2018 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.

Beginning with this roadmap (2018) we are treating arXiv-NG and "classic" arXiv development, business planning and governance as a single, unified effort. For technical development, some parts of arXiv will be rebuilt anew, while others will be transitioned in stages (utilizing "shims") to new infrastructure.

Much of the work planned for 2018 is essential infrastructure that is required in order to rebuild arXiv. *Many tasks involve dependencies and cannot be reprioritized if the overall effort is to move forward.*

Related documents: arXiv software releases || arXiv development blog posts

Technical

Switch over to New Database Infrastructure: This is a system operations priority.

> **Status:** Completed January 2018.

Author Notification of Reclassification: Add new features to provide authors with notification about reclassification of their paper.

> **Status:** Initial specification complete; policy questions resolved. Will be integrated into submission development.

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. Deferred for integration into development of user management system/interface.

Infrastructure: New infrastructure setup and configuration to support deployment of arXiv-NG services in the cloud.

- **Centralized Logging Infrastructure:** classic and NG systems deposit access and application logs in a centralized log store.

  > **Status:** Beta.

- **Cloud Deployment Infrastructure:** implementation of service deployment and networking infrastructure.

  > **Status:** Beta.

- **Backup & Recovery** finalization of backup and recovery strategy for NG.

  > **Status:** In progress.

Submission System: Implements user interfaces and APIs for submissions in the NG architecture.

- **Submission UI:** the submission UI is replaced with a Flask application that runs on the existing CUL web servers. Includes allowing users to view and rewrite notices of self-overlap prior to publish.

  > **Status:** Alpha.

- **Essential utility services:** isolating essential processes from the classic system. Includes fulltext extraction, classifier, overlap detection, upload sanitization, TeX/compilation service, etc. This lays the foundation for isolating the submission system as a whole from the legacy codebase.

  > **Status:** In progress.

- **Submission API:** a new interface to support programmatic submission of papers. This is part of the broader automated submission collaboration with partner organizations. May involve implementation of forthcoming version 3 of the SWORD protocol.

  > **Status:** In progress.

Accounts & Authorization: Implements mechanisms for user registration, authentication, and authorization in the NG architecture.

- **Login & distributed sessions:** authentication views are handled by a new accounts service running on CUL servers, integrated with a distributed session store. This makes it possible to deploy new services with protected views in the cloud.

  Status: Complete.

- **Authentication service:** implementation of a service for managing user and API client authentication in NG services.

  Status: Deployed as alpha.

- **Registration & profile management:** accounts service is extended to replace account registration views, including password reset.

  Status: In progress.

- **OAuth2 support:** accounts service supports OAuth2, which allows arXiv users to authorize API clients to perform actions on their behalf. Necessary for submission API.

  Status: Deployed as alpha.

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

- **Abstract page:** a new Flask application running on the existing CUL web servers replaces the abs view, with minimal design changes. This facilitates decoupling elements of display from the classic system.

  Status: Completed.

- **Additional browse views:** classification views (/archive, /list, /year) are provided by the browse application.

  Status: In progress.

APIs & API Gateway: Provides an integrated access point for programmatic consumption of arXiv content.

- **Public API gateway:** a public API gateway is available. arXiv users are able to register their application and receive an API token, which is used to access API resources. At this stage, we will expose search API, fulltext content, and automatically extracted cited references.

  Status: Alpha together with Search API.

- **RSS feed:** RSS feed is reimplemented as a standalone service, available via the API gateway. Deprecates the classic RSS feed.

  Status: In progress.

Access & Discovery: Search: Combined metadata and full-text search using Elasticsearch and Python/Flask as a standalone application.

- **Replace legacy search:** reimplementation of the existing search views (/find), including advanced search functionality, in a new search service backed by Elasticsearch.

  Status: Complete. First version released April 2018; ongoing improvements in progress.

- **REST API:** provides a search-backed RESTful JSON API for arXiv papers via the API gateway.

  Status: Deployed as alpha.

