Data Preservation

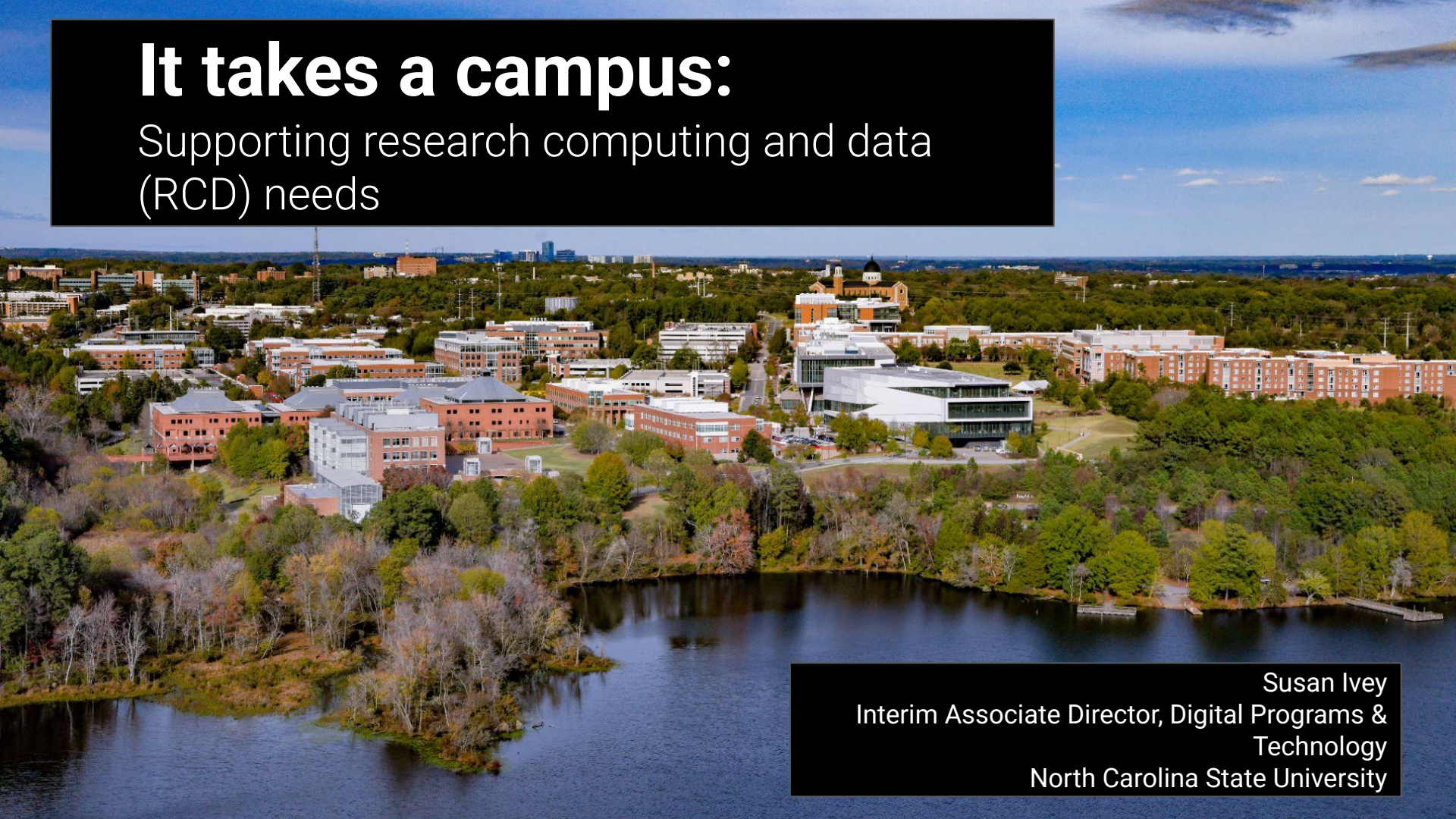
Alexandria, VA, July 10, 2025

Rachel Fleming
Susan Ivey
Mary Ellen Sloane

Moderator: **Tabitha Samuel**

It takes a campus:

