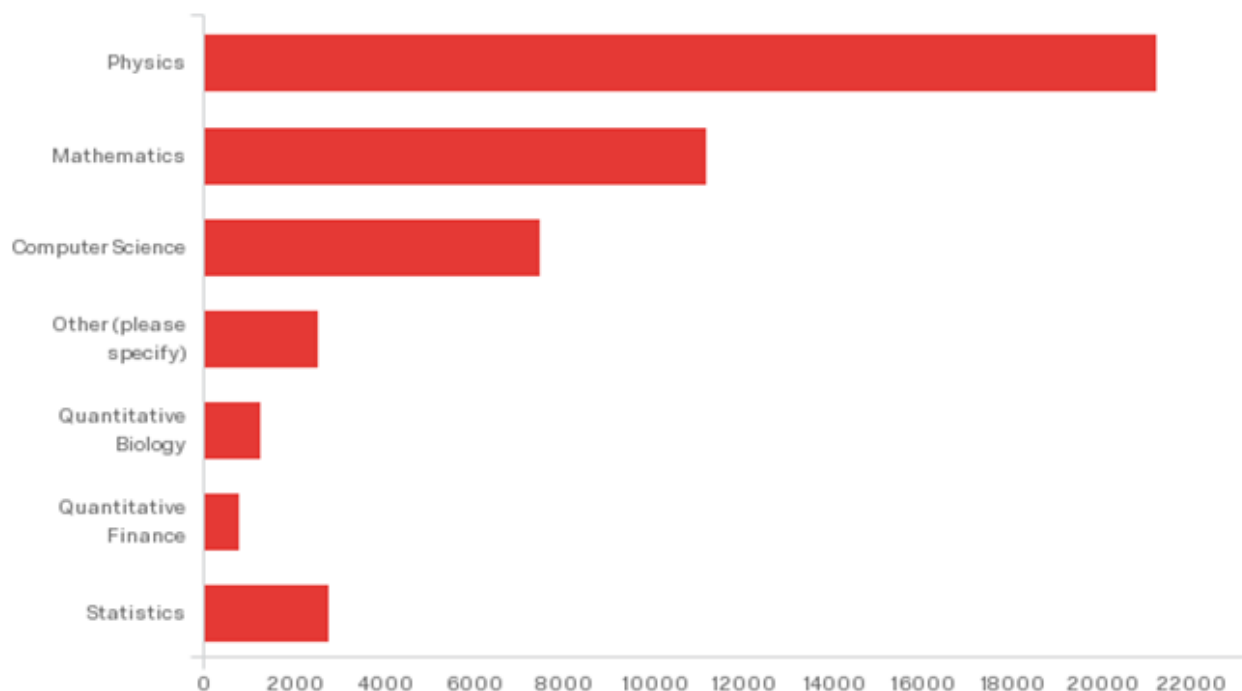
Almost 2,000 respondents checked the Other option to specify their main area of interest. The