

Data Core

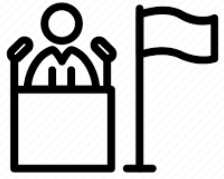
A secure computational enclave for sensitive data analysis



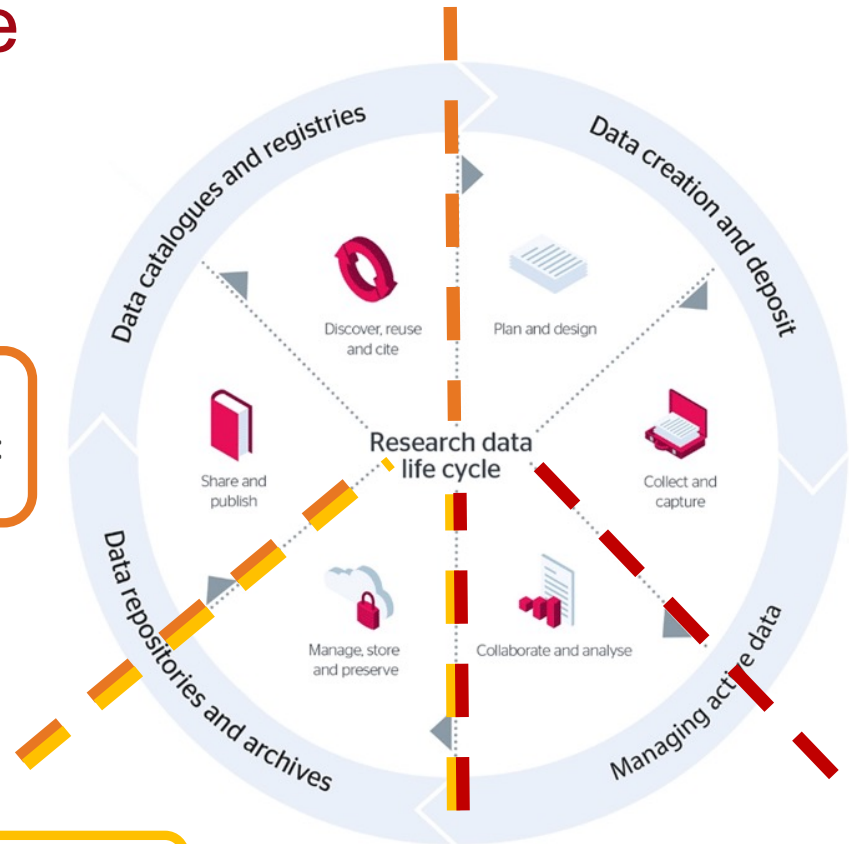
Sarah Ben Maamar,
Associate Director for Research Services
Terrie Rose Wheeler,
Samuel J. Wood library Director