Supporting research computing and data (RCD) needs



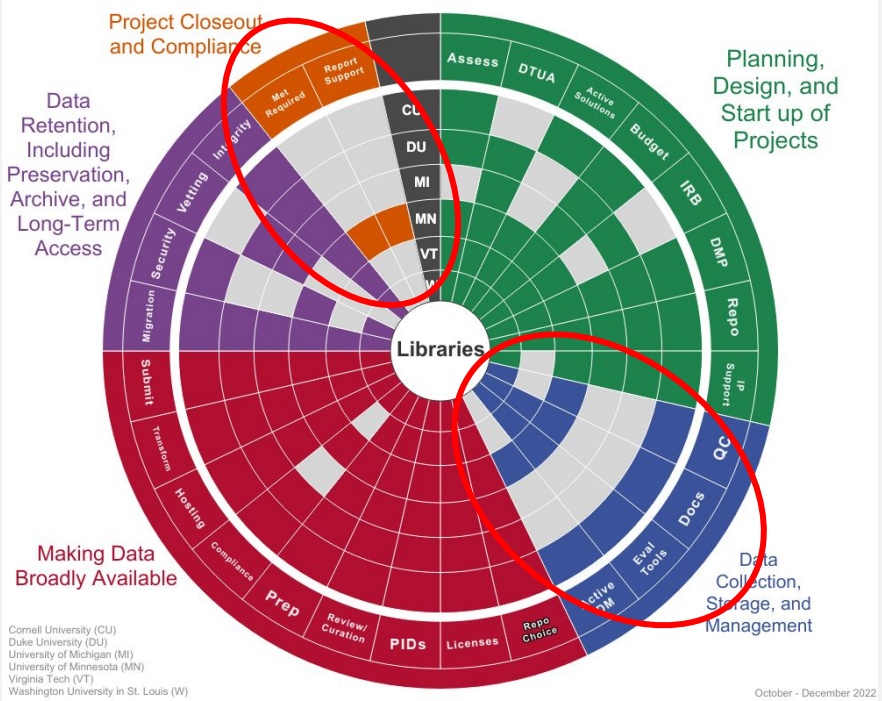
Susan Ivey
Interim Associate Director, Digital Programs & Technology
North Carolina State University

Modern research landscape

- Is increasingly technologically complex, subject to evolving regulatory nuances, and generating ever increasing amounts of data from a wide range of sources
  - Can you tell me how to get data from an IoT sensor off-campus into a database on our NC State storage and then serve it back to other researchers all over the world via a database-backed dashboard-style web application?
  - Can you tell me where to share 10TB of ocean reanalysis data to meet NSF expectations?
  - Can you advise on notification and consent language for users who are submitting images to a web application that we'd like to retain these images and use them to retrain our ML model?
- Involves costs that are difficult to predict and, sometimes, to control
  - How do I budget for a data manager or curator?
  - How do I budget for short-term software developer?
  - How do I budget for commercial cloud services?

Modern research landscape

- Data management and sharing involves units from across campus doing work that Libraries haven't traditionally taken on
  - Infrastructure for data storage
  - Custom tools for data movement during active phase research (in other words: software development)
  - Compliance monitoring



Petters, Jonathan, Shawna Taylor, Alicia Hofelich Mohr, Jake Carlson, Lizhao Ge, Joel Herndon, Wendy Kozlowski, Jennifer Moore, and Cynthia Hudson Vitale. Publicly Shared Data: A Gap Analysis of Researcher Actions and Institutional Support throughout the Data Life Cycle. Washington, DC: Association of Research Libraries, March 2024. https://doi.org/10.29242/report.radsgapanalysis2024.

US RCD landscape in 2025

- Still working to define the field
  - "[M]any people whose work aligns with RCD activities do not yet identify as RCD professionals or identify their work as research computing and data work"
- Staff wear many hats
  - "nearly all respondents report that they spend significant time on work that falls under multiple CaRCC Facings (researcher, software, data, systems, strategy & policy)"
- 2022 estimate: at least 5000 RCD professionals at academic institutions
- Campus Research Computing Consortium (CaRCC) and other orgs (e.g. Research Data Access and Preservation Association (RDAP), US-Research Software Engineer Association (US-RSE)) working to professionalize the space

Christina Maimone, Scott Yockel, Timothy Middelkoop, Ashley Stauffer, and Chris Reidy. 2022. Characterizing the US Research Computing and Data (RCD) Workforce. In Practice and Experience in Advanced Research Computing 2022: Revolutionary: Computing, Connections, You (PEARC '22). Association for Computing Machinery, New York, NY, USA, Article 27, 1–7. https://doi.org/10.1145/3491418.3530289

