RSpace and iRODS
Update, Plans and Opportunities

Rory Macneil, Research Space

Terrell Russell, iRODS

iRODS UGM
June 15, 2023
The Research Data Lifecycle

- **UCL Research Data Repository (for data)**
- **UCL Discovery (for papers / monographs)**
- **UCL Explore**
- Bibliographic / citations databases
- Specialist data repositories
- Commercial and/or generalist data repositories

**UCL-managed systems**

- **DMP Online**
- **Worktribe** n.b. Not all research projects are funded by external funders, some are simply undertaken informally, covered by salary

**Research Data Lifecycle**

- Experiments
- Clinical trials
- Images
- Literature
- Sensors
- Datasets
- Software Models
- Surveys & Interviews

+ metadata

Ideally using RDSS or Data Safe Haven, but may include:
- Laptop hard drives
- Departmental Storage
- UCL Filestore
- OneDrive
- Dropbox or other commercial cloud services
- Storage managed by collaborators
- B2Store (EU service)
- Many others!

UCL Research Publication Service (a.k.a. Symplectic Elements)

- UCL Research Data Repository (for data)
- Some specialist data repositories

- UCL Research Data Repository (for data)
- UCL Research Publication Service (for papers / monographs)
- Specialist data repositories
- Commercial and/or generalist data repositories

Myriad, Kathleen – RITS HPC facilities
National / International HPC facilities
Locally-installed software (a vast number of specialist software tools are used for data analysis)
Problem

Lack of data-centric tools

Lack of interoperability between tools

Siloed data

Impeding FAIR

Creating friction for researchers
RSpace: Data-centric digital research platform designed to interoperate with and connect research infrastructure.
RSpace @ UCL

The Research Data Lifecycle

- UCL Research Data Repository (for data)
- UCL Discovery (for papers / monographs)
- UCL Explore
  - Bibliographic / citations databases
  - B2Find (EUDAT Service)
  - Specialist data repositories
  - Commercial and/or generalist data repositories

- UCL Research Publication Service (a.k.a. Symplectic Elements)

Research Data Lifecycle

- Funding bid
- Data collection
- Data storage
- Data deposit
- Data analysis
- Article / Data Publication
- Preservation & Curation
- Discovery
- Access
- Planning
- Idea

- Worktribe
  - n.b. Not all research projects are funded by external funders, some are simply undertaken informally, covered by salary

Experiments
  - Clinical trials
  - Images
  - Literature
  - Sensors
  - Datasets
  - Software Models
  - Surveys & Interviews

Ideally using RDSS or Data Safe Haven, but may include:
  - Laptop hard drives
  - Departmental Storage
  - UCL Filestore
  - OneDrive
  - DropBox or other commercial cloud services
  - Storage managed by collaborators
  - B2Safe / B2Drop (EUDAT Services)
  - Many others!

Upcoming RSpace Integrations

UCL-managed systems
RSpace @UCL
Powering a FAIR research data/metadata flow at the institutional level
RSpace + iRODS

Together:

GUI and ecosystem + policy and auditing

Can provide an easy-to-use, highly connected, shared namespace with data discovery meeting all the components of the FAIR principles

This can be a useful, and extensible, platform/glue in RDM ecosystems being deployed and developed by both institutions and national infrastructure providers
Phase 1 Integration

Solving the ‘Broken Links’ Problem

RSpace - iRODS video
Phase 1 Integration
Leibniz Institute of Aging use case

“We are using the Yoda system from Utrecht for archiving. Yoda is the frontend and gives us a useful API and a system where we can freeze files (vault area). With iRODS icommands we can push large files into the archive.

Our goal is to connect our systems more closely in the future, including to store big files in iRODS / Yoda and link them in the ELN.” Fabian Monheim
Phase 2 integration:
Exposé RSpace metadata to iRODS for inclusion in broader ‘metadata lake’

- Use triggers to push metadata into iRODS, upon
  - Export of RSpace data to a repository
  - ‘Publication’ of RSpace data
  - Application of tags to RSpace documents

- A document title along with the user’s ORCID and ROR, possibly combined with a (contactable) author name that identifies the user outside RSpace, could be pushed to iRODS on creation of a document

- A set of constrained values allowed by a form in RSpace could constitute relevant metadata to push to iRODS
Phase 2 integration:
Metadata relating to files tracked by iRODS that are linked to in RSpace is exposed in RSpace

- Combining this with search in RSpace would make iRODS documents ‘findable’ in RSpace by their metadata. Potentially useful for ManGO users who use RSpace.

- The richer the meta data in iRODS, the more we get out of this – e.g., it might be beneficial for ManGO users who can add rich metadata to their iRODS documents.
Phase 2 integration:
Enhanced workflows from leveraging the RSpace/OMERO integration and planned iRODS/OMERO integration

- RSpace and OMERO could have access to each other’s metadata if they both stored things in the iRODS catalog
Phase 2 Integration

Terrell

“There is still much to consider”

“We need use cases and community discussion”
Opportunities

- RSpace + iRODS @ institutions (Leibniz Institute for Aging, Leibniz Institute for Natural Product Research, Leiden University)

- RSpace + iRODS in institutional RDM ecosystems (e.g. ManGO)

- RSpace + iRODS in Research Commons
"Bring together data with cloud computing infrastructure and commonly used software, services and applications for managing, analyzing and sharing data to create an interoperable resource for a research community”

Scott Yockel, University Research Computing Officer
Harvard University
Towards a Data Commons at Harvard
Generalist Research Commons

- Universities
  Towards a Data Commons at Harvard

- National research organizations
  Canada’s Digital Research Alliance as a Research Commons

- Supra-national research organizations
  EUDAT Collaborative Data Infrastructure
Core components of generalist Research/Data Commons Canada’s DRA

- Repositories
- Data management planning tool
- Storage
- PID provision
- Discovery
iRODS and RSpace in the Digital Research Alliance Research Commons
Core components of generalist Research/Data Commons

EUDAT’s CDI

- Repositories
- Data management planning tool
- Storage
- PID provision
- Discovery
iRODS and RSpace in the
EUDAT Controlled Data Infrastructure
iRODS and RSpace in the Harvard Commons
Some news and an invitation

- RSpace is going open source

- We would love to work with institutions, NRENS and projects using iRODS (and Dataverse, OMEROMERO and other RDM tools RSpace integrates with) to explore and develop the ideas presented above
Questions?