# IRODS AND SYNC & SHARE, GETTING THE BEST OUT OF BOTH WORLDS

RON TROMPERT

# Research Drive Recap

- Once upon a time there was……

FLEXIBLE RESEARCH DATA SHARING

SURF DRIVE

PERSONAL SYNC-AND-SHARE

SURF RESEARCH DRIVE

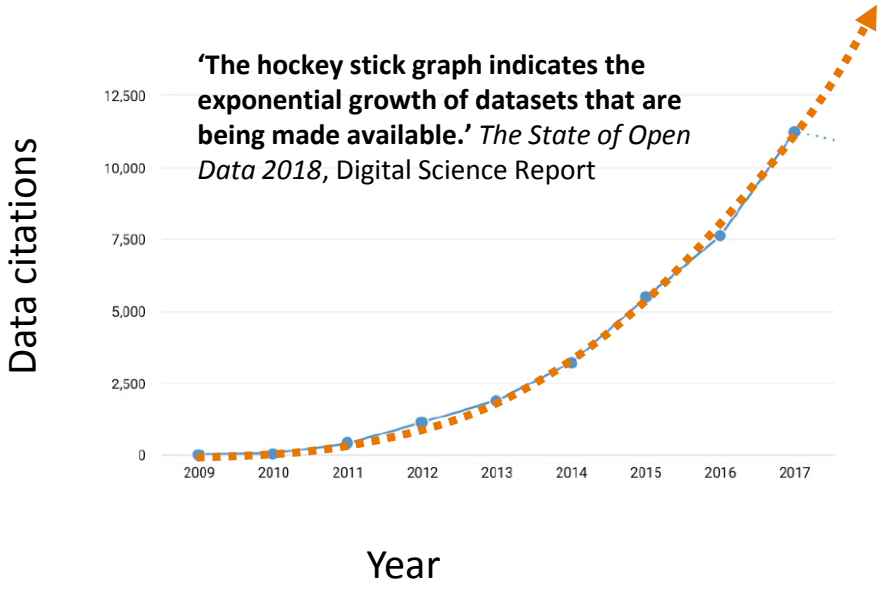
# Research Drive Recap

- Research Drive attracted a lot of attention from universities and other higher education
- Why?

# Research Drive Recap

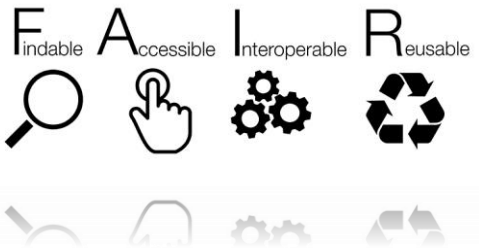
## *Lots of data*

**'The hockey stick graph indicates the exponential growth of datasets that are being made available.'** *The State of Open Data 2018*, Digital Science Report




**Research becomes more data-intensive and more interdisciplinary – and researchers need the right tools to do their job (in a way that complies with their institute's policies & guidelines)**

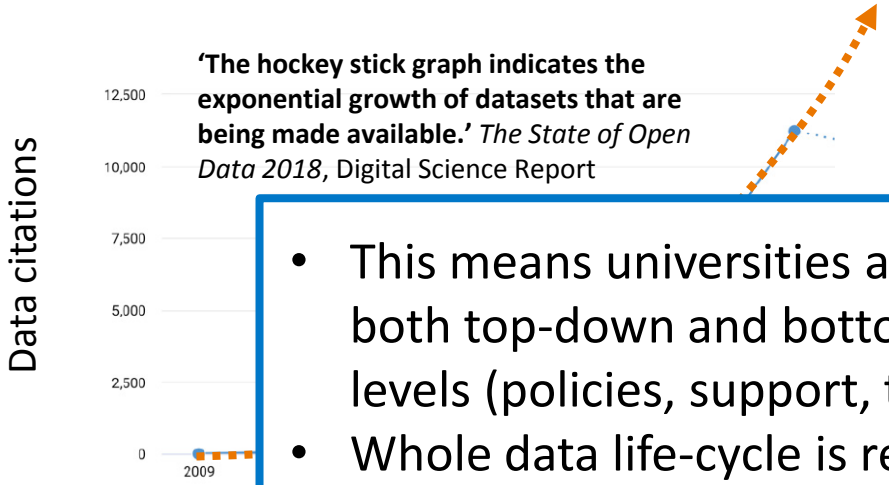
## *Lots of attention, lots of ambition*



**The FAIR Principles and Open Science are on the agenda of university boards, funders and the government.**

# Research Drive Recap

## *Lots of data*




**'The hockey stick graph indicates the exponential growth of datasets that are being made available.'** *The State of Open Data 2018*, Digital Science Report

