Subgroup Conference Call Agendas (April-May 2013)

Member Relations

During the 2010-2012 governance planning meetings, several ideas emerged as potential free services for members, for instance, submission-based data when arXiv's metadata structure is ready; institutional repository bulk download. Member institutions are also very keen on informing their scientists and researchers about arXiv's business model (e.g.,arXiv LibGuide to share with their faculty and students in related disciplines). 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. For 2013, within the context of Member Relations, we'd like to focus on the following issues:

- Often, we are advised to make it easier to find business info on the arXiv webpage (one idea proposed is to put links from abstracts or submission forms to 'sustainability' page). What are your thoughts and suggestions?
- How can we involve the members in documenting and sharing with us information about how arXiv is being used? What is working well for scientists that needs to be maintained? What are their unmet needs? What kinds of changes they want to see implemented?
- How can we promote the libraries' and research laboratories' role in arXiv's sustainability to scientists? Scientists know very little about how arXiv is run and how much the operation costs. How can we improve communication so that scientists will understand the role of libraries in maintaining and supporting arXiv? What is the role of the arXiv member institutions in this communication process?
- 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. What are your suggestions?

Members: Jim, Eva, Molly, Zhixiong

Conference Call: April 15, 8pm EDT (Zhixiong) & April 25, 9:30am EDT

Meeting Notes:

- Zhixiong: the National Science Library (Beijing) has put together a working group (arXiv Service Group). This group is composed of librarians from several different institutions. One goal for the year is to raise awareness of arXiv among researchers (especially in non-physics areas). They would appreciate any promotional material arXiv could provide.
- Zhixiong: another goal of the Service Group is to conduct a survey to discover how arXiv is used by researchers and students, the level of knowledge about arXiv, and what problems researchers may have with using arXiv. This may be of general value to arXiv.

Molly, Eva, Jim, Jaron, Oya:

- Highlights from subscriber news: we have invoiced 135 institutions who pledged as part of the campaign. We will invoice both new institutions that have appeared on the top 50 (and top 200), as well as institutions that did not respond.
- We discussed feasibility for a new class of membership for society publishers. This category in arXiv could give them a new identify and a new way of communicating scholarship, particularly as relates to society publisher communication with faculty. Faculty members like the arXiv project and are happy to contribute to contribute it, but faculty are still quite conservative with regards to wanting access to print documents -- in other words, they run the gamet of supporting open access as well as the old publishing guard. For instance, if we get AMS to be a sponsor, they might be to promote arXiv and connect with it and researchers in new and beneficial ways.
- Publishers recognize the value of arXiv to their members; AAS, for instance, would be very supportive. Other international societies would be likely to support arXiv. How are scientists in China (for instance) using arXiv?
- How can we encourage institutions to become sponsors? We have the acknowledgement banner and we agree that "Libraries" should be added to "University/College/etc." to indicate that arXiv support is library-funded (nearly all). Many do not know that their institutional LIBRARY supports arXiv.
- There is some discussion about the look of arXiv; for frequent users of the system the interface is clean and navigable; for others who are new the front-end seems rather outdated, particularly in terms of a search and discovery interface. Unless someone has an understanding of how arXiv works, they may 1) not "understand" that arXiv.org is the main arXiv page, and 2) this setup may deter NEW USERS OR SUPPORTERS -- thought not the old guard of scientists -- from using or supporting arXiv if they find it difficult to navigate. How can we facilitate a process to provide user feedback, and new usability options, to the arXiv SAB and other stakeholders? For instance, Eva described a faculty member who

