

arXiv Update - January 2017

During 2016, we witnessed another important milestone as [arXiv turned 25](#). 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. During 2016, the repository saw [113,380 new submissions and over 163 million downloads](#) from all over the world. arXiv has an international scope, with submissions and readership from around the world, and collaborations with U.S. and foreign professional societies and other international organizations.

From the users' perspective, arXiv continues to be a successful, prominent subject repository system serving the needs of many scientists around the world. However, under the hood, the service is facing significant pressures. The conclusion of the Scientific Advisory Board (SAB) and the Member Advisory Board (MAB) 2015 annual meetings was that, in addition to the current business model with a focus on maintenance, the arXiv team needs to embark on a significant fundraising effort, pursuing grants and collaborations. To accomplish this goal, we first focused on creating a compelling and coherent vision to be able to persuasively articulate our fund raising goals beyond the current sustainability plan that aims to support the baseline operation. We used the 25th anniversary of arXiv as an important milestone to engage us in a series of [vision-setting processes](#) including:

- Conducted a SAB/MAB survey to poll their opinions on future directions and priorities
- Implemented a user survey based on the initial findings of the vision-setting exercise ([survey results](#))
- Held a technical infrastructure workshop to brainstorm and refine a set of possible technology options for the re-implementation of all or parts of arXiv, and to provide a solid foundation for additional development
- Implemented a survey targeting arXiv's approximately 150 moderators to expand our understanding of their needs

Additional information: [arxivpub:arXiv@25 Review Strategy](#)

Key Findings of User Study

A user survey was conducted 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. 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. 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. 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.

The combination of multiple choice responses 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. 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, respondents 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.

Additional information: [arXiv@25: Key Findings of a User Survey](#)

Next Generation arXiv & Grant from the Sloan Foundation

Based on the conclusions of ten months of planning activities described above, we successfully secured a [\\$450,000 grant from the Sloan Foundation](#) to initiate the next-generation arXiv (arXiv-NG) initiative. The 18-month project will enable us to plan and begin to improve the architecture of this critical scientific infrastructure. We anticipate that a multi-phase design and development of a next-generation arXiv (arXiv-NG) will require approximately 3 years with an additional \$2+ million budget, including requirements specification, evaluation of alternate strategies and partnerships, design of a new system architecture, assessment, and deployment. As we are developing arXiv-NG, during the next three years we'll continue to rely on the existing system (Classic arXiv) and are committed to continue its robust services. Additional information: [Next Generation arXiv](#)

Another source of funding in support of arXiv-NG came from the [Allen Institute for Artificial Intelligence](#) (AI2). An initial gift of \$200,000 for 2017 will support a collaboration between the Cornell University Library and the Cornell Computing and Information Science (CIS). This donation will allow us to hire a Research & Innovation Fellow to collaborate with the arXiv team in designing and integrating a series of updated, research-oriented features for arXiv. The ultimate goal is integrating tools emerging from research into the production system to improve user and moderator experience.

arXiv's funding and governance for the current operation (Classic arXiv) is based on a membership program that engages libraries and research laboratories worldwide that represent the repository's heaviest institutional users. Currently, we have 201 [members](#) representing 25 countries. arXiv's sustainability plan is founded on and presents a [business model](#) for generating revenues. Cornell University Library (CUL), the [Simons Foundation](#), and a global collective of institutional members support arXiv financially. The [financial model for 2013-2017](#) entails three sources of revenues:

- CUL provides a cash subsidy of \$75,000 per year in support of arXiv's operational costs. In addition, CUL 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 (\$50,000 prior to 2016) 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 four tiers from \$1,500-\$3,000.

