RSpace + iRODS = Powering research infrastructures

Ander Astudillo, Rory Macneil, and Tilo Mathes

SURF and Research Space
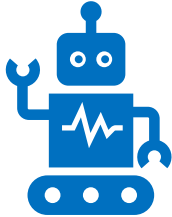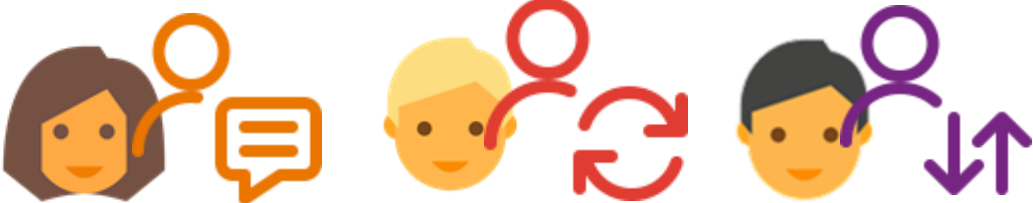
iRODS UGM 2024

Outline

iRODS as RDM enabler

Realizing enablement in Research Commons
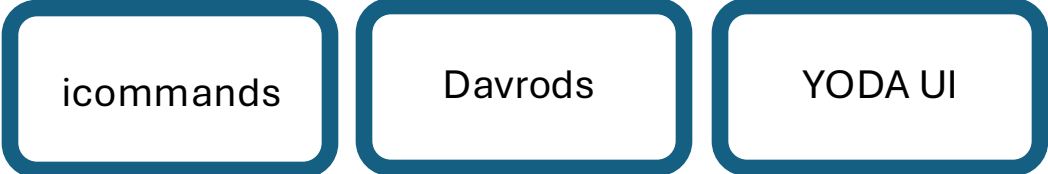
Current iRODS<>RSpace integration

Demo: Stage 2 integration

Towards a universal RDM frontend

An iRODS stack
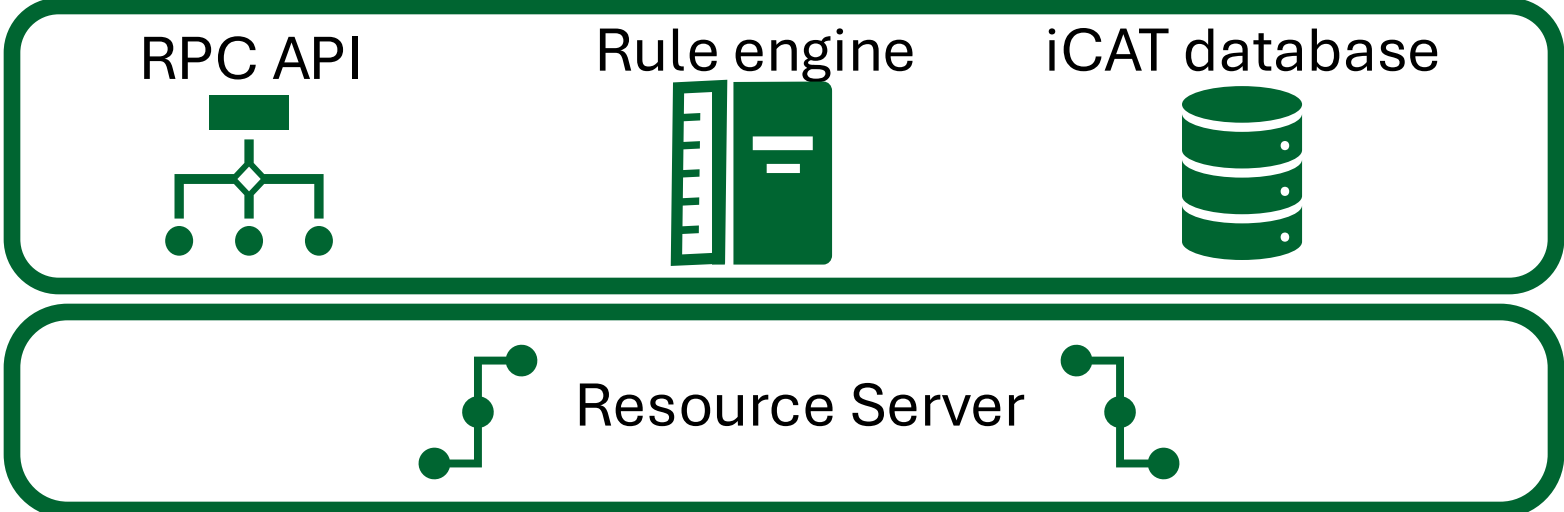
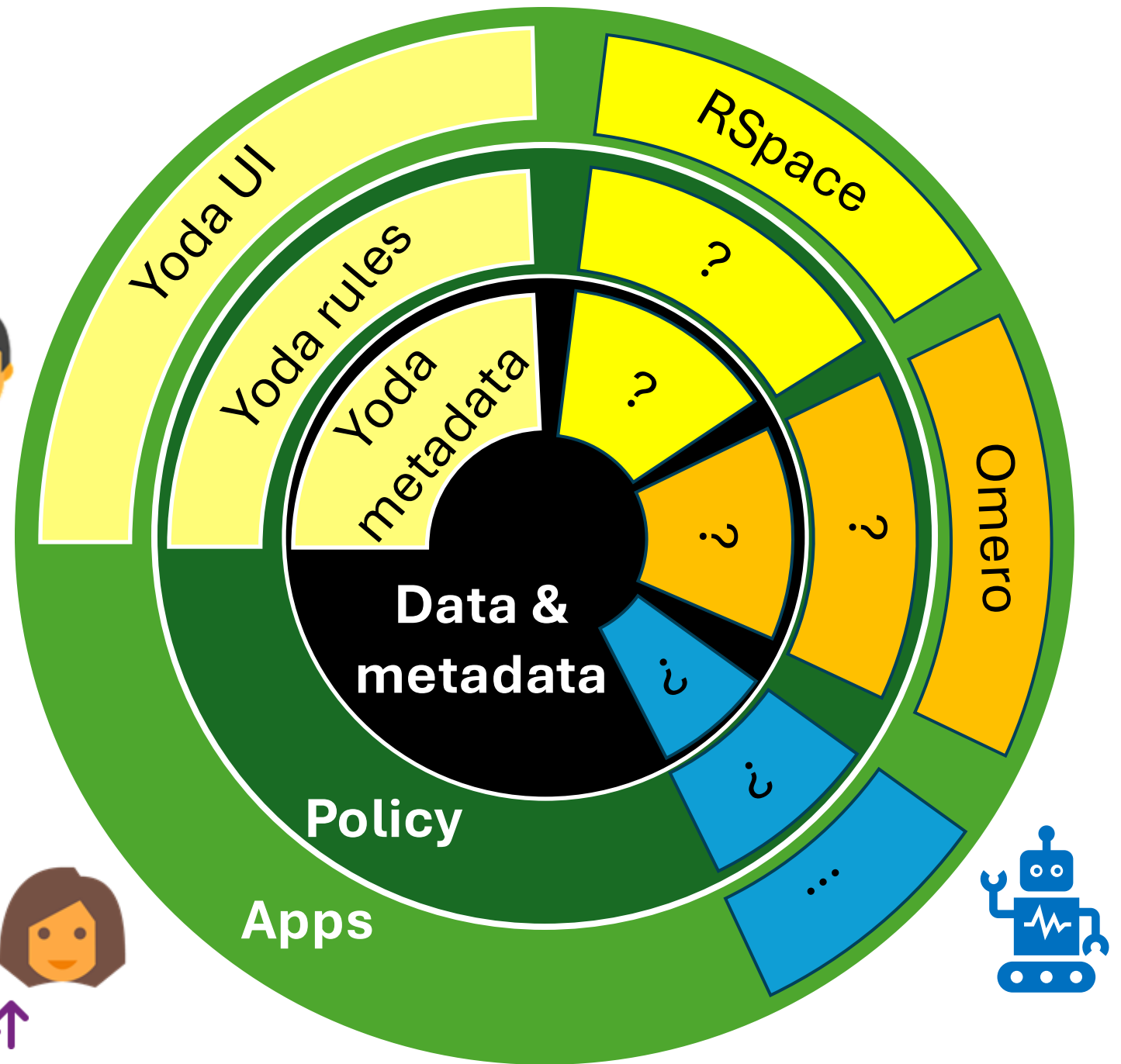**Interface layer**

icommands | Davrods | YODA UI

iRODS layer

RPC API | Rule engine | iCAT database

Resource Server
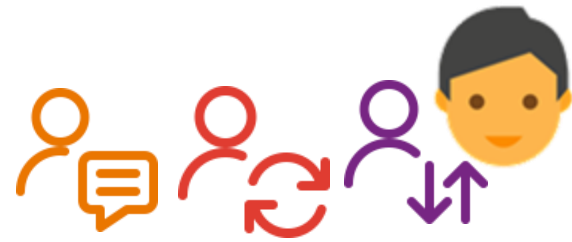
Storage layer

Storage resc. 1 | Storage resc. 2 | Storage resc. 3

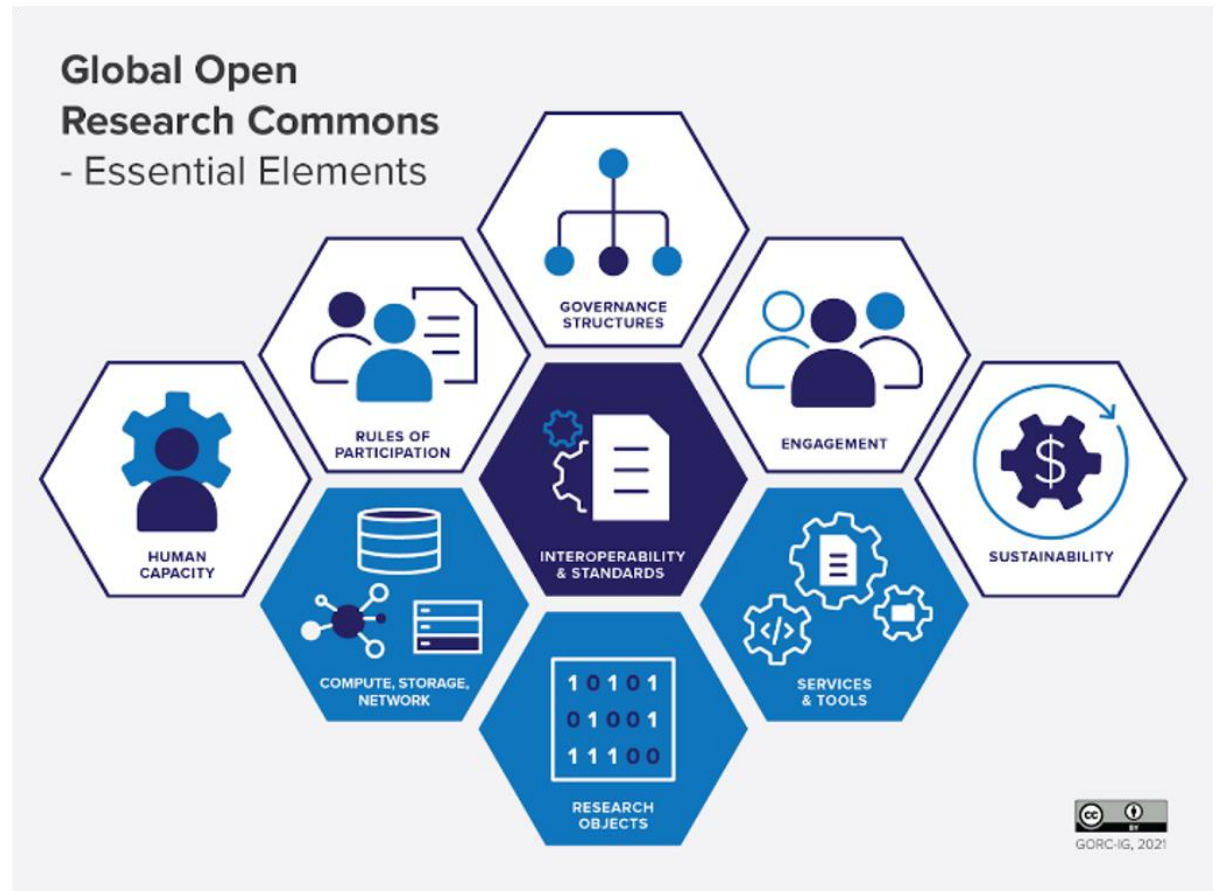Apps allow collaboration on a data management platform

Realizing enablement 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, Harvard University
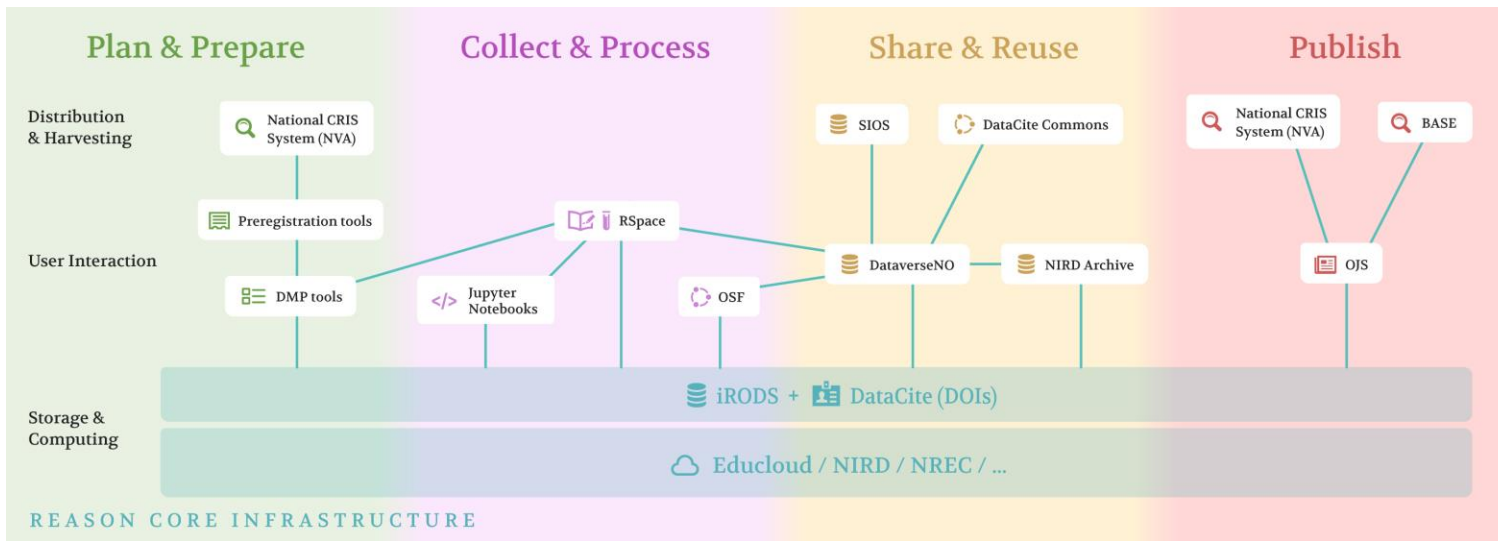

Global Open Research Commons - Essential Elements

GOVERNANCE STRUCTURES
RULES OF PARTICIPATION
ENGAGEMENT
HUMAN CAPACITY
INTEROPERABILITY & STANDARDS
SUSTAINABILITY
COMPUTE, STORAGE, NETWORK
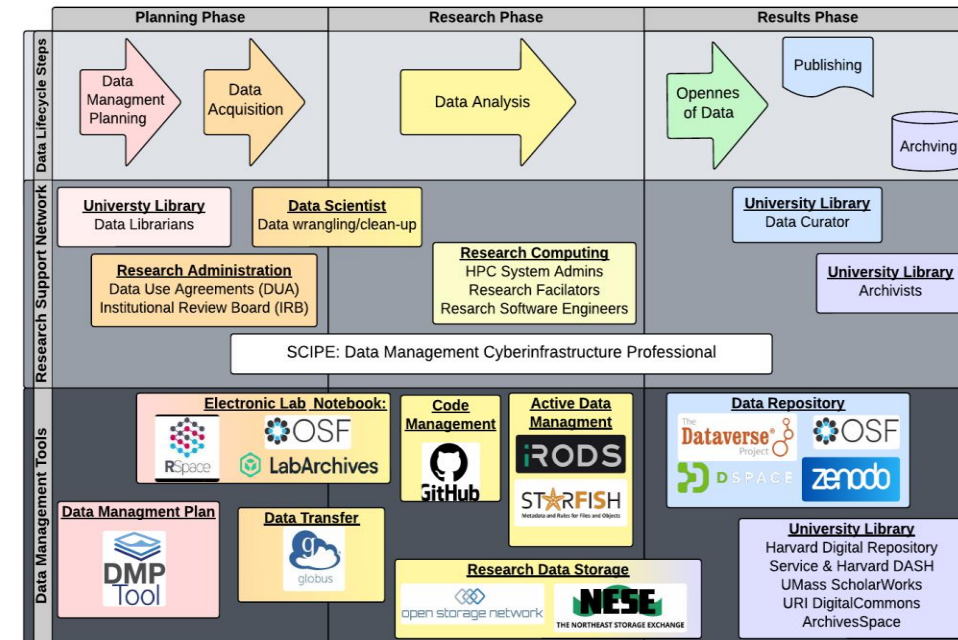SERVICES & TOOLS
RESEARCH OBJECTS

GORC-IG, 2021

Realizing enablement in Research Commons

REASON
Proposed Norwegian Research Commons
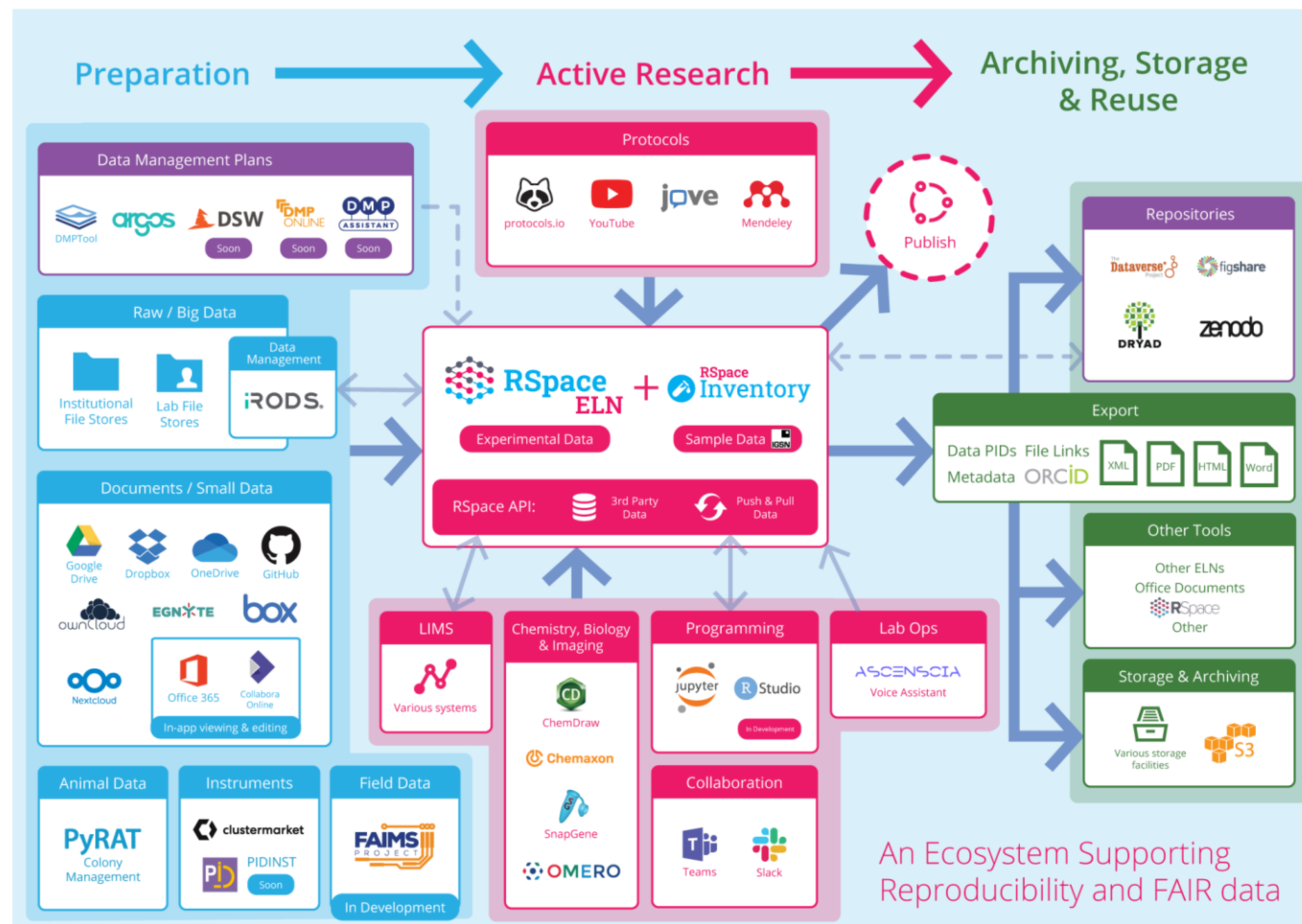

HARVARD Data Connect

Contextualized in
Research Data Lifecycle

Key issue: interoperability
- Storage ⟷ Apps (iRODS)
- Between Apps (RSpace)

RSpace and its ecosystem

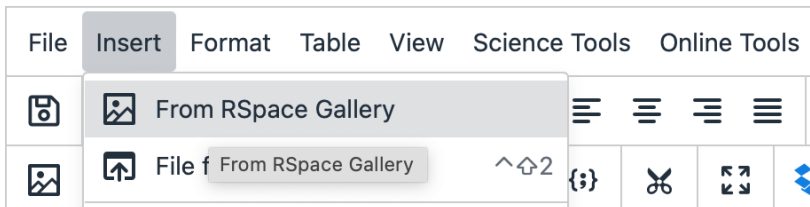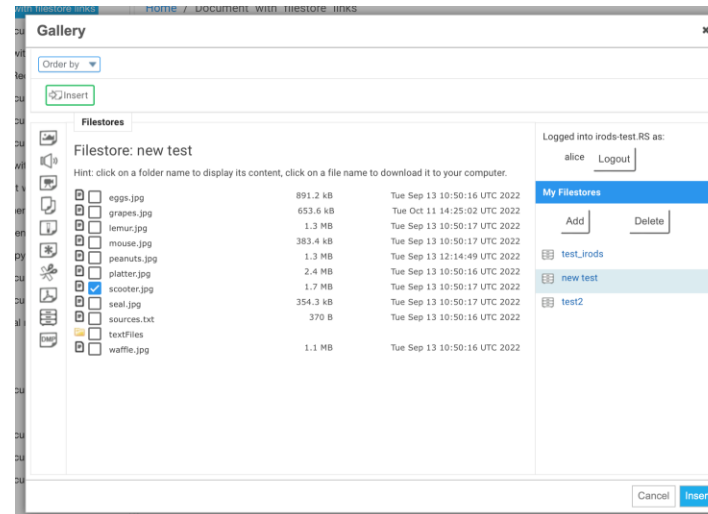An Ecosystem Supporting Reproducibility and FAIR data

Stage 1: RSpace <> iRODS integration

Existing functionality

- Administration of connection to iRODS
- Create links to iRODS objects in ELN documents
- Retrieve linked files from iRODS upon export from RSpace

Move/copy files to iRODS

MVP

We are here

iRODS & RSpace

UGM 2024

RDM Frontend vision

Researchers seamlessly work with internal <u>and externalized files</u> via efficient RSpace workflows that **synchronize metadata** and **maintain provenance** information

"[...] it would be of much interest to us if the [...] RSpace [...] gallery could act as a front end for iRODS to allow upload of data to iRODS (possibly even defaulting to storage in iRODS).

[...] to allow retrieval of data from the iRODS interface (outside of RSpace), it would be good if RSpace were to update a specific metadata field in the iRODS catalogue when files are linked to experiment entries, so that querying the metadata catalogue for a particular document identifier could retrieve all data associated with that document"

Dr. Sam Eyley, PI, KU Leuven

Stage 3: Metadata exchange between RSpace and iRODS

Provenance - Preserving links and version history between RSpace documents and externalised research data

Discoverability - Finding and accessing content in iRODS and RSpace via RSpace-managed tags (ontologies, custom tags, ...) and document IDs (Forms, Templates)

Approach

- Metadata schema exploration
- Programmatic vs. RO-Crate facilitated metadata exchange
- Working with Leiden U to address specific usecase(s)

Stage 4: Roadmap towards a universal RDM frontend

Q2/3 - Metadata transfer

- **Provenance** tracking and **discoverability** through synchronized metadata between iRODS and RSpace
- Explore **RO-crate** as a universal vessel for object metadata transfer

Q3/4 - RDM frontend usability

- Improved RSpace Gallery workflow, e.g. direct upload to iRODS, drag&drop, …
- *iRODS file store (metadata) search from within RSpace*

Functional improvements (tbd)

- *Active management workflows for iRODS files and metadata from within RSpace*
- *iRODS as default storage for RSpace files*

RSpace + iRODS: from project based use to research infrastructure

- *For institutions*
- *For Research Commons and Research Clouds*

RSpace:iRODS:Yoda workflow
Example: semi-automatic metadata enrichment

Researcher selects data to add to documentation in RSpace

Data is uploaded directly to iRODS for storage and discoverability

Metadata of RSpace document is passed along to iRODS

author, experiment ID, project ID, tags, instrument, …

Yoda metadata (e.g., ePIC,) is passed back to RSpace along with iRODS file ID

RSpace document receives link to data on iRODS
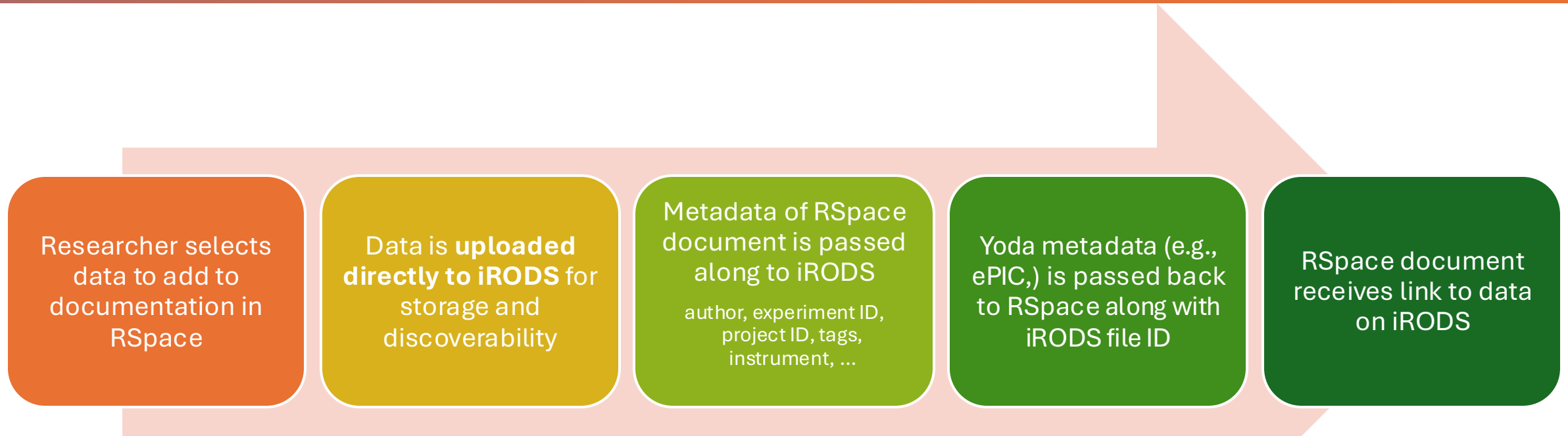
*Inspired by **Fabian Monheim** (Leibniz Institute on Aging (FLI), Jena)

FLI uses **iRODS**, **Yoda**, and **RSpace** in their research data ecosystem

NB Our open source transition is happening on June 26ᵗʰ!

Open Source model

- RSpace to be licensed under AGPL v3
- Research Space maintains RSpace on GitHub
- Research Space provides services to implement, adopt, and maintain RSpace

Collaborating with institutional partners

- Co-development of solutions with customers and partners (Leibniz Institute on Aging , Leiden, TU Delft,..)
- Shout out to Leonardo Lenoci!

Individual contributions

- Pitch ideas / share feedback & experiences / contribute code / support the community
- We welcome engagement with you!

Stay in touch with RSpace open-source!

Opensource community mailing list →

GitHub: https://github.com/rspace-os

Blog: https://www.researchspace.com/blog

Contact: opensource@researchspace.com