## *Lots of attention, lots of ambition*

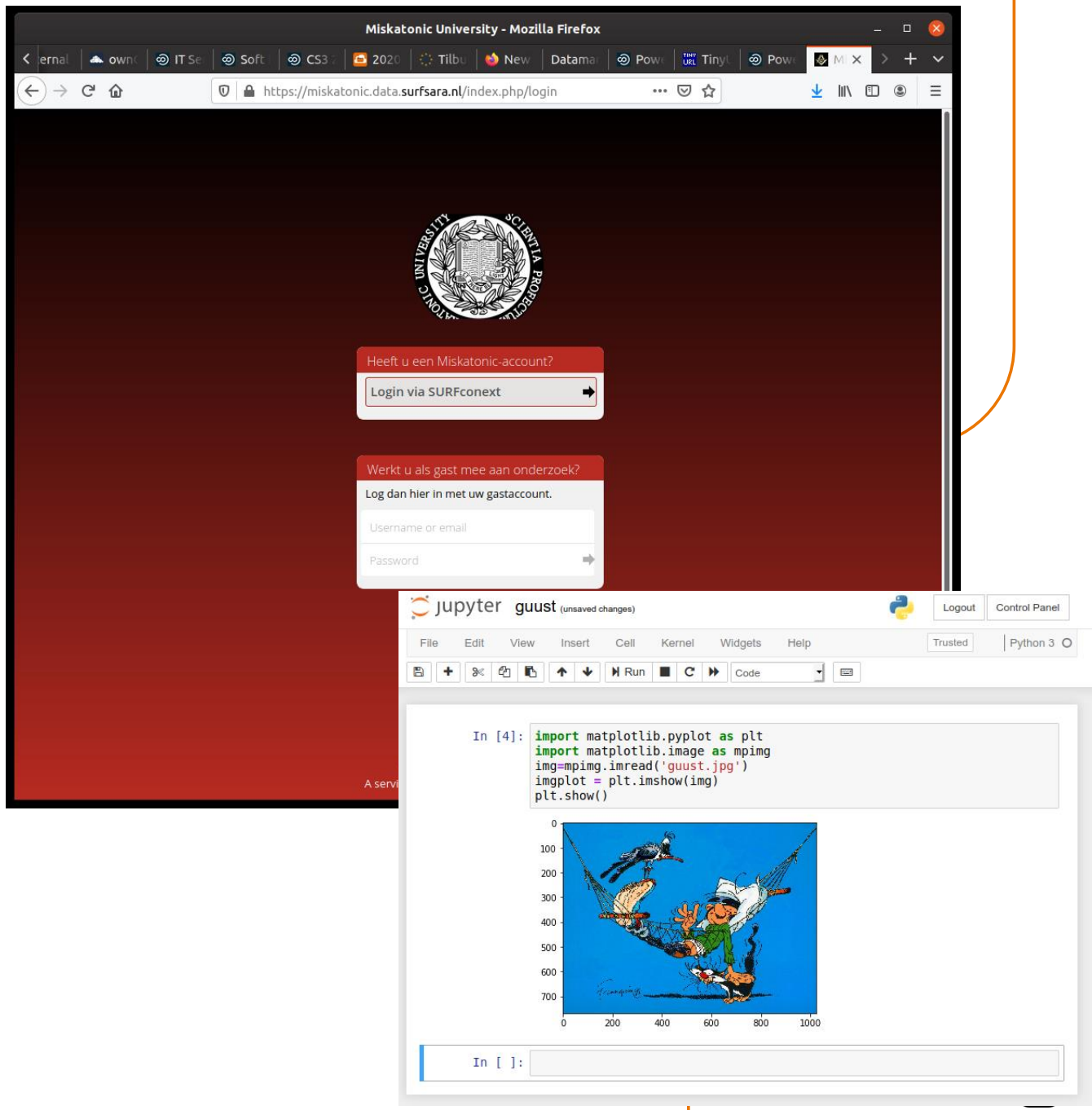
- This means universities and faculties experience a sense of urgency – both top-down and bottom-up – to offer better support for RDM on all levels (policies, support, technology, etc.)
- Whole data life-cycle is relevant, but long-term archival and publication of data are often seen as a priority.

**Research becomes more data-intensive and more interdisciplinary – and researchers need the right tools to do their job (in a way that complies with their institute's policies & guidelines)**

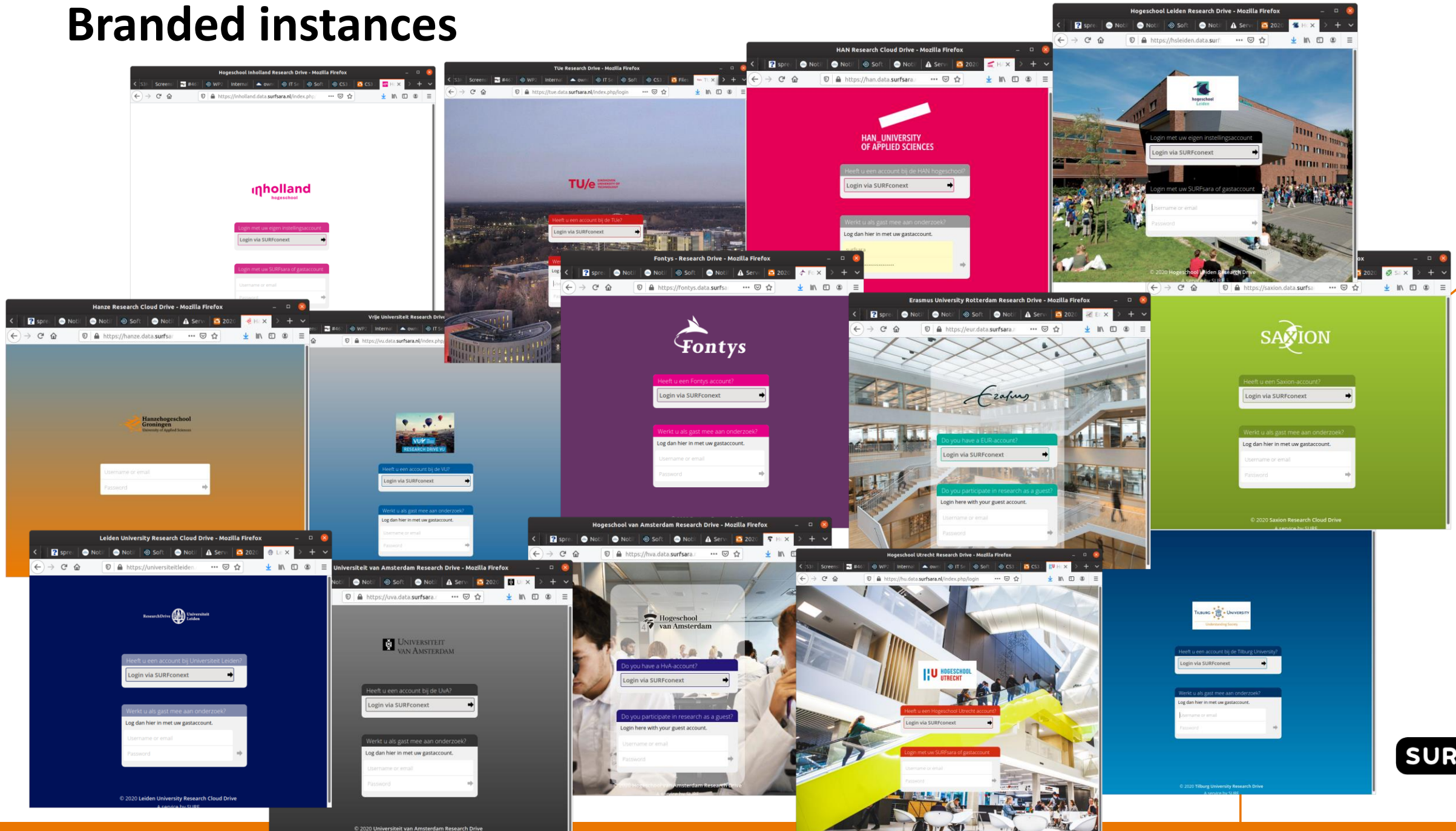
**The FAIR Principles and Open Science are on the agenda of university boards, funders and the government.**

