



1<sup>st</sup> International  
Conference of  
TWAS Young  
Affiliates Network

# Let's talk about preprints

Eduardo S. Fraga



---

Instituto de Física  
Universidade Federal do Rio de Janeiro

## Preprint: an old way to make scientific work public

- Although there is a lot of discussion on the revolution that preprints might bring, they have been used in physics for over 50 years!



Library staff (l-r) Julie Kelley, Mae West, Roger Thompson call attention to preprint rack in Library

Fermilab History & Archive Project

Source: The Village Crier Vol. 7 No. 31, July 31, 1975



- High energy physics (HEP) has a tradition of preprints that dates back to decades before the creation of the arXiv in 1991 (the paradigm for current preprint databases).
- Labs and universities sent printed copies of preprints by mail before they were published in journals.
- Every major HEP center had a preprint library and “publisher”.
- The web was developed at CERN in part to make the preprint distribution system more efficient.
- The arXiv appeared in an epoch in which most commercial editorial houses could not be at the internet (those were the rules at that time).

# Natural evolution after the internet – which makes it new!

arXiv ([arXiv.org](http://arXiv.org), previously [xxx.lanl.gov](http://xxx.lanl.gov))

preprint → e-print

- **Birth date:** August, 1991 ( > 25 years!) as a repository for preprints in physics.
- **Creator:** Paul Ginsparg.
- **Def.:** “highly-automated electronic archive and distribution server for research articles”.
- **Covered fields:** physics, mathematics, computer science, nonlinear science, quantitative biology and statistics.
- **Maintenance and operation:** Cornell University (previously, LANL) + 8 “mirrors”.
- There is a Scientific Advisory Board and Area Committees (most people from the US).
- Users can download papers for free and in several formats.
- **Registered** authors can submit papers and revised versions (**but all versions remain!**); there is a moderator system.
- There are RSS lists and automatic (daily!) lists by email.

## ***1. General & Editorial Principles***

- 1.1. arXiv provides an open-access repository of scientific research to authors and researchers worldwide.
- 1.2. arXiv is a scholarly communication forum informed and guided by scientists and the scientific cultures being served.
- 1.3. Access to arXiv content via arXiv.org is free to individual end users.
- 1.4. Individual researchers can deposit their own content in arXiv for free.
- 1.5. Criteria and standards for depositing content in arXiv are maintained by the Scientific Advisory Board, and deposit is governed by transparent and publicly posted policies and procedures.
- 1.6. arXiv serves the needs of researchers in physics, mathematics, computer science, quantitative biology, quantitative finance, and statistics. Any expansion into other subjects or disciplines must include scholarly community support, satisfy arXiv's quality standards, and take into consideration its operational capacity and financial requirements.
- 1.7. Whenever possible, arXiv adopts open-source software and relevant standards and best practices.

- 2010 - 2013: formulation of a long-term plan for its financial support.
- New model: sponsored by libraries and research centers which represent its major users + Cornell Univ. + Simons Foundation + donations.
- List, values, costs spreadsheet and other details on line (total transparency!).