NC State's Research Facilitation Service

Assessing the needs

- 2017 - Redesign of campus IT governance
- 2018 - Libraries & OIT create Research Data and Infrastructure Librarian
- 2019-2020 - Increased demand for RCD support; Libraries, OIT & ORI discuss need for more coordinated strategy
- 2020 - Planning team convened; report submitted to university leadership
- 2020 - CaRCC Capabilities Model
- 2021 - Focus Groups
- 2021 - Ithaka Big Data study

Challenges to providing RCD support at NC State

- Researchers aren't always aware of all available resources; Services are spread across several campus units with varying contact methods
  - Libraries: Data science consultation, data viz support, RDM support (planning, active, long term), OA and public access policies (sharing/repositories, metadata & PIDS, preservation, etc.)
  - Central IT: Data storage and transfer, high performance computing, VMs
  - Research Office: IRB, compliance, funded projects, commercialization
  - Core facilities
  - College and departmental IT
- Gaps! We were lacking RCD resources & support in key areas
- Complex technical needs must be coordinated with multiple service providers
- Difficult for researchers to estimate costs for budgets
- Changing policies, requirements, & restrictions imposed by the university & elsewhere - convenience to researchers isn't usually a top priority

Creation of the Research Facilitation Service

- 2021 - Received funding for 3 new positions
  - 2 positions by Provost Office, 1 position from Chancellor's Office, 1 reallocated in Libraries
- 2021 - Service design task force charged with envisioning the Research Facilitation Service (RFS)
- 2022 - Launched the RFS via a pilot with College of Sciences

We have chosen an iterative roll out approach, launching college by college, to sustainably grow and develop the service:

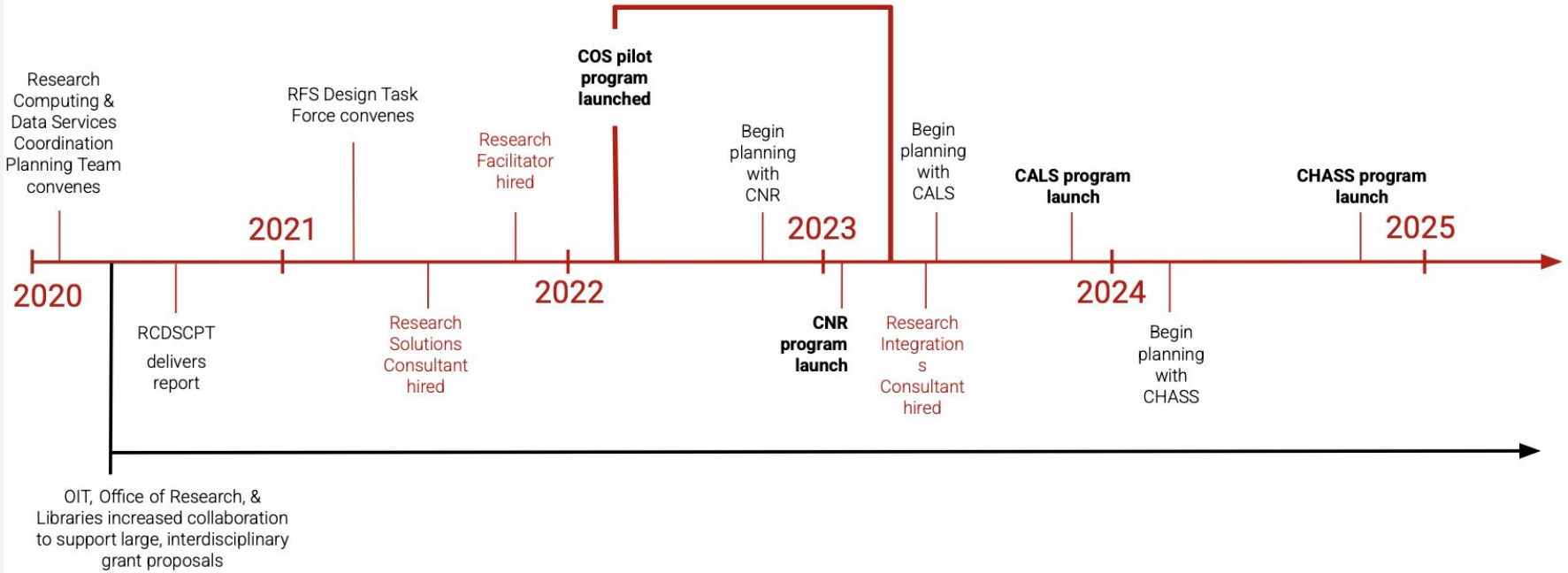
- Active with 4 colleges (Sciences, Natural Resources, Agriculture and Life Sciences, Humanities and Social Sciences)
- Collaborating with the Plant Science Initiative to pilot new models of support for large scale, interdisciplinary research

