CaRCC Data-Facing Call, 2025-02-04

Title/Topic

Insights from the Realities of Academic Data Sharing (RADS) initiative

Speakers

Speakers: Cynthia Hudson Vitale- Associate Dean for Technology Strategy & Digital Services, Johns Hopkins University; Fernando Rios - Research Data Management Specialist at the University of Arizona Libraries; Shawna Taylor - Project Manager, Open Science, at the Association of Research Libraries (ARL)

Abstract

Researchers from the Realities of Academic Data Sharing (RADS) initiative will provide an overview of the costs of research data sharing, highlighting the significant investments institutions make in services and infrastructure to support researchers. Although data management and sharing costs are being incurred by both the institution and researchers, early findings in new RADS research reveals that many NIH-funded researchers currently budget $0 for data management and sharing expenses, creating challenges as policy mandates take effect. The current, expanded RADS cohort builds on this work by further refining cost and expense models and detailing additional organizational models for research data infrastructure and services.

[URL of Project Site](#)

Table of Contents

Please Note:

- ***We will record this call and post shortly thereafter on [CaRCC's YouTube channel](#).***
- ***We expect all persons on the call to adhere to [Our Community Policies](#).***

Announcements

- Welcome to the Data-Facing Track of the CaRCC People Network!
    - [Join the People Network email lists](#). Join the [CaRCC Slack workspace](#).
    - [Track descriptions](#)

- The CaRCC People Network is your community for research computing and data professionals. Please see [this brief overview](#) and https://carcc.org/ for more details.

- If you have questions about CaRCC or are interested in becoming more involved, please contact:
    - df-coordinators@carcc.org for D-F-related activities, or
    - getstarted@carcc.org or getinvolved@carcc.org for other CaRCC work

- We expect all persons on the call to adhere to [CaRCC's Code of Conduct](#).

Sign-In (Name / Affiliation / Email)

Note: To help with our outreach efforts, please sign in with the suggested format (we will not sell/share your information per our [Privacy Policy](#)). Switch your editing mode to "Viewing" after sign-in, if you're worried about making accidental edits.

1. Moira Downey / NC State University / mcdowney@ncsu.edu
2. Amit Amritkar / Penn State / amit@psu.edu
3. Sajesh Singh / American Museum of Natural History / saingh@amnh.org
4. Milo Golub / Starfish Storage / milo@starfishstorage.com
5. Amy Koshoffer / University of Cincinnati / koshofae@ucmail.uc.edu
6. Jessica Eaton / Columbia University / j.eaton@columbia.edu
7. Deb McCaffrey / Arizona State University / dmccaff4@asu.edu
8. Ann McCranie / Indiana University Research Data Commons / amccrani@iu.edu
9. Vas Vasiliadis/ University of Chicago / vas@uchicago.edu
10. Eleanor Koehl, UCLA / ekoehl@oarc.ucla.edu
11. Juli Petereit / University of Nevada, Reno / jpetereit@unr.edu
12. Sean Smith / Rice University / mrsmith@rice.edu
13. Venice Bayrd / Montana State University / venice.bayrd@montana.edu
14. Rory Macneil / Research Space / rmacneil@researchspace.com
15. Terry Haygood / Jackson State University / terry.haygood@jsums.edu
16. Ross Hyman / University of Chicago / rhyman@uchicago.edu
17. Bobby Roybal / University of Nevada, Las Vegas / bobby.roybal@unlv.edu
18. Jim Leous / Consultant / jim.leous@gmail.com
19. Gabriel Smith / University of British Columbia / gabriel.smith@ubc.ca
20. Tom Lewis / University of Washington / tomlewis@uw.edu
21. Kristin McLean / Northwestern University / kristin.mclean@northwestern.edu
22. John Heimaster / Ohio State / heimaster.1@osu.edu
23. Manasvita Joshi/ Harvard/ manasvitajoshi@g.harvard.edu
24. Jaynal Pervez/ UT Health San Antonio/ pervez@uthscsa.edu
25. Robert McDermott / Fred Hutch Cancer Center / rmcdermo@fredhutch.org
26. Matthew Smith / University of British Columbia / matthew.smith@ubc.ca
27. David G. Cyrille / Stony Brook University / David Cyrille
28. Zoe Braiterman / Independent Consultant / braiterman.z@gmail.com
29. Mara Sedlins / Colorado State University / mara.sedlins@colostate.edu
30. Jeff Dusenberry / UMass Boston / jeff.dusenberry@umb.edu
31. Amy Schuler / Cary Institute / schulera@caryinstitute.org
32. Bob Freeman / Harvard University / robert_freeman@harvard.edu
33. David Merten-Jones / The Claremont Colleges Library / david.merten-jones@claremont.edu
34. Stephanie Vasko / Michigan State University / vaskoste@msu.edu
35. Alisa Kang / Georgetown university / alisa.kang@georgetown.edu
36. Brian Ensor/ George Washington / bensor@gwu.edu

