

2018-08-16 Meeting Summary

Date

16 Aug 2018 - 11:00-12:00 ET

Attendees

- Daina Bouquin, Harvard-Smithsonian Center for Astrophysics
- Mark Doyle, American Physical Society
- Alberto Accomazzi, Astrophysics Data System
- Thorsten Schwander, SLAC National Accelerator Laboratory, Stanford University
- [user-4c1fa, arXiv.org](#)
- [user-6ff8e, arXiv.org](#)

Agenda

- Welcome and introductions
- [IT AG mandate/charter](#)
- Overview of the arXiv-NG project and technical architecture
 - Martin + Erick will give a brief overview (5-10 minutes)
 - [Architectural documentation](#)
- Current areas of deliberation
 - Author metadata and disambiguation
 - Backup and recovery
- Discussion

Meeting Summary

- Welcome and introductions
- Review of the [IT AG mandate/charter](#)
 - The advisory group recommended keeping discussions mostly informal, permitting frank discussions about topics.
 - The group agreed on the creation of an email list as a mechanism for raising discussion points at any time between meetings, and that discussion points need not be constrained to technical subject areas; for example, this might include discussion of issues within the community, like the default license associated with a submission.
 - The arXiv team agreed to provide internal meeting minutes and a public meeting summary that the advisory group could review prior to dissemination. There was a recommendation to add a section about conflict resolution to the charter in the event that group participants disagree about discussion points or action items.
 - The group discussed the frequency and duration of our meetings. As a starting point, it may be appropriate to convene meetings on a quarterly basis or coincidental with progress on our roadmap, but the advisory group is also amenable to ad hoc meetings; a 1 hour duration for the meetings seems fine.
- Overview of the arXiv-NG project and technical architecture
 - The arXiv team provided a brief history of arXiv with respect to IT and the pressures the team has faced with the legacy codebase and its sustainability. These ongoing pressures pointed to a clear need to renew the codebase with better sustainability in mind. A formal technology review did not surface any off-the-shelf solutions that could easily be adapted to arXiv's needs; instead, they provided some inspiration for our technology choices and architectural decisions. The team described how arXiv "Next Gen" (NG) will be a modular, service-oriented architecture. Functional areas of the classic system will be renewed in situ with NG services, coexisting with the classic services until they are completely replaced.
 - The arXiv team discussed the progress that has already been made towards NG development: [architectural design](#), improved processes, code review, and deliverables (e.g. new search feature).
 - The arXiv team described the annual roadmap and presented a high-level milestones document.
 - The advisory group recommended caution and counsel around technical decisions, especially when it comes to turnkey solutions versus "roll-your-own" solutions. Turnkey solutions may be preferable if they reduce administrative overhead and turnaround time, even if monetary costs are higher.
 - In looking at recent trends in API development, the advisory group recommended exploring [GraphQL](#) as an example standard query language for APIs. In general, it is best to review use cases and to keep an open mind in considering what technologies best match needs. Modifications to existing APIs require a deliberate, thoughtful process.
- Author Metadata and Disambiguation
 - The arXiv team described the current (simple) data model for authors and the difficulty in meeting user needs around precision and recall with the current model. On the other hand, the simple structure has been a low barrier for entry, and arXiv would like to keep it that way if possible.
 - The advisory group indicated that author disambiguation is a complex problem; ORCID may replace anything that arXiv decides to do independently. Even with a standard like ORCID or a database like [GRID](#), it would be impossible to achieve 100% user compliance or matching accuracy, so annotation and hand curation would still be necessary. Authentication would also be an issue. Any solution should assume that coverage will be incomplete, and should take an annotation-based approach.
 - Inspire intends to migrate to and substantially adopt GRID moving forward, however it will be lengthy process of mapping current institutional records to GRID records and adapting granularity, etc. GRID appears to be the institution authority system around which various services in the HEP realm congregate.
 - The advisory group indicated that author disambiguation as well as institutional records are both subjects that would potentially deserve dedicated discussions among the technical advisory group.

- The advisory group recommended keeping this topic a priority for arXiv, since many groups working with arXiv depend on this data to maintain their bibliographic data.
- The advisory group indicated the possibility of data exchange (e.g. with INSPIRE)
- INSPIRE encourages any working group, collaboration, team, or task force with more than a handful of members and an extended lifecycle to use author.xml, which also supports ORCIDs. INSPIRE provides some tools and guidance and is also in a position to provide assistance in video chat.

Action items

- ☑ arXiv team creates listserv (arxiv-tag-l@cornell.edu) for group communications