



## 2022 SCOSS ANNUAL REPORT FORM

NAME OF THE REVIEWED INFRASTRUCTURE: arXiv

APPLICATION FORM: [arXiv Application Form](#)

YEAR IN THE CYCLE:

1/3

2/3

3/3

1. Discuss the status of the most important activities and milestones planned as specified in the Application Form (see section 6 of the Application Form). If there are important omissions or milestones that have not been achieved, please explain in the Additional comments section below.

You can keep the structure of your application or organize the milestones in main fields, such as Governance and Sustainability”, “Community Development”, “Technical Development”, “Pledges and Grants”.

### Activity/Milestone: (Fundraising and Sustainability)

Status: Through the work of our Community Engagement Manager, arXiv was able to successfully maintain relationships with and ensure continued membership and sponsorship from 178 institutions and organizations in 2022, while significantly increasing funding through pledges facilitated by our relationship with SCOSS. In year 1, arXiv received funding commitments via SCOSS from 62 institutions for a total of \$381,184 (€336,077) with \$122,504 in payments being recognized in 2022. This included memberships coming through relationships with consortia in Canada, France, and Finland.

In addition to our collaboration with SCOSS, our fundraising was successful due to:

- A strong, well-communicated strategy
- Membership and other outreach, including the new fee structure based on submissions
- Improved internal workflows

arXiv’s fundraising activities are not limited to library and consortia membership, but also include an affiliates program (professional societies and other organizations), a sponsorship program (corporate entities), and contributions from individuals. In 2022, arXiv held two “Giving Weeks,” organized and managed by the Community Engagement Manager, which yielded contributions from nearly 1,400 individuals totaling \$48,169.

Throughout the year, the Community Engagement Manager provided information to and had conversations with prospective and existing members, in addition to:

[2022 SCOSS Annual Report](#)



- hosting a membership meeting in June 2022
- organizing Open Access Week activities
  - o Panel Discussion: Trends in Peer Review of Open Access Preprints
  - o Lightning Talk: Making Research Papers Available to All
  - o promotion of lightning talk held at an arXiv member institution: Open for Climate Justice at The Ohio State University
  - o one of the Giving Weeks
  - o Twitter celebration
- creating and distributing quarterly membership emails □ preparing the arXiv Annual Report

### Activity/Milestone 2 (Cloud migration, Microservice infrastructure for submission and moderation)

Status:

Leadership and other staffing changes in 2021/2022 limited arXiv's ability to make significant progress towards the cloud migration and related system enhancements in 2022. However, following a roughly three-month gap in technical leadership, arXiv's new Technical Director joined the staff in mid-June 2022 and has reinvigorated the development team and articulated a plan for our technology modernization. In addition to stabilizing the existing system, an in-depth plan for arXiv's cloud migration is now in place and progress has been made on a number of fronts to set the stage for a fully cloud-based system:

- DNS service for arXiv is now entirely hosted in Google Cloud Platform (GCP)
- Our help system has been moved to GCP (<https://info.arxiv.org>), which will make it easier to substantially updating certain sections
- arXiv's main search function is hosted in GCP
- GCP hosts a near-real time clone of the main arXiv database housing all of our article metadata, and all associated article content
- The above has enabled us to implement essentially all the read-only portions of arXiv on GCP. We are now beta testing the arXiv home page, articles listings, abstract pages, and article downloads on GCP. We have already diverted a small percentage of our production downloads to GCP for test and evaluation purposes.

While making headway on our cloud migration, we have also continued to work on new user interfaces for volunteer moderators and the arXiv admin team. The moderator interface is in use and awaiting enhancements; as is, it has streamlined the daily tasks of moderators and improved the ability to coordinate their work. A working prototype of a similar interface for the arXiv administrative team has been developed, and we hope to release a mini-MVP to the team by the end of the first quarter of 2023.

Significantly, in October 2022, arXiv added a machine learning (ML) engineer to the staff who has been tasked with investigating and documenting the manual and automated processes that drive arXiv's submission and moderation systems and contribute to the daily



work of volunteers and staff alike. In the coming year, the ML engineer will play a key role as we enhance the interfaces described above by overhauling Q/A functionality, including automating many tasks that are currently done manually.

In addition, we took steps to make our web site and our help pages more accessible to people with visual impairments. We conducted considerable research on making our content available to people with impairments (see [Accessibility Report - arXiv info](#)), leading to a redoubling of our partnership with [ar5iv – Articles from arXiv.org as responsive HTML5 web documents](#) as responsive HTML5 web documents project to make all arXiv articles available via HTML.

As submissions and usage continue to rise (arXiv receives 15,500 submissions and handles 32 million downloads each month), it is increasingly vital that we improve stability and capacity, expand into new subject categories, and allow for easier integration with other services. In 2022 we were able to lay the groundwork that will ensure that arXiv is operating a best-in-class system that will serve its many stakeholders well in the years ahead.

### [Activity/Milestone 3 \(Product and Partnership management\)](#)

arXiv is happy to report that we have made meaningful progress in 2022 related to our ability to engage with others in the community.

A planned DOI implementation moved forward in January 2022, starting with every new arXiv paper being assigned a DOI through Datacite. Over the course of the year, we were able to register DOIs for all articles in the arXiv corpus, and our hope is that in the coming year we will enable DOIs for all versions of a preprint in arXiv.

While the ability to submit to our arXivLabs program (a framework for integrating community-built tools into the arXiv platform) was temporarily suspended in early 2022 due to our staffing challenges, we continued to roll out integrations throughout the year. Happily, in late 2022 we reopened our proposal form so that anyone can suggest an integration. [ArXivLabs](#) now features 11 integrations, 5 of which were added in 2022:

- ar5iv—converts arXiv articles to responsive HTML5
- Connected Papers—provides links to interactive visualizations of similar articles
- Hugging Face—provides links to open-source machine learning demos
- Iarxiv—allows readers to receive, by email, daily arXiv papers in several physics categories sorted by artificial intelligence according to their preferences
- Science Cast—links researchers’ videos to their arXiv papers

In 2022, arXiv continued its collaboration with Clarivate as they worked to create the Preprint Citation Index. While arXiv’s metadata and articles are available to anyone through our APIs, by coordinating with Clarivate as they developed their service, we were able to ensure that they could successfully import arXiv data without encountering issues or impacting our operations.



With the hiring of a full-time Program Director in fall of 2022, arXiv is now better situated to establish and maintain relationships with infrastructure and other organizations in the open access/open science space (e.g., CHORUS, Core, etc.) as well as representing arXiv at conferences and participating in webinars and panels.