Campus Guide to Accelerate Public Access to Research Data

[Forward letter from AAU President Barbara Snyder and APLU President Peter McPherson will provide additional history and context for the APARD effort and the motivation for the development of this guide.]

Introduction

"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…. 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."

--Report of the AAU-APLU Working Group on Public Access, November 2017

Ensuring broad based public access to research data is fundamental to advancing the research, education, and service missions of institutions of higher education. Public access to research data is, in fact, a continuation of the role of academic institutions conducting research and their function in creating and disseminating new knowledge for societal and economic benefit. As an element of open scholarship, public access to research data can help accelerate the pace of discovery and its application to societal problems, as well as heighten the visibility and reputation of the institution and its scientists and scholars. As stewards of taxpayer dollars and innovators in research, research institutions must meet, and make good faith efforts to exceed, public expectations and government mandates regarding access to the results of their research and scholarship.

Data access is also important to preserving scientific integrity, ensuring transparency and public trust in science, and facilitating the reproducibility of scientific results. Instituting and sustaining public access policies and practices will create more opportunities for other scientists to examine, test, evaluate, and validate the research methods, data, and scientific findings of research performed by their colleagues. The rapid availability of research outputs also opens the potential for new research collaborations that extend the work of a researcher or introduce new threads of inquiry.

While a consensus has begun to emerge among federal policymakers and many in the university and scientific community regarding the value of making research data publicly accessible, many barriers still exist to achieving this goal. Overcoming these barriers will require: a commitment of resources by both universities and federal research agencies; the development of new institutional data policies; the creation of new data repositories and enhanced data storage infrastructure; and a major cultural shift within universities, scientific disciplines, and specific university departments concerning how individual faculty members are evaluated, assessed, recognized, and rewarded. Additionally, greater coordination between campus stakeholders in data management will be essential to ensuring broad-based data

accessibility, from the university provost, senior research officer, and chief information officer to general counsels, compliance officers, librarians, faculty members and students.

This guide seeks to provide universities with resources, tools, and information to help to accelerate public access to data and to assist universities in crafting a consistent and uniform approach to data management on their respective campuses. It is also hoped that it will facilitate the development of standard research data management practices at AAU and APLU member campuses that will promote interoperability among institutions, and which will ensure that they are able to retain academic control of research and data products and sharing. Ultimately, it is the hope of AAU and APLU that this guide will help to facilitate that adoption of 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.

Why Public Access to Research Data is Important

Public access to research data is part of an evolving transformation to modernize how researchers and funders create openness and transparency in scholarship. Spurred on by the internet's expanding capacity for collaboration and sharing, the open science movement generates significant potential to accelerate the pace of discovery and its application to societal problems, and is increasingly seen as necessary to address growing concerns about research integrity and waning public trust in research findings.

Although the process of sharing research data impacts research practice for most disciplinary and interdisciplinary research and requires academic institutions to expand and create new research support systems, it is strongly aligned with priorities held by researchers and their home institutions to generate high quality research outputs of value to society, as well as support the scholarly community's and public's expectations of rigor and transparency in scholarship.

Accelerating Discovery and Innovation

Funders and the public seek return on their investments through timely release of research findings, data and other outputs. By providing public access to data and other research outputs, the best science is accessible as quickly as possible and can be used to accelerate scientific advances that address scientific and societal challenges. This confers benefits to researchers and institutions through increased visibility, citations and impact associated with publicly accessible research findings [include citation], and the rapid availability of research outputs opens the potential for new collaborations that extend the researcher's work or introduce new threads of inquiry.

Increasing Rigor and Public Trust

Research institutions, funders, disciplinary communities and the public expect scholarship to rigorous and defensible. The reproducibility crisis and concerns about other forms of research integrity have led to waning public trust and highlighted the need to increase transparency in scholarship. Federal agencies now see transparent sharing of well-documented data as key to addressing issues of research integrity. Transparency enables others to understand the context (goals), process (methods) and products (article, data, code, …) and evaluate the quality, relevance and limitations of research for the specific question at hand. In addition, by embedding the intent to share data and other research outputs in study planning

and design phase, actions to meet transparency expectations essentially create and reinforce a research process that increases quality of work.

