COGR, FDP, and ARL: Putting Numbers Behind Institutional Expenses for Public Access to Research Data
Presenters

- **Cynthia Hudson Vitale**
  Director, Scholars and Scholarship, Association of Research Libraries

- **Christi Keene**
  Assistant Director, Post Award Support Services, University of Chicago; FDP Finance Audit & Costing Committee Co-Chair

- **Melissa Korf**
  Director, Grants & Contracts, Harvard Medical School; FDP Research Compliance Committee Co-Chair

- **Wendy Kozlowski**
  Data Curation Specialist, Cornell University Library; Cornell University

- **Jim Luther**
  Interim Research Compliance Officer, Yale University; Strategic Advisor, FDP and COGR

- **Shawna Taylor**
  Project Manager, Realities of Academic Data Sharing (RADS) Initiative, Association of Research Libraries
Today’s Agenda

- Background (Cynthia)
- Federal Demonstration Partnership (Christi Keene & Melissa Korf)
- ARL (Shawna Taylor)
- Council on Government Relations (Jim Luther)
- Audience Poll (Jim Luther)
- Discussion (Wendy Kozlowski)
Public Access to Research Data

Background

2003 NIH Policy on Data Sharing
• Applies to grants > $500k plan for sharing and “timely release and sharing” of data- In effect until January 2023

2011 NSF Data Management and Sharing Policy

2013 OSTP Memo: “Increasing Access to the Results of Federally Funded Scientific Research”
• Each Federal agency with over $100 million in R&D expenditures must develop a plan to support increased public access to results of funded research inclusive of peer-reviewed manuscripts and articles and research data
• As of November 2021 all 20 Federal departments and agencies subject to the Memo are in compliance; several agencies with R&D below $100 million have also adopted these principles

2023 NIH Policy on Data Sharing (Updated)
A comparative analysis of 69 funding agencies, which included 33 federal funding agencies and 13 non-profit or private funding agencies.

62% (n=43) of the data sharing policies allow costs for implementing the policy as an allowable direct expense to funding. What activities were allowable varied among the data policies analyzed.
- For example, Wellcome’s data sharing policy indicates that any justified cost for delivering the plan will be considered.
- The National Science Foundation directorates were mostly in agreement - stating that the “costs of documenting, preparing, publishing or otherwise making available to others the findings and products of the work conducted under the grant” are allowable.

None of the policies provided example costs or budgets for data management and sharing.

None of the policies addressed covering expenses post-award.
Costing Issues For Public Access to Research Data – Horizon

- Expect additional NIH FAQs and guidance on what researchers should be considering or thinking about with regards to costs for data management and sharing

- Many higher education organizations are now looking critically at how to account for the institutional costs for public access to research data – so we can then advocate collectively for funding and resources
FDP mission:

Association of federal agencies, academic and nonprofit research institutions, and research policy organizations that work together to streamline the administration of federally sponsored research and foster collaboration to enhance the national research enterprise, while maintaining high standards of stewardship and accountability.

Researchers doing science, not administration
**What is the FDP?**

**FEDERAL**
- Federal agency participation is **critical**.
- Participation **encourages partnership and interaction** among federal agencies and supports effective research outcomes, awardee accountability, and communication.

**FACULTY**
- Provides a **unique forum** for principal investigators to engage with research administrators, IT specialists, and federal partners, and effect real change.

**TECHNICAL & ADMINISTRATIVE**
- Offers a **diverse** group of institutions and administrators who are devoted to improving research productivity and minimizing administrative burden.
- Prompt and thoughtful **feedback** is available on potential improvements in grants management, and a diverse **test bed** for implementation of new research grant-related requirements, processes, and policies.
## Faculty ThoughtExchange Summary

### Top 3 Areas of Concern

<table>
<thead>
<tr>
<th>%</th>
<th>Answer (Multi-select)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12%</td>
<td>Data Management Plan (DMP) Development</td>
</tr>
<tr>
<td>20%</td>
<td>Data Curation &amp; Metadata Curation (Data dictionary, etc.)</td>
</tr>
<tr>
<td>3%</td>
<td>Data Ingest and Loading</td>
</tr>
<tr>
<td>38%</td>
<td>DMP Monitoring &amp; Compliance - During Life of Award</td>
</tr>
<tr>
<td>46%</td>
<td><strong>DMP Monitoring &amp; Compliance - At Closeout &amp; Post-Closeout</strong></td>
</tr>
<tr>
<td>33%</td>
<td>Data Storage - During Life of Award</td>
</tr>
<tr>
<td>68%</td>
<td><strong>Data Storage - At Closeout &amp; Post-Closeout</strong></td>
</tr>
<tr>
<td>9%</td>
<td>Data Processing</td>
</tr>
<tr>
<td>11%</td>
<td>Publication Fees</td>
</tr>
<tr>
<td>53%</td>
<td><strong>Data Security</strong></td>
</tr>
</tbody>
</table>

What data sets do you currently use, develop, or acquire?