## Why does arXiv moderate submissions?

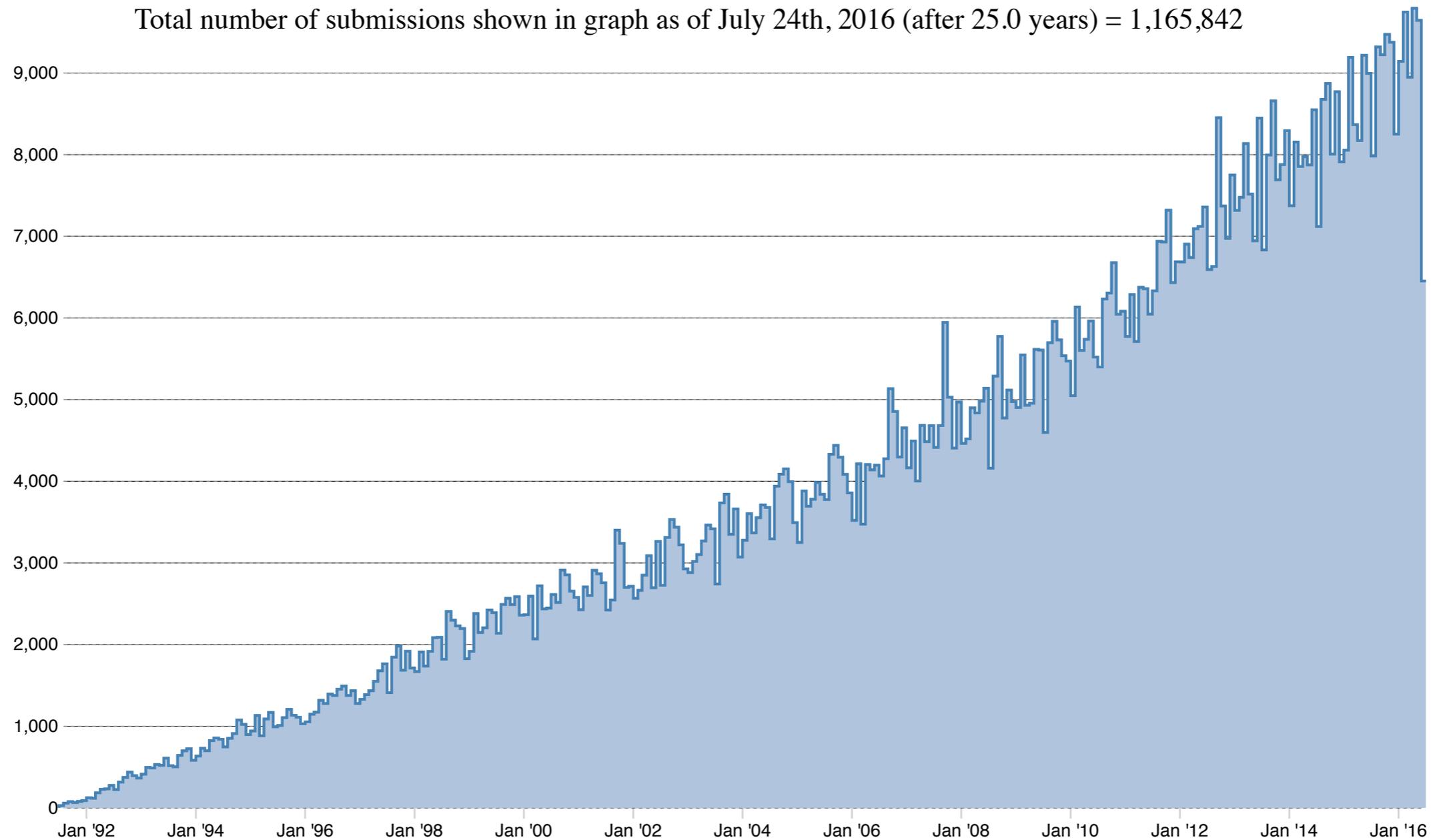
arXiv is distinct from the web as a whole, because arXiv contains *exclusively scientific research content*. Although arXiv is open to submissions from the scientific communities, our team has worked behind the scenes for a long time to ensure the quality of our content. Our policy is:

*arXiv is an openly accessible, moderated repository for scholarly papers in specific scientific disciplines. Material submitted to arXiv is expected to be of interest, relevance, and value to those disciplines. arXiv reserves the right to reject or reclassify any submission.*

Moderation helps to ensure that arXiv content is relevant to current research at much [lower cost](#) than conventional peer-reviewed journals, so we can continue to offer free access to the scientific community and the general public. Although our system may be imperfect, submissions that are determined to be inappropriate for arXiv may be still be posted on other sites or submitted to peer-reviewed journals.

