

arXiv Update - January 2019

Since its inception in 1991 with a focus on the high energy physics community, arXiv has significantly expanded both its subject coverage and user base. It provides open access to about 1.5 million e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance, Statistics, Electrical Engineering and Systems Science, and Economics.

In 2018, the repository received 140,616 [new submissions](#), a 14% increase from 2017. The subject distribution is evolving as Computer Science represented about 26% of overall submissions, and Math 24%. There were about 228 million downloads from all over the world. arXiv is truly a global resource, with almost 90% of supporting funds coming from sources other than Cornell and 70% of institutional use coming from countries other than the U. S.

We've Moved!

After a careful assessment process, we moved the arXiv operation from [Cornell University Library](#) (CUL) to [Cornell Computing and Information Science](#) (CIS). Cornell has hosted arXiv since 2001 when its founder, Paul Ginsparg, left the Los Alamos National Laboratory and joined the Cornell faculty in the Physics and Information Science departments. Running the service has always involved a collaboration between CUL and CIS, as Professor Ginsparg and other CIS faculty lead R&D efforts through several NSF grants and other funding sources and contribute to the development of moderation policies. CIS's mission and values provide an environment that supports arXiv as a scholarly scientific enterprise, with the need for optimum service delivery and infrastructure sustainability. As an academic unit, CIS not only spans core contemporary IT realms, such as networks, systems, robotics, and machine learning, but is also concerned with the human aspects of computing, such as human-computer interaction and human-centered design. For researchers and arXiv users, the stewardship transition did not affect any operation. But users should anticipate arXiv engaging in new partnerships to add new revenue sources, further strengthen arXiv's current business model, and invigorate community commitment to maintaining and developing arXiv as a public good. arXiv will continue to benefit from the expertise of the CUL staff, especially in the service areas of metadata, copyright, and scholarly communication policies and best practices. Please see the [arXiv Transition FAQ](#) to learn more about the impetus and process behind the move.

Next Generation arXiv

In 2017, we embarked on the [next-generation arXiv](#) (arXiv-NG) initiative with a cornerstone [grant of \\$450,000 from the Sloan Foundation](#) to improve the service's core infrastructure. Our strategy for the overall development and replacement of the infrastructure, is for incremental and modular renewal of the existing arXiv system, [Classic Renewal](#), rather than building an entirely new system and migrating to it. Based on an assessment of various technology components and partnership options, we finalized an architecture and have been implementing some key technology components in 2018. An integral part of the [arXiv-NG](#) initiative is to consider new organizational and staffing models to ensure continuity of operations as the sustainability principles are at the heart of the arXiv-NG initiative. As we continue developing arXiv-NG, we are committed to continue arXiv's robust services.

Sustainability Update

arXiv's operation and development is funded by grants and a membership program that engages libraries and research laboratories worldwide. Currently, we have [231 members](#) representing 27 countries. arXiv's [2018-2022 sustainability plan](#) includes the following sources of funding:

- Cornell University provides a subsidy of \$170,000 and makes an in-kind contribution of all indirect costs, which currently represents 37% of total operating expenses.
- The [Simons Foundation](#) contributes \$100,000 per year in recognition of CUL's stewardship of arXiv. In addition, the Foundation matches \$300,000 per year of the funds generated through arXiv membership fees.
- Each member institution pledges a five-year funding commitment to support arXiv based on institutional usage ranking (the annual fees are set in six tiers from \$1,000-\$4,400).
- To diversify funding sources, we also rely on grant funds from foundations and governmental agencies to sponsor new development projects and implement online fundraising campaigns to encourage giving by the arXiv's international user community.

We remain grateful for the support from the Simons Foundation that has encouraged long-term community support by lowering arXiv membership fees and making participation affordable to a broader range of institutions. This model aims to ensure that the ultimate responsibility for sustaining arXiv remains with the research communities and institutions that benefit from the service most directly. Finding new, sustainable funding avenues to support current operations and upgrades of Classic arXiv's aging code has been an important effort for the arXiv team. Following a \$450,000 grant from the Sloan Foundation in 2017, arXiv received \$322,000 from the Heising-Simons Foundation to contribute to the arXiv-NG initiative. As reflected in the [2019 budget](#), the arXiv team is taking an integrated approach to consider the current operational system (Classic arXiv) and the next-gen system as a unified program. This approach is essential as we recruit and retain staff who will need to be conversant with the old and new systems, and transitioning from one to the other. We are strategically expanding the core arXiv development team to bring new skills to the arXiv-NG project, while continuing to provide excellent support and maintenance of the production arXiv system.

