

NATURE NEWS  
Welcome to nature.com

This website uses cookies. By using this website and agreeing to this policy, you consent to our use of cookies in accordance with the terms of this policy.

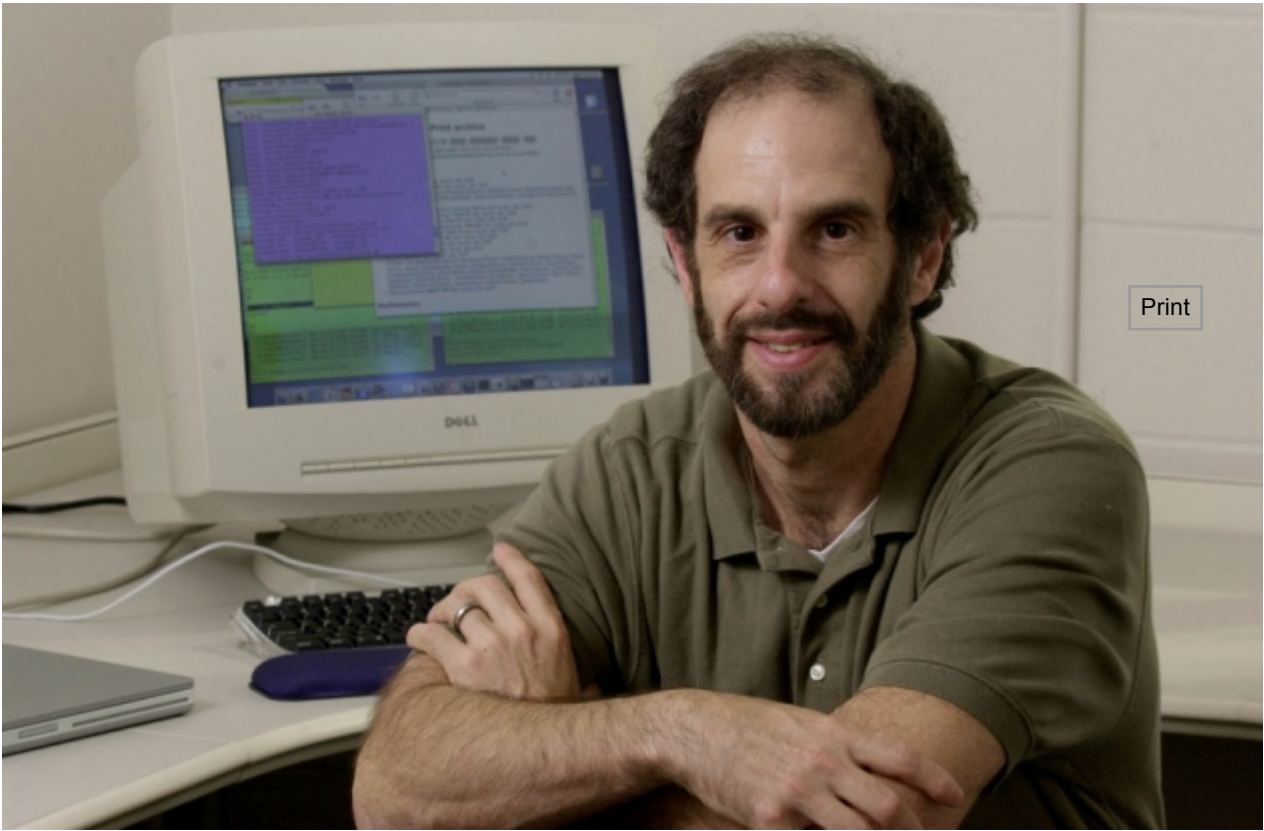
Accept and close

# The arXiv preprint server hits 1 million articles

**Website where scientists flock to upload manuscripts before peer review has doubled its holdings in six years.**

**Richard Van Noorden**

30 December 2014



*Courtesy of John D. and Catherine T. MacArthur Foundation*

Paul Ginsparg founded arXiv in 1991.

The popular preprint server arXiv.org, where physicists, mathematicians and computer scientists routinely upload manuscripts to publicly share their findings before peer review, now holds more than 1 million research articles.

The repository, launched as an 'electronic bulletin board' in August 1991, just before the dawn of the World Wide Web, took 17 years to accumulate half a million manuscripts, but has taken just 6 more to double its holdings.

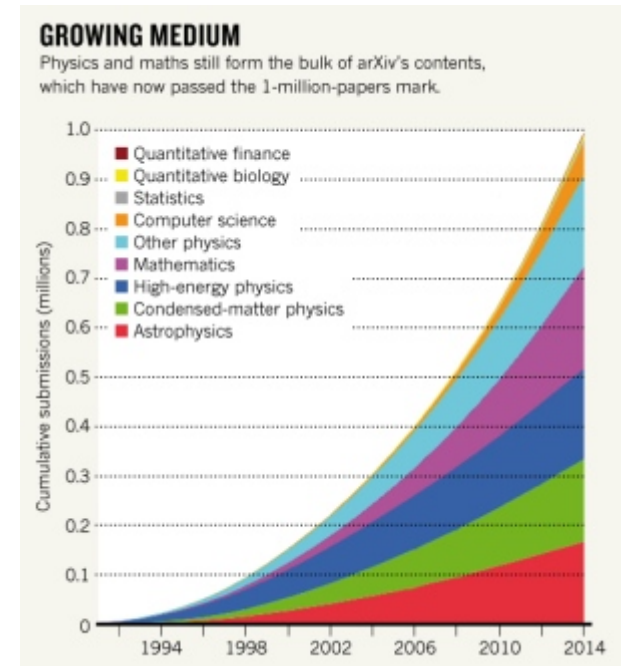
Researchers now submit around 8,000 articles to the arXiv each month — more than 250 a day, on average. The site's administrators make the raw, non-peer-reviewed manuscripts available in batches after a brief quality-control check, such as a cursory glance for appropriateness by one of 130 volunteer moderators, and automated filtering to check for text overlap with existing papers.

