

2017 arXiv Roadmap

The arXiv roadmap is a living document and communication tool to accommodate continuous prioritization throughout the year. Items are listed in approximate priority order, subject to change based upon consideration of input from arXiv stakeholders, assessment of new opportunities and initiatives as they arise, and progress on next generation arXiv development.

Technical

Moderator Web UI: Implement enhancements to the existing moderation Web UI to improve the proposal system based on moderator feedback. Also, enable moderators to take multiple actions and commit them together as a set of actions. Develop a new alternate moderator user experience to offer an efficient way for moderators to work on multiple submissions on a single responsive web page. The alternate moderator user experience will provide new paradigm for category management, improved proposal features, and an improved notification scheme to decrease email volume for moderators.

Status: Initial release June 2017; with subsequent minor improvements.

Moderation API: Define and implement a new API that provides a service endpoint for all essential operations necessary to support the moderation workflow. The API will be brought forward in to the new arXiv-NG architecture with the intention of it being able to support new applications and user interfaces in the future to support moderation in arXiv-NG.

Status: Completed.

Upgrade database infrastructure: This is a system operations priority.

Status: provisioning and initial configuration completed. Final configuration and tuning in progress with planned switchover in early 2018.

PHP/tapir to PERL/Catalyst: Complete the conversion of PHP code to provide modularity, consistency, and improved maintainability of key parts of arXiv codebase. Migrate functions away from old PHP/Tapir codebase and into Perl/Catalyst.

Status: Retired to reallocate effort towards reimplementing as Python/Flask application.

Improve submission error checking: Analyze submission process to identify improvements and minimize submission errors for authors.

Status: Analysis of existing system complete; re-implementation of submission in Python/Flask is in progress. Scope work for improving error handling and display not started.

Author notification of reclassification: Add new features to provide authors with notification about reclassification of their paper. Notifications will provide information about follow-up action that is required (i.e., what action in what timeframe).

Status: Initial specification in progress; implementation deferred to 2018 pending final policy decisions.

Auto-endorsement: Shift the auto-endorsement rules to work of a single white-list of recognized email domains. Expand the list of email domains that qualify for auto-endorsement. Develop an admin interface to update the auto-endorsement list.

Status: White list complete; implementation not started.

Submission system security audit: This is a systems operations priority.

Status: Completed for current codebase.

EESS: (Electrical Engineering and Systems Science) domain expansion

Status: Complete. Launched with 3 subcategories on 18 Sept 2017.

Econ: (Economics) domain expansion

Status: Complete. Launched with one subcategory on 26 Sept 2017.

Reference extraction: (promoted from Special Projects) Automatically extract references from papers.

Status: Implemented. Extraction for all new papers. (see [blog entry](#))

Reference linking: (promoted from Special Projects): Display linked references to users.

Status: API for reference links implemented. Display implementation deferred, pending consideration of user feedback and Scientific Director input. (see [blog entry](#))

Search: (promoted from Special Projects): Combined metadata and full-text search using Elasticsearch and Python/Flask as a standalone application.

Status: In progress.

Browse: Redeveloping abstract page and subject listing display (home page). Initial focus on abstract pages; to be implemented in Python/Flask.

Status: In progress.

Log Analysis: Redevelop general log analysis code, including annual institutional usage statistics; to be implemented in Python and AWS.

Status: Introduced late 2017; in progress.

Submission API: ACM submission special project evolved into multi-stakeholder initiative to design a new submission API in NG.

Status: Introduced late 2017; in progress.

NB: approximately 20% of developer effort is devoted to a range of activities that ensure that the service continues to run smoothly; this effort is variable from week to week and is not necessarily scheduled or reflected by discrete tickets in our task management system. Activities include (but are not limited to) the following:

- server maintenance and troubleshooting
- service monitoring
- analytics
- internal support requests
- ad hoc meetings and conferences

Given the relatively small size of the development team, which does not currently have a dedicated dev/ops engineer, every team member is responsible for some amount of operational work.

User Support & Moderation

Author curated links: Based on the findings of the [arXiv user study](#), improving support for linking research data, code, slides and other materials associated with papers emerged as an important service to expand.

Status: Not started; to be implemented in Python/Flask.

Publish physics category descriptions: Finalize input from moderators and publish the descriptions on [arXiv.org](#). This work started in 2016 and we hope to complete early 2017.

Status: In progress; awaiting review.

Overlap policy: Implement the SAB Sept 2014 decision that cases of self-overlap get bounced to users before publishing.

Status: Not started. Dependent on submission and overlap tools currently in development.

Update Help pages: Assess which section of the Help pages to update in 2017 (submission, moderation, reading/downloading, etc). Rewrite those sections. Identify straggling pages and pages that could be slated for removal.

Status: Completed for 2017.

Systematize arXiv's individual archive pages: The "home" pages for individual archives within arXiv have grown organically over time and are inconsistent with each other and in some cases outdated.