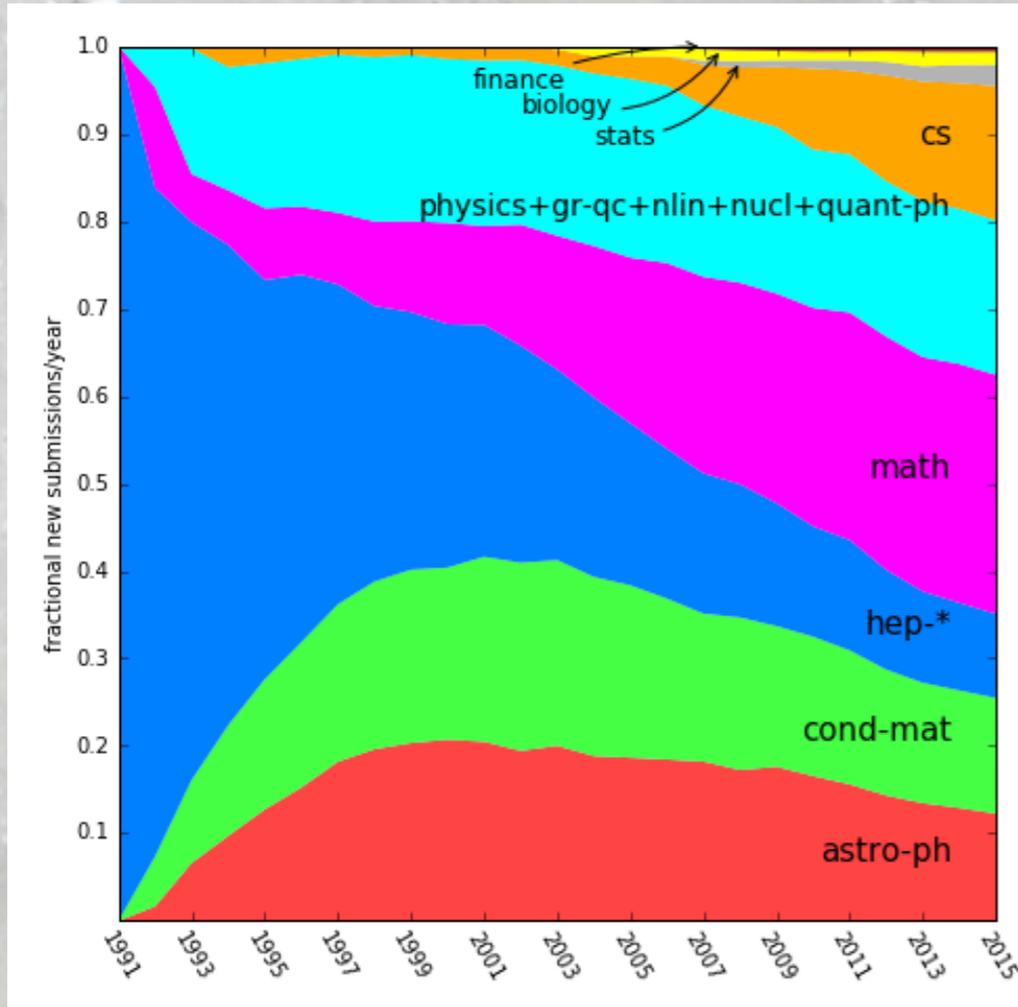
# Some impressive numbers

## arXiv monthly submission rates [[CSV](#)]

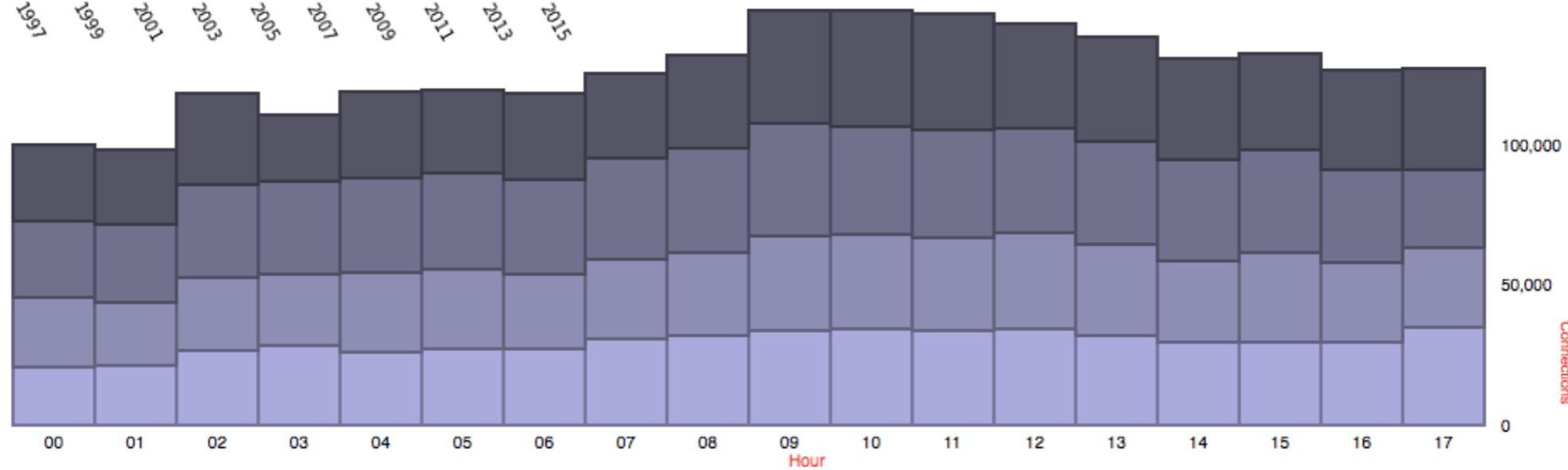