# Research Drive Recap

- Sync-and-Share Service with focus on the needs of the research community
  - Owncloud-based
  - Features
    - Your own branding
    - Sharing outside of the organisation.
    - Multiple AUTHN methods
    - Integration with external storage systems
    - Functional accounts supporting data stewardship
    - Quota as large as you can afford
    - Collaborative editing
    - Jupyter
    - Pay per TB and not per user
- More details, see last years talk: [https://tinyurl.com/rtvpvhe](https://tinyurl.com/rtvpvhe)



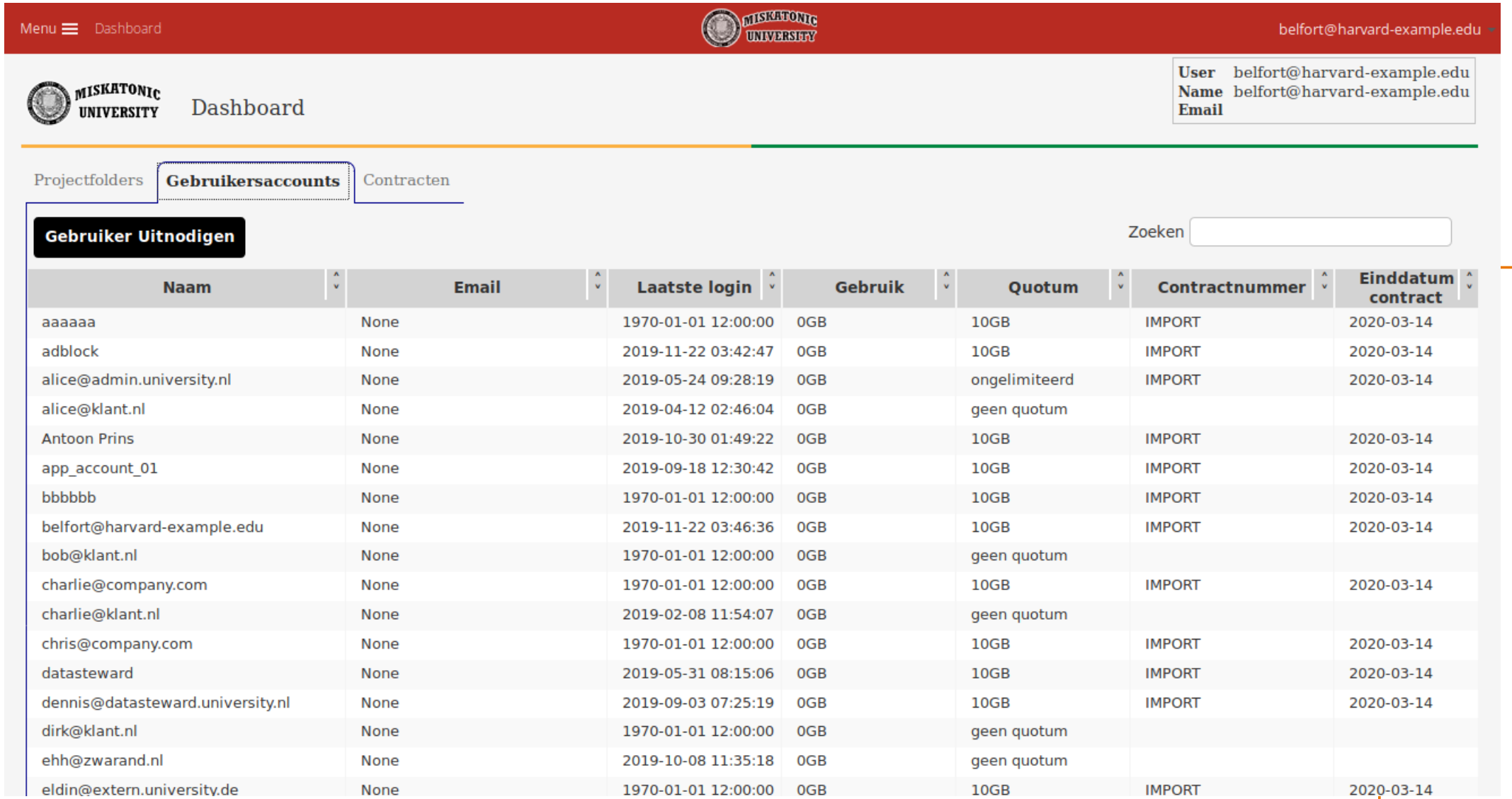
# Branded instances



# Research Drive Roles

- University purchases an amount of storage capacity (cake)
- Hands out parts (piece) of it to faculties
- Faculties hand out parts of their  amount of storage to people responsible for project data
- Different roles and capabilities are needed
- Supports internal billing