12.12.23
sbm4003@med.cornell.edu

Research Data Lifecycle



Share & publish;
Discover, reuse and cite:
Data Catalog



Manage, store and preserve:
Data Retention tool



Collaborate and analyze:
Data Core & Software Hub

Objective 1:
Build a DM
program

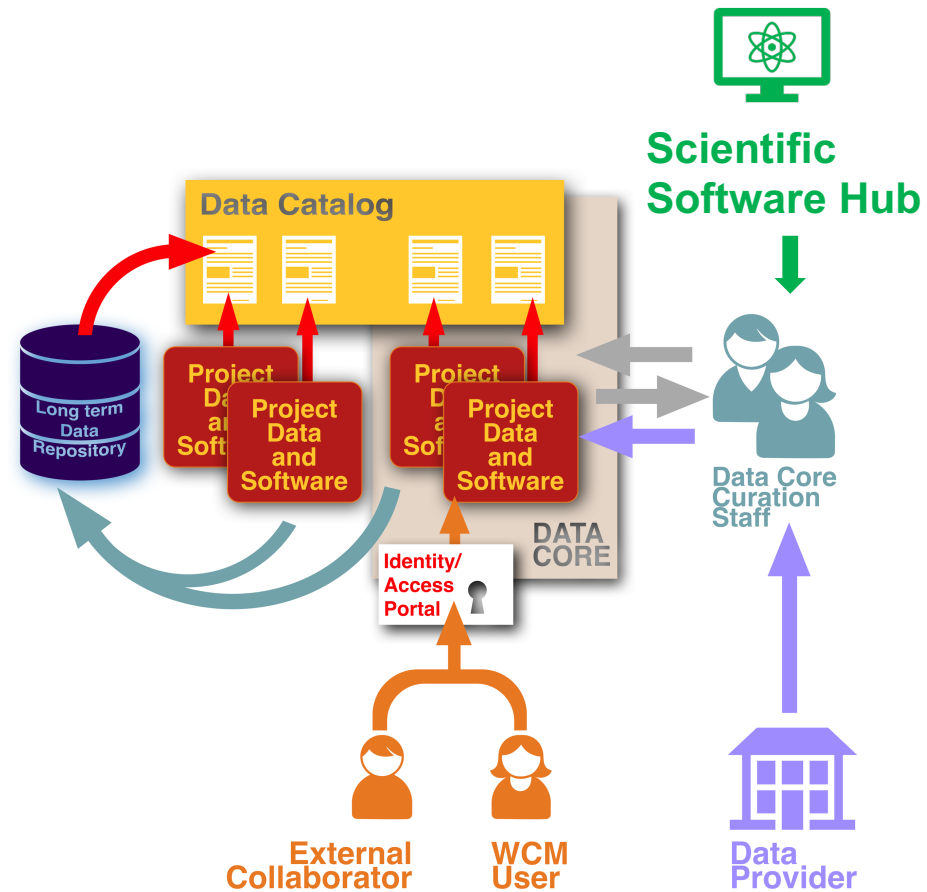
Objective 2:
Engage & FAIR
principles

Objective 3:
Storage
solutions

Kaye, John & Bruce, Rachel & Fripp, Dom. (2017). Establishing a shared research data service for UK universities. Insights the UKSG journal. 30. 59-70. [10.1629/uksg.346](https://doi.org/10.1629/uksg.346).

WCM Data Management Services

- Librarians Curate Data
- Data Core
- Scientific Software Management
- Data Catalog
- Data Retention – started July 2022
- Data Management & Sharing Plan
- Data Repository
- Data Integrity



Objective 1:
Build a DM
program

Objective 2:
Engage & FAIR
principles

Objective 3:
Storage
solutions

Some history

- Institutional demand for a secure enclave
- Sensitive big datasets
- PHS
- Compute and storage
- One to multiple users
- Collaborations
- Department subsidy