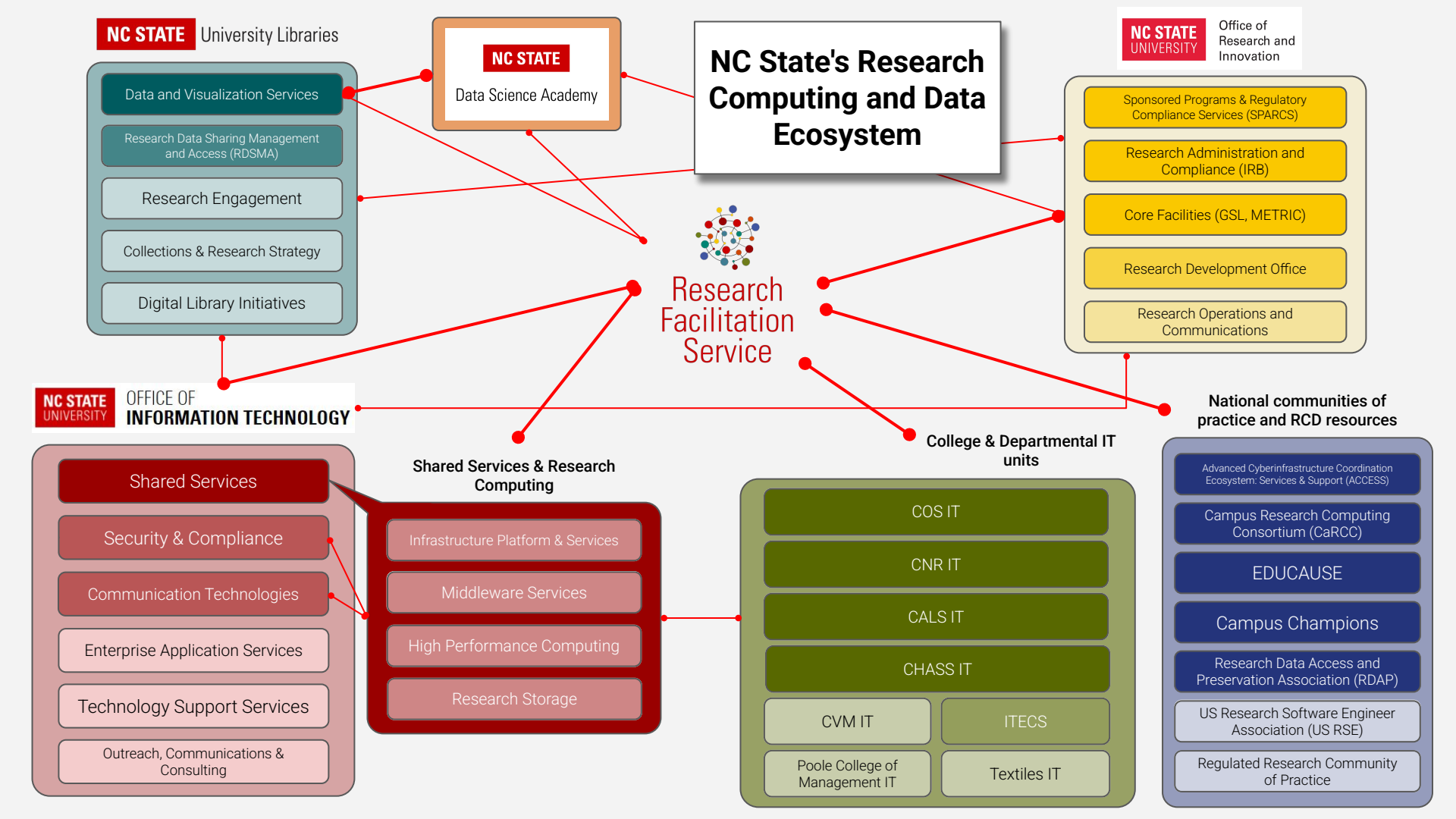
What is research facilitation?

