NECA NORTHEAST CONFERENCE ON COLLEGE COST ACCOUNTING

2021 NECA Conference

Public Data Management & Sharing – Costing Considerations

Presenter: Jim Luther, Associate VP Finance & Research Costing Compliance Officer, Duke University

September 14, 2021 (2:10 p.m. -2:55 p.m.)

Public Data Management & Sharing – Costing Considerations

- **Description:** Starting in 2013, with OSTP's release of the "Holdren memo" that directed institutions to make the results of federally funded research available, institutions and federal partners have been struggling with "How" to do this, including topics related to rigor and reproducibility, tenure and recognition, infrastructure implications, numerous data security issues, etc. etc.

- Today, based on discussions with various advocacy groups (AAU, COGR, APLU, FDP, etc.) that have occurred over the past 8 years, **we will focus on the "Costing" aspects of how institutions can/should consider funding this expanding expense**; this will include discussions related to direct and indirectly charging these costs. Critical to this is the continued partnership with federal entities who also acknowledge the difficulty inherent in undertaking this critical initiative so important to the continued success of research in the US and around the world.

Agenda

- Setting the stage - Overview
- National Issues and Background
- Duke University Example (from APARD meeting in 2018)
- Costing Issues and Discussion

NOTE:

Background – FYI
(Provided as context but not to be discussed in detail)

Overview – A Quintessential Costing Issue

- Big $
- Aspects of both direct and indirect charging
  - Requires complex decision-making to allocate
- Evolving science, processes and regulatory environment…
- **Complex internal control environment**
  - Multiple purchasing mechanisms likely
  - Implications on pre, post, and after end of award (e.g. data storage)
  - Central and departmental costs
- **Multiple cost pools: Library, DA, GA, O&M, Equipment, & Base**
- **Lifecycle is broad and complex**

Cost Implications: Lifecycle Public Data Access Activities

#	Activity	Timing	Sponsor Pay			Institution Pay				External Repository	Other
	Lifecycle Public Data Access Activities	Timing (Pre-Proposal submission, Life of award, Post-Closing)	Direct Charge to Sponsor as direct line item or via Service Center	Separate Supplement / Companion Award (with different period of performance) for Data Storage after period of performance	Budget Line Item for Data Only that is Paid/Accrued at end of Award for Future Data Costs (would require OMB UG approval)	Service Center (likely subsidized by institution but charged to project)	Institution Pay (Admin Capped)	Institution Pay (Uncapped: O&M or Library)	Institutionally Supported Repository	Publisher / Discipline / Sponsor / Professional Society / One Time (Fig Share)	Institution Covers Cost but then Charges User
1	DMP Development	PRE - PROPOSAL	No								
2	Data Curation & Metadata Curation FAIR, Data dictionary, etc.	LIFE (SOME PRE)	Some sp… allow							N /	
3	Data Ingest	LIFE	Ho…								
4	DMP Monitoring & Compliance through life of award through closeout	LIFE	Hopefully								
5	Data Storage (during life of project)	LIFE	Probably yes								
6	Data Processing	LIFE	Probably yes								
7	Data Storage (post-closeout for publication)	POST/LIFE	Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance								Contrary to open access principles depending on implementation (NEEDS MORE DISCUSSION)
8	DMP Monitoring & Compliance - post closeout	POST	Probably No								
9	Data Storage (post-closeout for DMP Compliance)	POST	Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance								
10	Cold Data Storage (post-closeout / last resort)	POST	Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance								
11	Publication Fees (often based on size and duration of data)	POST	Probably yes								
12	Data Security (PHI, HIPAA, Export Controls, FISMA, student data and IP)	PRE, LIFE & POST	Varies								