Secure



Secured



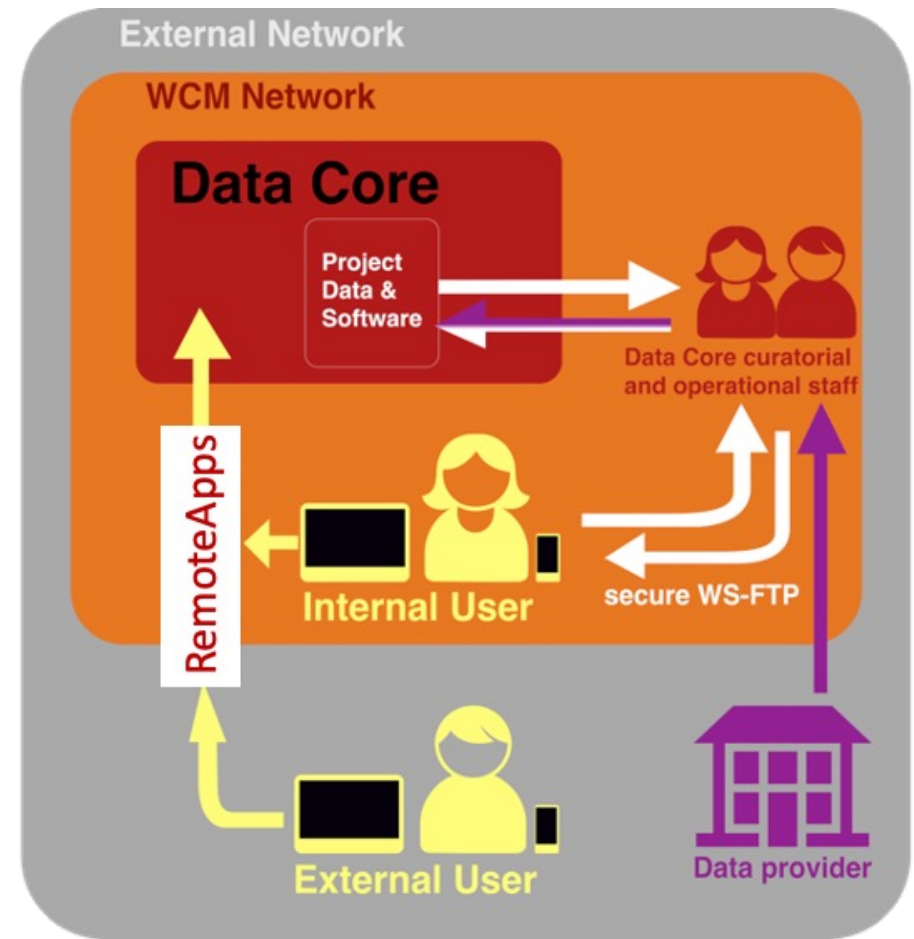
Collaborative



Flexible

| <u>Volume</u> | <u>PIs</u> | <u>Users</u> | <u>Projects</u> |
|---------------|------------|--------------|-----------------|
| Current | 62 | 312 | 83 |
| Total | 82 | 679 | 181 |

WCM IRB and NYP prefer Data Core for ePHI



Design and Implementation of a Secure Computing Environment for Analysis of Sensitive Data at an Academic Medical Center.

Oxley PR, Ruffing J, Champion TR Jr, Wheeler TR, Cole CL.

AMIA AnnuSympProc. 2018 Dec 5; 2018:857-866.

Collaborative

DATA CORE SECURE PROJECT SERVICES

Data Core Service Catalog

View our service catalog, where you can see prices for various configurations of a Data Core project.

SERVICE CATALOG

Request a new Data Core project

This will take you to our form to request a new project.

REQUEST NEW PROJECT

Report an incident with Data Core

If you cannot log in, you have inadvertently deleted a file, or your Data Core project is not behaving as expected.