Key Accomplishments and Plans for 2019

Since we started the arXiv sustainability initiative in 2010, an integral part of our work has been assessing the services, technologies, standards, and policies that constitute arXiv. Here are some of our key accomplishments from 2018 to illustrate the range of issues we have been trying to tackle. Please see the [2018 Roadmap](#) for a full account of our work.

- Our development team has improved various search, browse, and accounts features as we [reimplement](#), test, [refine](#) and [continue to improve](#) the arXiv platform. We also made significant progress reimplementing the submission user interface, API, and backend services with a new data architecture. Wherever possible, new software components are developed in [public repositories](#) and released under permissive open source licenses. Information about software releases can be found [here](#).
- We continue to improve facilities for administrators and moderators in order to streamline their workflows, and to improve clarity and transparency of arXiv communications. During 2018, the arXiv team made improvements to the auto-endorsement system and created a moderator policy page. They also identified areas to improve help pages on user facing policies.
- We refined our overall process for prioritization and [getting things done](#) in arXiv-NG and continued to develop and implement methods for soliciting user input and feedback, particularly to guide arXiv-NG decision making, including a protocol for remote usability testing, and applied it to the development of a new moderation interface.
- Last July, arXiv added two new subject areas: [Theoretical Economics and General Economics](#)

- Our user support team updated our internal policies and practices documentation, filling in areas where policies have evolved. Our goal is to provide clear, consistent, established arXiv policies. Going forward we will be expanding and reorganizing policy documentation for moderators and submitters.
- With funding from the [Allen Institute for Artificial Intelligence](#) (AI2), we conducted a pilot to investigate the role of arXiv Labs in facilitating lightweight and more sustainable path for collaboration with Cornell Computing and Information Science through a [reference extraction and linking project](#).
- We explored the implications of the General Data Protection Regulation (GDPR) and started to work with the Cornell experts in supporting related questions and requests from arXiv users.
- We continued to assess and refine arXiv's organizational and governance model and held an annual meeting for the Scientific Advisory Board (SAB) and Member Advisory Board (MAB) to discuss IT development priorities, financial state, moderation tools and policies, and fundraising strategies.
- As we witness the proliferation of preprint services, we collaborated with bioRxiv and ASAPbio in hosting a meeting of nascent preprint services to start discussing collaboration opportunities.
- We strengthened the daily oversight of arXiv and the arXiv NG development process of the arXiv team with a modified [organizational structure](#) and welcomed new team members. We also established a new external Technology Advisory Group to advise the IT leads.

The [2019 Roadmap](#) includes our goals as we strive to improve the technical infrastructure, moderation system, user support, and the sustainability framework. Additional information about arXiv is available on [arXiv Public Wiki](#), including annual budgets, reports, roadmaps, and [arXiv.org blog](#). As it is evident from this update, changes are afoot for arXiv, all aimed at ensuring that the scientific scholarship repository remains free, open, essential and sustainable.

arXiv relies on a wide network of individual scientists who contribute their expertise and time. Paul Ginsparg, arXiv's founder, continues to make significant contributions through his involvement in the daily moderation process as well as developing and running some applications to help automate the quality control process. Composed of 165 subject experts, the worldwide network of arXiv moderators is an essential component of arXiv's success. They verify that submissions are topical and of interest to the scientific community, follow accepted standards of scholarly communication, and are classified in the appropriate subject categories. We salute their efforts that are crucial to our operation. We remain grateful for strong support from our member organizations, [Simons Foundation](#), and essential contributions from arXiv's advisory groups as they consistently provide us with input as representatives of scientific and library communities. We would also like to thank [Sloan Foundation](#), [Heising-Simons Foundation](#), and [Allen Institute for Artificial Intelligence](#) for their generous support of the arXiv-NG initiative.

arXiv Management Team

Oya Y. Rieger (Program Director), Steinn Sigurdsson (Scientific Director), Jim Entwold (Operations Manager), Martin Lessmeister (IT Lead), Janelle Morano (Community Engagement & Development), Erick Peirson (Lead Software Architect)

[Website: arXiv Public Wiki](#)

[Contact email: support@arXiv.org](#)

[Twitter: @arXiv](#)

[Updates: If you are interested in getting updates from the arXiv team and have not yet signed up for the mailing list, send an email message to: \[arxiv-support-updates-L-request@cornell.edu\]\(mailto:arxiv-support-updates-L-request@cornell.edu\). Leave the subject line blank and the body of the message should be a single word: join](#)