2018-2022: Sustainability Plan for Classic arXiv

Introduction to arXiv

Started in August 1991 by Paul Ginsparg, arXiv.org is internationally acknowledged as a pioneering digital archive and open access distribution service for research articles. The e-print repository, which moved to the Cornell University in 2001, has transformed the scholarly communication infrastructure of multiple fields of physics and plays an increasingly prominent role in mathematics, computer science, quantitative biology, quantitative finance, and statistics. arXiv is an essential component of scientific communication for many researchers worldwide in order to rapidly and widely disseminate their findings, establish priority of their discoveries, and seek feedback to help improve their work. 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. Please see the recent arXiv update for additional information.

The sustainability model lays out a business model for arXiv including anticipated expenses, potential revenue streams, value propositions, and communication strategies. The plan entails only the regular operation of arXiv—in other words, what we call "keeping the lights on." It should be seen as a baseline operational budget, as it does not factor in additional expenses required for R&D or new development projects such as the next-gen arXiv (arXiv-NG) initiative.

What is arXiv's Financial Model for 2018-2022?

Since 2010, Cornell's sustainability planning initiative has aimed to reduce arXiv's financial burden and dependence on a single institution, instead creating a broad-based, community-supported resource. 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 broad 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.arXiv's funding and governance for the current operation (Classic arXiv) is based on a membership program engaging libraries and research laboratories worldwide that represent the repository's heaviest institutional users. The financial model for 2018–2022 entails four sources of revenues:

- Cornell provides a cash subsidy of $170,000 per year in support of arXiv's operational costs (including the Program Director salary and benefits). In addition, Cornell 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 Cornell'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 to $4,400.
- Grants funds from foundations and agencies (e.g., Sloan Foundation and Heising-Simons Foundation) to assist with special projects such as the arXiv-NG.

What is arXiv-NG, and how is it funded?

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. With significant input from the Scientific Advisory Board (SAB) and the Member Advisory Board (MAB), during the 2015 annual meetings, the arXiv team reached the conclusion 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. Based on the conclusions of 10 months of planning activities in 2016, the arXiv team has 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 three 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 develop 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.

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 and Cornell Computing and Information Science (CIS). This donation will allow us to hire a Research and 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 to integrate tools emerging from research into the production system to improve user and moderator experience.

In addition to supporting Classic arXiv through the membership model, Cornell University will continue to raise funds through grant proposals to agencies and foundations in order to fund the arXiv-NG initiative. Also, there will be some reliance on development reserve funds as needed. Please see the recent arXiv update for additional information.

What is the tier model for 2018–2022?

Tier	Previous Model Fee	Current Fee
Tier 1: 1–25	$3,000	$4,400
Tier 2: 26–50	$3,000	$3,800
Tier 3: 51–100	$2,500	$3,200
Tier 4: 101– 150	$2,000	$2,500
Tier 5: 151–200	$1,500	$1,800
Tier 6: 201+	$1,500	$1,000

Note: The goal of Tier 6 is to provide an equitable and affordable fee to allow participation. Although we focus our membership drive on the top 200 institutions, we would like to encourage other libraries to contribute.

What is the purpose of the membership tier and fee revision?

arXiv has not adjusted its member fees, which currently represent about 45% of its operating income, since they were initially set in 2012. The rationale for increasing the arXiv membership fees is as follows:

- Given the erosive effect of inflation, a $3,000 fee in 2012 is worth $2,587 in 2016. In other words, total arXiv member fees are worth about 85% of their 2012 value.
- arXiv's user support and development costs have increased due to the incorporation of unforeseen or underestimated expenses, including staffing. The new lines include Scientific Director (vacant, 0.4 FTE), CTO (0.30 FTE), paid moderation for physics.gen-ph, and 1.5 FTE additional developer time (including UX; see arXiv 2016 organization chart).
- The operational cost upsurge reflects the increase in user submissions and the proportional increase in workload for administrators. As illustrated below, the annual submissions since 2012 have increased by 35%. During 2012–2016, the number of submissions grew from 84,000 to 113,400.
- The new tier and fee structure aims to be more equitable as the number of downloads from the top 20 institutions corresponds to 22% of the total downloads, while the sum of their contributions corresponds to less than 10% of all member contributions.
- In addition to the Cornell's indirect contributions, the University"s direct financial contribution will increase to $170,000, including the salary of Program Director (.40 FTE), which is not considered an indirect expense line.
- To simplify the application of the fee structure for both member institutions and for arXiv, institutional tier rankings will be assigned based on a three-year rolling average of use (institutional downloads).
- Without an increase in the member fees, arXiv would run a deficit of $450,000 for 2018–2022, an average of $90,000 per year. If Cornell is unable or unwilling to absorb increased operating deficits, arXiv support and development spending will need to be cut or the service will need to be transitioned to a new host.

What is the value proposition of the arXiv Institutional Membership Program for libraries?

User feedback tells us that arXiv continues to be the most successful and prominent subject repository system serving the needs of scientists worldwide. 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 (see the survey findings As you assess the value proposition of supporting arXiv, please bear in mind that a Tier 2 contribution of $3,800 is roughly equivalent to the subscription cost of one highly ranked physics journal or one APC charge in a journal like Nature Communications (with an APC of $5,200), New Journal of Physics (with an APC of $2,080), Physical Review X (with an APC of $2,900), or Journal of Physics D: Applied Physics(with an annual subscription cost of $5,740). arXiv ranks #1 in the Ranking Web of Repositories and arXiv sections rank as five of the top ten publications in Google Scholar's Top Publications in Physics and Mathematics . Aside from the knowledge that you are part of a shared investment in a culturally embedded resource providing unambiguous value to a global network of scientific researchers, here are some other reasons we would like you to become members:

- Your membership fees are essential to garner the $300,000 per year matching funds from the Simons Foundation. These funds encourage long-term community support by keeping arXiv membership fees low, making participation affordable to a broader range of institutions.
- You will be able to participate in arXiv's ongoing governance through the Member Advisory Board, which provides input for project prioritization, new service offerings, financial planning, use of discretionary funds, future technical developments, and policy decisions.
- You will have access to enhanced institutional use statistics.
- Each member library will receive public acknowledgement of their financial support both on arXiv.org and locally to the institution via an IP-generated banner. We would like to expand our strategies to make sure that scientists are aware of the role of libraries in ensuring arXiv's sustainability. Please send us your ideas to support@arxiv.org!
- Greater integration with, and implementation of, Open Access scholarship as the arXiv team works closely with related and complementary initiatives such as SCOAP3, Open Archives Initiative (OAI) , ORCiD, Center for Open Science, and Confederation of Open Access Repositories (COAR).

Have you considered basing the membership model on submissions instead of downloads?

One of the reasons for using download statistics as the basis of the arXiv tier model is the challenges associated with tracking submissions from a particular institution to arXiv. The author metadata arXiv collects is not sufficiently consistent to support systematic submission analysis. Manual analysis of a single month of submissions in 2013 indicated a similar ranking in submission- and download-based data. During the arXiv-NG initiative, arXiv expects to undertake metadata remediation to improve the authorship data for existing submissions (including full ORCID implementation). In the current version of arXiv, it would be difficult to support a reliable submission-based tier model. While we do require submitters to register their institutional affiliation when they register an account with us, and most submitters are authors in the papers they submit (though not necessarily first authors), the relationship between the submitter and the submission's author metadata is still rather weak. We allow submitters to supply affiliation information in the author metadata, but it's neither enforced nor widely used. This makes it especially difficult to report on the institutional affiliation of co-authors. In order to implement a submission-based tier model, we will need to expand arXiv's data model to better integrate with ORCIDs to start. In some cases (like large collaborations), it would still be difficult to expect submitters to provide complete authorship and affiliation data; here it would be helpful have tools to extract this information from the full-text automatically.

