

2013 arXiv Roadmap



The content on this page has moved

The content on this wiki page has been moved to <https://confluence.cornell.edu/x/K6ZRF>.

This page is no longer kept up to date.

Technical

Complete development issues from 2012-09-21 Scientific Advisory Board meeting - The SAB agreed to a number of changes in the way arXiv should send emails to moderators and control user selection of categories. Changes also add the ability for moderators to comment on and put submissions on "hold" without the need for arXiv admin interaction, a first step toward empowering moderators through more direct facilities. Work was started in 2012 and will be completed in early 2013. **Status as of 12/13: Completed**

Complete move of servers to virtual machine (VM) infrastructure - In 2012 we migrated two out of three of arXiv's server machines to new VMs. We will complete the move by migration our main web server to pair of load-balanced VMs managed by IT@Cornell. This arrangement will support scaling to additional web front-ends as necessary. As part of this transition we will formalize the update and maintenance process for these machines, and remove access by non-CUL or IT@Cornell staff. **Status as of 12/13: Completed**

Change email handling to support arXiv admin and moderation ticketing system - The "Request Tracker" software has been selected and we will rework email filtering software to use this instead of the current email-based arXiv admin workflows. **Status as of 12/13: Completed**

Add automatic classification checks to submission system - We will use classifier software developed by Paul Ginsparg to classify incoming submissions according to our category scheme. Where these automatic classifications differ significantly from the user-selected classifications we will add a warning to the moderator alerts. **Status as of 12/13: In discussion with Paul Ginsparg to get software**

Improve tools and interfaces to support moderators - We will work with the SAB and moderators to define and develop better tools that allow moderators to interact more directly and efficiently with the arXiv system and administrators. The overarching aim is to make best use of available moderator effort by making the work of moderators as quick and convenient as possible, consistent with achieving the quality goals and following policies set out by the SAB. **Status as of 12/13: Work-in-progress**

Implement new category aliases for cs/math/stat and add a new category for q-fin - In the past the creation of category aliases (e.g. math.IT/cs.IT) and associated re-categorization of articles has required a mix of manual DB edits and one-off scripting. We will develop tools to safely do bulk edits of this sort. Testing is also required to ensure that having articles where the primary classification does not match the old-style id as a result of these new aliases is handled correctly everywhere. **Status as of 12/13: Not done**

Improve author identifier support and data export - Add basic support for ORCID and other author identifiers associated with arXiv accounts. Add periodic data dumps for all public authorship data. **Status as of 12/13: Not done**

Improve dataset support - Review use of and experience with the Data Conservancy pilot and then either discontinue or improve interaction. Decide on a medium-term strategy for data and consider assigning DataCite DOIs for ancillary files. **Status as of 12/13: The Data Conservancy pilot was discontinued. A brief report on our experience can be found at the DSPS Blog. We will continue to support modest sized datasets and other materials using the existing ancillary files mechanism (http://arxiv.org/help/ancillary_files), and will ingest data from the pilot as ancillary files to support it long-term. We have subscribed to the EZID service and will assign DataCite DOIs to ancillary files.**

Security and login, email privacy - There are significant deficiencies which should be addressed. Issues include: all password entry and authenticated interactions should occur via https; domain based cookies should not be sent to mirror sites; and user email addresses should be more carefully protected. **Status as of 12/13: Not done**

Alerting system - The email alerting system remains popular but the mechanisms for subscription management are outdated. Users should be able to see and control their subscriptions from their user account page. New software will be developed to replace the extremely old and hard to maintain software implementing the current email subscription system. **Status as of 12/13: Not done**

Add MathJax to arXiv so that users see the TeX math rendered - We would like to add MathJax to arXiv so that users see the TeX math rendered in the browse user-interface, and in the submission system. Administrators and moderators should see math rendered in /admin/ and /mod/ screens. Restrict rendering to title and abstract fields, will not add to pages with lists yet (search results, listings pages). **Status as of 12/13: Completed.**

User Support and Moderation

Establish a new Physics Subject Advisory Committee - Physics moderation has become overly dependent on a few active physics moderators, especially Paul Ginsparg. The current Subject Advisory Committee in physics is dormant. The arXiv Scientific Advisory Board is engaged in an effort to identify additional moderators and seek new leadership within the arXiv physics community. **Status as of 8/13: Completed. See Physics Advisory Committee page.**

Explore the advantages and practicability of having an arXiv Scientific Director - Work with arXiv Scientific Advisory Board to define the role of a Scientific Director, and explore with stakeholders the desirability and feasibility of such a position. **Status as of 12/13: Search underway to fill interim Scientific Director position.**

Define new tools and interfaces for moderators - Work with existing moderators to gather and refine requirements for new tools and interfaces to support their work. Existing processes are in need of enhancement or replacement as we upgrade moderation processes and work to improve moderator efficiency and convenience. **Status as of 12/13: In progress. See "Improve tools and interfaces to support moderators" in Technical section above.**