- encourages his students to deposit McS thesis to arXiv; could these students be a user-testing/feedback group for us, as potentially new scientists?
- Is there a growing market for data tools to make freely available data searchable and usable?
- Is data a major challenge for arXiv (due to metadata, storage, etc)? Does arXiv intend to hold on to the data/ancillary material forever, along with the articles? Purdue is going for track (sp?) certification for CRL (Purdue research repository); one of the issues to address is how long will the institution preserve the data, esp. as much of it will deteriorate over time. In addition, researchers might not want to reuse any of the data (despite government mandates prompting them to do so) given all of the variables in each data set.
- While research data is somewhat marginal for arXiv -- which might be natural as the biggest user groups are HEP and astrophysics, both disciplines that typically pay attention to data management and run big data repositories. On the other hand, the more formal sciences like math have different needs and may not be as well-organized.
- arXiv partnered with Data Conservancy (Johns Hopkins) for a pilot, where arXiv acted in a front-end for DC for researchers to deposit data; the conclusion was that most of the data sets were small, and were the kind of ancillary files that are already found in arXiv.
- How will the new OSTP mandate affect arXiv? arXiv has been approached by at least one publisher to discuss "partnership" in order to fulfill the repository requirement for the project. Compliance issues will reside with the publisher.
- In Europe, Open Science is becoming reality because European Union is supporting it. For instance, OpenAIRE project, https://www.openaire.eu has garned a lot of attention. EU requires everyone who has received Framework 7 EU funding (and that is quite a lot of people) to publish the resulting papers OA in institutional repositories. What makes this different from previous similar attempts is that EU has real bureaucratic resources to check whether* {}this requirement is fulfilled, and is able withdraw funding. *What does this mean, in practical terms? For example, at the University of Helsinki this means we must make our research database TUHAT compatible with OpenAIRE technical specificatons so that they can harvest the relevant metadata. In an ideal case, our institutional repository should include full text OA documents, but we are accepting links to arXiv instead. EU not stopping at OA pubs, data is next to be OA. In Finland they are working on the next step, an accompanying metadata catalog. More info: http://www.csc.fi/sivut/tta/national-reseach-data-project

ACTION ITEMS:
1. Discuss and investigate a new class of membership for scientific and scholarly societies;
2. Work with SAB and other MAB members to discuss usability studies for the arXiv interface (landing page);
3. Monitor the discussion about OSTP to see if they will impact arXiv's core mission.

Financial Planning

- We would like to diversity of revenue sources in order to raise additional funds to allow us to expand arXiv's staffing (especially the current minimal IT staffing configuration), lowering the current tier structure to encourage more organizations to become members, and support R&D projects. What are your suggestions? One of the ideas we would like to discuss is adding a Give button on arXiv home page to invite donations from users (examples). What are the pros and cons of this idea?
- 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 (for instance, funds needed for closing the business versus development funds - also a part of the reserve funds to create an endowment & the interest can be used for reducing membership fees). Also, we need to determine at what point of reserve fund accumulation (total reserve balance) we will consider to lower the annual membership fees.
- The current annual institutional fees are in the $1,500-$3,000 range, which may be high for some developing countries or institutions that are lower in the use ranking (below 200). Shall we add a lower fee to encourage these institutions to support arXiv?
- Some of our members feel that it makes sense to increase the annual fee for the top users as the current model does not appear fair (see the usage chart). What do you think? Should we add a higher top tier pricing for the organizations that are listed high in the use list?

arXiv 2013 Budget & Pledge Summary (May'13): arXiv snapshot_May2013.pdf

Members: Uwe (Esther), Yuri, Philip

Conference Call: May 10, 9am EDT (Uwe) - another one with Yuri & Philip TBD

Meeting Notes:

- Question about why arXiv may need more revenues. Because the current budget is based on operational costs and does not factor in adding new features, revamping software, etc.
- Question about IT staffing. A few months ago, arXiv team decided not to migrate the browse component of arXiv to Invenio. There are a number of factors behind this decision. The main reasons for re-evaluation were slow progress by our team in the migration, an understanding that replication of arXiv functionality in Invenio involved rather more coding that we had earlier imagined, and pressure from our advisory board to address other issues. Instead of this large-scale migration we have decided to refocus our efforts on continuing to develop arXiv within its existing technology infrastructure and will reassess our strategy to use Invenio as platform for arXiv next year.
- Esther stated that German scientists like arXiv lean and simple, operated at low cost. We need to make a very clear case why extra money would be required. She said that if there were a "give button" on arXiv it would make sense to earmark and carefully communicate the use to which this money would be put. Better to have a specific project in mind that general support.
- Esther cautioned us that because we announced that we have a sustainability model in place, we need to be careful about not confusing scientists that we still need more funds. Also, also we don't want scientists or member of consortia members to think that we are approaching them for funding twice. For instance, TIB is contributing on behalf of German universities and scientists.
- Question: Are there any institutions with rank>200 who would be willing to pay if there were a lower tier? We have not asked so we don't know. However we are aware that some institutions are finding even the lowest tier not affordable due to their budget challenges.
- If we decide to introduce a lower tier, we may engage new members but we'll also reduce contributions from the current members that are ranked lower than 200
- Germany would like to have a more "fair" model (higher fees for the top tier uses) but understands the challenges in complicating the current system by introducing another detail that would require individual negotiations with the top users. Oya suggests that for next year's fees, we might include for very heavy users (top 10-20) an invitation to contribute additional funds due to their high use.
- Philip sees the potential to increase the fees, because arXiv is still a great value at current fees. He also asked how the disciplinary coverage in arXiv has changed since physics was the main user group; are we paying close enough attention to disciplinary differences?

Action items:

- Cornell will draft a reserve funds proposal as it wil be easier for MAB members to give feedback (rather than commenting on the general reserve-fund related issues articulated in the meeting agenda).

Collaboration with Publishers and 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. 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). Please see the September 2011 planning meeting summary about the outcomes of a collaboration discussion with a group of publisher and society representatives. We would like to review, prioritize, and move our initial meeting recommendations into action. What are your recommendations?
- With regard to recent U.S. legislation that mandates scientists and researchers make their federally-funded data and publication available via open access, what are the possible (and desired) role that arXiv can play?

Members: Catriona, Carol, Tim, Diane, Tommy

Conference Call: May 21, 10:30am EDT

Meeting Notes:

- Interest in developing surveys:Scientists. Members across multiple MAB subgroups are interested in talking to scientists' about their use of arXiv. One suggestion is to survey non-physicists about arXiv in an effort to better understand their obstacles. We all agree it is important not to conflate "scientists" into one lump group because of both disciplinary differences and use-case scenarios (i.e. some scientist are users of arXiv, others are readers of arXiv, submitters to arXiv, etc.)
- A publisher survey before the next publisher-conference call may help develop the agenda and get a sense of emerging priorities, ideas that are no longer relevant, etc.
- How are new fields of study introduced into arXiv? It's largely been a practical question in terms of analyzing submission data. Before new fields are considered for arXIv, both MAB and SAB will need to be on board.
- OSTP/open access: *We discussed the OSTP mandate and arXiv's role/position in this newest U.S. open access initative. Given our role as an international rapid, pre-print service, we wish to remain true to our core mission; at the same time, we understand that publishers and agencies might look to arXiv as a repository solution. We are currently working with one publisher to investigate this idea, but our goal in working with any publisher to to investigate and potentially develop a framework that is generalizable across multiple publishers/societies. We agree a process /solution cannot be U.S.-centric and will would require significant changes in metadata and possibly compliance. For many, arXiv + scholarly journals= green open access, and this combination may be attractive across other related-fields. Does arXiv wish to collaborate with paid journals for "green open access" in order to provide an alternative to gold open access?
- What's the current outlook for the system architecture? *Currently, we are pursuing a modular approach to update the system. Our priority is the moderation system, but we will refocus architecture discussion for next year and beyond as it relates to metadata, cross-linkin, DOI, etc.
- Supplementary material: *We have concluded our trial with Data Conservancy. We learned that very little of material being deposited was "large" data. It was mostly small data sets similar to the kind that arXiv has always accepted. We will start assigning EZIDs to data sets.
- Research data has really changed so at the call we will need to spend time reviewing the developments in this area.
- Author identification is still a priority. Simeon Warner is on the board of ORCID. We learned that at LANL, if a researcher is presenting at a conference/publishing, they now get an internal ID to attach to the preprint and deposit to repository. LANL is interested in DOI/author verification issues.
- Statistics: aggregate data versus alt-metrics. The idea of knowing what articles are used during the lifecycle of research is interesting. It is our policy not to provide any download counts for individual article, even to authors of the article. That door is already opened on individual article metrics.

