



AAU-APLU Public Access Working Group Report and Recommendations

November 29, 2017

In this era of open scholarship, greater access to research findings and data, especially when grounded in the FAIR principles (findable, accessible, interoperable, reusable), has proven to be an important way to accelerate scientific progress and advance innovation to better serve the public good. Although there is general agreement about the value of increased public access to data, ensuring such expanded access will require a significant culture shift at universities and among their faculty, thoughtful and carefully crafted new government policies and practices, and investment in the infrastructure required to make data publicly accessible.

Success will require overcoming some major obstacles. To overcome these barriers, universities and federal agencies must work together to respond effectively to the growing demand among scholars and the public to have broader access to each other's data, algorithms, and other digital products of publicly funded scientific research. Universities will need to create the infrastructure required by the public access mandates of the federal agencies funding their research so that data collected to support federally funded research can be shared, to the extent possible, with the public. Faculty will have to come to understand that the data they create with federally funded research is not "their" data alone, and therefore they will need to adapt their views concerning data sharing. At the same time, federal agencies will need to fund the costs associated with making these data widely available and provide consistent and clear policies, compliance guidelines, and definitions across agencies to minimize the burden on researchers and institutions. Ensuring that research data are more accessible clearly has tremendous potential to fuel scientific analysis and discovery by making data more open to scrutiny, re-analysis, and extension.

There is a danger, however, that in striving to meet demands for public access, each university will respond with its own blend of homegrown and commercial solutions, each federal agency will develop its unique standards and policies for what constitutes public access, and commercial data management providers with different goals and profit incentives may build services that are costly and inaccessible to universities and the public. Fortunately, by committing to a set of shared principles and minimal levels of standardization across institutions and agencies, we can help minimize costs, enhance interoperability between institutions and disciplines, and maximize the control institutions can exert over how they ensure access to publicly funded scholarship.

This brief report, prepared by a joint working group on public access convened by the Association of American Universities (AAU) and the Association of Public and Land-grant Universities (APLU), makes recommendations and summarizes actions that universities and federal agencies can take to advance public access to data in a viable and sustainable way.

Recommendations for Federal Agencies Implementing Public Access Data Requirements

The AAU-APLU Public Access Working Group proposes the following recommendations to the executive branch and for federal agencies as they develop and implement new policies and guidelines for public access to data. These are offered in support of the following broad **goals**:

- Providing public access to research data in the most useful ways to society;
- Minimizing the administrative burden on agencies, universities, and researchers;
- Allowing exceptions for privacy, security, and intellectual property (IP) concerns;
- Prioritizing data quality and its rigorous evaluation as a foundation in preparing, documenting, and releasing data;
- Balancing the substantial costs of data access against the benefits of access;
- Recognizing that data types and accessibility needs vary across disciplines, requiring a flexible approach; and
- Considering the community of interest and duration of usefulness for the data in question and making retention and access requirements clear.

The Working Group puts forth the following recommendations to advance these goals and thus benefit the public, government, universities, and research communities:

1. All federal agencies should strive to harmonize their policies, documentation (e.g., data management plan, data use agreement terms, and data sharing certifications), and compliance approach to public data access to the greatest extent possible to minimize the administrative requirements and costs for both agencies and funding recipients.
2. Agencies should continue to support universities' retention and free accessibility of data obtained with federal funding, in analogy to the treatment of IP. Third-party vendors who may be contracted by universities and other funding recipients to assist with data access should not be granted an exclusive license to such data or other exclusive rights to make such data available in the future.
3. Agencies should provide clear information on how compliance with data sharing requirements will be monitored, evaluated, and enforced, being mindful of the need for institutions and agencies to manage constrained resources. This includes providing clear guidance on when the compliance period for data sharing expires. In addition, we strongly recommend that once a project is completed, agency research staff oversee compliance with data management plans since they are in the best position to work with PIs and universities in providing an ongoing assessment of the value of specific data and overseeing and managing ongoing data access commitments.
4. Agencies should provide clear information on expectations regarding what data do and do not need to be shared, and, at least initially, agree upon a minimum standard across agencies regarding the data subject to public access requirements. Specifically, we recommend that the minimum standard be defined as data that are essential to understanding and reproducing peer-reviewed publications and conference proceedings, i.e., that are reasonably expected by the discipline to be accessible for re-analysis. Additional requirements could be added at the discretion of the program to facilitate specific goals, e.g., where the data outputs are a main goal of the funding, where disciplinary conventions call for other kinds of data outputs, or where a negative result may not be published but still could be useful for other researchers or the public.
5. The length of availability of data and the possible embargo of availability should be specified in the award documents based on the set standards created for each discipline where available and should depend on the specific costs and benefits associated with public access to the data. Agency expectations for data access after the funding period has ended should be specific and finite in duration and should be accompanied by additional funding if they require more than a nominal cost.
6. Agencies should recognize that before an institution publicly shares data, it must be evaluated for quality and well-documented to ensure the data adds value to the scientific discussion, can be understood and effectively reused, and can withstand public scrutiny.