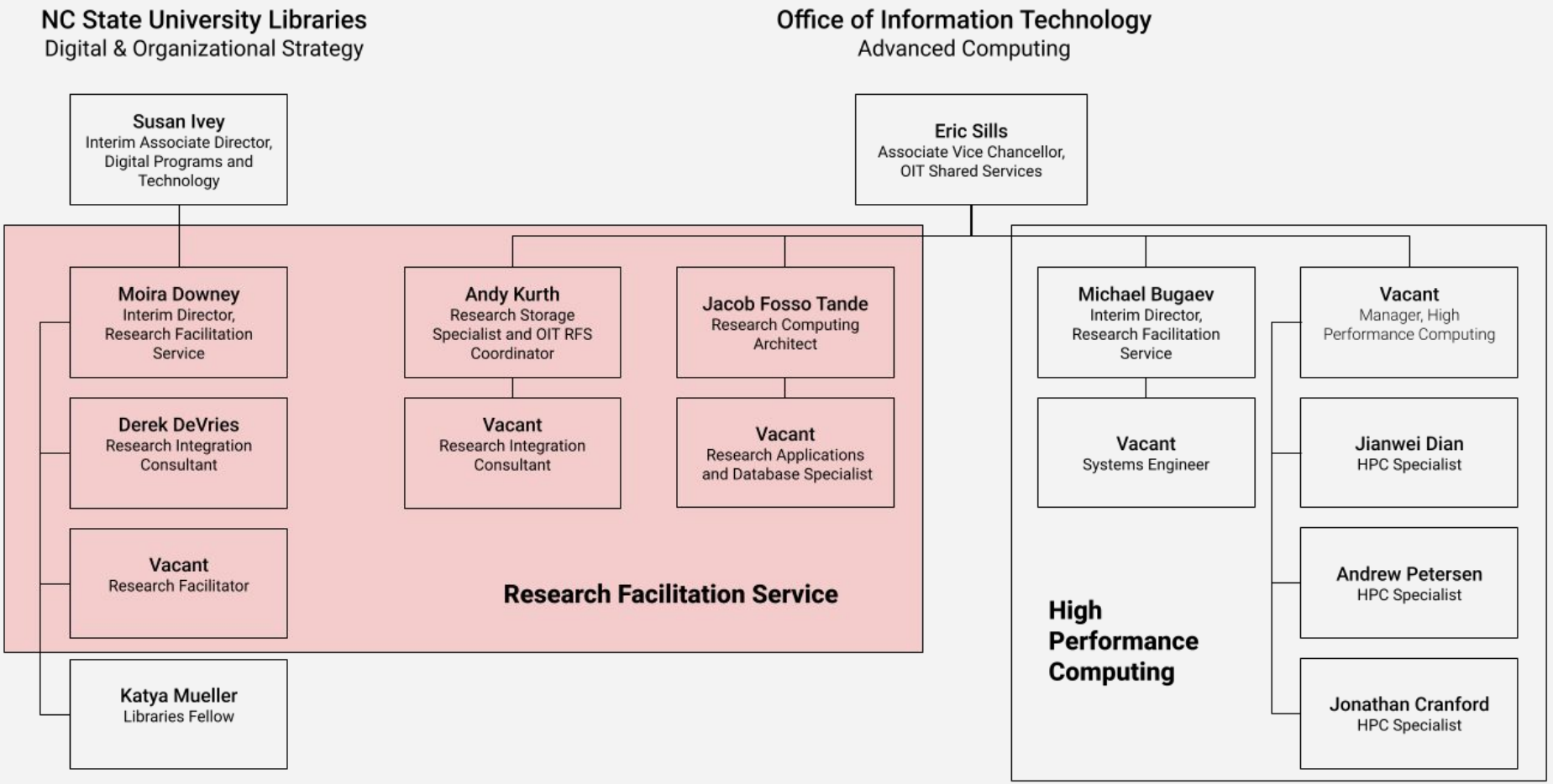
- "RC facilitators serve as proactive and personalized guides, helping researchers identify and implement computational approaches that result in the greatest impact to their projects"
- "RC facilitators act at the front line of building relationships between computing providers and research communities with specific goals and activities that exceed the ability of traditional documentation-and-help-desk models"
- Facilitator activities:
  - Outreach within the research community to promote awareness of computational resources and their potential impacts
  - Engagement with researchers to understand their needs and advise on computational strategies
  - Ongoing support of researchers executing projects on computing resources
  - Education and training of researchers regarding computing capabilities, best practices, and specific skills
  - Liaising researcher connections
  - Advocating for the needs of researchers to inform research IT design and institutional support