Action items:

- arXiv will arrange a follow-up meeting with publishers to discuss developments since the Sept. 2011 meeting. Many publishers have expressed an interest and willingness to be part of arXiv's future plans and this call will be an opportunity to discuss new initiatives, like open access mandates;
- Develop a subgroup to coordinate feedback from publisher surveys' to provide information and feedback about new collaborations
- Develop a sub-group across MAB and SAB to draft and coordinate another feedback survey

Research and Development Agenda

- 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. How can we move beyond the current sustainability model that focuses on operations to a strategic plan for arXiv's further development and innovation? Beyond the 2013 Roadmap, we would like to set a 3-5 year strategic plan for arXiv (encompassing all aspects of arXiv including user support, IT, moderation, etc.). Do you recommend any strategies or tools in setting priorities and assessing collaboration models? Do you know of any successful models we can learn from? We also need to find additional funds but we want this goal to be driven by an agenda.
- Based on your knowledge of arXiv, its principles, and the repository ecology, do you have any recommended R&D areas that we should be focusing on?
- 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. We would like to develop a methodology to guide us make decisions about potential working collaborations with outside groups. Do you have any recommendations?

Members: Tommy, Mackenzie (Oya & Simeon from arXiv)

Conference Call: April 22, 11:30am EDT

Meeting Notes:

- Tommy presented six questions to consider: Ideas for improving the arXiv interface; adding new functions; who can submit to arXiv; considering a peer review function; information exchange between arXiv and local IRs; ORCID implementation.
- To identify an R&D agenda, Mackenzie suggests that we should think of a set of categories and then see what can be rolled into the operation and addressed in the short term and requires real R&D. Possible categories: Interface, author identity, infrastructure scalability, mining arXiv.
- Discussion of the OSTP OA mandate and questions about whether arXiv could be a fee-based submission venue to meet the mandate - what added value might make this worthwhile. Mackenzie stresses that what is needed is a system for getting credit when one deposits to an open access system. For instance, can a scientist only post on arXiv with a DOI and use this as an academic achievement in support of his/her tenure process (not requiring a peer-reviewed publication)? Can we build some article level metrics into arXiv? This is an R&D area for consideration.
- One idea to consider is positioning arXiv as a preservation system and the possibility of developing for-fee-services for journals wanting to archive on arXiv, or even whether we might charge depositors.
- Tommy wonders if arXiv should promote itself as a way to meet green OA mandates and if and how it will meet the requirements emerging in different countries (compliance issues).
- Mackenzie recommends that when we enter a new collaboration, we need to assess the costs and do not enter in a partnership without a clear benefit for arXiv - whether it is visibility or additional revenues. She wonders if we should work out ways to charge for our services.
- Mackenzie questions whether we can develop a set of mining tools and monetize by providing cost-recovery services (similar to the data mining tools HathiTrust Research Center is providing) We can avoid conflict with free open-access to corpus by still providing that free. Mackenzie suggests Sloan Foundation as a potential funding sources. Many research groups and publishers are interested in data mining tools.
- Mackenzie says that supporting research data is a big issue and that we must be careful not to get distracted from our open access article mission. Possible collaboration with Dryad would be interesting.