

Summary of Comments / Themes from discussions on the AAU-APLU Workshop

Written Up by: Jake Carlson (University of Michigan) and Sherry Lake (University of Virginia)

The AAU-APLU will be holding a 1.5 day workshop for research universities interested in drafting institutional plans for public access to research data at the end of October. This is an invitation only event, but we see this workshop is an opportunity to influence the creation or augmentation of the kinds of investments, connections and supports that we as data librarians would like to see at our respective institutions. Therefore, we sought to engage the larger data librarian community for their thoughts and ideas on what our collective goals should be for this workshop.

We asked data librarians to take a look at the [AAU-APLU report on ensuring public access to federally-sponsored research data](#) that inspired this workshop, and [the initial description](#) of the workshop. We then asked that they share their thoughts and feedback in a Google Document and in an online discussion that took place on October 23, 2018.

The following is a summary of the points that were made from the Google Document and the online discussion:

How could the current process of making data publicly accessible be improved upon (generally, not just at the institution level)? What would be needed? [Talking Points]

- More specificity and clarity on what needs to be shared from funding agencies: all data collected from the research project? just the data that underlie published findings? something in-between?
 - What are the sharing requirements for assessment data (particularly from education based research projects), IRB reviewed data and other specialized data types?
 - More uniformity from funding agencies would be helpful too. Some funding agencies, including the NIH, are discussing all research output rather than just focusing on data and publications.
- More and better funding, tools and staffing for ensuring compliance and providing support for making data publicly accessible. It's not clear that any stakeholder (including the funding agencies themselves) have the resources needed to follow through on data sharing mandates.
 - Costs of data sharing, stewardship and preservation need to be acknowledged and allowance should be made for including these costs in proposals.
 - Could funds explicitly dedicated solely to the purpose of managing, sharing and preserving data be incorporated into grants?

- How could the ongoing costs (beyond the life of the funding) of data sharing, stewardship and preservation be accounted for?
- Consistency by agencies in defining what to be shared, how DMPs are evaluated - take advantage of tools created by Libraries (JHU review Rubric & DART rubric)
- Agencies highlighting that grant funds should be used for data curation and sharing - not taking away from the baseline request, but adding more money for data management.

- Better defined roles, responsibilities and processes for stakeholders in making data publicly accessible from funding agencies, but also from university administration.
 - Push some responsibility for compliance back on funding agencies.
 - Make funding agencies pay for the cost for data management - have it built-in in funding amount & not subtract from the amount awarded.

- Criteria for evaluating the quality and documentation of data to be sure it has value and can be understood and used by others before it is publicly released.
 - This would be challenging to scale particularly at current levels of funding and staffing.
 - Leverage the work of Data Curation Network to share evaluation of data.
 - Leaving this type of data review up to institutions to do would likely result in highly variable outcomes.
 - Professional societies could collaborate with institutions to set minimal standards for documentation and reusability for institutions to use in the evaluation process.
 - Disciplinary repositories generally define what metadata, documentation and other materials they want as a part of the data deposit process.

- A level playing field. Publishers have a role to play in making research data publicly accessible as one of many stakeholders, but they should not be given ownership or exclusive rights over data. Instead, ownership and decision making control over the data should be retained by the institution (and/or researcher(s) who produced the data).
 - Publishers were not included in the AAU-APLU report and were not invited to this workshop.

What do we want from university administrators? (i.e. What do we want our fellow attendees to know and take away with them when the workshop is over)? [Goals and Objectives]

- We should expand our thinking around public access to research data beyond compliance to federal regulations and recognize that ready access to high quality data is about advancing science and is increasingly important to research in all fields. Therefore, providing public access to data needs to be built into research and publication practices and supported by institutions.