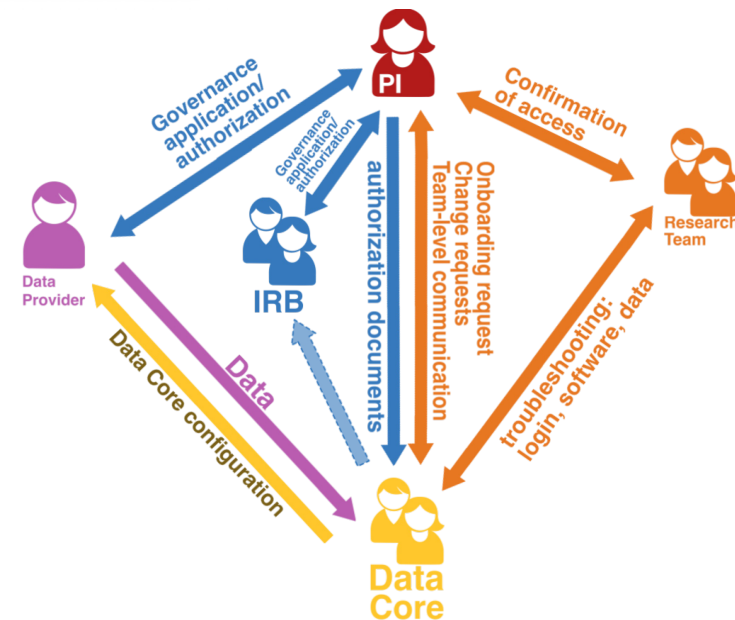
REPORT INCIDENT

Request a change to your Data Core project

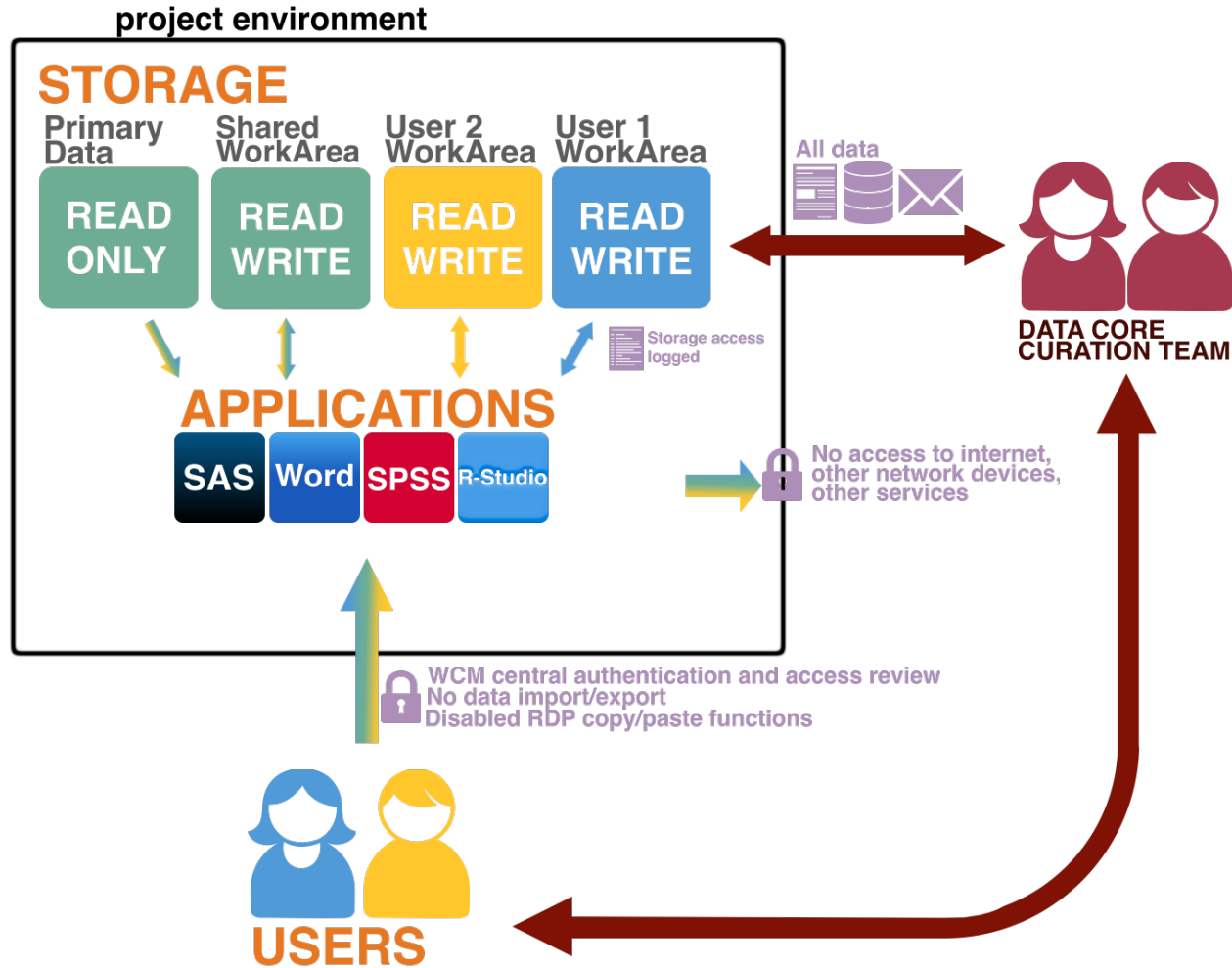
Add users to project, request data import/export, or install new software.