# Dashboard



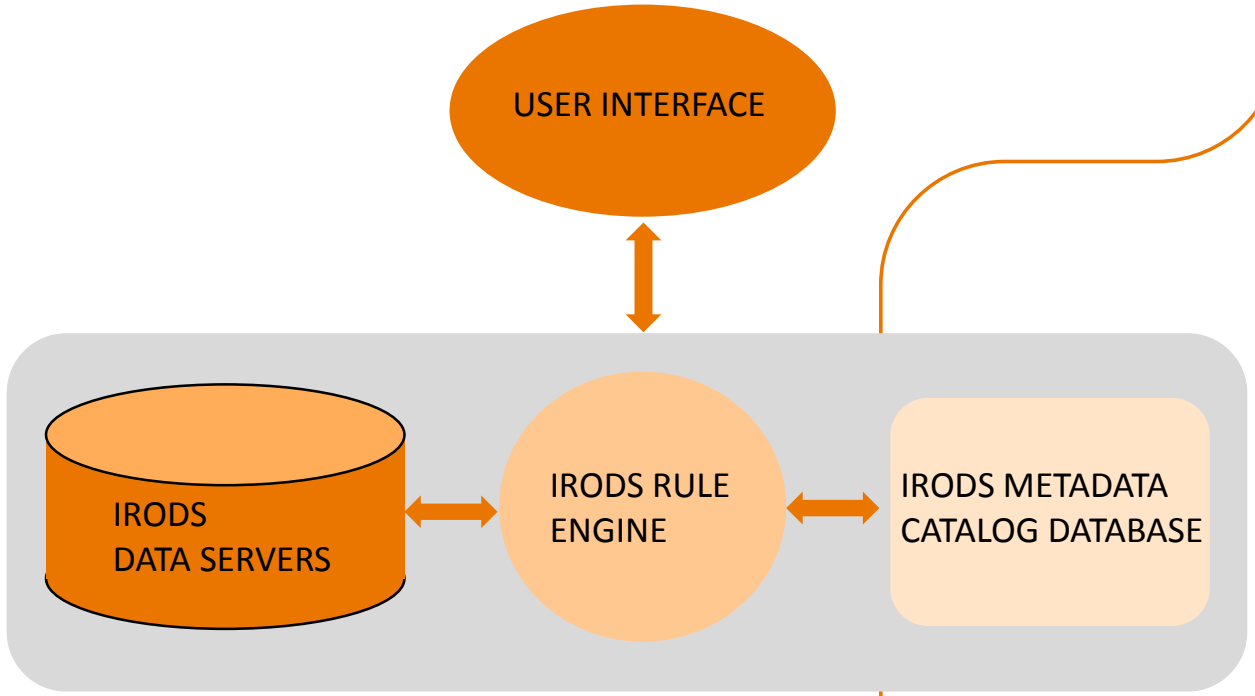

| Naam | Email | Laatste login | Gebruik | Quotum | Contractnummer | Einddatum contract |
|---|---|---|---|---|---|---|
| aaaaaa | None | 1970-01-01 12:00:00 | 0GB | 10GB | IMPORT | 2020-03-14 |
| adblock | None | 2019-11-22 03:42:47 | 0GB | 10GB | IMPORT | 2020-03-14 |
| alice@admin.university.nl | None | 2019-05-24 09:28:19 | 0GB | ongelimiteerd | IMPORT | 2020-03-14 |
| alice@klant.nl | None | 2019-04-12 02:46:04 | 0GB | geen quotum | | |
| Antoon Prins | None | 2019-10-30 01:49:22 | 0GB | 10GB | IMPORT | 2020-03-14 |
| app_account_01 | None | 2019-09-18 12:30:42 | 0GB | 10GB | IMPORT | 2020-03-14 |
| bbbbbb | None | 1970-01-01 12:00:00 | 0GB | 10GB | IMPORT | 2020-03-14 |
| belfort@harvard-example.edu | None | 2019-11-22 03:46:36 | 0GB | 10GB | IMPORT | 2020-03-14 |
| bob@klant.nl | None | 1970-01-01 12:00:00 | 0GB | geen quotum | | |
| charlie@company.com | None | 1970-01-01 12:00:00 | 0GB | 10GB | IMPORT | 2020-03-14 |
| charlie@klant.nl | None | 2019-02-08 11:54:07 | 0GB | geen quotum | | |
| chris@company.com | None | 1970-01-01 12:00:00 | 0GB | 10GB | IMPORT | 2020-03-14 |
| datasteward | None | 2019-05-31 08:15:06 | 0GB | 10GB | IMPORT | 2020-03-14 |
| dennis@datasteward.university.nl | None | 2019-09-03 07:25:19 | 0GB | 10GB | IMPORT | 2020-03-14 |
| dirk@klant.nl | None | 1970-01-01 12:00:00 | 0GB | geen quotum | | |
| ehh@zwarand.nl | None | 2019-10-08 11:35:18 | 0GB | geen quotum | | |
| eldin@extern.university.de | None | 1970-01-01 12:00:00 | 0GB | 10GB | IMPORT | 2020-03-14 |