7. Agencies should continue to promote public access rules for data that are consistent with rules protecting human subjects (e.g., IRB and HIPAA), proprietary information, national security, and controlled unclassified information (CUI). At the same time, agencies must carefully balance the goal of providing public access to data with other government-wide and agency-specific data access controls.
8. Federal agencies should look to universities and the broader research community to collectively set standards and guidance on retention, security, and data use terms. Federal agencies should then use these standards as a basis to harmonize their own policies.
9. Agencies should clarify and continue to explicitly note in their calls for proposals that costs to support a program's requirements for data accessibility are allowable as direct or indirect charges in research program budgets.
10. Data access rules from agencies should emphasize adherence to the broadly accepted FAIR principles (data should be findable, accessible, interoperable, and reusable).

Guidance for Research Universities

In light of governmental mandates and the scientific benefits of making data accessible to the public, universities will need to adopt new institutional policies, procedures, and approaches that actively support and promote research data sharing, while at the same time ensuring rigor in the research process and the veracity of its intellectual outputs. Standard research data management practices will need to be developed and adopted across research universities to promote interoperability among institutions and retain academic control of data products and sharing. In support of these goals, we propose that the following **actions** be taken both within and across research universities to manage the growing demand for public access data:

Within a university:

1. In the absence of sponsor guidelines on data access and long-term stewardship, focus on ensuring access to data sets associated with a peer-reviewed scholarly publication and/or which are otherwise delineated in the grantee's data access management plans. Institutions may also need to consider other cases that require long-term data stewardship, such as where the discipline has a reasonable expectation of the data being accessible for future re-analysis by researchers in their discipline or other disciplines, or specific research projects for which data acquisition and dissemination are a primary goal.
2. Develop supporting guidance, services, and workflow systems to guide researchers and support consistent and rigorous data sharing practices within the institution. This includes providing clear guidance to researchers aimed at ensuring uniformity in their data access plans depending upon their discipline and the specific nature of their research. Appendix A provides examples of resources, tools, and guidance for making publications and data publicly available. Considerations should include support for study design, data collection, data quality review, study and data documentation, and depositing and archiving data.
3. Create the capacity for researchers to develop adequate documentation and descriptive metadata following applicable standards to support data sharing. Universities should ensure that this metadata is being created and that it enables researchers to identify a data set and the context of its creation, record and manage its IP, and carry out data stewardship practices.
4. Establish principles and an efficient process by which researchers hand off the stewardship of the data to the institution, where applicable, to store the data or document where data are stored if it is not the researcher's home institution.
5. Establish granular access controls for research data sets that consider how they enable sharing, when that sharing can occur, and with whom (e.g., individuals, a team, an academic unit, a disciplinary community, entities approved via a restricted access proposal process).

6. Develop the capacity to recommend appropriate trusted digital repositories to their researchers for deposit of their data sets, which may be supported by institutions, disciplines/societies, academic consortia, or government agencies. Where no proper repository exists, the institution will need to develop its own data repository service.
7. Implement policies to support public access to data. Policies to support data sharing should be consistent with an institution's research policies, such as those governing the institutional review board, controlled unclassified information, and conflict of interest.
8. Consider promotion and tenure review processes that acknowledge and reward researchers who publish their data sets, given that review and publishing of digital research data has become a mainstay in the research methods of many disciplines and is mandated by federal research sponsors as a required scholarly product.
9. Develop and require the use of core identifiers for data sets as follows:
 - Require a *unique researcher identifier*. We recommend using the platform independent ORCID system. Other identifiers may be useful as long as they are platform independent and are not ID systems that are recognized by that platform only.
 - Require a *unique digital object identifier* (DOI) that associates the data set with the specific project from which it was produced and/or the scholarly works that the data supports. DataCite and EZID are examples of DOI registration providers.
 - Adopt the use of unique, platform independent *funding organization* and *institutional identifiers*. The DUNS number is one system that could be considered to identify research institutions. By establishing such identifiers, institutions can benefit from creating unique identifiers that associate institutions with their principal investigators, research projects, and data sets.