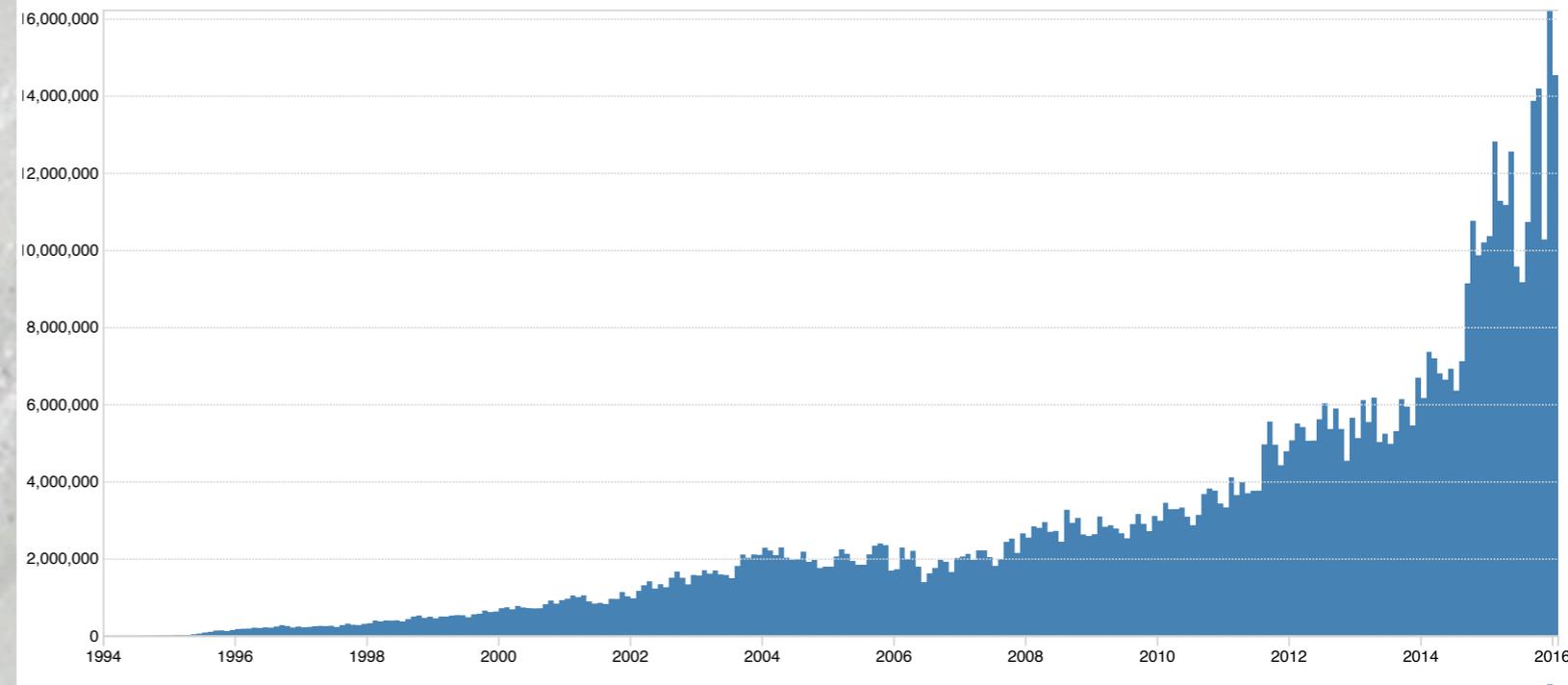


**Blue:** Number of new submissions received during each month since August 1991.  
Hover over the graph to see the exact count for a given month.