A B C D E F G H I J

Draft

In-Process

#	Activity	Timing
	Lifecycle Public Data Access Activities	**Timing (Pre-Proposal submission, Life of award, Post-Closing)**
1	**DMP Development**	PRE - PROPOSAL
2	**Data Curation & Metadata Curation FAIR, Data dictionary, etc.**	LIFE (SOME PRE)
3	**Data Ingest**	LIFE
4	**DMP Monitoring & Compliance through life of award through closeout**	LIFE
5	**Data Storage (during life of project)**	LIFE
6	**Data Processing**	LIFE
7	**Data Storage (post-closeout for publication)**	POST/LIFE
8	**DMP Monitoring & Compliance - post closeout**	POST
9	**Data Storage (post-closeout for DMP Compliance)**	POST
10	**Cold Data Storage (post-closeout / last resort)**	POST
11	**Publication Fees (often based on size and duration of data)**	POST
12	**Data Security (PHI, HIPAA, Export Controls, FISMA, student data and IP)**	PRE, LIFE & POST

Cost Implications: Lifecycle Public Data Access Activities

Sponsor Pay			Institution Pay				External Repository	Other
Direct Charge to Sponsor as direct line item or via Service Center	Separate Supplement / Companion Award (with different period of performance) for Data Storage after period of performance	Budget Line Item for Data Only that is Paid/Accrued at end of Award for Future Data Costs (would require OMB UG approval)	Service Center (likely subsidized by institution but charged to project)	Institution Pay (Admin Capped)	Institution Pay (Uncapped: O&M or Library)	Institutionally Supported Repository	Publisher / Discipline / Sponsor / Professional Society / One Time (Fig Share)	Institution Covers Cost but then Charges User
No								
Some sp[illegible]rs allow							N / A	Contrary to open access principles depending on implementation (NEEDS MORE
Hopefully								
Hopefully								
Probably yes								
Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance								
Probably No								
Probably no unless feds allow								

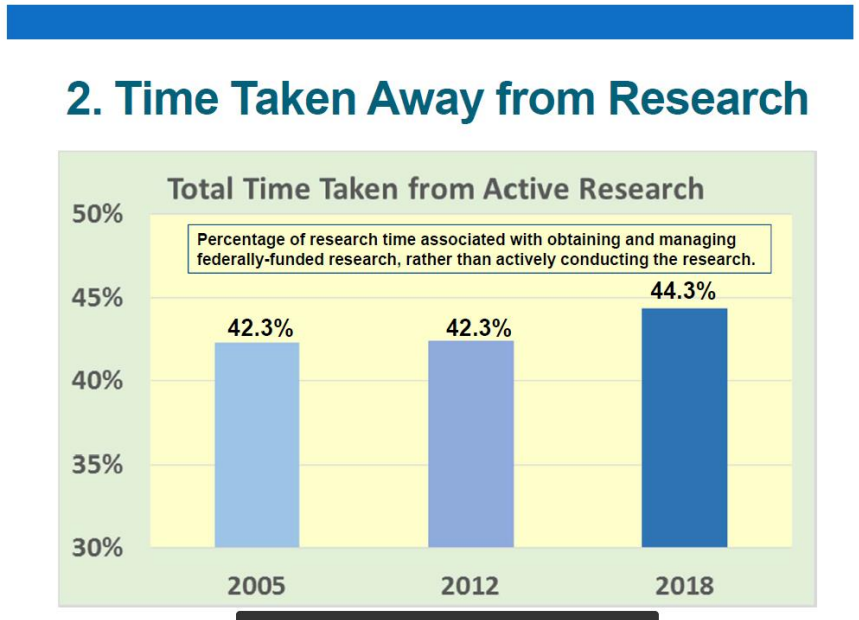
B C D E F G H I J

Draft

FEDERAL DEMONSTRATION PARTNERSHIP

Redefining the Government & University Research Partnership

- **Survey Description:** Meeting federal requirements for **resource and data sharing**
- **Quantitative Responses:**
  - **Prevalence:** **49%** experienced time taken away from research by this responsibility.
  - **Substantial Workload:** **38%** of those experiencing this responsibility reported "some" to "very much" time taken away from active research by the responsibility.