Status: Deferred for more systematic site map review and development process.

Process Improvement

Stakeholder communication improvements:

- Improve practices for prioritizing and ensuring that requests do not get lost in the action or become stale.

Status: Implemented quarterly Roadmap and backlog reviews.

Develop and implement stakeholder engagement plan:

- Plan and implement regular communications for various stakeholder groups.

Status: Implemented quarterly stakeholder reports; revamped arXiv blog and distributed quarterly blog digests; established twitter account.

Develop and implement user engagement plan:

- Develop and implement methods for soliciting user input and feedback, particularly to guide arXiv-NG decision-making.

Status: Implemented arXiv-usability-I email list to announce polls/surveys and solicit subjects for usability testing. Conducted two surveys to inform design decisions. Developed protocol for remote usability testing and applied to development of new moderation interface.

Continuous prioritization of development priorities:

- Modify the planning process so that the Roadmap becomes living document that can incorporate changes on a quarterly basis. Evolve development team tactical process to manage backlog and support tickets effectively and enable continuous evaluation of priorities and resources to maximize impact.

Status: Implemented.

- Coordinate the communication between the SAB, the arXiv development and admin teams to ensure a transparent process that engages key stakeholders in a new prioritization process.

Status: Implemented with the development of the 2017 Roadmap.

arXiv Workflow and Moderation analysis: Embark on a fresh analysis of workflows for moderators, appellate moderators, and Chairs. Identify and implement process improvements. This will inform both classic arXiv and arXiv-NG. Improve communication to moderators regarding system updates, process improvement, guidelines for moderation, and recruitment. Expand documentation on roles, relationships, and workflows between moderators, appellate moderators, committees, and Chairs.

Status: Expanded internal documentation for moderators regarding specific policies. Integrated classifier tool into the proposal system. Conducted design sessions around workflows and interfaces to inform future improvements.

Business Model, Governance, and Partnership

Recruit a Scientific Director: In 2014, we created the Scientific Director position to provide intellectual leadership for arXiv and appointed Dr. Chris Myers as the interim Scientific Director to test and refine the role. After Chris Myers's departure in April 2016, we have started to revise the job description with the goal of filling the position in 2017. The Scientific Director (.5 FTE) will collaborate with the arXiv staff and the arXiv's Scientific Advisory Board in providing intellectual leadership for arXiv's operations.

Status: Completed. After Chris Myers' departure in April 2016 as the interim Scientific Director, we took a few months to review the original job description and expectations and Chris's assessment of how the position played out in reality. Accordingly, a revised job description was developed and announced. We are pleased to announce that Professor Steinn Sigurdsson accepted the position and will start in September, working remotely from Penn State. Coordinated by Oya Rieger, the search committee included Dave Morrison (SAB), Carol Hoover (MAB/SAB liaison), and Eberhard Bodenchatz.

Continue the membership drive & identify new funding sources: Ensuring a broad international network of supporters requires an ongoing effort in order to develop new outreach strategies and revenue sources. We will continue our efforts in increasing awareness about sustainability issues and the arXiv business plan. Also, we will raise additional revenue streams through online fund raising campaigns and the promotion of the new sponsorship program.

Status: In progress with successful online fund raising efforts.

Create a business model for 2018-2022: This is the final year of our [5-year business plan](#). We have started planning for the next 5-years during 2016 and will finalize and announce the new model, including a revised tier structure.

Status: The new business model was finalized and rolled out in Spring 2017: [2018-2022: Sustainability Plan for Classic arXiv](#)

Raise funds to support arXiv-NG project: In addition to raising funds to support the current arXiv operation, during 2016, we successfully secured two grants in the amount of \$650,000 to initiate the next-generation arXiv initiative, which is envisioned to be a 3-year project with a total budget of \$2+ million. Therefore it is essential that we continue writing grants to raise the necessary funds for successful and timely completion of the project. Click for more information on arXiv-NG initiative.

Status: In progress - proposal submitted to Heising-Simons Foundation

arXiv Next Generation (arXiv-NG)

Based on the conclusions of ten months of planning activities during 2016, the arXiv team 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 will 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 to integrate tools that emerge from research into the production system to improve user and moderator experience.

Special Projects (under consideration)

A number of special projects from previous roadmaps are being incorporated into next generation arXiv planning and development.

The current 5-year business plan represents a baseline maintenance scenario. It was developed based on an analysis of arXiv's baseline expenses during 2010-2012. It does not factor in any new functionality requirements or other unforeseen resource needs. Although a development reserve was established to fund such expenses, it is not sufficient to subsidize significant development efforts through surplus funds.

ACM preprint loading into arXiv and submission API collaboration

Selected interoperability initiatives

- Improve ORCID integration
- Reference linking (promoted to User Support & Moderation)

Other initiatives

- Mobile app for moderation