What are the expenses and revenue sources for 2018–2022?

The draft budget factors in support only for the routine, daily arXiv operation. It does not include resources needed for significant improvements, emergency interventions (server failure), special projects, or research and innovation. This budget does not consider the nascent arXiv-NG initiative as the new development initiative needs to be supported by an additional fund raising effort. As we transition from Classic arXiv (current system) to arXiv-NG in approximately 2020, the annual expenses will need to be adjusted, taking into consideration the staffing needs of the new architecture.

arXiv Summary Budget	2017	2018	2019	2020	2021	2022
Revenue						
Member contributions	$400,000	$500,400	$500,400	$500,400	$500,400	$500,400
Simons Foundation Annual Commitment	$100,000	$100,000	$100,000	$100,000	$100,000	$100,000
Simons Foundation Matching Fund	$300,000	$300,000	$300,000	$300,000	$300,000	$300,000
Cornell Direct Contribution & Leadership [1]	$75,000	$170,000	$170,000	$175,000	$175,000	$180,000
Online donation program & gifts	$50,000	$50,000	$50,000	$50,000	$50,000	$50,000
Total Revenue	$925,000	$1,120,400	$1,120,400	$1,125,400	$1,125,400	$1,130,400
Expenses						
Direct Expenses	$925,000	$1,044,600	$1,066,103	$1,097,551	$1,124,792	$1,157,851
Indirect Expenses						
College and dept administration and staff support [2]	$240,500	$271,596	$277,187	$285,363	$292,446	$301,041
Facilities [3]	$101,750	$114,906	$117,271	$120,731	$123,727	$127,364
Total Indirect Expenses	$342,250	$386,502	$394,458	$406,094	$416,173	$428,405
Contribution to Reserve	$0	$75,800	$54,298	$27,849	$608	-$27,451

Notes:

1. In 2018, Cornell's direct financial contribution will increase from $75,000 to $170,000, including the salary of Program Director (.40 FTE) as a bridging strategy until we identify additional revenue sources. The MAB recommended that the Program Director's salary be considered a direct cost due to its prominent and necessary role. Prior to 2018 this position has been included with indirect costs.
2. University-negotiated indirect cost rate for administration & staff support (HR, finance, facility & IT); based on direct costs (less Cornell in-kind staff contribution) at 26%.
3. University-negotiated indirect cost rate for maintenance, custodial & other facility-related costs; based on direct costs (less Cornell in-kind staff contribution) at 11%.

What is the staffing configuration for Classic arXiv?

Position Description	Staff	2018	2019	2020	2021	2022
User support (management + staff + student)	1FTE Entwood, 1FTE Weiskoff, 1FTE Salguero, 0.5FTE Goldweber, 0.6FTE students	4.1	4.1	4.1	4.1	4.1
IT (management + staff)	1FTE Lessmeister, 1FTE Caruso, 0.5FTE Fielding, 0.5FTE Perez, 0.50FTE Barker (CAC)	3.5	3.5	3.5	3.5	3.5
Leadership: Program Director , Scientific Director, CTO	0.4FTE Rieger , 0.4FTE Scientific Director, 0.3FTE Payette	1.1	1.1	1.1	1.1	1.1

		0.45	0.45	0.45	0.45	0.45
Total FTE:		9.15	9.15	9.15	9.15	9.15

Please see the recent arXiv update for up-to-date information.

What funds are generated through the online giving campaigns?

arXiv has run three week-long online giving campaigns to-date: the first campaign in September 2015 generated about $16,200 in donations, the second in June–July 2016 about $31,475 in donations, and a campaign in February-March 2017 generated around $31,700. We aim to raise at least $50,000 per year through online giving campaigns. A secondary goal of the online giving campaigns is to remind arXiv users that although arXiv.org is free to use, it is not free to produce.