# Moving up higher in the stack

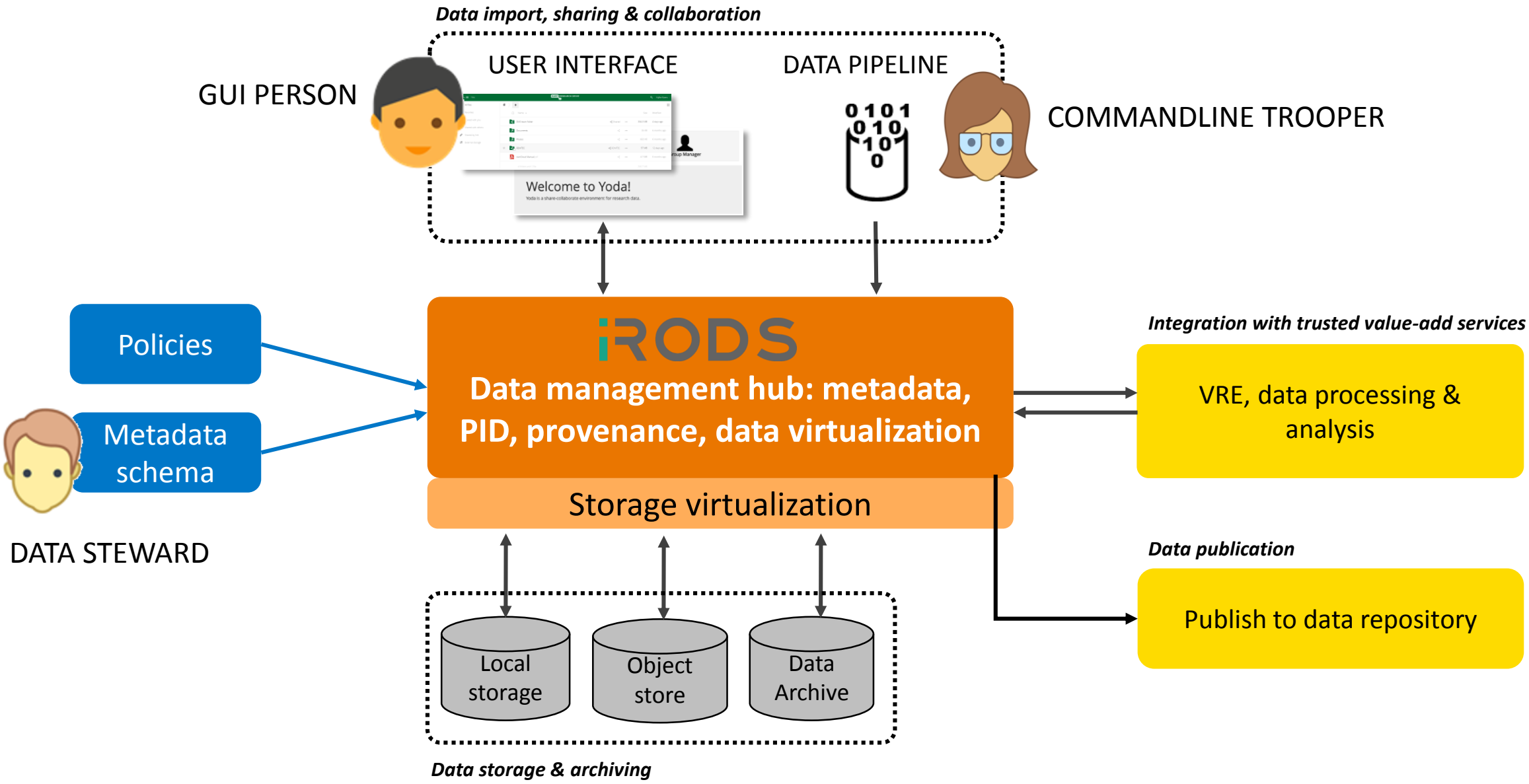
- Sync-and-Share is not just about storage
  - Collaborative editing
  - Research interfaces
    - Jupyter, Spark, **iRODS**
  - Data repositories

# iRODS

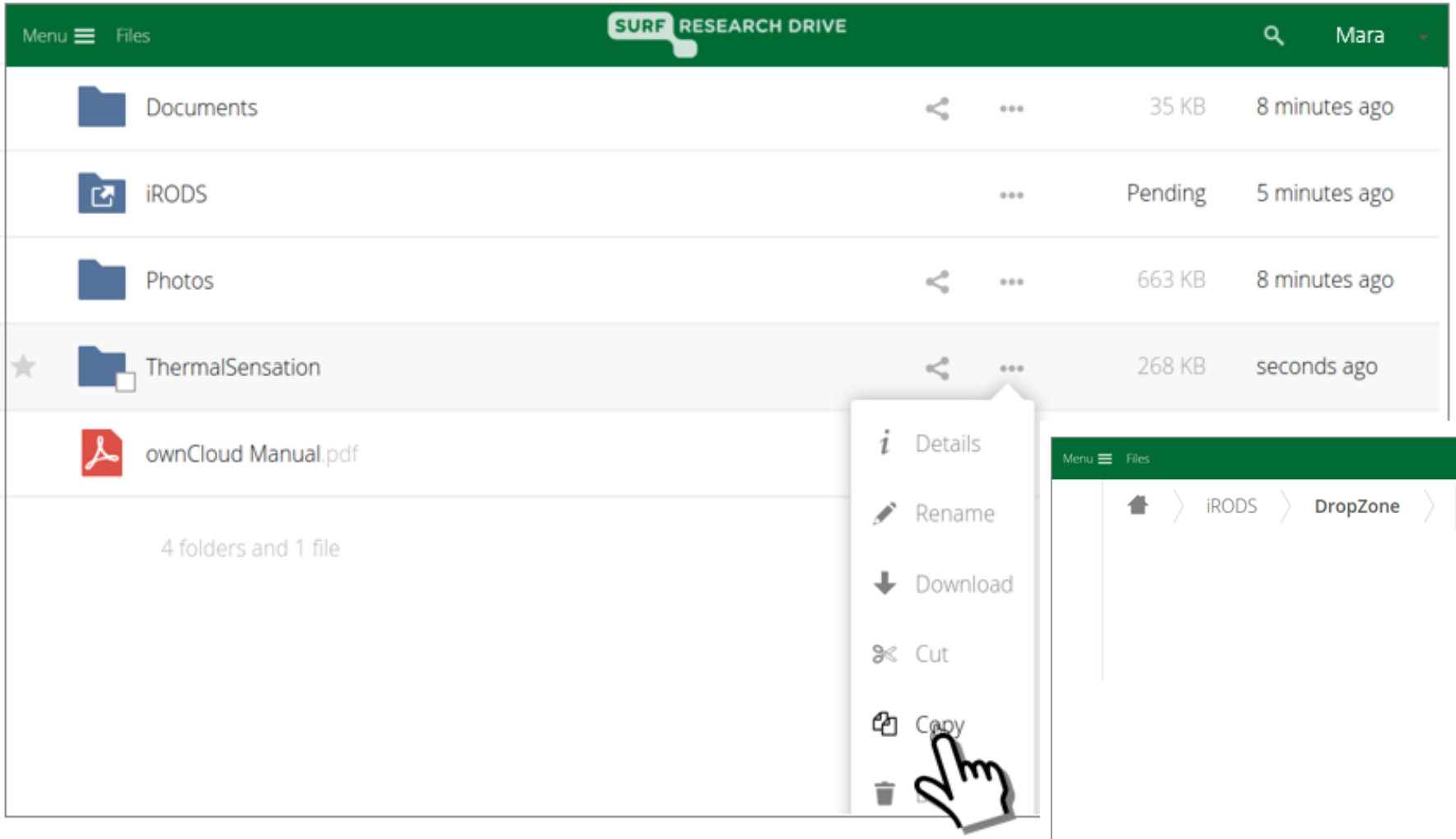
- Data Management Software
  - Metadata
  - Storage virtualization of different types of storages
  - Logical namespace across storage locations
  - Rule engine to automate data management according to defined policies
  - Federation of iRODS instances