Collectively across universities:

1. To promote practical implementation of public access to data from federally funded research, research universities should work with each other and collectively with federal agencies to promote interagency harmonization of data sharing guidelines and compliance requirements that minimize the burden on researchers and institutions.
2. To enable appropriate public access to data, research universities should create common guidance for researchers and institutions on IP matters, addressing when the institution owns a data set, engages in licensing to a researcher who uses it in his/her research or educational activities, or licenses it to other parties who ostensibly will use the data set for research or educational purposes. For data sets subject to funders' public access policies, the default status is that data is to be shared publicly in accordance with the specified agency policy. If a data set is not subject to a funder's clearly defined public access policy, then access is governed by the terms and conditions agreed upon for that research project by the researcher, research institution, funding organization, and any other pertinent data access plan, policy, regulation, or law.
3. To support interoperability and sharing, universities should work together to define the requirements and methods to provide core identifiers for data products, such as data set, researcher, and funder identifiers. Where pertinent, we encourage institutions to work with interested associations, academies, non-profit research organizations, service providers, and other related entities in developing and implementing a standardized set of identifiers.
4. To accelerate development of credit systems and promotion and tenure policies, universities should work together to define citation and impact evaluation methods to recognize the work of researchers who publicly share their research data as a scholarly product and the subsequent impact of sharing these data.
5. To promote consistency in a disciplinary or scholarly area, universities and their researchers should work with professional societies, government agencies, and other organizations to define disciplinary and other types of standards for data sharing.

APPENDIX A

Data Management Resources

Purpose: The purpose of this appendix is to provide universities with resources, tools, and guidance for making publications and data accessible. The list is not comprehensive.

Select Resources for Data Management Best Practices in Research Universities

- DataONE Best Practices: <https://www.dataone.org/all-best-practices>
- Digital Curation Centre - RDMGuidance Webpages (UK and select US universities): <http://www.dcc.ac.uk/resources/policy-and-legal/rdm-guidance-webpages/rdm-guidance-webpages>
- The UK Data Archive - Create and Manage Data: <https://www.ukdataservice.ac.uk/manage-data>
- Stanford University Libraries - Data Best Practices: <https://library.stanford.edu/research/data-management-services/data-best-practices>
- Johns Hopkins University Libraries - Data Management Resources:
 - <http://dms.data.jhu.edu/data-management-resources/>
- University of Minnesota Libraries - Managing Your Data:
 - <https://www.lib.umn.edu/datamanagement>

Resources on University Public Access Policies and Procedures

- Shieber, S., & Suber, P. (2015). *Good Practices for University Open-Access Policies* (Vol. 1). Retrieved from <http://cyber.harvard.edu/hoap/sites/hoap/images/Goodpracticesguide-2015.pdf>. This is a guide to good practices for university open-access (OA) policies. It's based on the type of policy first adopted at Harvard, Stanford, MIT, and the University of Kansas. Policies of this kind have since been adopted at a wide variety of institutions in North America, Europe, Africa, and Asia, for example, at affluent and indigent institutions, public and private institutions, research universities and liberal arts colleges, and at whole universities, schools within universities, and departments within schools.
- Utah State University has comprehensive [data management services](#), which are organized around Data Management Plans (DMPs). A description of their [procedures](#) and [data workflow](#) are outlined in this [paper](#) and [presentation](#).
- Two reports from the University of Michigan:
 - [An Analysis of Data Management Plans from the University of Michigan](#) contains the findings of a content analysis of 100 data management plans (DMPs) from grants submitted to the National Science Foundation (NSF) written by researchers at the University of Michigan (U-M). The intent behind this analysis was to better understand how researchers at the U-M have interpreted and responded to the NSF's requirement to submit a DMP that describes how they will manage, share and archive the data generated over the course of their funded research.
 - [Data Curation Priorities and Activities: A Report from a Researcher Engagement Event at the University of Michigan](#) contains the results of an event designed to identify researcher priorities, needs in curating data and the activities they take in support of curation.

Selected Institutional Repositories with Data Sets

- Harvard Dataverse. Retrieved from <https://dataverse.harvard.edu/>
- Stanford Digital Repository | Stanford University Libraries. (n.d.). Retrieved from <https://library.stanford.edu/research/stanford-digital-repository>
- UNC-Chapel Hill - Carolina Digital Repository. (n.d.). Retrieved from <https://cdr.lib.unc.edu/>

- Cornell University - eCommons. (n.d.). Retrieved from <https://ecommons.cornell.edu/>
- Purdue - PURR - Home. (n.d.). Retrieved from <https://purr.purdue.edu/>
- Johns Hopkins University - University Data Archives. Retrieved from <https://archive.data.jhu.edu/>
- University of Minnesota Libraries - Data Repository of U of M. Retrieved from <https://conservancy.umn.edu/handle/11299/166578>