REQUEST CHANGE

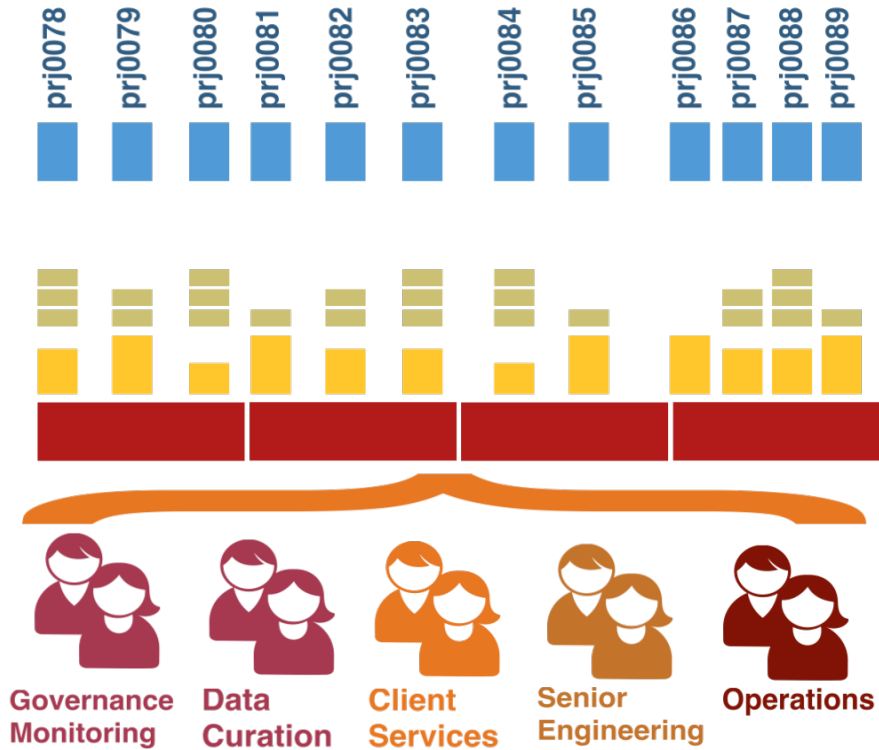
- The New York INSIGHT Clinical Research Network (CRN) COVID study
- NLP project (Llama 2) automated clinical note summarization
- Pediatric Epilepsy LHS project



Data Core environment

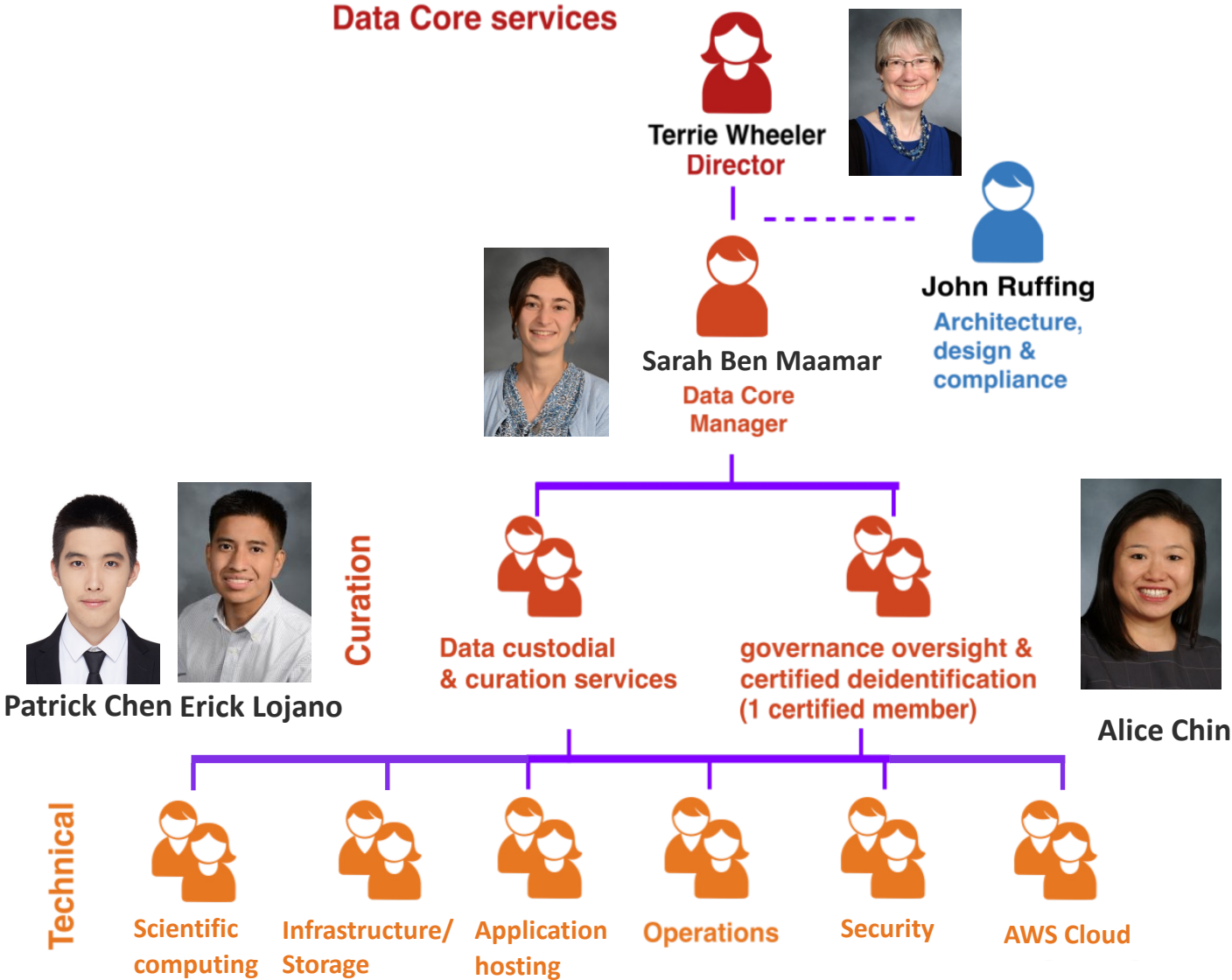


Data Core actors

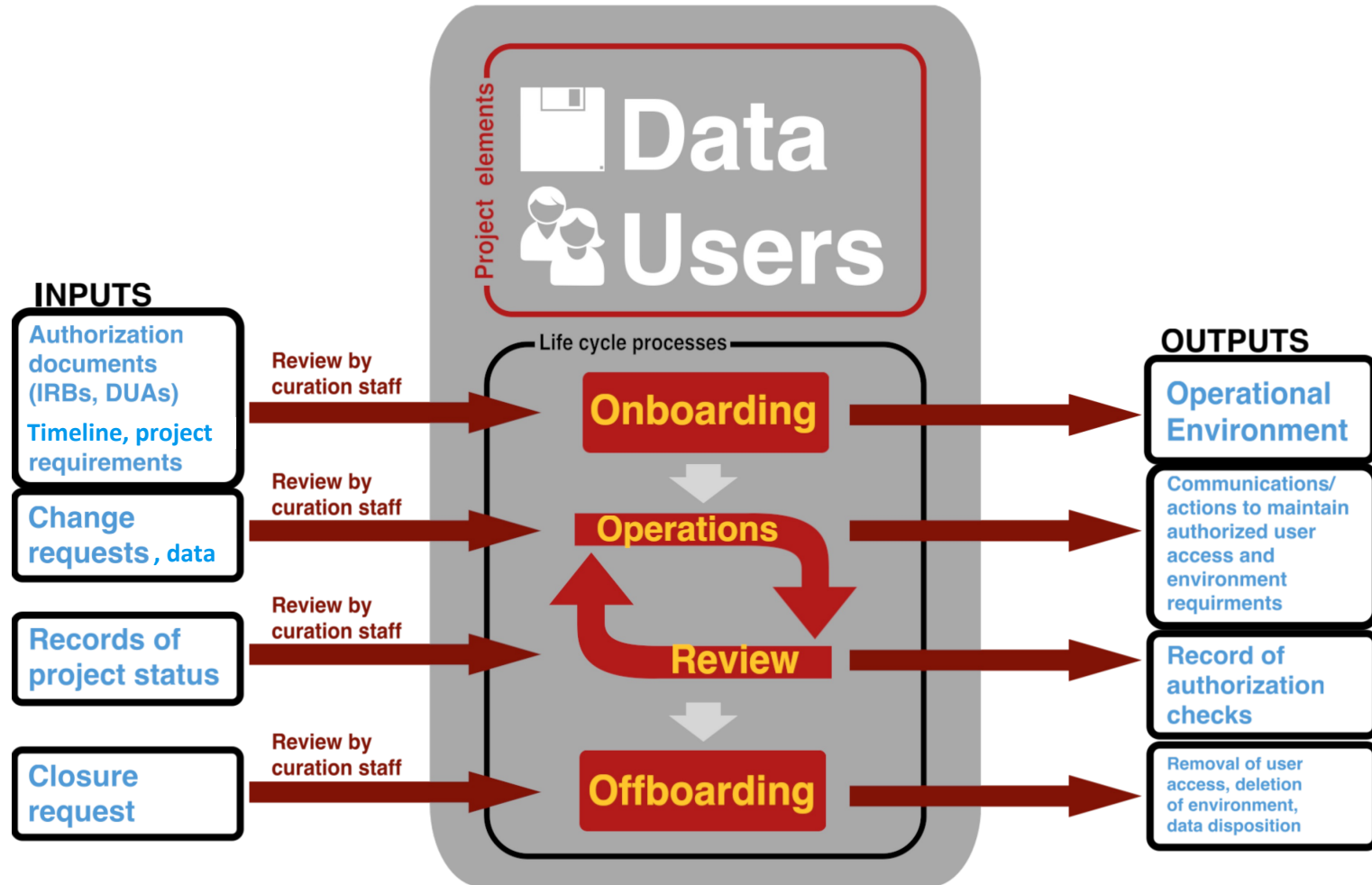


- Infrastructure/Storage team (2)
- Security team (2)
- Operations team (2)
- AWS Cloud team (2)
- Application hosting (1)
- ITS service architect (2)
- Scientific computing team (4)
- Librarians (4)

Data Core team



Data Core operations

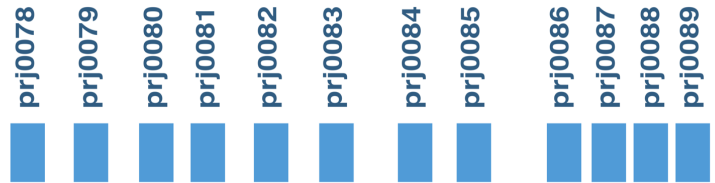


Features and options

- Integrated with Service Management System (ServiceNow)
- Data Core admin web application, “MARIGOLD”
- Cost recovery:
 - # users, Software, Compute
- Grant applications
- Basic licensed statistical software
- Option to add:
 - More memory
 - More CPU
 - GPU
 - Databases (postgresSQL)
 - Storage
 - Additional licensed software
- Connected to Data Catalog, Retention tool (deep archive)



Data Core charge back model



Projects



Software

Storage

Hosting (servers)



Staffing

Monthly cost for a project is typically:

\$0 - 85

\$2.50 - 25/100GB

\$250 (32CPU, 128GB RAM)

\$250/user

TOTAL:

\$502.5 - 610

Status

- Center for Medicaid Services (CMS) certification:
 - **On-premise & AWS Cloud environments**
- Cloud initiatives:
 - **AWS (Windows, Linux, GPU)**
 - **Azure – testing phase**
- Packages repositories:
 - **CRAN (R), active**
 - **Python, in progress**

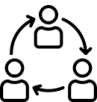




Challenges

- Moving target 
- Training 
- Engagement 
- Cost 

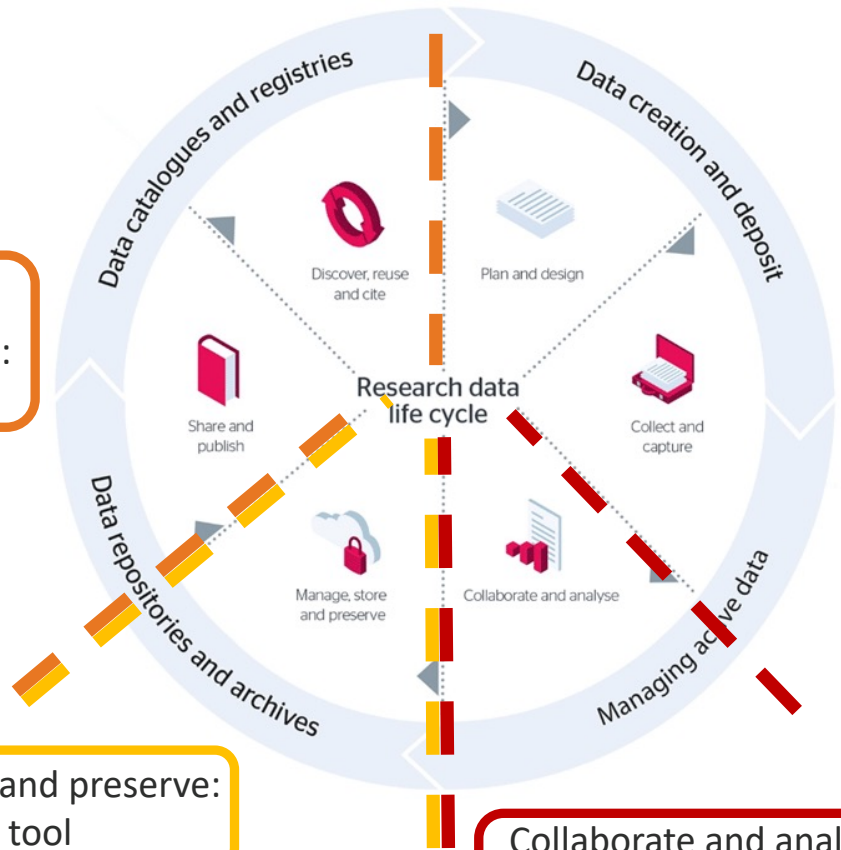
Share & publish;
Discover, reuse and cite:
Data Catalog

And successes

- Collaboration 
- Library/ITS 
- Highly motivated staff 
- Leadership
- Help from researchers

Manage, store and preserve:
Data Retention tool

Collaborate and analyze:
Data Core & Software Hub



Objective 1:
Build a DM program

Objective 2:
Engage & FAIR principles

Objective 3:
Storage solutions

Kaye, John & Bruce, Rachel & Fripp, Dom. (2017). Establishing a shared research data service for UK universities. Insights the UKSG journal. 30. 59-70. [10.1629/uksg.346](https://doi.org/10.1629/uksg.346).

Thank you!

Samuel J. Wood Library

Sarah Ben Maamar,
Associate Director for Research Services
sbm4003@med.cornell.edu

Terrie R. Wheeler,
Library Director
tew2004@med.cornell.edu





**Weill
Cornell
Medicine**