Michael, Lauren, and Bruce Maas. "Research computing facilitators: The missing human link in needs-based research cyberinfrastructure." Research bulletin. Louisville, CO: ECAR (2016). https://library.educause.edu/resources/2016/5/research-computing-facilitators-the-missing-human-link-in-needs-based-research-cyberinfrastructure

Designing and launching the RFS




NC STATE University Libraries
Data and Visualization Services
Research Data Sharing Management and Access (RDSMA)
Research Engagement
Collections & Research Strategy
Digital Library Initiatives
NC STATE
Data Science Academy
NC State's Research Computing and Data Ecosystem
NC STATE UNIVERSITY
Office of Research and Innovation
Sponsored Programs & Regulatory Compliance Services (SPARCS)
Research Administration and Compliance (IRB)
Core Facilities (GSL, METRIC)
Research Development Office
Research Operations and Communications
Research Facilitation Service
NC STATE UNIVERSITY
OFFICE OF INFORMATION TECHNOLOGY
National communities of practice and RCD resources
Shared Services
Security & Compliance
Communication Technologies
Enterprise Application Services
Technology Support Services
Outreach, Communications & Consulting
Shared Services & Research Computing
Infrastructure Platform & Services
Middleware Services
High Performance Computing
Research Storage
College & Departmental IT units
COS IT
CNR IT
CALS IT
CHASS IT
CVM IT
ITECS
Poole College of Management IT
Textiles IT
Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS)
Campus Research Computing Consortium (CaRCC)
EDUCAUSE
Campus Champions
Research Data Access and Preservation Association (RDAP)
US Research Software Engineer Association (US RSE)
Regulated Research Community of Practice


NC State University Libraries
Digital & Organizational Strategy
Office of Information Technology
Advanced Computing
Susan Ivey
Interim Associate Director, Digital Programs and Technology
Eric Sills
Associate Vice Chancellor, OIT Shared Services
Moira Downey
Interim Director, Research Facilitation Service
Andy Kurth
Research Storage Specialist and OIT RFS Coordinator
Jacob Fosso Tande
Research Computing Architect
Derek DeVries
Research Integration Consultant
Vacant
Research Integration Consultant
Vacant
Research Applications and Database Specialist
Vacant
Research Facilitator
Research Facilitation Service
Katya Mueller
Libraries Fellow
Michael Bugaev
Interim Director, Research Facilitation Service
Vacant
Manager, High Performance Computing
Vacant
Systems Engineer
Jianwei Dian
HPC Specialist
Andrew Petersen
HPC Specialist
High Performance Computing
Jonathan Cranford
HPC Specialist

What we do, broadly speaking

What we do

About the RFS

The RFS is available to NC State's research community to consult with researchers on various topics related to research computing and data (RCD) and connect researchers to appropriate RCD service providers, remaining engaged to help ensure needs are met. Given our staff's computing and data skill sets, we are also available to provide hands-on support to develop and implement various technical solutions on a case-by-case basis.

Below is an outline of the facilitation areas the RFS typically engages in and five anonymized examples. If something resonates with you and your research, please contact us to learn more about how we can help.

Areas of support

- Publication of research outputs to meet FAIR, open access, and reproducibility needs and requirements, with specific expertise in **research data and code**
- Research Data Management (RDM) and data storage support:
  - Tier 1: RDM education and connecting to related campus services, including lab or research group **consultations on RDM best practices** (virtual and in-person) and collaboration with the Institutional Review Board, OIT's Information Security, Risk & Assurance, and ORI's compliance units
  - Tier 2: Hands-on assistance in building or configuring one-off solutions to **automate currently manual data workflows**
- Proposal development and project planning support, including **boilerplate language** and consultation for available resources at NC State, coordination with the Libraries' Data Management Plan Review Service, and collaboration with College Research Offices to **help researchers budget for IT resources** during proposal planning
- Research Computing, including coordinating with NC State's **High Performance Computing** unit, on-boarding to various **nationally allocated computing systems**, and, increasingly, support for use of **commercial cloud options**

Sample projects

As Director of the Genomic Sciences Laboratory, we are constantly looking to improve our methods in data processing and transfer of Next Generation Sequencing Data to users, both internal and external. Would like to have a general discussion of our methods for moving data and making available to users. [W]e are trying to identify a suitable method of storage and data delivery at low costs. A possible cloud solution might be best, but we don't know. Really, we're looking for an evaluation of how we might continue to do things in a more modern and better way, but keeping overhead costs low.