Call Notes

Slides

📕 20250204 - RADS Initiative.pdf

Notes

- [RADS expense report](#)
- Support for research data management spread across various units on campuses; it has been difficult to understand the full cost for data management and sharing because this support is so distributed
- RADS initiative designed to better understand the full costs
- Initially scoped to research underwritten by small number of specific agencies
- Deliberately asked questions not just about storage, but also about staffing and labor involved in data curation and sharing
- Focused on activities in 5 general areas/phases that mirror data/research lifecycle
- Found that average cost per funded project for data sharing was $30K – about 6% on average of total award
  - As award amount increased, data management and sharing (DMS) expenditures also increased, however, smaller awards spent a proportionally larger amt of the award on DMS expenses
- Most researchers had 0.25 FTE dedicated to DMS activities
- Classified institution services into 4 categories: Research offices, Libraries and archives, IT/Compute (may include things like information security, etc), and Institutes and centers
- How do DMS expenses scale? Total institutional costs ranged from $800K to $6M, averaging around $2.5M – tend to think this is an undercount
- DMS costs have no been standardized, or recognized as such; some of these costs are already occurring and to varying degrees already had associated costs
- [RADS gap analysis report](#)
  - Most researchers do most data sharing activities on their own, not regularly engaging with support offered by the institution
- IT offices support DMS across the research lifecycle ([visualization](#)) – (all of the project [visualizations are here](#))
- Opportunities for IT units:
  - expand support for data security
  - enhance quality control support
- Opportunities for Libraries:
  - raise awareness of data licensing support
  - promote repository & PID services
  - collaborate with research offices

- RADS list of DMS activities can help understand how to budget for costs
- RADS2: meeting with prospective NIH researchers quarterly to record DMS costs in real time
  - In meeting with NIH researchers, have found that they are budgeting $0 for DMS costs (this finding is in alignment with FDP pilot project)
    - From interviews with researchers, two things seem to be driving the $0 budgeting: 1) how researchers think about DMS activities – researchers see data sharing as routine and have a hard time extrapolating that activity out; 2) data sharing is new and some researchers think about it as supplemental to what they're doing
    - Opportunity to change the narrative that budgeting for DMS activities doesn't take away from spending on the science, but is integral to it
  - 11 institutions now participating in the second round of research
    - Expanded cohort focuses on institutional diversity (both in research expenditures and student demographics)
    - Also interested in varied approaches to RDM services
    - Expanded the list of funders
    - Also looking at a narrower band of administrators
    - Asking new questions of researchers, specifically around data size, file types, data sources and data access
- University of Arizona:
  - Data services (curation, sharing, workshops, consultations, etc) for 10+ years; launched data repository in 2020
  - Goal for joining RADS was to quantify resources needed for DMS
  - RADS has provided resources and expertise for data gathering and has acted as a good conversation starter for campus discussions
  - Also joined the [Data Curation Network](#) – useful knowledge sharing network & advocacy organization

Questions
- Expenditures of 6% of the award seem surprising if pre-award budgeted amounts for RDM are near $0. Were expenditures solely spent from the grant, or did institutions bear some/most of those costs?
  - We asked about expenses, not about what was budgeted. We did ask how expenses were paid - the majority of allowable costs were from grant funds, followed by department or college funds. (p 15 of the report shows fund sources)
- It's interesting that the IT departments do most of the DMS services in most of the institutions. Do you have insight about why that is?
- Was any data gathered to compare DMS costs by domain?
  - Yes – this is in the expense report (p 16); leveraged NSF and NIH APIs to pull down grant information and categorized according to specific disciplinary areas of interest (had to omit some due to low response rates)
  - Have done away with this in RADS2