Move arXiv communications administration to an issue tracking system - We have begun the process of shifting all communications with users and moderators to a web-based issue tracking system (Request Tracker). Service desk functions managed by arXiv administrators include answering user questions, troubleshooting technical problems, soliciting and following up on moderator input, responding to and resolving moderation appeals, etc. These functions are currently carried out using multiple arXiv.org email mailboxes. A modern issue tracking system will provide us not only with better tracking, but built in reporting and issue classification tools for better management of this labor intensive aspect of arXiv administration. **Status as of 12/13: Completed.**

Governance

Form the first MAB - Make MAB appointments in January '13 and start working with the group. **Status: Done**

Finalize SAB bylaws - The SAB is reviewing a revised draft of its bylaws. A final version of these will be adopted in early 2013, along with the appointment of a Board Chair. Also under consideration is the question of appointing a scientific director — a part-time position to provide intellectual leadership from the perspective of the scientific community. **Status: Done**

Test and refine the operation of the new governance model - The arXiv principles aims to clarify the authority, responsibilities, and constraints of CUL, MAB, and SAB. Ironing out problems and developing a working system may require some time to test and observe the inner operation of the governance model. Currently there is not a clear mechanism to settle potential conflicts among the three chambers. The first strategy implemented in 2013 was to create a roadmap for arXiv so that everyone knows the CUL team's priorities and goals. Also we've appointed representatives from SAB to MAB and from MAB to SAB in order to facilitate communication and idea exchange. Having ex officio member representation in each group will also facilitate information sharing and developing a common understanding of the respective goals. **Status: Work-in-progress**

Further refine the benefits of being a member for participating institutions - During the governance planning meetings, several ideas emerged as potential free services for members; however, implementing these features may require significant staff time. Therefore these services need to be considered in the context of arXiv's current maintenance and development priorities. After the MAB is formed, in order to better understand demand for services, CUL will invite ideas/proposals and review them with MAB and SAB. **Status: Work-in-progress**

Continue the membership drive - We are very encouraged with the five-year pledges received so far. We want to increase the number of arXiv member institutions to create a large and international network of supporters. Another goal is to be able to reduce the institutional membership fees in the future (current annual institutional fees are in the \$1,500-\$3,000 range). **Status: Work-in-progress**

Develop reserve fund policies - The purpose of the arXiv reserve fund is to support unexpected expenses to ensure a sound business model. Currently, arXiv has a reserve fund of approximately \$100,000 that accumulated during 2010-2011 due to unexpected staff vacancies and other savings (2012 expenses have not been factored in yet). The 2013-2017 budget projections assume that we will be able to add \$50,000-\$100,000 per year to the reserve funds. We need to develop policies about how contingency funds will be used and how the account will be structured. **Status: Completed**

New Partnerships & Communication

Define a R&D agenda and seek external funds to advance arXiv - So far, CUL's sustainability planning efforts focused on arXiv's operational budget to support the core services and arXiv's strengths in order to stay mission-centric. One of the key goals ahead of us is to define a research agenda for arXiv. We'll seek input from the advisory boards, users/scientists at large, and information scientists. We will develop a methodology for making decisions about arXiv's R&D projects and partners. **Status: Not enough attention was given to this goal as there were more pressing issues**

Continue the dialogue with publishers/societies - In celebration of the arXiv's 20th anniversary, on September 23, 2011 Cornell University Library (CUL) hosted a meeting at Cornell with the representatives from several publishers and societies that are interested in Cornell's sustainability planning efforts. The meeting report provides a synopsis of the discussion and recommends next steps for continuing this dialogue. We will resume this investigation and discuss the feasibility and desirability of establishing a research and innovation collaboration in support of arXiv. This effort is envisioned to entail a separate funding stream (created by participating publishers and societies) from the operational budget, which includes resources for the routine and core services currently provided by the arXiv team (including essential updates). **Status: Held a conference call in September to continue information exchange**

Define and communicate measures of success - To address this dynamic process, one goal over the next several months is to create an assessment model to help CUL continue to fine-tune the sustainability model. An assessment plan will also help identify unforeseen developments and making course adjustments to the service and collaboration model. Working with SAB and MAB, we will develop an assessment framework with three key components. Desired outcomes and success measures to measure progress (for instance, dynamics of the governance model, level of financial support, enhancements to arXiv, improvements to moderation system, etc.); Assessment model to gauge and report success based on the identified outcome measures; Plan a five-year review process to enable CUL to conduct a comprehensive self-evaluation with input from the Simons Foundation, MAB, SAB, and other key partners. **Status: Incomplete**

Develop guidelines for engaging in new partnership - arXiv is approached on a regular basis by outside groups asking for advice or special services. Sometimes the assistance requested is minor, but often it would require devoting some amount of staff effort. We need guidelines for determining under what circumstances we will allocate resources in order to collaborate with an outside organization. Working with the arXiv Boards, develop policies that guide consideration of potential working collaborations with outside groups. **Status: Work-in-progress**

Enhance communication with users (scientists) - Share opinions received from scientists and users with SAB and MAB. How about seeking input from other scientists, arXiv users? What is working well for them that needs to be maintained? What are their unmet needs? What kinds of changes they want to see implemented? CUL needs to have a systematic way of gathering user feedback and this info ought to be shared with MAB and SAB (both groups are also important channels of input as they represent other groups). **Status: Work-in-progress (discussed at SAB meeting)**