I am a postdoc [and] I would like to have an account on HPC for analyses. Please can you put me in touch with the HPC office for access.

I'm working on my dissertation which involves human participation. … I wanted to discuss the IRB process with someone to better understand the types of information needed, as well as any concerns with my approach.

We are starting to … sequence student-generated samples as part of the biotechnology and genetics courses we offer. [The process] generates a lot of data! ~100-300 GB per run, for example. We've been using Google Drive and external hard drives. Google Drive has been great for storage of information and quick sharing...but not the best solution.

Sample projects cont.

I have a new faculty member who is coming from the University of Illinois. She has around 20TB of data that she would like to transfer to NCSU via Globus. I need help with correctly classifying her data and identifying the most appropriate data storage solution based on the classification level. I'm really hoping she can use the research storage, which I have been advocating for our faculty to use more.To provide some context, the research datasets are derived from proprietary NielsenIQ data, which she indicated that NC State is also licensed to use through an institutional agreement with the Kilts Center of the University of Chicago (the data provider). The data cover the US consumer goods sector and contain sensitive information collected from retailers. The datasets are currently stored on secure servers at the National Center for Supercomputing Applications (NCSA) at the University of Illinois.

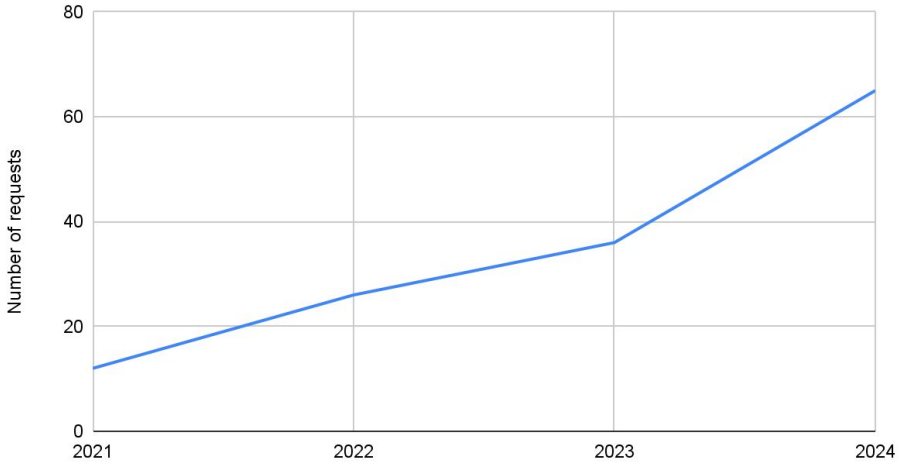
Engineering librarian: Could we do a shared consult please? A faculty member is interested in: 1) General walkthrough of how to deposit data in Dryad, 2) Discussion of metadata and documentation of shared data, 3) How to know what data to share and what to do with data that isn't shared

I would like to schedule a meeting and ask for a workshop on developing a data management plan for a 5-year CDC project. I have started working with the DMP tool and would like to ensure I am filling it out with best practices. I am also not very familiar with data storage at NC State (I started in May) and would like to discuss best practices for storing data (audio recordings, data transcriptions, and other participant consent forms, and participant ID files).