Institutional Example (2018)

- Light microscopy: 30-100Gb/experiment, 100 experiments/researcher, 20-30 researcher/yr. Projection: 300Tb/yr
- CryoEm: Potential storage needs of ~400Tb/yr

Size / Timeframe	Annually	5 Years (one time)	7 years (one time)	Perpetual (one time)
	$0.515/GB	$2.58/GB	$3.61/GB	$12.88/GB
100 GB	$51	$258	$361	$1,288
512 GB	$263	$1,320	$1,848	$6,594
1,024 GB (1 TB)	$527	$2,641	$3,696	$13,189
5,120 GB (5 TB)	$2,636	$13,209	$18,483	$65,945
51,200 GB (50 TB)	$26,368	$132,096	$184,832	$659,456
102,400 GB (100 TB)	$52,736	$264,192	$369,664	$1,318,912

Excludes:
Curation
DMP Support
Tech Support

APARD

2017 – 2020: AAU-APLU Public Access Working Group (PAWG) Report and Recommendations

https://www.aau.edu/accelerating-public-access-research-data

August 6, 2021

Accelerating Public Access to Research Data

PUBLIC/OPEN ACCESS

The Association of American Universities (AAU) and the Association of Public and Land-grant Universities (APLU) have collaborated and led national discussions to improve public access to data resulting from federally funded research. The current Guide to Accelerate Public Access to Research Data builds on many prior efforts and is consistent with national and global open science efforts as well as international declarations, such as the Sorbonne declaration on research data rights.

- In 2016 AAU and APLU formed a working group to examine issues relating to public access to the research results. This working group examined how to improve public access to data resulting from federally funded research. In 2017, the group issued a report with a series of recommendations to universities on how to increase public access to research data on their own campuses and how they also work together to advance these efforts. The group also made recommendations concerning how federal agencies could help facilitate sharing of research data at universities.

- In 2018, APLU and AAU hosted a National Science Foundation (NSF) funded workshop (NSF #1837847) that convened cross-institutional teams from 30 universities with the goal of developing campus-specific strategies for making data resulting from federally funded research publicly available. The two associations issued a report chronicling learning from the workshop.

Conversation Series:

Join AAU and APLU for a series of conversations with campuses that are advancing public access to research data. We will explore as a community strategies that align with the recommendations in the Guide to Accelerate Public Access to Research Data.

LEARN MORE

APARD and Other Recent Activities

1

Association of American Universities
Inquiry · Innovation · Impact

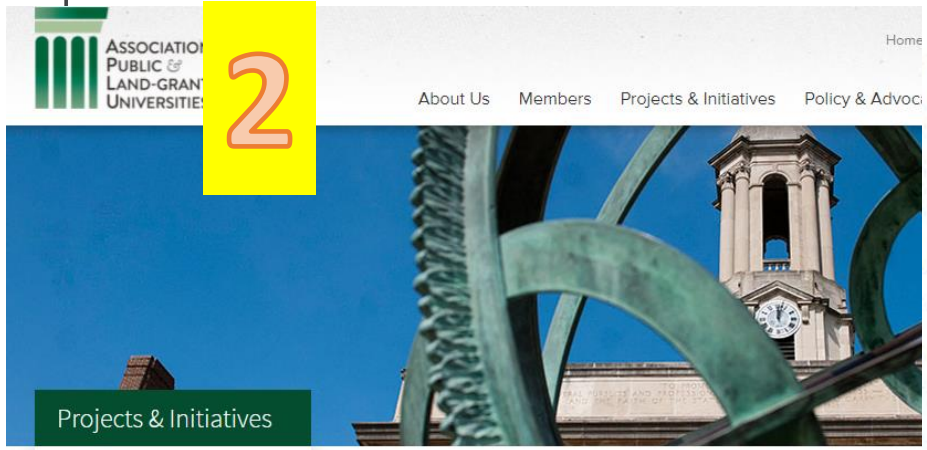
ASSOCIATION OF PUBLIC & LAND-GRANT UNIVERSITIES

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

2



Kacy Redd, Ph.D.
Associate Vice President, Research & STEM Education
202-478-6022