Guidelines for Public Access (Federal and International)

- Federal plan analysis:
 - <http://datasharing.sparcopen.org/>
 - <https://libraries.mit.edu/scholarly/research-funders/>
- International guidance:
 - The **TOP Guidelines** provide guidelines to enhance transparency in the science that journals publish. The TOP Guidelines Committee, sponsored by the Center for Open Science, maintains an information commons for transparency standards, serves as an advisory group for journals and funders, evaluate guidelines' effectiveness, and manages guideline updating to maximize quality and interdisciplinary applicability. Retrieved from <https://cos.io/our-services/top-guidelines/>
 - **OECD Principles and Guidelines** for Access to Research Data from Public Funding. These
 - Guidelines provide broad policy recommendations to the governmental science policy and funding bodies of member countries on access to research data from public funding. They are intended to promote data access and sharing among researchers, research institutions, and national research agencies, while at the same time, recognizing and taking into account, the various national laws, research policies and organizational structures of member countries. Retrieved from <http://www.oecd.org/sti/sci-tech/38500813.pdf>
 - Tsoukala, V., Angelaki, M., Kalaitzi, V., Wessels, B., Price, L., Taylor, M. J., & Smallwood, R. (n.d.). Policy Recommendations for Open access to research Data in Europe. *Royal Netherlands Academy of Arts and Sciences Lorenzo Bigagli (National Research Council of Italy. RECODE* provides recommendations for a policy framework to support open access to European research data. Retrieved from <http://recodeproject.eu/>

Data Management Tools

- The **Open Science Framework** (OSF) is a free, open source web application that connects and supports the research workflow, to collaborate, document, archive, share, and register research projects, materials, and data. The OSF was developed by the non-profit Center for Open Science. Researchers can use the OSF to manage their projects and collaborations or register their studies. Institutions can use OSF for Institutions to provide their researchers with a free, open source scholarly commons. Retrieved from <https://osf.io/>
- **Dat** is a grant-funded, open-source, decentralized data sharing tool for versioning and syncing changes to data. Retrieved from <https://datproject.org/>
- **DMPTool** - The Data Management Plan Tool (DMPTool) is an online application that helps researchers create data management plans. The DMPTool provides detailed guidance and links to general and institutional resources and walks a researcher through the process of generating a comprehensive plan tailored to specific DMP requirements. Retrieved from <https://dmptool.org/about>
- **Dataverse Project** is an open source web application to share, preserve, cite, explore, and analyze research data. Each dataverse contains datasets, and each dataset contains descriptive

metadata and data files (including documentation and code that accompany the data). The Institute for Quantitative Social Science (IQSS) collaborates with the Harvard University Library and Harvard University Information Technology organization to make the installation of the Harvard Dataverse openly available to researchers and data collectors worldwide from all disciplines, to deposit data. Retrieved from <https://dataverse.org/>

- **EZID** is a service of the California Digital Library, a division of the University of California Office of the President. It creates and manages long-term, globally unique identifiers for data and sources <https://ezid.cdlib.org/>
- **ORCID** (Open Researcher and Contributor ID) is a nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors. Retrieved from
- **DataCite** is not-for-profit organization which aims to improve data citation by assigning persistent identifiers, such as digital object identifiers (DOIs), to data sets <https://www.datacite.org/>

Directories of Open Access Journals

- [Directory of Open Access Journals](#) is a community-curated online directory that indexes and provides access to open access, peer-reviewed journals.
- [Sherpa/RoMEO](#) aggregates publisher open access policies from around the world and provides summaries of self-archiving permissions and conditions of rights given to authors on a journal-by-journal basis.

Directories of Data Repositories

- Open Access Directory - Directory of data repositories. (n.d.). Retrieved from http://oad.simmons.edu/oadwiki/Data_repositories
- OpenDOAR - Directory of data repositories. (n.d.). Retrieved from <http://www.opendoar.org/countrylist.php?cContinent=North%20America>
- OpenAIRE - A network of Open Access repositories, archives and journals that support Open Access policies. Retrieved from <https://www.openaire.eu/>
- re3data.org. The Registry of Research Data Repositories (re3data.org) is an Open Science tool that offers an overview of existing international repositories for research data. Retrieved from <https://www.re3data.org/search>

Training Resources

Thank you to [Utah State University](#) for this list.