# Spreading into new fields, monthly downloads and daily access on a Sunday...



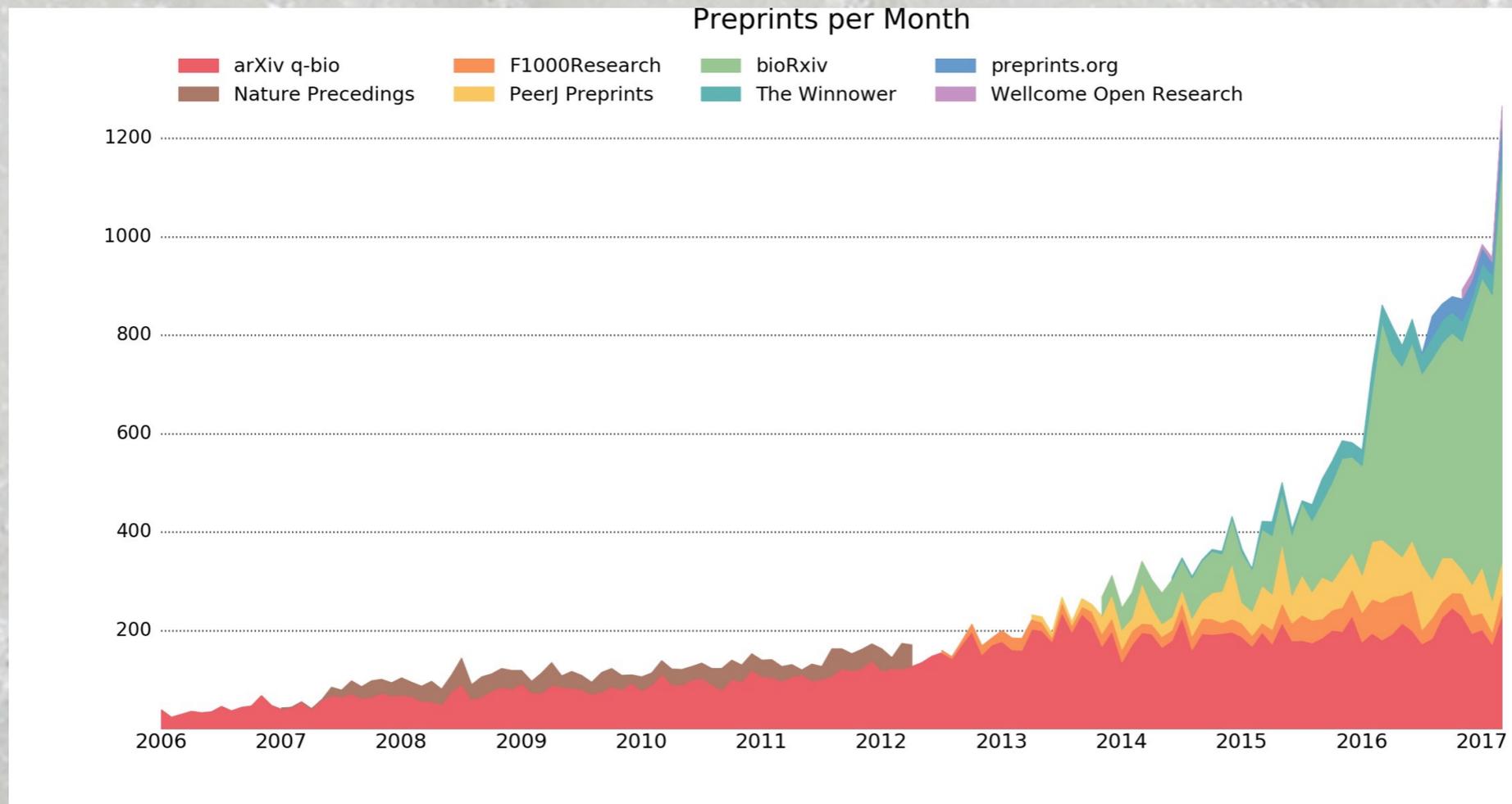
arXiv monthly download rates [CSV]