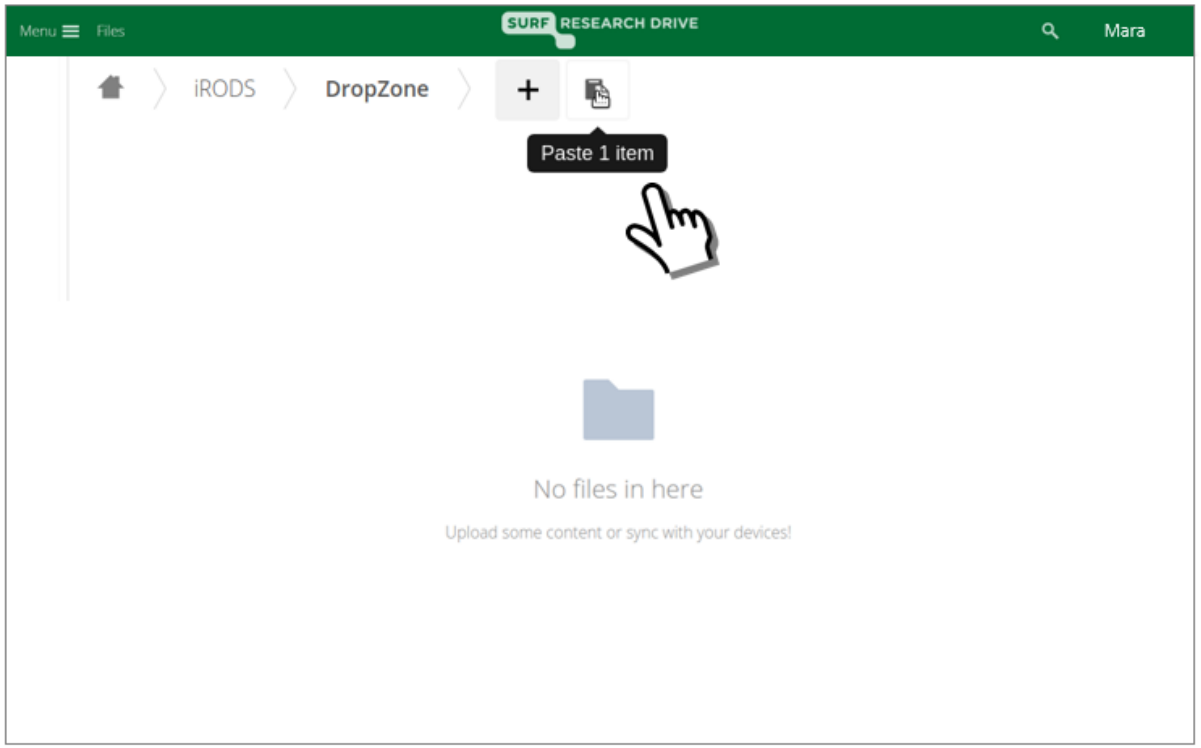


# An RDM framework that addresses the needs of all actors