<table>
<thead>
<tr>
<th>%</th>
<th>Answer (Multi-select)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40%</td>
<td>Institutionally provided</td>
</tr>
<tr>
<td>26%</td>
<td>Sponsor provided</td>
</tr>
<tr>
<td>38%</td>
<td>Data acquired through purchase</td>
</tr>
<tr>
<td>44%</td>
<td>Data acquired through your activities</td>
</tr>
<tr>
<td>32%</td>
<td>Data sharing (with consortium or cooperative agreement)</td>
</tr>
<tr>
<td>6%</td>
<td>Other</td>
</tr>
<tr>
<td>22%</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
Faculty ThoughtExchange Themes

• Feels like an “Unfunded mandate”
• Uniform cross-agency requirements (costing, etc.) to reduce burden
• Funding - who pays
• Long-term funding for storage and curation
• Clear guidance on how to budget these costs
• Explicitly recognize data collection, transformation and documentation as direct costs
• Clear regulations / concern about grad students, etc. being required to do this and not PI
• Repositories
• NIH Deadline
• Culture change
FDP Data Sharing & Management Working Group

Research Compliance & Finance, Audit, & Costing Committee collaboration

Data management expert

Cost analysis expert

Federal representative
FDP Data Sharing & Management Working Group Priority Issues

Collaborate to promote data management and sharing best practices by reducing administrative burden and ensuring researchers are able to cover associated costs.

Explore costing models

Quantify data management and storage costs through the lifecycle

Guidance on drafting data sharing plans and budgeting associated costs in proposals
Association of Research Libraries (ARL)  
Realities of Academic Data Sharing (RADS)  
Initiative

NSF EAGER ##2135874:  
Completing the Lifecycle:  
Developing Evidence Based  
Models of Research Data  
Sharing
Where are researchers sharing their research data and what is the quality of that metadata?

How are researchers making decisions about why and how to share research data and what costs do they incur?

What is the cost to the institution to implement federally mandated public access to research data policies?
ARL RADS
Workflow - Institutional Infrastructure

Objectives

1. Conduct a landscape analysis of existing costing frameworks.
2. Use buckets/categories to define “public access” within the larger costs of data management.
## ARL RADS

### Workflow - Institutional Infrastructure

<table>
<thead>
<tr>
<th>FUNCTION (Data Lifecycle* Stage)</th>
<th>CATEGORY</th>
<th>SUBCATEGORY</th>
</tr>
</thead>
</table>
| **ENVISION** Review of the overall strategies and drivers of an organization’s research data program. | Data Governance* Structure | • Identification of Goals and Roles  
• Data vision and/or **data policy**  
• Data management value proposition  
• Data management organization  
• Value of data (quantitative or qualitative)  
• Legal and regulatory **compliance**  
• **Data quality** (including Trust and Certification)  
• **Data privacy**  
• Data ethics |
| **Community Engagement** | • **Stakeholder community(ies)**  
• Communication with stakeholder community(ies)  
• Interactions with other organizations  
• **Cross-community engagement** (across domains and sectors)  
• Inclusivity in interactions |
ARL RADS
Workflow - Institutional Infrastructure

Objectives

3. Consider activities to make (meta)data FAIR.
1. For each activity:
   a. What services and infrastructure does your department provide for this activity?
   b. Who is paying for the services?
   c. What is the designated labor cost to run these services or provide the infrastructure?
   d. What is the cost of the services (per usage/per hour/monthly)?
ARL RADS

Workflow - Researchers

Objectives

1. Examine funded researcher activities to make research data publicly accessible.
2. Five disciplines: environmental science, materials science, psychology, biomedical sciences, & physics.
3. Costing questions included from the institutional infrastructure survey.
ARL RADS
Workflow - Researchers

Survey Questions

1. Direct vs. indirect costs?
2. Costs after the end of project lifecycle?
3. Were any costs identified in the Data Management Plan?
4. Analysis by discipline, institution, and as individual case studies.
Mission Statement
COGR is an association of 200 leading research universities, affiliated medical centers, and independent research institutes. We are the national authorities on the financial and regulatory infrastructure, and the corresponding compliance requirements associated with managing federal research grants and contracts within research institutions. We provide information, analyses, advice, policy perspective, and historical context to our members in the areas of research administration and compliance, financial oversight, and intellectual property. COGR communicates the viewpoint and concerns of its members and fosters productive relationships between the research community and federal policymakers, advocating for innovation and change that avoid unnecessary regulatory burden.
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 Indirect cost pools: Library, DA, GA, O&M, Equipment, & Base
- Lifecycle is broad and complex
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.