HOME > PROJECTS & INITIATIVES > RESEARCH, SCIENCE & TECHNOLOGY > PUBLIC ACCESS TO

Public Access to Research Data

APLU Impact

- **APLU and the Association of American Universities (AAU) are working with res improve public access to federally funded research data.**
- **Drawing on insights from leading institutions, APLU and AAU have issued seve steps the federal government and public universities can take to increase pub data.**

3a

Accelerating Public Access to Research Data

WORKSHOP REPORT

This report provides the summary outcomes from a 2018 workshop supported by the National Science Foundation, Association of Public and Land-grant Universities (APLU), and the Association of American Universities (AAU) , which brought together 30 institutional teams of research officers, librarians, information technology officers, general counsel, and faculty. These teams developed institutional plans to make publications and data more accessible.

Kacy Redd, Katie Steen, Sarah Nusser, Toby Smith, Tyler Walters, Jeff Chasen, James Luther, and James Reecy

3b

Guide to Accelerate Public Access to Research Data

NIH Releases New Policy for Data Management and Sharing

Background – FYI
(Provided as context but not to be discussed in detail)

- Today, nearly twenty years after the publication of the Final NIH Statement on Sharing Research Data in 2003, we have released a *Final NIH Policy for Data Management and Sharing.*
- We hope it will be a **critical step in moving towards a culture change, in which data management and sharing is seen as integral to the conduct of research.**
- Responsible data management and sharing is good for science; it maximizes availability of data to the best and brightest minds, underlies reproducibility, honors the participation of human participants by ensuring their data is both protected and fully utilized, and provides an element of transparency to ensure public trust and accountability.

COGR
Council On Governmental Relations

NIH Policy Notices & Supplemental Information

Background – FYI
(Provided as context but not to be discussed in detail)

- *Released October 29, 2020, Effective January 25, 2023*
  - NOT-OD-21-013 - Final NIH Policy for Data Management and Sharing
    - **Two main requirements (1) the submission of a Data Management and Sharing Plan (Plan); and (2) Compliance with the approved Plan.**
  - NOT-OD-21-014 – Supplemental Information to the NIH Policy for Data Management and Sharing: **Elements of an NIH Data Management and Sharing Plan**
  - NOT-OD-21-015 – Supplemental Information to the NIH Policy for Data Management and Sharing: **Allowable Costs** for Data Management and Sharing
  - NOT-OD-21-016 – Supplemental Information to the NIH Policy for Data Management and Sharing: **Selecting a Repository** for Data Resulting from NIH-Supported Research

COGR
Council On Governmental Relations

Background – FYI
(Provided as context but not to be discussed in detail)

Other Costing Models are in Development

(links included at the end)

Guidance on Costs

The National Institute of Standards and Technology (NIST) has released a Research Data Framework (RDaF) that has both a research data ecosystem and data lifecycle approach. The RDaF core in Appendix E can provide insight to campuses trying to estimate where costs might be accrued.

Kaiser, D. L., & Hanisch, R. J. (2021). Research Data Framework (RDaF): Motivation, Development, and a Preliminary Framework Core. Available at https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-18.pdf.

Life Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs provides a framework for cost-effective decision making for biomedical research data preservation, discoverability, and use. The appendices include salary ranges for relevant jobs for the data life cycle (Appendix C), soft costs for digital preservation (Appendix D), and a template to map cost drivers (Appendix E).