Meeting Sponsor Requirements

Of course, an important obligation of sponsored funding is to meet requirements to share research data when confidentiality, proprietary information and national security can be protected. While the federal mandate to provide public access to research data has not involved much compliance activity to date, major funders are now providing specific about expected behaviors for data sharing during proposal, sponsored project, and post-award phases. Any research institution knows that failing to systematically comply with common contractual requirements will impact the capacity of their researchers to be successful in seeking funds and potentially create onerous administrative burden for the institution.

[Quotations from *Open Science by Design*, NASEM 2018]

Recommendations to Accelerate Public Access to Research Data at Research Universities

[Each recommendation is followed by a set of initial actions, a selection of case examples, and a listing of additional resources. We are actively seeking illustrative institutional and organizational programs, policies, and resources that can be highlighted as case studies, examples, and/or additional resources.]

1. Establish an Effective Planning, Consultation, and Decision-making Structure

The foundational step toward increasing public access to research data is the clear articulation by the highest levels of campus administration of the importance to the institution of a culture of sharing research data. Further, a critical element of developing integrated and effective institutional research data planning, management, and sharing (aka “data stewardship”) policies, practices, and procedures is ensuring that all relevant campus stakeholders are involved in the creation of a multi-year implementation plan. Public access to research data is an issue that impacts a broad spectrum of communities and interests at any college or university, ranging from academic affairs, research, and information technology to faculty, students, librarians, data stewards, data scientists, legal counsel and many others. Universities are well positioned to address many of the technical and ethical challenges involved in making data public, especially where data privacy is a major concern. Therefore, it is important to establish a structure for coordinating the planning, communication, consultation, and decision-making that will capture and consider the perspectives, priorities, and concerns of the various units and individuals on campus most affected by the research data stewardship practices and procedures being developed. Senior leaders should also clearly define the roles and responsibilities of the participating campus stakeholders, including clearly defining who will manage ("own") the planning and development process.

[Quotation from campus administrator or the importance of creating an effective, coordinated process to develop a public access policy.]

DRAFT - ACCELERATING PUBLIC ACCESS TO RESEARCH DATA

Initial Actions

- Have President/Provost articulate – and regularly re-affirm - a strong commitment by the institution to prioritize and support Open Scholarship
- Identify existing campus organizations and structures that would be central to the planning, development, implementation, and governance of research data stewardship policies and practices
- Establish an inclusive, institution-wide body (i.e., "Coordinating Committee") comprising relevant stakeholders to develop the campus research data policy (including faculty and students)
- Articulate a clear charge to the Coordinating Committee regarding its roles, responsibilities, and measures of success/progress
- Identify the lead unit and individual to coordinate the planning and development process
- Identify adequate resources to support the planning and development process

Case Examples

[The following are a sample of case study examples. We welcome additional case study examples in support of the recommendation.]

Cross-campus Coordinating Committee (*Virginia Tech*)

Advisory Group - made up of members who can represent the interests of the Provost, the Office of the Vice President for Research and Innovation, the Dean of the University Libraries, and the Vice President of Information Technology.

Public Access to Research Data Committee - to fully explore public access to data at Virginia Tech, and to create a policy recommendation and/or a strategic implementation plan for better supporting researchers in sharing their data. Membership of this Committee should include representation from: 1) Research Compliance; 2) Division of Information Technology; 3) University Libraries; 4) Commission on Research; 5) Institutes (at least 1); 6) Faculty; 7) Research Faculty; 8) Graduate School; 9) University Legal Counsel; 10) Advanced Research Computing (in addition to other Division of IT representation)

High-level "charge" to Cross-campus Coordinating Committee (*University of New Hampshire*)

The charge to the working group will be to review relevant policies and recommend processes and tactics to ensure compliance with funder requirements for public access to research data. Actions that we identified as a high priority for this group include the following:

1) Ensure PIs are complying with Data Management Plans (DMPs) and sponsor public access requirements. Establish a process to ensure compliance. Capture DMPs that get submitted electronically with proposals.
2) Define university policy/guidance for considering retention/reusability/appraisal criteria for research data.
3) Establish the ability to mint Digital Object Identifiers (DOIs) through the UNH library (only 1 other university out of 30 at the workshop did NOT issue DOIs).
4) Develop guidance on data workflow for PIs. Get workflow integrated into DMPs.
5) Identify long term storage solutions for unpublished data (which doesn't have same public access requirement currently from federal agencies).