https://www.cogr.edu/sites/default/files/NIH%20Data%20Management%20-%20v1.1_0.pdf
NIH Policy Notices & Supplemental Information

- Released October 29, 2020, Effective January 25, 2023
    - 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
<table>
<thead>
<tr>
<th>#</th>
<th>Activity</th>
<th>Timing</th>
<th>Sponsor or Pay</th>
<th>Institution Pay</th>
<th>External Repository</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lifecycle Public Data Access Activities</td>
<td>Timing (Pre-Proposal submission, Life of Award, Post-closing)</td>
<td>Direct Charge to Sponsor as direct line item or via Service Center</td>
<td>Separate Supplement / Companion Award (with different period of performance) for Data Storage after period of performance</td>
<td>Institution Only that is Paid/Accrued at end of Award for Future Data Costs (would require DMB/UG approval)</td>
<td>Service Center (likely subsidized by institution but charged to project)</td>
</tr>
<tr>
<td>2</td>
<td>Data Curation &amp; Metadata Curation FAIR, Data dictionary, etc.</td>
<td>Life (Some Pre)</td>
<td>No</td>
<td>Some Sponsor Allow</td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>3</td>
<td>Data ingest</td>
<td>Life</td>
<td>No</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>4</td>
<td>DMP Monitoring &amp; Compliance through life of award through closeout</td>
<td>Life</td>
<td>No</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>5</td>
<td>Data Storage (during life of project)</td>
<td>Life</td>
<td>Probably yes</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>6</td>
<td>Data Processing</td>
<td>Life</td>
<td>Probably yes</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>7</td>
<td>Data Storage (post-closeout for publication)</td>
<td>Post/Life</td>
<td>Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>8</td>
<td>DMP Monitoring &amp; Compliance - post-closeout</td>
<td>Post</td>
<td>Probably No</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>9</td>
<td>Data Storage (post-closeout for DMP Compliance)</td>
<td>Post</td>
<td>Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>10</td>
<td>Cold Data Storage (post-closeout / Last resort)</td>
<td>Post</td>
<td>Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>11</td>
<td>Publication Fees (often based on size and duration of data)</td>
<td>Post</td>
<td>Probably yes</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
<tr>
<td>12</td>
<td>Data Security (PII, HIPAA, Export Controls, FSMA, student data and #)</td>
<td>Pre, Life &amp; Post</td>
<td>No</td>
<td></td>
<td></td>
<td>N/</td>
</tr>
</tbody>
</table>
# Life-Cycle Costing Considerations

<table>
<thead>
<tr>
<th>#</th>
<th>Activity</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DMP Development</td>
<td>PRE - PROPOSAL</td>
</tr>
<tr>
<td>2</td>
<td>Data Curation &amp; Metadata Curation FAIR, Data dictionary, etc.</td>
<td>LIFE (SOME PRE)</td>
</tr>
<tr>
<td>3</td>
<td>Data Ingest</td>
<td>LIFE</td>
</tr>
<tr>
<td>4</td>
<td>DMP Monitoring &amp; Compliance through life of award through closeout</td>
<td>LIFE</td>
</tr>
<tr>
<td>5</td>
<td>Data Storage (during life of project)</td>
<td>LIFE</td>
</tr>
<tr>
<td>6</td>
<td>Data Processing</td>
<td>LIFE</td>
</tr>
<tr>
<td>7</td>
<td>Data Storage (post-closeout for publication)</td>
<td>POST/LIFE</td>
</tr>
<tr>
<td>8</td>
<td>DMP Monitoring &amp; Compliance - post closeout</td>
<td>POST</td>
</tr>
<tr>
<td>9</td>
<td>Data Storage (post-closeout for DMP Compliance)</td>
<td>POST</td>
</tr>
<tr>
<td>10</td>
<td>Cold Data Storage (post-closeout / last resort)</td>
<td>POST</td>
</tr>
<tr>
<td>11</td>
<td>Publication Fees (often based on size and duration of data)</td>
<td>POST</td>
</tr>
<tr>
<td>12</td>
<td>Data Security (PHI, HIPAA, Export Controls, FISMA, student data and IP)</td>
<td>PRE, LIFE &amp; POST</td>
</tr>
</tbody>
</table>
# Life-Cycle Costing Considerations