National Academies of Sciences, Engineering, and Medicine. (2020). Life-Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs. Available at https://www.nationalacademies.org/our-work/forecasting-costs-for-preserving-archiving-and-promoting-access-to-biomedical-data

Examples of Developing University Models for Data Sharing

Duke University's Approach

Duke University Overview

- Research base: $1B ($650M is from federal)
- Number faculty receiving federal funding: ~1,000
- Very decentralized management structure

- Campus Library: primary provider of service in this space
- SOM: has a number of solutions driven by their specific needs in this space (e.g. clinical, PHI, discipline)
- Technology Organizations: Technology organizations are discrete but overlap across both campus and health

Campus Overview

- **2013** - OSTP Public Data Access Requirement
- **2015** - NSF rejects Duke research proposal due to insufficient DMP, specifically due to lack of plans to deposit data
- **2015** - University Libraries, Arts & Sciences, and Office of Information Technology (OIT) each request significant funding increase for research data storage from Provost
- **2015** - Provost seeks white paper and charge for *Faculty Working Group for Digital Research Data Services*

Campus Overview

- **2016** Recommendations by *Digital Research Data Services Faculty Working Group*:
  - **Full-time dedicated Research Data Specialists and Repository Ingest Specialists** to support data management planning, compliance, public access, and retention requirements (**four positions** created and filled in January 2017);
  - **Funding to support baseline**, minimum levels of computing and digital storage for research data projects available to all faculty projects;
  - **Funding to support a minimum level of research storage for long-term archiving** and preservation of faculty research data in the Duke Digital Repository (DDR) that cannot be deposited into a discipline-based repository, or to meet other requirements

School of Medicine

Background – FYI
(Provided as context but not to be discussed in detail)

- Vivli – Duke joins in September 2018
  - Vivli has established a **data sharing platform** that includes an independent general access data repository that is available for **searching for data from clinical trials** from all sponsors that can be hosted, shared and accessed

- The Duke Clinical Research Institute's (DCRI) SOAR initiative is based in clinical research data sets:
  - SOAR™ (Supporting Open Access for Researchers) is a collaboration between the Duke Clinical Research Institute (DCRI), academia, and private industry to **open clinical research data for the benefit of the broader research community.**

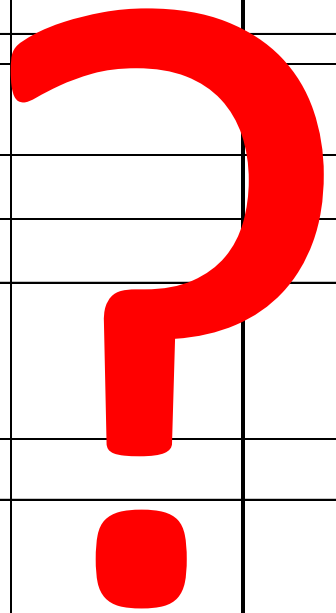
Duke University Challenges

Background – FYI
(Provided as context but not to be discussed in detail)

- Practical integration of data management planning, compliance, public access, and retention requirements in a manner that is cost-effective and fully supportive of the faculty research objectives.
- Scale of digital storage to meet the pressing demand and exponential growth of scientific data in digital formats.
- **Costing aspects** of this requirement, whether they be funded by the sponsor or the institution, will have significant budgetary implications.
- A rapidly evolving landscape as sponsors, funding agencies, institutions, and publishing entities actively modify and shift their policies; this uncertainty causes challenges for planning and allocation
- For research involving human participants, the complexity of rules and regulations involving protected health information, deidentification, and anonymization, which are very relevant as we incorporate pragmatic clinical trial designs and the concept of learning health seeks to bridge the gap between research and patient care