In 2016, Cornell raised approximately \$515,000 through membership fees from 201 institutions and the total revenue (including CUL, Simons Foundation direct contributions, and online fund raising) is around \$1,015,000. We remain grateful for the support from the Simons Foundation that 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. One of the strategies has been encouraging libraries and research labs around the world to support arXiv by becoming members. In addition, after a successful pilot in 2015, we repeated the online fund raising campaign during 2016 and raised an additional \$32,000. We intend to continue our online fund raising efforts through a couple of calls each year. Seeking to increase the representation of Indian institutions among arXiv's membership, Anurag Acharya, the chief scientist behind Google Scholar, donated membership fees on behalf of three Indian libraries. This model, in which sponsors provide membership fees for institutions which can't afford them, is one which we hope to expand.

As 2017 is the last year of the current [5-year business model](#), working with MAB, the arXiv team has started the 2018-2020 business planning process. We are anticipating some minor changes to the existing model including a modest increase to annual membership fees to accommodate for increasing costs (as the rates have been flat during the last five years), adding a new top tier for institutions that are heavy users of arXiv, and using an average based on last several years' usage in calculating annual tiers. We plan to share the 2018-2022 business model the latest by March 2017 to allow sufficient time for getting feedback and preparing member libraries for 2018 fees.

Classic arXiv: Key Accomplishments and Plans for 2017

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 2016 to illustrate the range of issues we have been trying to tackle. Please see the [2016 Roadmap](#) for a fuller account of our work.

- The worldwide network of arXiv moderators--over 150 subject experts - are an essential contributors to arXiv's success as 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. The focus of technical work in 2016 was improvement of facilities for administrators and moderators in order to streamline their workflows, and to improve clarity and transparency of arXiv communications. A new set of features was added to the admin interfaces that allow the admins to work with category proposals on submissions.
- 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 fund raising strategies.
- We continued following the new developments in regard to public access mandates and related compliance issues to understand arXiv's role vis-à-vis the evolving scholarly communication ecology. We completed a preliminary assessment of how we can enable one-stop deposit and compliance reporting in support of new public access requirements such as the one from the UK ([HEFCE/REF](#)). One of the goals of the arXiv-NG initiative is to create an open architecture to allow other service providers to develop APIs as overlay services, for instance to simplify scientists' publishing process by enabling submission of an article to a journal at the same time as it is uploaded to arXiv.
- After Chris Myers's (interim Scientific Director) departure in April 2016, we've started to assess the position and are in the process of revising the job description. Dr. Myers was instrumental in evaluating the position and proving useful feedback to help us refine the set of responsibilities and expectations. We aim to fill the position in early 2017.
- On the staffing front, in order to strengthen the daily oversight of arXiv, during 2016 we transitioned to a team model that has two full-time managers, one for IT (Martin Lessmeister, Lead Programmer) and the other one for user support (Jim Entwood, Operations Manager). As Simeon Warner, who is a long-term contributor to arXiv, shifted his attention to other CUL projects, we appointed Sandy Payette as arXiv's first CTO. Our team is further strengthened with the addition of Chloe McLaren (Membership Program Coordinator) and Gail Steinhart (Program Associate). Here is the current team model: [arXiv Sustainability Initiative](#)

The [2017 roadmap](#) includes our goals within the scope of the current business model with a focus on Classic arXiv. Additional information about arXiv's sustainability model is available on [arXiv Public Wiki](#), including budgets, annual reports, roadmaps, and other relevant information. Throughout the year, Gail Steinhart, Coordinator of arXiv-NG, is planning to keep you informed and provide updates on progress.

As it is evident from this brief update, changes are afoot for arXiv, including new funding sources and a planning for a three-year overhaul, all aimed at ensuring that the scientific scholarship repository remains free, open, essential and sustainable. 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.

Cornell University Library, arXiv Team

Oya Y. Rieger (Program Director), Jim Entwood (Operations Manager), Chloe McLaren (Membership Program Coordinator) Sandy Payette (CTO), Gail Steinhart (Program Associate)

Contact email: support@arxiv.org

[arXiv Public Wiki](#)

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. Leave the subject line blank and the body of the message should be a single word: *join*