6) Identify procedures for preservation, public access, and repository solutions.
7) Deploy active storage repository, connect to larger workflow from 4.
8) Identify policies, practices and procedures to ensure data integrity:
    a. Define requirements for integrity controls on public access data.
    b. Ensure suitable security controls exist for all storage environments
    c. Develop standards/processes to vet [security/copywrite] external repositories to ensure compliance synergies with UNH and funding agencies.
    d. Identify processes for data integrity QA/QC at each workflow stage and during transfer to repositories. Provide guidance for researchers on QA/QC methodologies.
9) Develop processes for training UNH researchers (faculty, staff and students) in emerging principles of data management i.e. data is a valued university asset.
10) Ensure that UNH data, which has been publicly archived/published/shared by our researchers, is recorded in the Data Catalog in the Scholars' Repository.
11) Ensure shared data citations are visible and accessible on faculty profiles, and that faculty receive credit during Promotion and Tenure (P&T) processes.
12) Coordinate with other DMSC Working Groups, in particular the Data Coordination and Communication Committee, for establishing a communication plan.

Additional Resources

[The following is a sample of an additional resource. We welcome additional resources in support of the recommendation, for example "charges" to institution-wide bodies.]

- https://info.hsls.pitt.edu/updatereport/2013/july-2013/public-access-open-access-outlining-the-key-differences/

2. Establish Institutional Research Data Policies - Align Practice with Policy

Developing and implementing campus-wide research data stewardship practices must rest on the complementary foundations of: 1) ease of use and compliance by faculty/researchers; and 2) ease of access by external researchers and the general public. Underlying those practices and procedures must be an institutional research data policy that defines public access to research data, is actionable, and reflects, to the greatest extent possible, existing (and evolving) standard operating procedures across disciplines and funders. The policy should strive to impose the least amount of administrative burden on faculty/researchers and staff and be mindful of the variations in research data stewardship practices across disciplines and sponsors. The policy should be clear and concise and use – defining where necessary – common terminology (i.e., research data, data management, data sharing, curation, time period of data storage) that will be understood by all relevant members of the campus community. The policy should emphasize the importance and benefits of managing research data from "Day 1" (ideation to pre-proposal to proposal to award to post award), and link to practical guidelines/solutions to help with this effort. Development of the policy should include regular consultation with faculty/researchers, administrators, and service providers to ensure the policy is relevant to and implementable across the institution.

[Insert Quotation]

Initial Actions

- Begin a collaborative and inclusive process to develop a concise research data policy that clearly defines the roles and responsibilities of individuals and organizations across the campus.
- Connect the research data policy to the research data workflow developed by the Coordination Committee
- Apply existing (or develop new) mechanisms to integrate faculty/researcher and administrator perspectives into the policy development process
- Define a comprehensive strategy around awareness, compliance, and monitoring regarding the campus research data policy

Case Examples

[Case studies will be developed. We welcome case study examples in support of the recommendation.]

Minnesota has a research data management policy worth mention: https://policy.umn.edu/research/researchdata

Virginia Tech Policy 130015 on Ownership and Control of Research Results is currently undergoing revision and will say more about research data management

Additional Resources

[The following is a sample of an additional resources. We welcome additional resources in support of the recommendation.]