Cost Implications: Lifecycle Public Data Access Activities

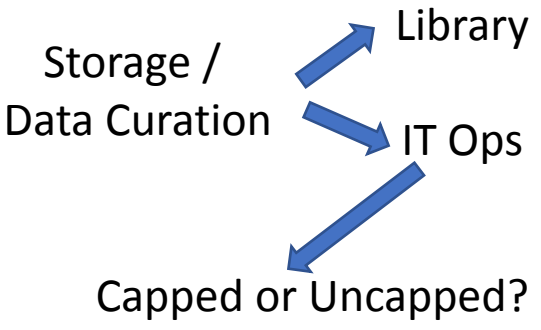
#	Activity	Timing	Sponsor Pay			Institution Pay				External Repository	Other
	Lifecycle Public Data Access Activities	Timing (Pre-Proposal submission, Life of award, Post-Closing)	Direct Charge to Sponsor as direct line item or via Service Center	Separate Supplement / Companion Award (with different period of performance) for Data Storage after period of performance	Budget Line Item for Data Only that is Paid/Accrued at end of Award for Future Data Costs (would require OMB UG approval)	Service Center (likely subsidized by institution but charged to project)	Institution Pay (Admin Capped)	Institution Pay (Uncapped: O&M or Library)	Institutionally Supported Repository	Publisher / Discipline / Sponsor / Professional Society / One Time (Fig Share)	Institution Covers Cost but then Charges User
1	DMP Development	PRE - PROPOSAL	No							N / A	Contrary to open access principles depending on implementation (NEEDS MORE DISCUSSION)
2	Data Curation & Metadata Curation FAIR, Data dictionary, etc.	LIFE (SOME PRE)	Some sponsors allow								
3	Data Ingest	LIFE	Hopefully								
4	DMP Monitoring & Compliance through life of award through closeout	LIFE	Hopefully								
5	Data Storage (during life of project)	LIFE	Probably yes								
6	Data Processing	LIFE	Probably yes								
7	Data Storage (post-closeout for publication)	POST/LIFE	Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance								
8	DMP Monitoring & Compliance - post closeout	POST	Probably No								
9	Data Storage (post-closeout for DMP Compliance)	POST	Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance								
10	Cold Data Storage (post-closeout / last resort)	POST	Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance								
11	Publication Fees (often based on size and duration of data)	POST	Probably yes								
12	Data Security (PHI, HIPAA, Export Controls, FISMA, student data and IP)	PRE, LIFE & POST	Varies								



Draft
In-Process

Costing Issues and Discussion – Storage Only

Example

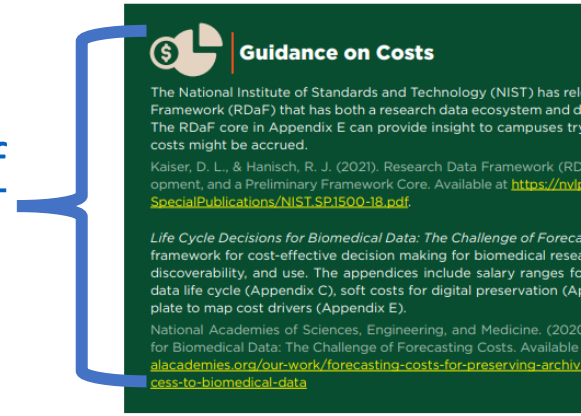


1) Internal Budget Management ⟷ Reimbursement
2) Validity of Controls to support - Direct costing
   a) Breadth of types of costs
   b) Breadth of procurement methodologies
   c) AAR: Data storage and other Life Cycle costs can be expended anywhere and from multiple different sources/methods
3) Regs → Switching cost pools … capped pools
4) Costing implications of service centers… research computing facilities (storage, processing)
5) Library "caps" – Realistically how much could you recover?
6) Faculty pressure → Waive F&A for "Cloud Computing" expenditures (e.g. remember Genomic Arrays)
7) Creation of new Cost pool?

Resources

Background – FYI
(Provided as context but not to be discussed in detail)

- APARD
  - https://www.aau.edu/accelerating-public-access-research-data
  - https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Intellectual-Property/Public-Open-Access/AAU-APLU-Public-Access-Working-Group-Report.pdf
  - https://www.aplu.org/projects-and-initiatives/research-science-and-technology/public-access/
  - https://www.aplu.org/projects-and-initiatives/research-science-and-technology/public-access/workshop-on-public-access-report-aplu-aau-2019.pdf
  - https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Public%20Access/AAU%20APLU%20Guide%20to%20Accelerate%20Public%20Access%20to%20Research%20Data.pdf
- COGR
  - https://www.cogr.edu/nih-data-sharing-and-management
- Other
  - https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-18.pdf
  - https://www.nap.edu/download/25639
- FDP → In-process *(to be discussed at next week's meeting)*



Discussion

&

“Admiring the Problem”