Chat Comments

13:11:28 From Shawna Taylor (she/her/hers) to Everyone:
https://www.arl.org/wp-content/uploads/2024/02/2024.02.07-RADS-expense-report.pdf

13:21:41 From Shawna Taylor (she/her/hers) to Everyone: And ecommons for IRs! Sorry, can't chat and advance slides at the same time

13:23:48 From Venice Bayrd (Montana State Univ.) to Everyone: Expenditures of 6% of the award seem surprising if pre-award budgeted amounts for RDM are near $0. Were expenditures solely spent from the grant, or did institutions bear some/most of those costs?
13:29:04 From Cynthia Hudson Vitale to Everyone: We asked about expenses, not about what was budgeted. We did ask how expenses were paid - the majority of allowable costs were from grant funds, followed by department or college funds.
13:30:34 From Cynthia Hudson Vitale to Everyone: Page 15 of the report shows the funds source
13:31:50 From Venice Bayrd (Montana State Univ.) to Everyone: Thank you. I may have missed the context of the budget reference in the abstract

13:27:59 From Cynthia Hudson Vitale to Everyone:
https://www.arl.org/news/realities-of-academic-data-sharing-rads-initiative-releases-gap-analysis-of-researcher-actions-and-institutional-support/

13:29:42 From Cynthia Hudson Vitale to Everyone:
https://public.tableau.com/app/profile/cynthia.vitale8121/viz/InformationTechnology-ServicesInfrastructureforPublicAccesstoResearchData/rads_it

13:44:57 From Deb McCaffrey to Everyone: It's interesting that the IT departments do most of the DMS services in most of the institutions. Do you have insight about why that is? (I used to work at UMich)
13:48:06 From Deb McCaffrey to Everyone: And at UMinn, lots like all the departments help with everything.

13:47:07 From Moira Downey to Everyone: https://datacurationnetwork.org/

13:51:20 From Shawna Taylor (she/her/hers) to Everyone: Here's a link to all of the visualizations - https://public.tableau.com/app/profile/cynthia.vitale8121/vizzes

13:51:47 From Juli Petereit to Everyone: Hi everyone! Just a quick announcement—we're hosting a Data Science Conference in Reno, NV, February 18-20. If you're nearby, we'd love to have you join us! It'll be a great opportunity to connect, learn, and share ideas. Feel free to

reach out if you're interested!
https://www.unr.edu/bioinformatics/training-events/data-science-conference

13:52:20 From Milo Golub to Everyone: Was any data gathered to compare DMS costs by domain?
13:55:09 From Shawna Taylor (she/her/hers) to Everyone: Pg 16 -
https://www.arl.org/wp-content/uploads/2024/02/2024.02.07-RADS-expense-report.pdf
Milo Golub: 👍

13:53:39 From Venice Bayrd (Montana State Univ.) to Everyone: I have to jump off a little early — thank you for sharing this work!

13:57:07 From Deb McCaffrey to Everyone: https://carcc.org/rcdcm/

13:57:46 From Alisa Kang to Everyone: We divide DMS by data size and complexity. Smaller and simpler DMPs are handled by the libraries, while larger data sizes (>1TB) and DMPs that also involve computation are handled by IT.
Cynthia Hudson Vitale: 👍

13:58:27 From Jim Leous to Everyone: There's a whole section in the DCM which deals with information/data management.
Cynthia Hudson Vitale: 🎉

13:58:47 From koshofae to Everyone: Great talk!! Thank you for this work

13:58:48 From Tom Lewis to Everyone: Lovely session.. Thank you

13:58:50 From David Merten-Jones to Everyone: Thank you!

13:58:54 From Juli Petereit to Everyone: Thank you

13:59:02 From Milo Golub to Everyone: Thank you!

Connection Details

Dial by your location

+1 346 248 7799 US (Houston)

+1 669 900 6833 US (San Jose)

Join by Skype for Business: https://utah.zoom.us/skype/8240518198

Cursor Garden ↘ *"Don't leave your cursor in a seedy alleyway"*