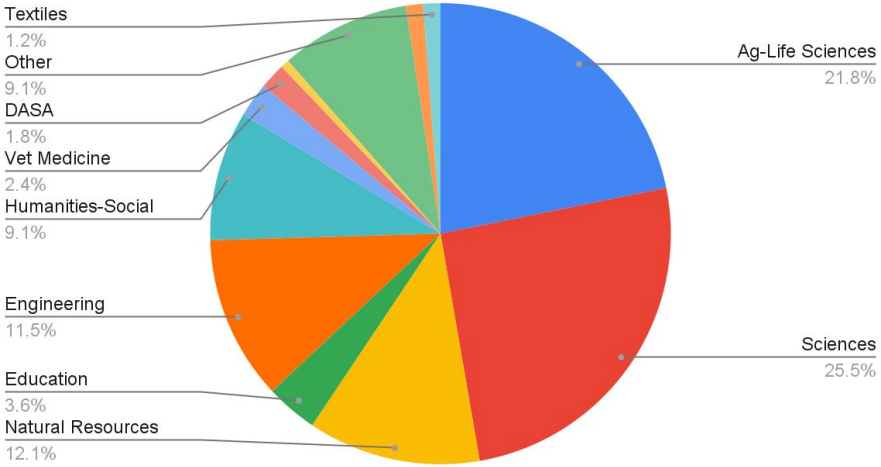
I am completing an application for secondary data and need to know information related to how the data can be stored and accessed virtually for my research team. I will likely need specific details for the application because I do not know how to answer their questions related to data security and storage.

RFS historical statistics

Requests over time



Distribution of requests by college



Reflecting on the past and looking forward

- Strength in partnerships!
  - Coupling RDM and research computing/ research IT
  - How can we creatively fill gaps in resources (both people and infrastructure)
    - To be clear, we are very underfunded and understaffed
- Navigating organizational change
  - New Chancellor
  - New Vice Chancellor for Research and Innovation
  - New (Interim) Associate Vice Chancellor for RCD
- The state of higher ed and the research enterprise
- Ongoing changes to federal agency public access & open access (OA) policies
- Training students– AI, computational skills, data science expertise

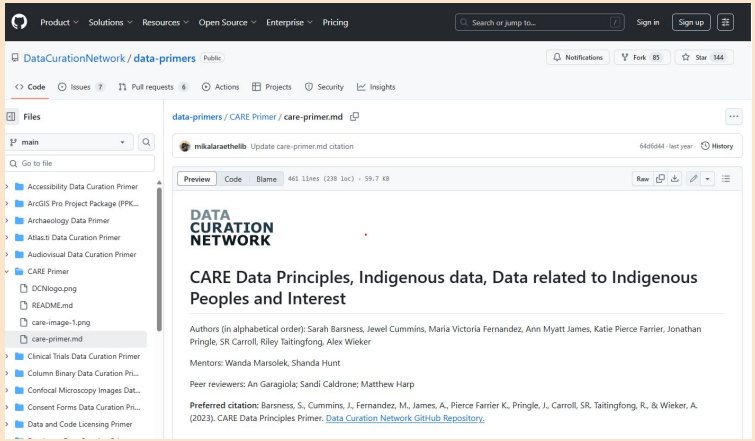
Thanks!

Susan Ivey - slivey@ncsu.edu
Interim Associate Director, Digital Programs & Technology, North Carolina State University Libraries

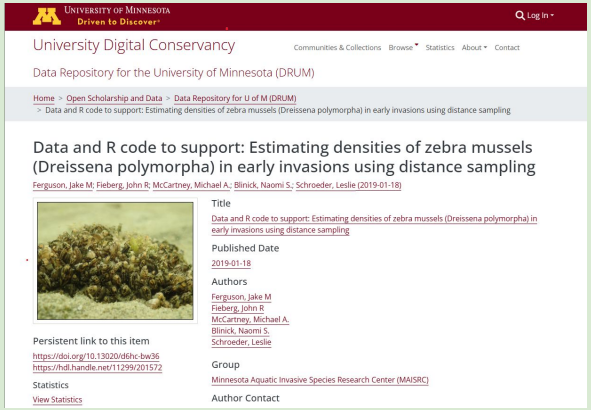
3 THINGS TO KNOW ABOUT THE DATA CURATION NETWORK

I. The Data Curation Network (DCN) is a membership organization of institutional and non-profit data repositories whose vision is to advance open research by making data more ethical, reusable, and understandable.

II. CARE - Data related to Indigenous Peoples and Interest



III, Directory of Datasets Curated by the DCN



3 THINGS TO KNOW ABOUT THE GENERALIST REPOSITORY ECOSYSTEM INITIATIVE

I. The vision of the Generalist Repository Ecosystem Initiative (GREI) is to develop collaborative approaches for data management and sharing through the inclusion of generalist repositories in the NIH data ecosystem. GREI also aims to better enable search and discovery of NIH-funded data in generalist repositories.

II

GREI Zenodo Community

III

NIH
The Dataverse® Project
DRYAD
figshare
MENDELEY
OSF
Vivli
zenodo