- https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs
- University of Minnesota Research Data Management Policy

3. Develop an Effective Research Data Stewardship Implementation Plan

Once an effective planning, messaging, and implementation, structure has been established, the campus can turn to the task of crafting an effective and implementable plan to provide support for research data stewardship, one that recognizes existing institutional and funder policies and practices. The plan should focus, at least initially, on the research data underlying scholarly publications, and should include guidance for researchers and administrators encompassing the entire research data life cycle – i.e., collection, management, sharing, curation, maintenance, and storage. While acknowledging the differences among academic disciplines and scientific fields in how research data is collected, managed, and stored, campuses should strive to plan for practices and procedures that would be applicable across the institution, regardless of discipline or field. It is strongly recommended that campuses avail themselves of the practices and procedures already developed by a number of institutions, scholarly societies, and other organizations (See: Additional Resources below) to help inform their efforts. The development of research data stewardship practices and procedures should include regular consultation with relevant academic and administrative communities impacted by such practices.

[Insert Quotation]

Initial Actions (to be conducted by Coordinating Committee)

- Develop a high-level workflow for research data stewardship to serve as a framework for consideration of human, technical, and financial resource issues
- Agree on common terminology around research data stewardship (e.g., what data falls under policy, what is meant by "data lifecycle")
- Inventory existing research data stewardship policies and practices on campus and across funders
- Establish a plan to develop research data stewardship practices and procedures that build upon existing activities, as appropriate

Case Examples

[Case studies will be developed. We welcome case study examples in support of the recommendation.]

Additional Resources

[We welcome resources in support of the recommendation.]

4. Provide Services and Support: Expert Resources

The successful implementation of effective research data stewardship practices and procedures rests, in part, on identifying, marshalling and, if necessary, identifying the appropriate *expertise* to make it as easy as possible for faculty, staff, and students to follow the established practices and procedures. It is critical for campuses to recognize at the outset that no single person or department understands or is able to implement all aspects of research data stewardship. Therefore, a broad array of individuals from across the campus must be assembled to ensure the successful application of research data stewardship practices and procedures. The needed expertise will come from a range of perspectives – e.g., research librarians, data stewards, data scientists, software and technology developers and programmers, research administrators, information technology officers, data security and privacy experts, and more – to ensure the quality, integrity, security, and accessibility of the research data being shared and managed.

[Insert Quotation]

Initial Actions

- Utilize the high-level research data stewardship workflow developed by the Coordinating Committee to serve as a framework to identify the required expertise
- Identify current expertise that can be deployed to promote and implement effective data stewardship practices
- Identify gaps in expertise needed to ensure that the campus has the technological capabilities to implement effective research data stewardship practices and procedures, and the capacity to assist, guide, and train faculty, staff, and students in appropriate research data stewardship
- Begin to implement research data stewardship practices that include the use of personal Persistent Identifiers and data set Object Identifiers

Case Examples

[Case studies will be developed. We welcome case study examples in support of the recommendation.]

Additional Resources

[The following is a sample of an additional resources. We welcome additional resources in support of the recommendation.]

- Odom Institute – UNC Chapel Hill
- *The Data Curation Network is a useful approach to help address gaps in expertise through sharing of curation expertise across institutions. https://datacurationnetwork.org/*
- *VT University Libraries provides a variety of research data management expertise and liaises with other groups around campus. This linked paper gives a good overview.*
- *VT University Libraries also provides undergraduates experiences in research data management that enhance our services to the community. See this linked paper*

5. Provide Services and Support: Technical Resources

The successful implementation of effective research data stewardship practices and procedures rests, in part, on identifying, marshalling and, if necessary, building the appropriate *technological resources* to make it as easy as possible for faculty, staff, and students to follow the practices and procedures being developed. An overarching goal is to create a campus system that enables scientists, researchers, students and staff to: 1) obtain and utilize personal and object identifiers (e.g., PIDs, DOIs, RORs; Grant IDs); 2) create research data management plans for proposal submissions and/or to satisfy campus data stewardship protocols, using forms that could be interpreted and acted upon by software systems; 3) curate the data and documentation generated by research projects; 4) locate and utilize suitable repositories (on or off campus) to store and maintain the data underlying published research; and 5) streamline the creation of contracts, Data Use Agreements, and other forms of data governance that could help accelerate data access and sharing with critical data service providers. Campuses should also consider how to leverage external resources (e.g., discipline-based or other external repositories), as well as partnerships with other institutions, government agencies, non-profit and for-profit providers, and consortia to share resources and capabilities.