The site reached more than 1 million papers on 29 December, after administrators returned from holidays and updated the server with manuscripts submitted after business hours on Christmas Eve (24 December).

Judging by the running count of articles currently displayed on the arXiv's home page, the manuscript that bears the landmark 1-millionth identifier is 'A well conditioned and sparse estimate of covariance and inverse covariance matrix using a joint penalty', which was submitted at 7:34:19 GMT on 26 December by Ashwini Maurya of

#### Related stories

- Preprints come to life
- Researchers aim to chart intellectual trends in arXiv



(Click to enlarge)

Michigan State University in East Lansing. But in fact, the site's millionth article cannot be pinpointed so precisely, says arXiv founder Paul Ginsparg, a physicist at Cornell University in Ithaca, New York. The count is actually a slightly fuzzy estimate owing to the way submissions are indexed and because the occasional duplicate or junk submission creeps in — something that can now be spotted by screening software but was easier to miss in the early days of the site.

- ArXiv at 20

**More related stories**

### **The start of a revolution**

The arXiv's massive collection of free manuscripts now runs alongside the slower system of publishing peer-reviewed manuscripts in scientific journals. But it was all a distant prospect when Ginsparg, then at the Los Alamos National Laboratory in New Mexico, founded an electronic service to share preprint articles for “a few hundred friends and colleagues working in a subfield of high-energy physics”, as he recounted in a 2011 *Nature* article written for the arXiv's twentieth anniversary<sup>1</sup>. Since its inception, the server has broadened its reach to cover numerous other fields in physics, as well as mathematics, computer science, statistics and the quantitative aspects of finance and biology.

The popularity of the arXiv server has varied among disciplines. Some physicists were reluctant at first to share their results before formal, peer-reviewed publication, but came around to the idea when they realized the benefits of a quick way to publicize their work. For example, the discovery of a class of iron-based superconductors in 2008 brought a host of condensed-matter experimentalists to the site, “won over by the need to stake precedence claims and get their results in front of theorists”, Ginsparg wrote in his 2011 piece.

Nowadays, many important findings are posted first at the site. When the reclusive Russian mathematician Grigori Perelman proved the Poincaré conjecture (a statement about the nature of three-dimensional spaces that had resisted proof for almost a century), he posted his papers only at arXiv.org, and nowhere else. (Perelman later declined a Fields medal for the work). Last year, the site inspired an imitator in biology, bioRxiv.org, launched by Cold Spring Harbor Laboratory Press in New York.

The arXiv is expanding ever faster. On 19 December, it announced that it would make its paper identification numbers one digit longer to cope with expected spikes exceeding 10,000 submissions per month. It now receives more than 10 million download requests a month.

None of this comes for free, but it is still relatively cheap. Projected annual costs for staff and servers ran to US\$885,987 in 2014, less than \$10 per paper added. Much of that is funded by member institutions — after a plea put out by Cornell University Library in 2010 — and by the Simons Foundation, a private foundation based in New York.

Researchers are also mining the arXiv repository to study how scientists communicate their work. Earlier this month, Ginsparg and Daniel Citron, a graduate student in physics at Cornell, reported on how often scientists re-use the text of other papers<sup>2</sup> by analysing overlap of 7-word phrases in some 757,000 articles published at arXiv.org from 1991 to 2012. Other researchers are mining arXiv articles to chart trends in the popularity of scientific ideas, much as the digitization of Google Books has allowed humanities researchers to spot the incidence of particular phrases in English literature.

One million articles is a natural milestone, but physicists might prefer other numerical landmarks, Ginsparg says. “The significance of 1,000,000 is just the base-10 accident that we happen to have 10 fingers, whereas some would argue that  $2^{20} = 1,048,576$  is a more important number,” he says. “Also it’s the number of bytes in a megabyte.”

Ginsparg will not have to wait long — at current rates of growth, the site should hit that point by the summer.

*Nature* doi:10.1038/nature.2014.16643

## References

---

1. Ginsparg, P. *Nature* **476**, 145–147 (2011).

Show context

Article PubMed ISI ChemPort

2. Citron, D. T. & Ginsparg, P. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1415135111> (2015).

Show context

## Related stories and links

---

#### From nature.com

- **Preprints come to life**  
12 November 2013
- **Researchers aim to chart intellectual trends in arXiv**  
24 February 2012
- **ArXiv at 20**  
10 August 2011
- **European Research Council funds arXiv - a taste of changes to come**

#### From elsewhere

- **arXiv.org**

---

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

**Comments for this thread are now closed.**

#### Comments

**Subscribe to comments**

There are currently no comments.

**See other News & Comment articles from *Nature***

**Nature** ISSN 0028-0836 EISSN 1476-4687

© 2016 Macmillan Publishers Limited, part of Springer Nature. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER