A team approach to enabling streamlined generation, aggregation, management, and reuse of primary data

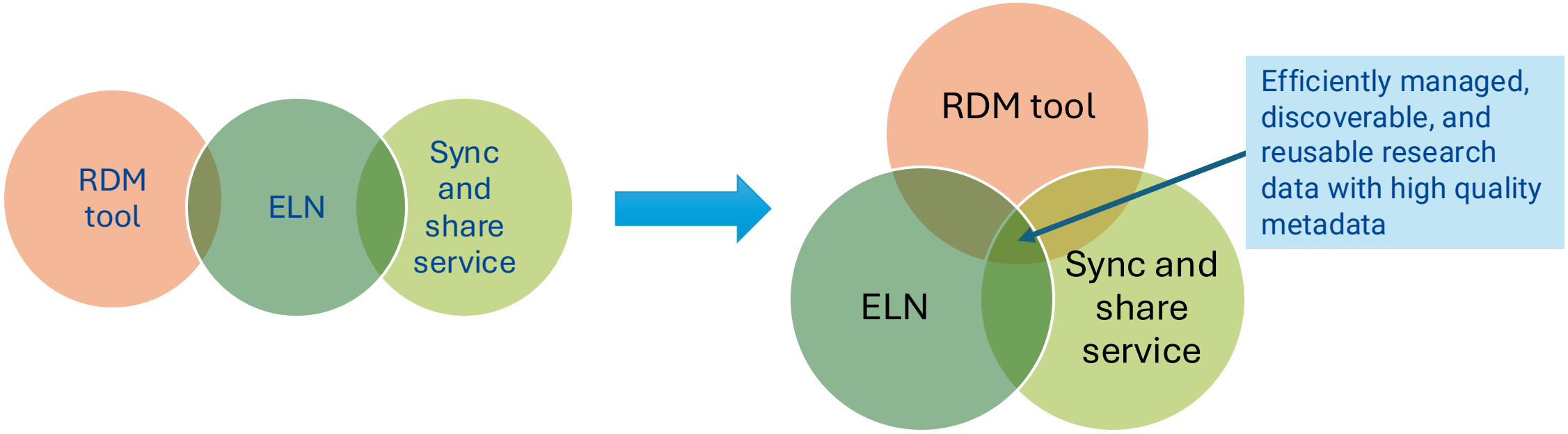
iRODS UGM 2025

Rory Macneil (Research Space)
Terrell Russell (iRODS Consortium)
John Martin (University of North Carolina at Chapel Hill)

Outline

- Existing bottlenecks to efficiently managed, FAIR data
- Making RDM a team sport
  - Why? Enable development of scalable, generalizable solutions
  - How? Leverage individual strengths of tools and services
- Assembling a team: iRODS, RSpace, Dataverse
- Opportunities

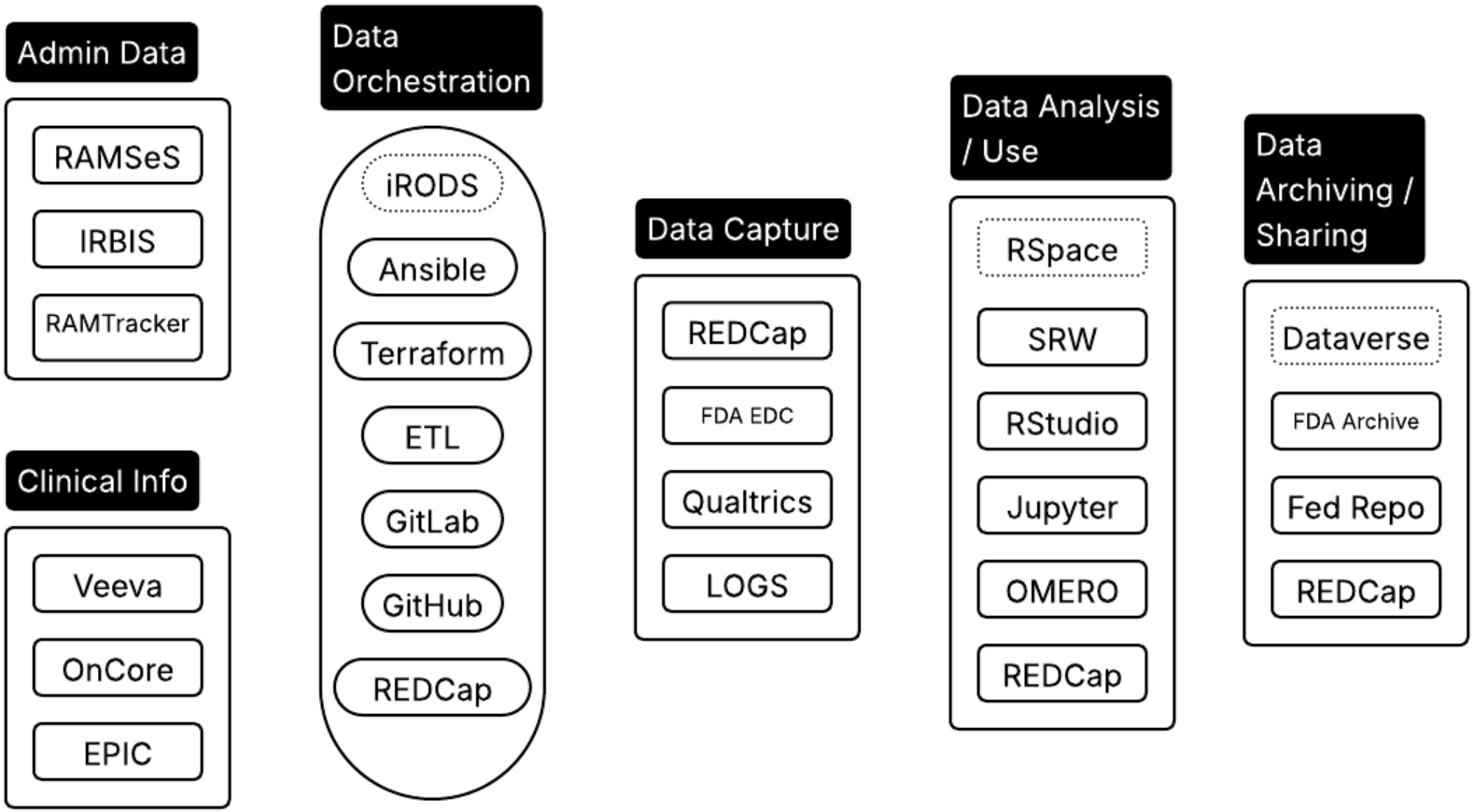
Research data management currently is not concerted



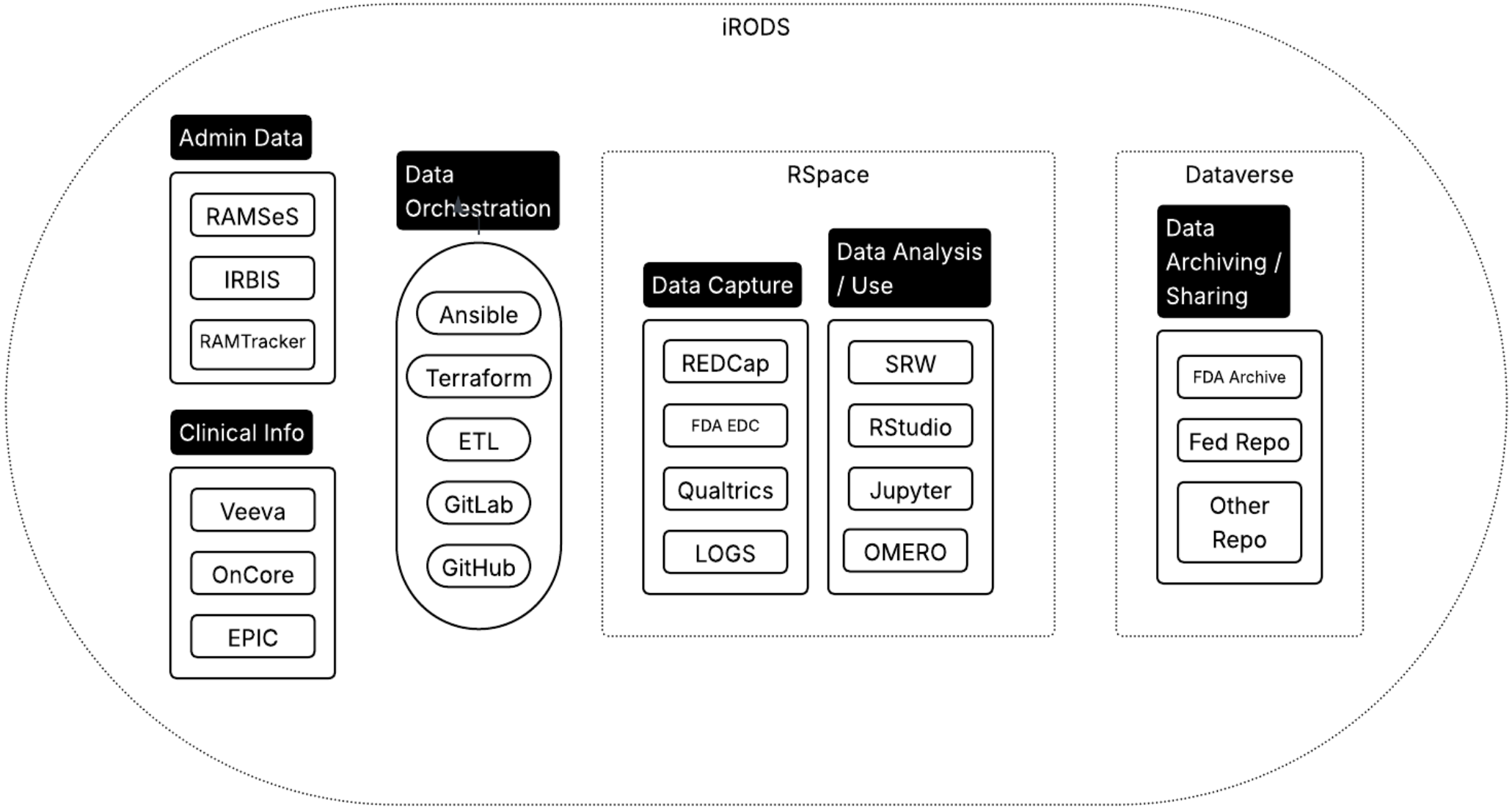
Tools and services are often set up to work in isolation or in sequence with little overlap, rather than coordinated around a shared goal.

If tools/services work together they can create and mutually benefit from efficiently managed FAIR research data, and **address issues of dark and invisible data**.

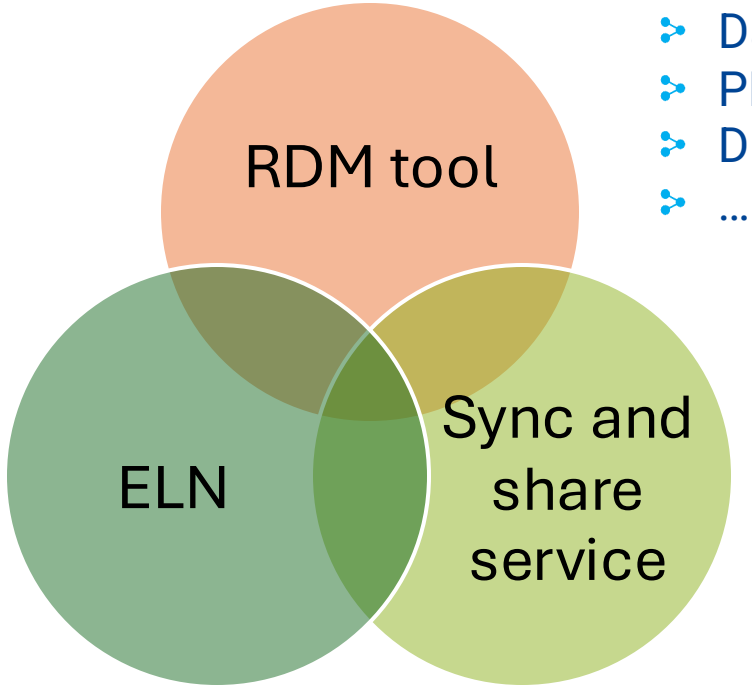
What is...


Admin Data
RAMSeS
IRBIS
RAMTracker
Clinical Info
Veeva
OnCore
EPIC
Data Orchestration
iRODS
Ansible
Terraform
ETL
GitLab
GitHub
REDCap
Data Capture
REDCap
FDA EDC
Qualtrics
LOGS
Data Analysis / Use
RSpace
SRW
RStudio
Jupyter
OMERO
REDCap
Data Archiving / Sharing
Dataverse
FDA Archive
Fed Repo
REDCap

… and what could be?



Our vision:
Concerted research data management starting from creation, to active use, to preservation and publication



- Data curation
- PIDs
- Discoverability
- …

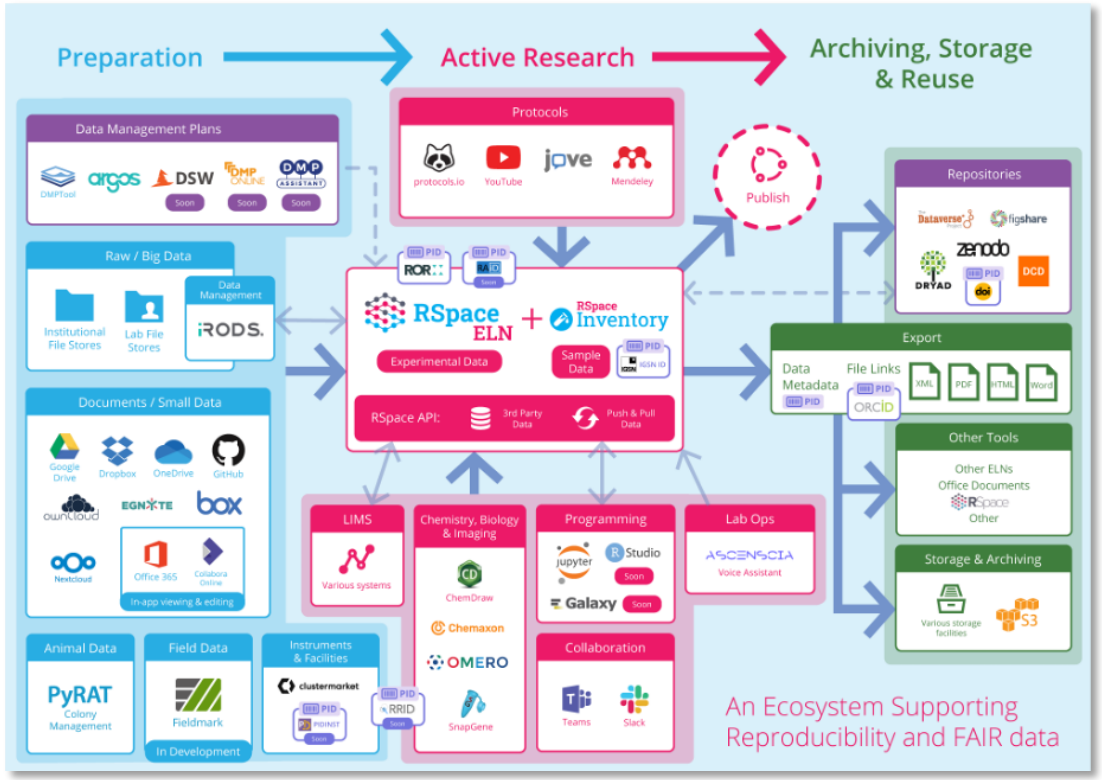
- Context
- Scientific record
- Provenance
- …

- Active data management
- HPC orchestration
- Optimised storage
- Automation
- …

Specialised tools and services bring in different aspects and values for research data management

Our team

- iRODS: file and metadata management, automation
- Dataverse: open source generalist repository providing publication, enabling reuse
- RSpace: RDM 'glue orchestrating primary data and provenance'

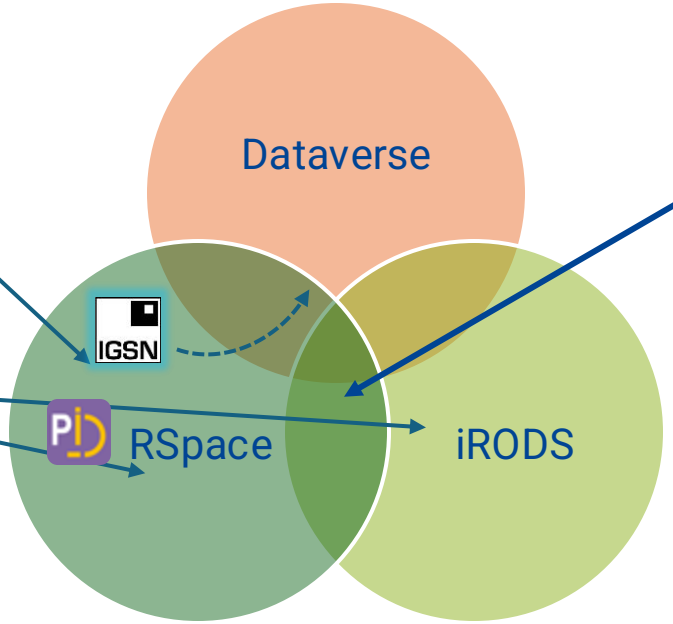


Team RSpace/iRODS/Dataverse

Streamlined generation, aggregation, management, and sharing of primary data

- Field studies
- Sample data/metadata

- Data from instruments
- Instrument/process metadata

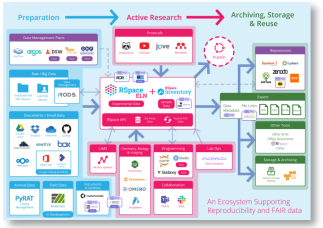


FAIR Data:
Discoverable/Reusable
PIDs: DOI, IGSN, PIDInst, …
Instrument and workflow metadata
Provenance / links to original data

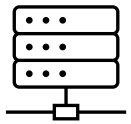
- Data exchange with computing resources
- Workflow metadata and provenance
- Validation and verification

Concerted research data management with iRODS, RSpace and other RDM tools

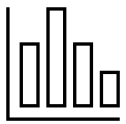
Researcher



Experimental context, physical sample management, primary data



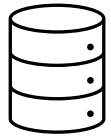
HPC



Data

Curate and manage data

Data Steward



iRODS

✨ Data, metadata stored and synchronised automatically on iRODS with reference to physical sample

The Dataverse® Project

RDM ecosystem

- Data linked to physical samples using globally unique persistent identifier and experimental context is accessible to all services, researchers and data stewards via iRODS
- Through RDM tool data stewards curate research data across the institutions
- Through RSpace researchers provide experimental context, manage their physical samples and associated data and exchange FAIR data with other tools and services (e.g. repositories, DMP tools)

Opportunity 1:
Reusing Dataverse data in RSpace

✅ RSpace already provides export to Dataverse

👉 **Implement workflow to import data from Dataverse to RSpace for reuse**

✨ Improved reproducibility via a practical mechanism for recreation of experimental procedures and workflows for experiments that have been documented in RSpace and in some cases experiments documented elsewhere whose data has been exported to Dataverse.

Opportunity 2:
Metadata Interoperability Framework

👉 Create a flexible metadata interoperability framework that facilitates seamless exchange of research metadata between disparate systems.

✨ Investigate applying existing community metadata schema (e.g. DataCite, RO-Crate), and develop, where necessary, affordances to propagate metadata between RSpace, Dataverse, iRODS, and other integrated platforms (e.g., Yoda, ManGO).

✨ Explore the support for PIDinsts to track and propagate provenance metadata of primary data production from research instruments.

Opportunity 3: REDCap integration for Research Data Synchronization

👉 **Implement REDCap integration as a demonstration case for real-time research data sharing and synchronization across the modular architecture.**

✨ Develop standardized APIs and event-driven mechanisms that enable automatic propagation of data and metadata changes between REDCap, RSpace, and the broader ecosystem.

✨ The implementation will serve as a template for integrating additional research tools of the active research phase, showcasing how disparate systems can build and maintain FAIR research data flows while supporting diverse research workflows.

Opportunity 4:
Cross-System Change Notification

👉 **Design and implement the infrastructure required to ensure timely notification and synchronization of data and metadata changes across integrated systems.**

✨ Explore existing or develop a new message queue architecture with standardized event schemas, subscription mechanisms, and secure notification protocols that allow RSpace, iRODS, Dataverse, and other connected tools to remain aware of relevant changes without tight coupling.

Extending the team – opportunities for FAIR data

Yoda

- Metadata templates
- Automation
- Controlled sharing with internal and external partners globally
- Data publication with automatic PID minting
- …

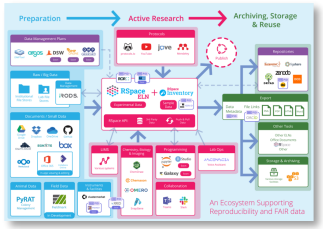
ManGO

- Metadata templates
- Automation
- Controlled sharing with internal and external partners globally
- Automated metadata extraction
- …

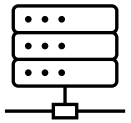
Research Space is part of the OSNL Yoda proposal

Extending the team – opportunities for RDM solutions

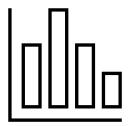
Researcher



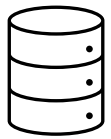
Experimental context, physical sample management, primary data



HPC



Data

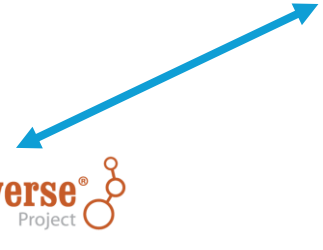


iRODS

Curate and manage data

Data Steward

✨ Data, metadata stored and synchronised automatically on iRODS with reference to physical sample



RDM ecosystem

- Data linked to physical samples using globally unique persistent identifier and experimental context is accessible to all services, researchers and data stewards via iRODS
- Through RDM tools like Yoda, Mango, etc., data stewards curate research data across the institutions
- Through RSpace researchers provide experimental context, manage their physical samples and associated data and exchange FAIR data with other tools and services (e.g. repositories, DMP tools)

Join us

- Have you solved these already?
- Are you working on these now?
- Does this sound like your future?

Thank you!

iRODS®

RSpace

The Dataverse Project

Instantiate the principle and take advantage of the potential of Vertical Interoperability

Ensure seamless passage of data and metadata **across research tools involved in every stage of the research lifecycle**, from the initial stages of research planning, to active staging and analysis, to archiving and deposit of research outputs