top categories were astrophysics (726) and astronomy (653).
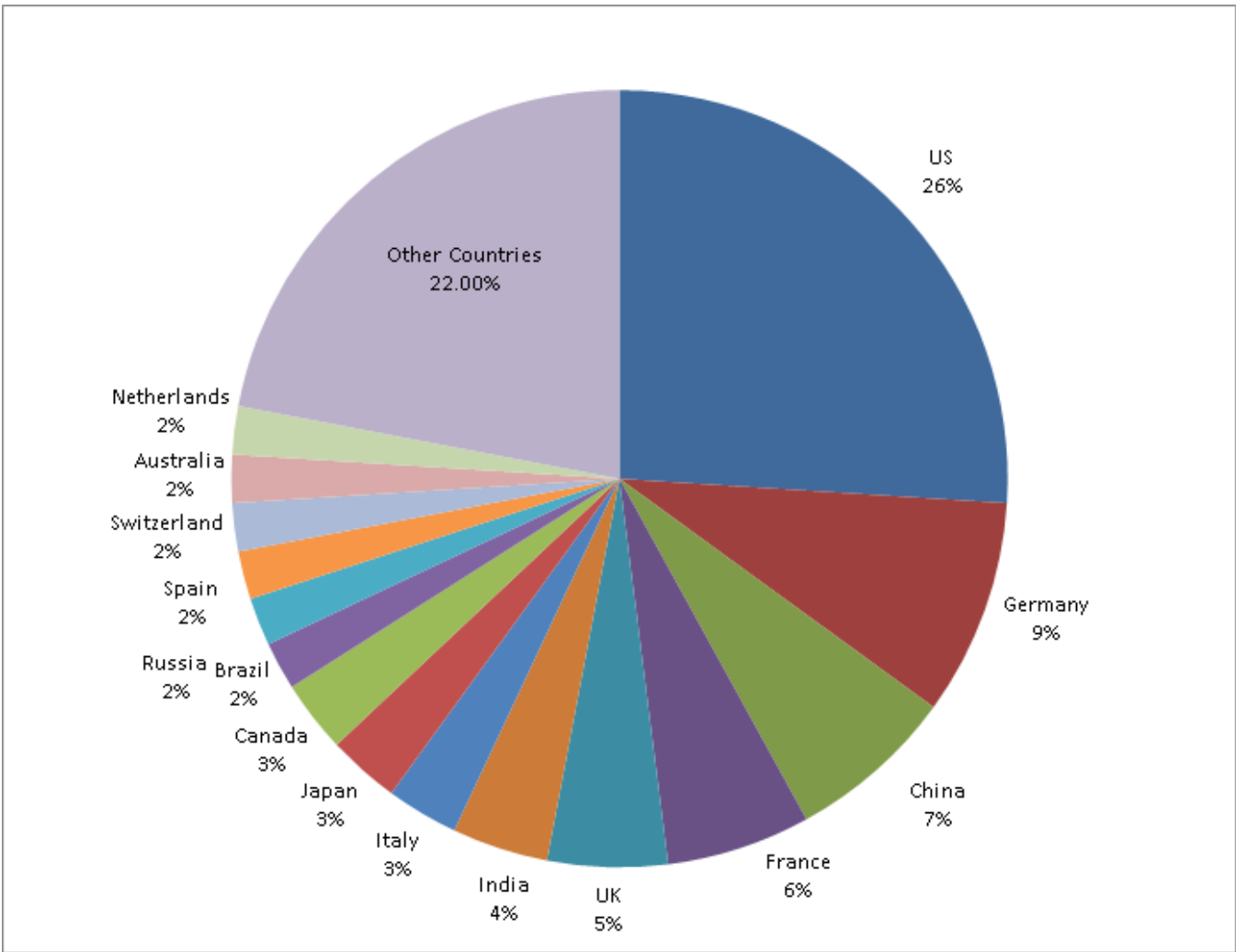
I have been using arXiv for:

Answer	%	Count
0 - 2 years	19.54%	6470
3 - 5 years	28.96%	9592
6- 10 years	25.44%	8425

11 or more years	26.06%	8632
Total	100%	33119

My age is:

Answer	%	Count
younger than 30 years	37.42%	12364
30 - 39 years	31.27%	10332
40 - 49 years	13.76%	4545
50 - 59 years	9.30%	3073
60 - 69 years	5.77%	1908
70 years and over	2.47%	817
Total	100%	33039

Q6 - My main place of work is located in:

Other Countries: 1% or less representation each from 113 countries

APPENDIX B: OPINIONS ON ARXIV'S CURRENT SERVICES & FUTURE DIRECTIONS

How important is it to improve on the following CURRENT arXiv services?

Question	Very important & important	Somewhat important	Not important & should not be doing this	No opinion
Improve search functions to allow more refined results (e.g., narrow down results by additional search terms, filter by publication year or institutional affiliation, etc.):	70.38%	19.34%	6.14%	4.13%
Improve support for submitting research data, code, slides, and other materials associated with a paper (e.g., I want to be able to upload my datasets/machine- readable tables with my article):	41.95%	22.64%	14.03%	21.37%
Improve support for linking research data, code, slides, and other materials associated with a paper (e.g., I want to be able to link to my slides on SlideShare):	40.65%	25.20%	17.70%	16.45%
Improve support for submitting research papers by updating the TeX engine:	39.36%	23.17%	16.71%	20.76%
Improve the email alert system so that readers can customize their settings and choose to receive alerts about specific sub-topics:	37.85%	26.48%	20.25%	15.42%
Improve the trackback mechanism (linking papers back to blogs and commentaries that cite thos papers):	36.52%	29.50%	20.30%	13.67%
Simplify the submission process by providing clearer instructions and simpler language:	32.45%	22.55%	25.20%	19.80%

How important is it to develop the following NEW arXiv services? -

Question	Very important & important	Somewhat important	Not important & should not be doing this	No opinion
Add direct links to papers in the references (support reference extraction):	63.04%	26.89%	5.78%	4.29%
Offer citation export in formats such as BibTeX, RIS, etc.:	57.68%	23%	10.95%	8.37%
Enable extraction for the BibTeX entry for the arXiv citation:	55.54%	23.82%	9.72%	10.91%
Provide Citation Analysis tools (examining the frequency and pattern of a paper's citation):	52.95%	27.08%	14.28%	5.69%
Support compliance with public/open access mandates (funding agency policies that require research results to be made public) by allowing final versions of papers to be submitted with information such as funding sources and grant numbers:	42.06%	26.21%	13.68%	18.05%
Enable submitting an article to a journal at the same time as it is uploaded to arXiv:	39.28%	23.09%	25.23%	12.40%
Offer a rating system so readers can recommend arXiv papers that they find valuable:	36.28%	21.76%	35.56%	6.40%
Enable linkages (interoperability) with other repositories (e.g., run by libraries), so that a paper accepted by arXiv is accepted at the same time by the other repositories:	35.25%	28.14%	17.25%	19.36%
Develop an annotation feature which will allow readers to comment on papers:	34.89%	23.62%	34.08%	7.41%

Where do you go to find arXiv papers? Please choose all that apply:

Answer	%	Count
Go directly to arXiv.org (arXiv homepage)	79%	22804
ADS	14%	4144
Inspire	13%	3773
Google Scholar	35%	10016
Google search engine	50%	14440
arXiv email alerts	14%	4086
Other search engines	5%	1402
Subject gateways for arXiv, such as the Math Front	4%	1203
Other (please specify):	9%	2662

If you have used the arXiv homepage for finding papers, how easy is it to navigate?

Answer	%	Count
Very easy	14.85%	3916
Easy	32.05%	8450
Somewhat easy	25.20%	6644
Somewhat difficult	21.60%	5696
Difficult	5.02%	1324
Very difficult	1.27%	336
Total	100%	26366

If you have used the arXiv homepage, how do you usually navigate our main page? Please choose all that apply.
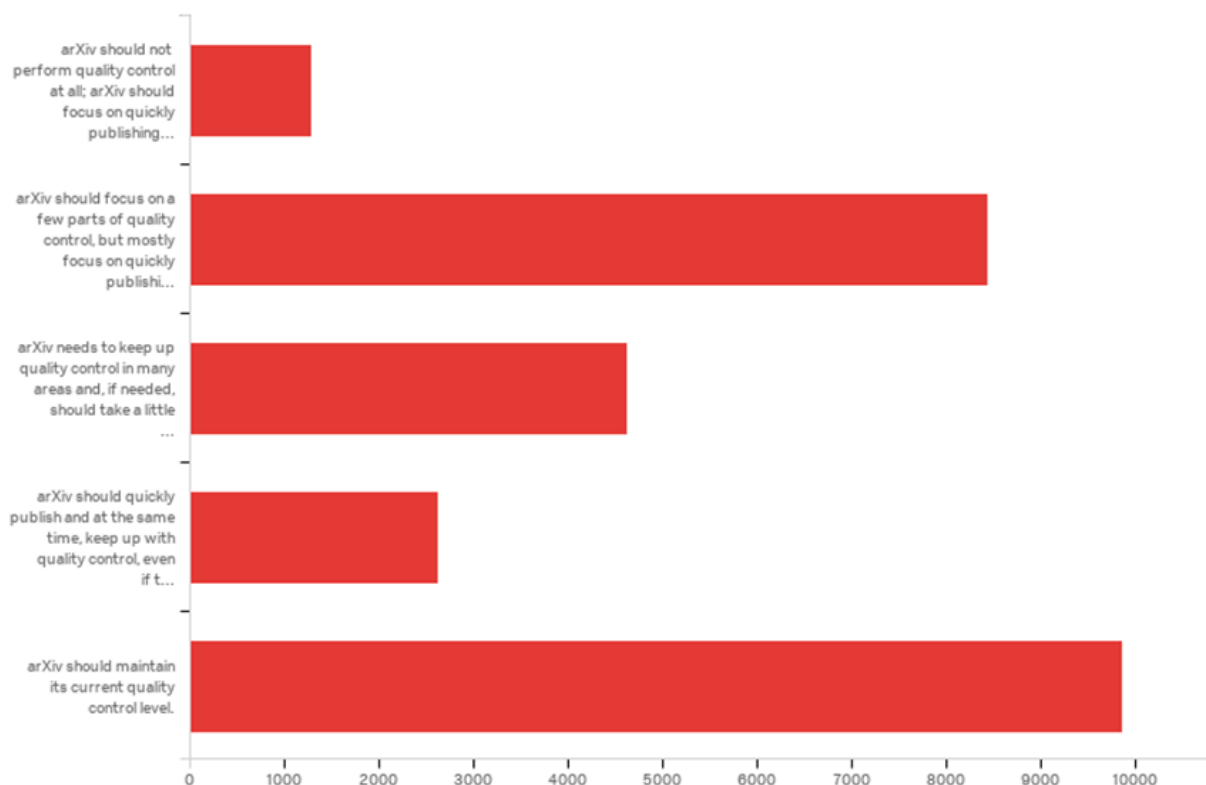
Answer	%	Count
Go to link "new" or "recent" under a particular category	63%	16503
Use arXiv search engine and enter a specific arXiv-id, author name, or search term	63%	16478
Receive daily mailing list, and then look for a particular article on the search field	14%	3692
Other, please explain:	3%	853

How important are the following CURRENT quality control measures?

Question	Very important & important	Somewhat important	Not important & should not be doing this	No opinion
arXiv checks papers for text overlap: an author's use of too much identical text from other authors' papers, without making it clear that the text is not their own material, i.e., "plagiarism":	77.41%	14.66%	4.96%	2.96%
arXiv makes sure submissions are correctly classified (the subject categories are included on the arXiv homepage):	64.38%	25.32%	7.01%	3.29%
arXiv keeps out (rejects) papers that don't have much scientific value:	60.02%	19.14%	15.49%	5.35%

arXiv checks papers for too much text re-use from an author's earlier works, i.e., "self-plagiarism" (reuse of identical content from one's own published work without citing):	57.77%	24.64%	14.08%	3.51%
arXiv checks papers for format-related problems (line numbers in text, missing references, oversize submissions, etc.) and asks authors to fix them before they are announced:.	55.00%	29.83%	11.51%	3.66%
arXiv moderates the scientific content of trackback (links to blogs and commentaries) before permitting the link to be added:	39.60%	26.31%	17.59%	16.50%

Please choose any ONE of the following statements that you agree with the most:

Overall, how satisfied are you with arXiv?

Answer	%	Count
Very satisfied	52.92%	14770
Satisfied	42.43%	11841
Somewhat satisfied	3.55%	990
Somewhat dissatisfied	0.54%	150
Very dissatisfied	0.15%	42
No opinion	0.42%	116
Total	100%	27909

Which of the following BEST describes your opinion of how arXiv needs to move forward?

Answer	%	Count

arXiv should focus on its main purpose, which is to quickly make available scientific papers. This will be enough to hold up the value of arXiv in the future.	71.94%	19865
arXiv should expand its main mission, and spend more time and resources to provide new services. This is necessary to hold up the value of arXiv in the future.	19.59%	5410
No opinion	8.47%	2340
Total	100%	27615