- [MANTRA: Research Data Management Training](#) - MANTRA is a free online course for those who manage digital data as part of their research project. It is managed by the University of Edinburgh. Modules include data protection, rights, and access; sharing and licensing; and metadata and curation.
- [Data Management Short Course for Scientists](#) - ESIP (Earth Science Information Partners) in cooperation with NOAA and the Data Conservancy have compiled the resources for this course.
- [Coursera: Research and Data Management and Sharing](#) - A five-week course designed to provide an introduction to research data management and sharing.
- [Disciplinary RDM Training](#) - This site lists Research Data Management materials specific to five areas (performing arts, archeology and social anthropology, health studies, psychology, and geosciences, social sciences & clinical psychology). Other courses are listed as well. Maintained by the [Digital Curation Centre](#) of the U.K.
- [DataONE Education Modules](#) - DataONE provides several lessons in PowerPoint format available for download that can be incorporated into teaching materials. Also available are webinars and screencast tutorials.

Projects/Initiatives/Organizations/Networks of Interest to Universities

- [Center for Open Science](#) is a non-profit technology company providing free and open services to increase inclusivity and transparency of research. Projects include: Open Science Framework
 - The Center for Open Science (COS) launched [OSF Preprints: The Open Preprint Repository Network](#) in September 2016. OSF Preprints is powered by the SHARE database and API, including search results from arXiv, PeerJ, Research Papers in Economics, and others, in addition to the branded preprint services hosted by COS (PsyArXiv, SocArXiv, and engrXiv).
- [CHORUS](#) (Clearinghouse for the Open Research of the United States) is an effort by the publishers that went live in July 2014. CHORUS enables users to find and access the articles on the **publishers' sites**. It uses DOI and FundRef New articles are continuously added and are freely accessible to the public, either immediately on publication or after a designated embargo period.
- [DataCite](#) is an international not-for-profit organization which aims to improve data citation. The primary means of establishing easier access to research data is by assigning persistent identifiers, such as digital object identifiers (DOIs), to data sets.
- [Data Curation Network](#) is a Sloan-funded project that aims to conceptualize and develop a “network of expertise” model for U.S. academic libraries to collectively provide data curation services to support digital research data deposit into repositories for open access and reuse. The project includes the University of Minnesota (lead), the University of Michigan, the University of Illinois, Cornell University, Penn State University, Washington University in St. Louis, Duke University, and Johns Hopkins University.
- [OASPA | Open Access Scholarly Publishers Association](#) is a trade association that was established in 2008 in order to represent the interests of Open Access (OA) journal publishers.
- [SHARE](#) is a higher education initiative whose mission is to maximize research impact by making research widely accessible, discoverable, and reusable. To fulfill this mission SHARE is building a free, open, data set about research and scholarly activities across their life cycle.
- [SPARC: Advancing Open Access, Open Data, Open Education](#) is an international alliance of academic and research libraries developed by the Association of Research Libraries in 1998 which promotes open access to scholarship.

Members of the AAU-APLU Public Access Working Group

- Lisa Lynch (co-chair), Provost and Maurice B. Hexter Professor of Social and Economic Policy, Brandeis University
- Sarah Nusser (co-chair), Vice President for Research, Iowa State University
- Sandra Brown, Vice Chancellor for Research and Distinguished Professor, University of California, San Diego
- Jeff Chasen, Associate Vice Provost, Integrity & Compliance, University of Kansas
- Debasish Dutta, Chancellor and Distinguished Professor of Engineering, Rutgers University-New Brunswick
- James Hilton, Vice Provost for Academic Innovation, Dean of Libraries, and Arthur F. Thurnau Professor and Professor of Information, University of Michigan
- Farnam Jahanian, Interim President, Carnegie Mellon University
- David Lifka, Vice President for Information Technologies and Chief Information Officer and Director, Cornell Center for Advanced Computing, Cornell University
- James Luther, Associate Vice President Finance & Compliance officer, Duke University
- Mark McLellan, Vice President for Research & Dean of Graduate Studies, Utah State University
- Daniel Reed, Vice President for Research and Economic Development, University of Iowa
- Peter Schiffer, Vice Provost for Research and Professor in Applied Physics, Yale University
- Tyler Walters, Dean, University Libraries, Virginia Polytechnic Institute and State University
- Susan Wentz, Provost and Vice Chancellor for Academic Affairs and Professor, Vanderbilt University
- Brad Wheeler, Vice President for Information Technology & CIO; Professor, Information Systems, IU Kelley School of Business, Indiana University

Association Staff:

- Kacy Redd, Assistant Vice President, Science & Mathematics Education Policy, Association of Public and Land-grant Universities
- Tobin Smith, Vice President for Policy, Association of American Universities
- Jessica Sebeok, Associate Vice President for Policy, Association of American Universities
- Jackie Bendall, Director, Research Compliance & Administration, Council on Governmental Relations