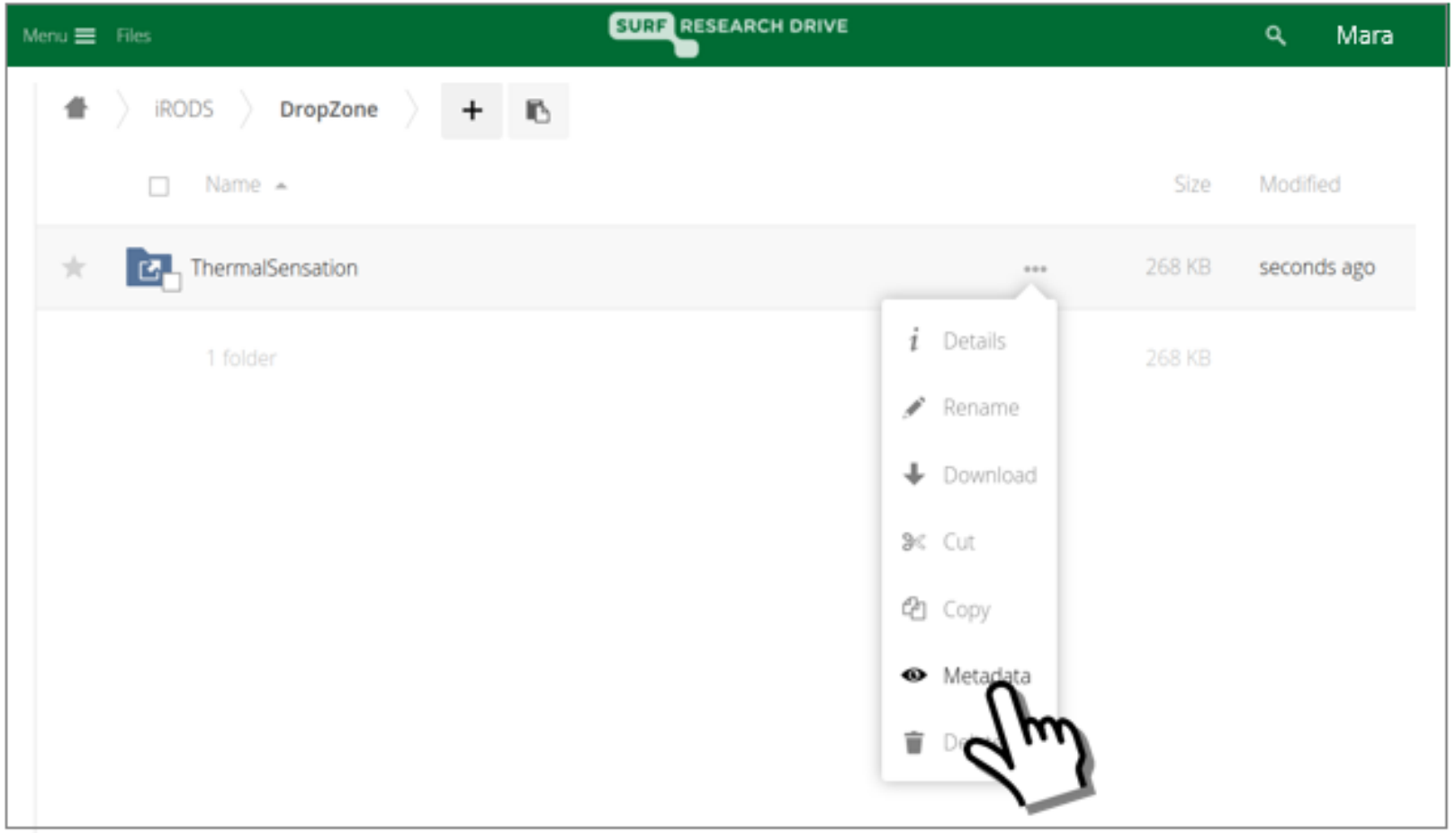


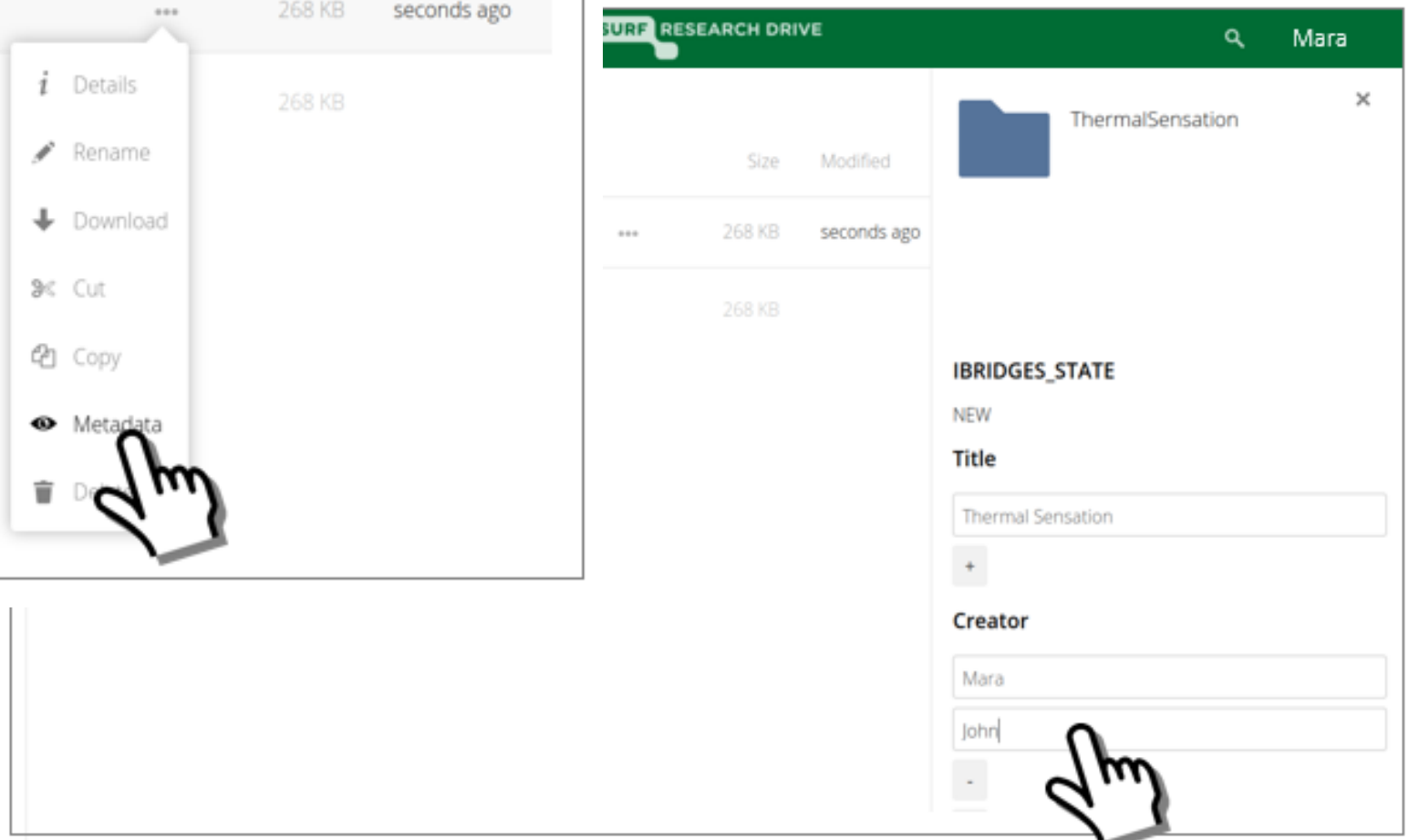
# Researcher copies folder into iRODS drop zone