Initial Actions

- Utilize high-level workflow developed by "Coordinating Committee" for data stewardship sharing to serve as a framework for consideration of *technical resource* issues
- Identify current *technical resources* that can be employed to implement effective data management and data sharing
- Identify gaps in *technical resources* needed to ensure that the campus has the technological capabilities to implement effective data stewardship practices and procedures, and the capacity to assist, guide, and train faculty, staff, and students in appropriate data stewardship activities
- Create a centralized web-based portal that will link users to the various campus administration systems (research, IT, libraries, repositories) critical to data stewardship processes
- Identify potential external partners to share technical and technological resources and associated costs

- Facilitate institutional membership in ORCID, DataCite, CrossRef or other member-based identifier infrastructures providers

[Insert Quotation]

Case Examples

[Case studies will be developed. We welcome case study examples in support of the recommendation.]

Additional Resources

[The following is a sample of an additional resources. We welcome additional resources in support of the recommendation.]

- Duke University – MyResearchHome
- Cornell University – Cornell Data Storage Finder
- Clowder: https://clowder.ncsa.illinois.edu/
- Brown Dog: https://browndog.ncsa.illinois.edu/
- Illinois Data Bank // http://research.depository.duke.edu
- http://privacy.ucsd.edu // DMPTool // EZDMP
- SPARC* Landscape Analysis: The Changing Academic Publishing Industry – Implications for Academic Institutions
- Data Repository for the U of M (DRUM): https://conservancy.umn.edu/handle/11299/166578
- UMN Select Digital Storage Options tool (based on Cornell's Data Storage Finder) -- https://it.umn.edu/services-technologies/comparisons/select-digital-storage-options

6. Provide Services and Support: Financial Resources

Underlying the effective implementation of research data stewardship practices and procedures is identifying and securing the financial resources necessary to support the expert and technological requirements to put those protocols into practice. A thorough examination of the costs involved in campus-wide research data stewardship – both inward facing (faculty/researcher use) and outward facing (public access) - and how those costs are allocated, is critical to successful implementation. Based on the analyses of Expertise and Technological Resources, each campus needs to determine the initial and annual, sustaining investments required to successfully implement research data stewardship practices and procedures. Such costs will include policy management, compliance tracking, IT costs, library costs, storage and curation costs, personnel, etc. A number of resources and templates currently exist that could provide templates or guidance to campuses developing research data stewardship costing models (See Additional Resources below).

[Insert Quotation]

Initial Actions

- Develop a multi-year costing model for the staff and infrastructure necessary to implement effective research data stewardship practices across the institution

- Begin to direct/encourage researchers to include budget lines in proposals for research data curation and storage – separate from the research budget – where allowed by funders
- Develop one or two pilot efforts to gain a clearer understanding of the component costs of effective research data stewardship programs
- Examine the cost and benefit of using domain or other external repositories, including cloud-based, compared to on-campus repositories
- Explore potential consortia approaches to research data storage and management as ways to control costs

Case Examples

[Case studies will be developed. We welcome case study examples in support of the recommendation.]

Additional Resources

[The following is a sample of an additional resources. We welcome additional resources in support of the recommendation.]