- Consideration for public access to data should extend beyond the “big data” and/or “data science” frames.
- Data are the building blocks of research and come in many forms, not just numbers. Software can be an integral part of a data set and therefore merits consideration in data access and preservation discussions.
- Recognition that it is a lot of work and a big ask to ensure that data are discoverable, accessible, usable and preserved, particularly for smaller, less resourced universities. This work requires an investment of time, resources and funding that does not currently exist yet.
 - Need to recognize that it is not just a matter of sharing data; it is a matter of ensuring that research data are of sufficient quality and are in a state where they can be easily discovered, understood, trusted and reused.
 - Practically speaking there is probably no way that any one university could ensure that this work is being done adequately.
- At the same time these are investments that are worth making. Administrators should have an understanding that supporting public access to data could be a benefit to the university. Making data publicly accessible is worth investing in and considering beyond what is required by funding agencies and others.
 - Plans are good, as are policies, but action is what is needed. How will services and support be developed around public access to research data? A governing body of some kind will be needed to enact and oversee institutional data policies.
- Recognition that many librarians / libraries are actively supporting public access to research data. Give the librarians/libraries recognition and support for doing this, as in good policies, institutional “mandates” i.e, including a data management training requirement, and funding for more people.
- Consideration for incentivizing / rewarding / supporting researches to invest their time, effort and resources into creating data sets of value that can be consumed by others.
 - Producing high quality data sets and making them publicly available should count in faculty promotion and tenure reviews.
 - Supporting public access to research data will require attention, services and support throughout the research lifecycle.
 - Also understanding that data may have an expiration date and need to be deaccessioned at a future date.
- An understanding that the issues and challenges of making data publicly accessible span across universities. Therefore, universities ought to consider developing cross-institutional partnerships to develop and provide shared infrastructure and services. Regional, consortial and potentially national efforts in investment and coordination should be a part of this conversation.

- We will need to recognize, address and overcome feelings of competition between institutions for scarce resources to create and sustain shared infrastructure. We will need new and more egalitarian models.
 - This may also be an opportunity to develop communities of peer review.
 - Data discovery may be a good place to start (including data citation)
- Recognition that giving publishers or other for profit companies ownership or exclusive licenses to research data is not beneficial to scholarship, even when they create products that may be of interest to the university (research platforms, analytics tools, etc.).
 - Publishers have a role to play, but it should not be a dominant one. Profit should not drive public access to data.
 - Open access to data (to the extent that is appropriate) will be critically important to realize the value of public access to data (public access and open access are not the same thing).

What could be done by university admin to further public access to data sets? [Possible next steps]

- Goal of this workshop: “accelerate the development & implementation of institutional plans providing public access to data” For institutions there is a need to create, and or update institutional data policies. ALL institutions should have data policies.
- However, plans and policies alone will not be sufficient and in fact may lead to a false sense of accomplishment. The real test will be in creating systems and structures that can support the work it will take to make research data publicly accessible. How can plans be made actionable and maintained across the university?
 - Roles and responsibilities over research data should be clearly defined and understood across the institution.
- Therefore, there is a need to work with other administrators who support research operations within their university to consider not just what infrastructure and policies would be needed for making data publicly accessible, but what support would be needed from whom to make this all function smoothly. Elements to consider would include:
 - Governance and governance structures
 - Expectations of the institution for data retention / sharing.
 - And policies / support MUST include and account for the public sharing of data per funder/publication requirements.
 - Reward systems, particularly Promotion and Tenure
 - Clearly statements on ownership and rights over the data
 - Classifications of different types of research data and how these classifications would pertain to stewardship, security, sharing and preservation
 - Controlled data sharing / Data Use Agreements (for sensitive data)

- Guidance / training (librarians can do this at the local level!)
- Definitions for terminology used to avoid ambiguity.

- Look for places where this work can be incorporated into existing systems and potentially could be automated.
 - Perform compliance checks on Data Management Plans - do the links to shared data sets given in final reports given to the funding agency resolve correctly?
 - Make Data Management Plans open, or at least available to the agencies and individuals who provide services and support for data.

- Work with their peers (and librarians) across universities to consider shared infrastructure, standard practices and policy norms for data (although common guidance on IP is somewhat controversial).
 - Recognize this is being done to some extent already and other institutions who feel not up to speed, can leverage what has already been done.
 - [Data Curation Network](#) as one possible example

- Research how other countries, the UK, Canada and Australia in particular, have addressed making research data publicly available to citizens as a means to inform how the USA might address these challenges.

- Consider how we will continue these discussions on supporting public access to data after the AAU-APLU workshop both within and across our institutions.