- **Classic authority/ORCID views:** search application has new views that replace the authority record (https://arxiv.org/a/warner_s_1.html) and ORCID (https://arxiv.org/a/0000-0002-7970-7855.html) views in the classic system. Integrated with the author name authority service.

  Status: In progress.

- **Faceted search:** the search interface is extended with facets, providing the ability to refine search results based on metadata.

  Status: Deferred to 2019.

- **Full text search:** support for full-text search is well documented, and improvements are made to the UI to give users' visibility and control over this functionality.

    Status: *Deferred to 2019.*

Publication: Services to support the publication process and core metadata to ensure that arXiv papers and metadata are protected from loss, and that they remain uncorrupted and accessible.

- **Author name authority service:** transforms the existing name authority mechanism into a more robust name authority service that provides URIs for unique author identities, supports mappings between author identities and external identifiers (e.g. ORCID), and provides mechanisms for mapping duplicate entries onto each other.

    Status: *In progress.*

- **Institution authority records:** name authority service is extended to support an institution record type, and back-filled with affiliation information already present in the classic database.

    Status: *Deferred to 2019.*

- **Email notifications:** authenticated users can configure their e-mail subscriptions in an announcement service.

    Status: *Deferred to 2019.*

Enhancement & Discovery: Metadata: Backend support for peripheral metadata, including relations to external resources (e.g. journal articles, datasets, code, etc).

- **External links service**: implements a backend service for storing and serving secondary relational metadata about arXiv papers, with provenance. Will support author-curated links, and eventually assume responsibility for DOI and JREF metadata.

    Status: *Deferred to 2019.*

Reference Linking: arXivLabs project to display linked references to users.

    Status: *Complete. First version released October 2018; ongoing improvements in progress (see also [blog entry](#)).*

Public code repositories: Provide access to open source components of arXiv codebase as they become available.

    Status: *Complete: [List of arXiv's public repositories](#)*

See also: [arXiv Software Releases](#)

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

Publish Physics Category Descriptions: Finalize input from moderators and publish the descriptions on [arXiv.org](#).

    Status: *[In progress](#); awaiting review.*

Help Page Updates: Ongoing project to update arXiv help pages.

> *Status:* *In progress.*

Refine User Account Policies: Update internal documentation on user account policies and identify ways to automate actions.

> *Status:* *In progress.*

Moderation

Position Descriptions for Volunteer Roles: Develop position descriptions for various arXiv volunteer roles to be used in recruitment.

> *Status:* *Completed. Description of Volunteer Moderator role posted.*

Reorganize Moderator Guides: Update moderator documentation. Reorganize the expanded guidelines for moderators.

> *Status: Completed March 2018.*

Business Model & Governance

Society-arXiv Collaboration: Consider adding a new membership category for societies to support arXiv and contribute to its governance. A preliminary proposal will be drafted by an MAB/SAB subgroup to be discussed by the advisory boards.

> *Status: We'll have a pilot implementation during the 2018 MAB/SAB annual meeting, pilot implementation plan*

arXiv Governance and Business Model Assessment: As arXiv is going through a classic renewal process to modernize the architecture, continue to explore the current business and governance model's effectiveness and durability to ensure that arXiv is stewarded to the future in a successful manner.

> *Status: arXiv is moving to CIS in January 2019, Transition FAQ*

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.

Supporting public access mandates: The arXiv public access subgroup has been formed to explore the feasibility and desirability of expanding arXiv's features to support the public access requirements of research funders, and institutions' ability to monitor the compliance of their researchers. The subgroup's input will be crucial as we seek to understand if and how supporting those requirements might align with arXiv's existing goals and priorities.

> *Status:* *In progress (exploratory).*

Videos for Moderator Training: Improve the moderator training experience with short videos on using the moderator tools and dealing with common scenarios.

Preprint Service Providers Summit: Collaborate with ASAPbio and bioRxiv to bring together the nascent preprint service providers from several different disciplines to discuss a range of curation and maintenance issues of common interest and to foster information sharing and standard practices.

> *Status: Completed January 2018.*