- Data Curation Network
- Rocky Mountain Advanced Computing Consortium
- Campus Research Computing Consortium (CaRCC)
- https://www.ukdataservice.ac.uk/manage-data/plan/costing
- https://wiki.dpconline.org/index.php?title=Digital_Preservation_Business_Case_Toolkit
- *NASEM Report "Forecasting Costs for Preserving, Archiving, and Promoting Access to Biomedical Data" (https://www.nationalacademies.org/our-work/forecasting-costs-for-preserving-archiving-and-promoting-access-to-biomedical-data)*

7. Build Capacity and Skills for Research Data Stewardship

Building the capacity and skills among faculty/researchers and graduate students is a critical element to successfully implement research data stewardship practices and procedures. Skills development includes creating a communications program to inform the campus of the research data policy, and the infrastructure and support services underpinning the policy, as well as educating faculty/researchers and training students on the most effective research data stewardship practices and procedures. Faculty/researcher education should begin as early as the recruitment (interviewing) and on-boarding of new hires, as well as workshops, seminars, and/or consultations at the laboratory level or through one-on-one interactions. The training of graduate and undergraduate students could occur through stand-alone modules and workshops and/or by incorporating research data management curriculum into existing responsible conduct of research training and other trainings offered by the institution.

[Insert Quotation]

Initial Actions

- Determine administrative roles and responsibilities regarding research data stewardship education and training

- Develop consensus of research data stewardship training priorities and goals
- Identify specific opportunities to educate and inform faculty/researchers – especially Early Career scholars – of the research data policy and the importance of effective research data stewardship practices. Such opportunities could include the hiring/on-boarding process, department level seminars, workshops, information sessions, and other professional training opportunities
- Identify research data stewardship education and training opportunities for graduate (and undergraduate) students, (e.g., Responsible Conduct of Research modules)
- Develop one or two case studies, using campus-based examples, highlighting exemplary research data stewardship practices, including descriptions of the outcomes/benefits of the practices employed.

Case Examples

[Case studies will be developed. We welcome case study examples in support of the recommendation.]

Additional Resources

[The following is a sample of an additional resources. We welcome additional resources in support of the recommendation.]

- CaRCC's Research Computing and Data Professionals: Job Elements and Career Guide -- https://carcc.org/wp-content/uploads/2019/01/CI-Professionalization-Job-Families-and-Career-Guide.pdf
- Many units across VT help to build capacity for better research data management, including the Libraries – we can provide example presentations as is useful, for PDN sessions, guest lectures in courses, and for SIRC
- Paper on building capacity for field projects, *(https://datascience.codata.org/articles/10.5334/dsj-2019-043/)*

8. Achieve Buy-In to the Research Data Policies and Research Data Stewardship Practices and Procedures

A significant challenge to successful implementation of a campus research data policy is achieving "buy-in" by faculty/researchers regarding the necessity and benefits of practicing effective research data stewardship methods. While many look to altering the promotion, tenure, and review processes to place new or more emphasis on research data stewardship, institutions may be able to take a number of earlier steps to realizing better research data stewardship. Communications should emphasize how the systems, services, and support being established across the institution are designed to minimize the burden and otherwise make the adoption of new research data stewardship practices and procedures as easy and seamless as possible. Messaging should also emphasize how effective research data stewardship practices and procedures can contribute to scholarly excellence, increase the value and visibility of one's research, make scientific collaboration easier and more efficient, etc. Institutions should seek opportunities to engage with faculty, students, and other researchers within individual research laboratories, centers, or institutes, as well as within departments, schools, and disciplines.

[Insert Quotation]

Initial Actions

- Create a research data stewardship communications and marketing plan.
- Develop materials to support that plan.
- Identify research data stewardship "champions" to share "success stories" that support institutional data sharing goals and who would have a positive impact on faculty/researchers; champions could be on-campus or off-campus, leaders in specific disciplines, or meet other criteria
- Identify or create opportunities to highlight/celebrate exemplary research data stewardship practices newly adopted by faculty/researchers (recognition, awards)
- Establish a study committee or other mechanism to explore how research data stewardship could be incorporated into annual review, promotion, and tenure processes (see DORA)

Case Examples

[Case studies will be developed. We welcome case study examples in support of the recommendation.]

Additional Resources

[The following is a sample of an additional resources. We welcome additional resources in support of the recommendation.]

- *Cambridge University Data Champions program*

Achieving Cultural Change

[This section will be developed after the virtual workshop. It will focus on establishing institutional data policies, aligning practices to policies, building awareness and ownership, achieving buy-in, and the role of federal agencies.]

Conclusion