Total number of connections = 2,277,087 (+66 local & administrative connections)  
 Current local time is *Sunday, March 19th, 2017 18:40:39 EDT*



# Spreading into new fields: bio sciences coming with full speed!



- ASAPbio: a scientist-driven initiative to promote the productive use of preprints in the life sciences.
- Several funding agencies have changed their policies surrounding using preprints, and others are considering doing so.
- “The NIH encourages investigators to use interim research products, such as preprints, to speed the dissemination and enhance the rigor of their work.”
- Wellcome Trust: “As of January 2017, we will permit researchers to cite preprints, or pre-peer reviewed manuscripts, in their grant applications and end-of-grant review reports.”

# Preprint: a daily practice in physics and now spreading to other fields

- In a large number of research centers there is a preprint number registry (prior to submission to a journal). A practice that run for decades. A “quality stamp” before publication!
- Preprints submitted to a database (such as arXiv) are seen seen by virtually everyone working in the field and generate: comments, criticisms, suggestions, citation complaints, new collaborations, more visibility for peripheral groups, etc.
- A common practice is sending to a journal only after this feedback from the community.
- Since the PDFs are open access, even published papers are usually read as preprints.



HEP :: HEPNAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

## HEP Search High-Energy Physics Literature Database

Use "find" for SPIRES-style search ([other tips](#))

Brief format  [Easy Search](#) [Advanced Search](#)

[find j "Phys.Rev.Lett.,105"](#) :: [more](#)

### HOW TO SEARCH

SPIRES syntax is (mostly) supported (requires "find")  
find a richter, b and t quark and date > 1984  
find j phys.rev.,D50,1140 or j hep.0903,112  
find eprint arxiv:1007.5048 (Note the plots available on the detailed record)  
find fulltext "quark-gluon plasma" (Note new "fulltext" operator)  
find a ellis and refersto a witten (Note "refersto")  
find a kane and citedby title SUSY and topcite 200+ (Note "citedby")

New techniques:  
1985 richter quark multiplicity  
arXiv:1007.5048  
citedby:author:ellis -refersto:author:witten  
author:randall | author:sundrum cited:450->1350

Additional Help:

### HEP

Additions  
Corrections  
Search Tips  
FAQ  
Topcites: annual | recent  
Reviews  
HEP Citesummary  
Tools

### INSPIRE

About INSPIRE  
INSPIRE Help Central  
Blog  
Twitter  
feedback@inspirehep.net

### RESOURCES

ADS  
arXiv  
HepData  
INIS  
PDG  
PDG review of online resources

## Open science, open data, open code

INSPIRE grew out of the Stanford Public Information Retrieval System (SPIRES), a database started at SLAC National Accelerator Laboratory in 1969 when the internet was in its infancy.

After Tim Berners-Lee developed the World Wide Web at CERN, SPIRES was the first US-hosted website.



1<sup>st</sup> International  
Conference of  
TWAS Young  
Affiliates Network

## Preprint x publishing in refereed journals

- In spite of their tremendous success, preprints did NOT substitute the traditional publication system as a means of legitimation. We are very conservative and cautious.
- A paper can be relevant, read and cited by a multitude of people and still face difficulties to be published by a journal for many reasons. With reliable preprint repositories, it will at least be there, for the community. But is that enough?
- The refereeing system has, necessarily, a great deal of prejudice propagation. And can also be a high barrier to boldness. And this spills over into the way our agencies (or we) evaluate and rank researchers and research institutes.
- Preprints did not come to substitute publication in refereed journals. Recall: they have been around for half a century or more! e-prints have been around for 25 years! But maybe they can help us improve the ways we communicate and evaluate science.