

Preprint website plans revamp

But users are wary of major changes to arXiv repository.

BY RICHARD VAN NOORDEN

A multimillion-dollar funding drive is being readied to transform arXiv, the vastly popular repository to which physicists, computer scientists and mathematicians flock to share their research pre-prints openly.

But the results of an enormous user survey published this week suggest that researchers are wary of drastic changes to a site that has become an essential part of the infrastructure of modern science.

Last year, the site served up around 139 million downloads, and it now holds more than 1.1 million free papers. But it is being sustained by fragile code, donations from libraries and a charitable foundation and the good will of about 150 or so volunteer moderators, says the site's programme director, Oya Rieger. With its 25th anniversary approaching in August, arXiv's advisory teams of scientists and librarians are considering a plan that involves raising US\$2.5 million to \$3 million to modernize the platform. That will sit on top of its \$1-million annual budget for staff and servers.

To attract support from donors, arXiv's operator, Cornell University Library in Ithaca, New York, is hoping to come up with a "compelling vision", Rieger says.

Scientists seem to love arXiv: 95% of the survey's 36,000 respondents said that they were very satisfied or satisfied with it. And most want to keep it just the way it is, although perhaps with some modernization. They were enthusiastic about the possibility of tweaks to

improve the site's search functions, and about allowing references to be hyperlinked directly to research papers, for example (see 'What do arXiv users want?'). Some wanted the site to broaden into new subject areas, such as chemistry — although such expansion would require the recruitment of scientists who are willing to moderate the manuscripts, notes David Morrison, chair of arXiv's scientific advisory board.

SOCIAL FORUM

When asked whether arXiv should embark on more transformational changes, respondents gave mixed answers. In particular, some questions focused on whether it should develop into a social forum that allows scientists to comment on papers or leave ratings. A few social-media sites have already been built around the repository for just such purposes — such as SciRate and Arxiv Sanity Preserver — and some argue that the site itself should begin to incorporate such functionalities. "ArXiv should be more dynamic — allowing readers to filter the wheat from the chaff," says Alán Aspuru-Guzik, a quantum chemist at Harvard University in Cambridge, Massachusetts. But one-third of respondents said that this wasn't important or that arXiv shouldn't be doing it. Only 34% voted in favour of such changes.

That response points to a tension between

researchers who want to see the site incorporate aspects of open review, and those who want it to stick to its core mission of allowing rapid exchange of scholarly papers, says Rieger. There were hints of a generational divide, with those aged under 30 more in favour of allowing comments. But even those who wanted a more social site said that they were keen to avoid a commenting free-for-all, Rieger adds.

"The message was more or less 'stay focused on the basic dissemination task, and don't get distracted by getting overextended or going commercial,'" says Paul Ginsparg, a physicist at Cornell University who launched arXiv in 1991 as a pre-World-Wide-Web-era bulletin board.

CHECKS AND BALANCES

Ginsparg notes, however, that arXiv's users sometimes don't know what they want until they get it. Researchers said that they liked the quality control now built into the site, including checks of papers for text overlap with other reports (potential plagiarism), classifying papers into the correct subject areas and rejecting work that has little scientific value. "These are for the most part things that users never actually requested," Ginsparg says. In the past 5 or so years, he has introduced automated machine-learning code that filters through the more than 9,000 papers submitted each month and flags up potential issues to human moderators.

In September, arXiv's advisory boards will meet to draw up a road map for progress and to discuss how to get the funds needed to modernize the site. The site is currently sustained by member institutions (mainly libraries, but also some research funding agencies) and by the Simons Foundation in New York. But some discussions have been held with other potential contributors such as the US National Science Foundation. It is also possible that publishers or scientific societies could be asked to contribute, says Rieger.

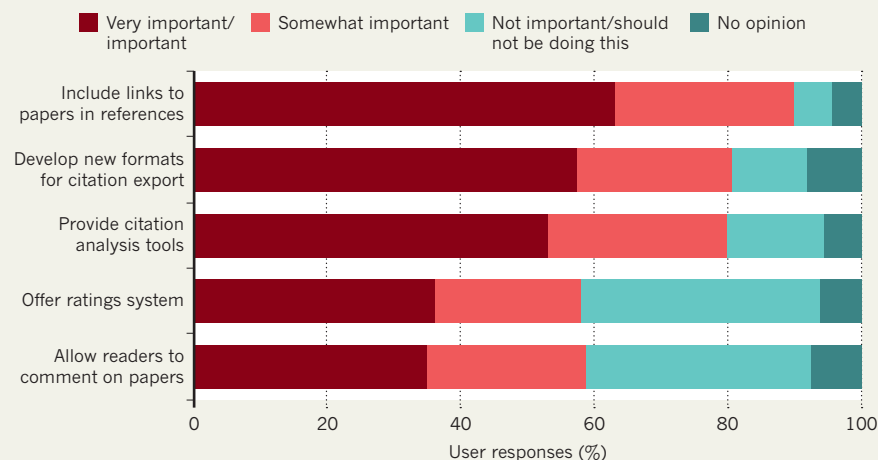
She adds that the site will need to be careful to remain objective. "We want to make sure that arXiv continues to be a neutral, trusted service," she says. ■

CORRECTION

The science-workforce graph in 'China by the Numbers' (*Nature* **534**, 452–453; 2016) erred in stating that China's population is 1.3 trillion. It is more than 1.3 billion.

WHAT DO ARXIV USERS WANT?

The preprint repository got high marks overall from 36,000 respondents to a survey, but there was no consensus over whether the site should add social-media functionalities.



SOURCE: ARXIV