## Cost Implications: Lifecycle Public Data Access Activities

<table>
<thead>
<tr>
<th>Sponsors Pay</th>
<th>Institution Pay</th>
<th>External Repository</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Charge to Sponsor as direct line item or via Service Center</strong></td>
<td><strong>Separate Supplement/Companion Award (with different period of performance for Data Storage after period of performance)</strong></td>
<td><strong>Budget Line Item for Data Only that is Paid/Accrued at end of Award for Future Data Costs (would require OMB-UG approval)</strong></td>
<td><strong>Service Center (likely subsidized by institution but charged to project)</strong></td>
</tr>
<tr>
<td><strong>Institution Pay (Admin Capped)</strong></td>
<td><strong>Institution Pay (Uncapped: O&amp;M or Library)</strong></td>
<td><strong>Institutionally Supported Repository</strong></td>
<td><strong>Publisher/Discipline/Professional Society/One Time (Fig Share)</strong></td>
</tr>
<tr>
<td>Institution Covers Cost but then Charges User</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some sponsors allow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Hopefully</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hopefully</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probably yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probably No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probably no unless feds allow booking an estimate (see UG) or they provide a separate award with different period of performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Contrary to open access principles depending on implementation (NEEDS MORE DISCUSSION)
**Charge:** Develop expertise, share with other associations, and conduct other mission-appropriate activities so that COGR can engage NIH DMS Policy

**Deliverables:**
1. **Advocacy** to engage NIH DMS Policy
2. **Education and resources** to COGR members – Identify compliance requirements and key challenges to be shared with COGR members in various contexts (FAQs, Effective Practices, Federal/NIH guidance, Education for Readiness, etc.)
3. **Cost of Compliance** survey and report
NIH Policy for Data Management and Sharing

- Current Workgroup Priorities
  1. NIH Data Sharing and Management – Briefing Sheet(s)

  1. Institutional Readiness Assessment
     - Key Players & Stakeholders
     - Key Activities
     - Practical Preparation for Implementation

  1. Detailed Analysis of NIH Notices Advocacy, Institutional Support, & Costing Implications
NIH Policy for Data Management and Sharing

- Briefing Document Targeted to Senior Institution Leaders to Assist in Awareness and Catalyst for Action
- Components
  - Summary
  - Applicable Policies
  - Background
  - Faculty & Institution Implications
  - Examples
  - References
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 Indirect cost pools: Library, DA, GA, O&M, Equipment, & Base  
- Lifecycle is broad and complex  
- Internal Operations / Budget Management  

Reimbursement
Attendee Poll: Activities or priorities for research data costing?
What is your current role?

- Library administrator: 22
- Librarian: 3
- IT administrator: 1
- IT staff: 1
- Other: 5
1. Consult on Data Management Plans (DMP)
2. Monitor DMP for compliance through life of award through closeout
3. Monitor DMP for compliance - post closeout
4. Curate data
5. Train personnel on data workflows and protocols
6. Address/collect data use agreements
7. Create metadata
8. Develop data documentation (Data Dictionary, README files, etc.)
9. Provide active data storage (during life of project)
10. Archive data (post closeout for publication or DMP compliance)
11. Pay data publication fees (often based on size and duration of data)
12. Manage and archive data with security needs (PHI, HIPAA, Export Controls, FISMA, student data, and IP)
13. Track data citations impact
14. Monitor terms of use
15. Assist others with research data reuse
16. Create persistent identifiers (DOIs, ORCIDs, etc)
17. Assist with licensing (Creative Commons, MIT, etc)
18. Assist with copyright or intellectual property impacts
19. Assist researchers with the IRB process
Top four selected: 1) Consult on Data Management Plans 2) Train personnel on data workflows and practices 3) Create Persistent Identifiers 4) Assist with copyright or intellectual property impacts
What are your top three public access to data activities you are most interested in understanding expense/cost ranges? (select three):

Top four selected: 1) Archive data post award closeout 2) Manage and archive data with security needs 3) Monitor DPM for compliance through lifecycle of award 4) Train personnel on data workflows and protocols
Discussion / Q&A Time

Are there additional “buckets” that your institutions are putting resources to, related to public access, that were NOT in the poll?

Are there any services in the poll list that you did not understand or would not have connected to “public access”?

Are there any services or activities that you do connect to public access that we left OUT of the poll?

Have you implemented any activities at your University that you feel specifically address costing concerns in any of these public sharing areas?
Thank You!

- Cynthia Hudson Vitale; cvitale@arl.org
- Christi Keene; ckeene@uchicago.edu
- Melissa Korf; melissa_korf@hms.harvard.edu
- Wendy Kozlowski; wak57@cornell.edu
- Jim Luther; jluther@thefdp.org
- Shawna Taylor; staylor@arl.org