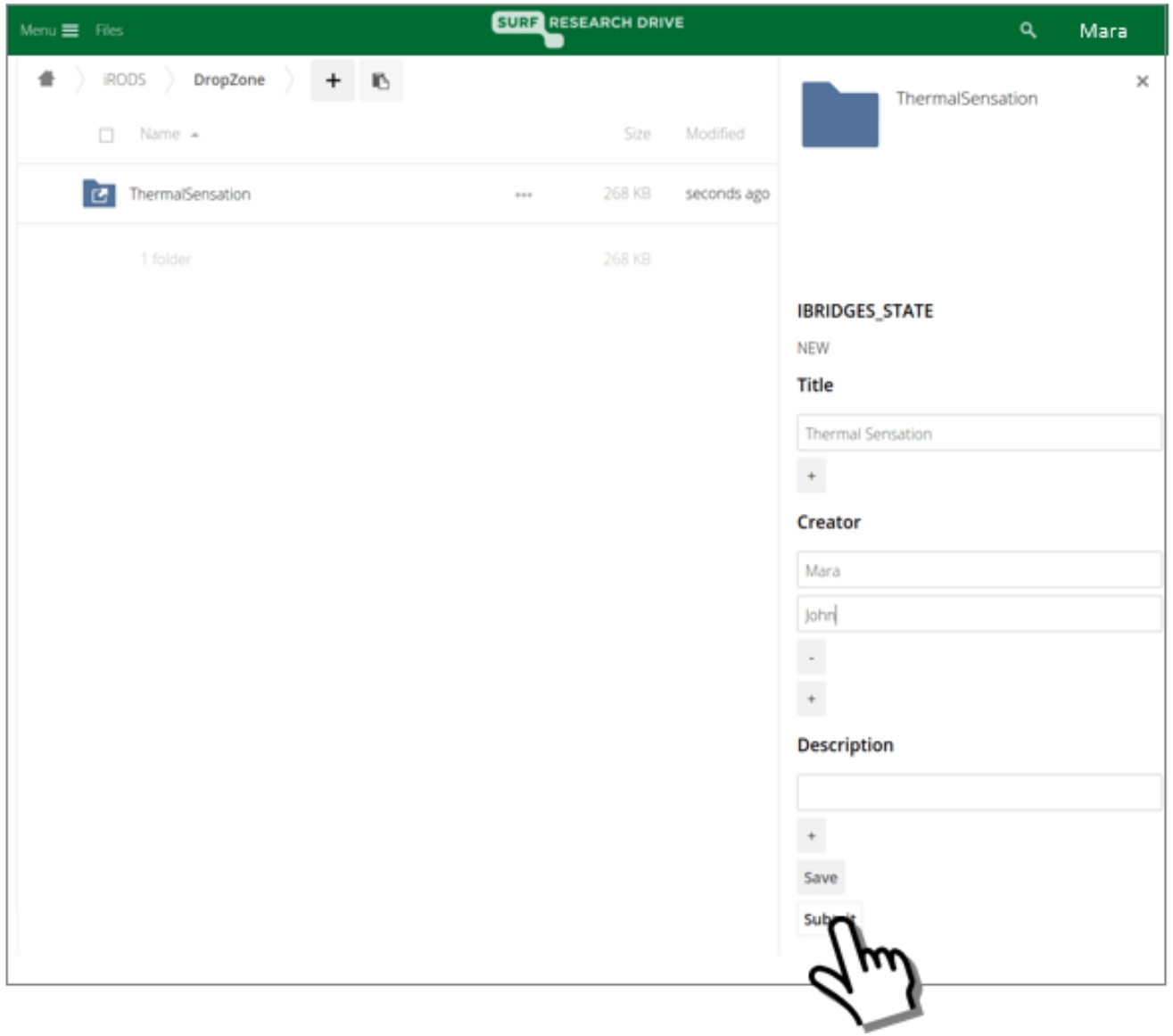



# Researcher selects folder for submission and proceeds to add metadata

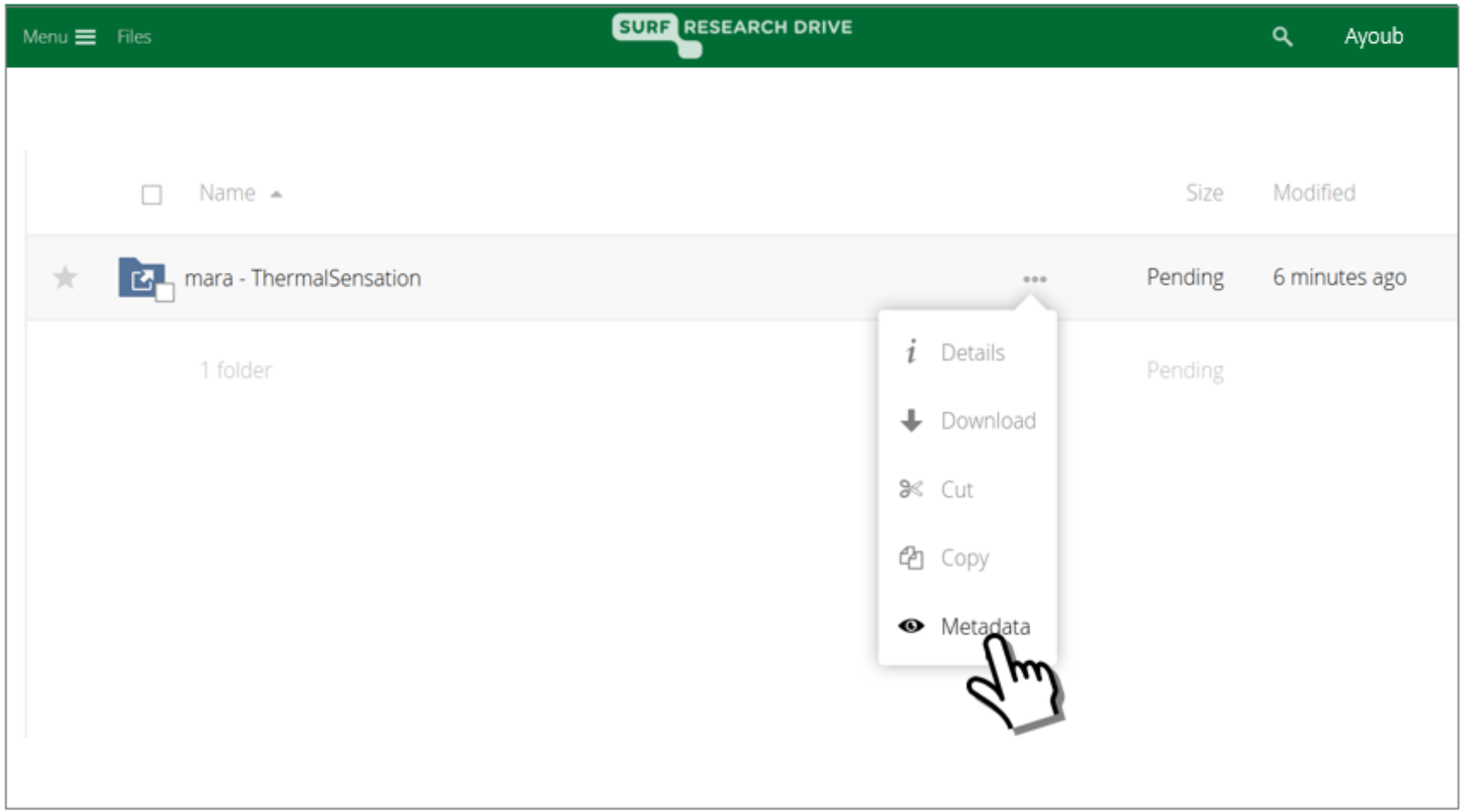


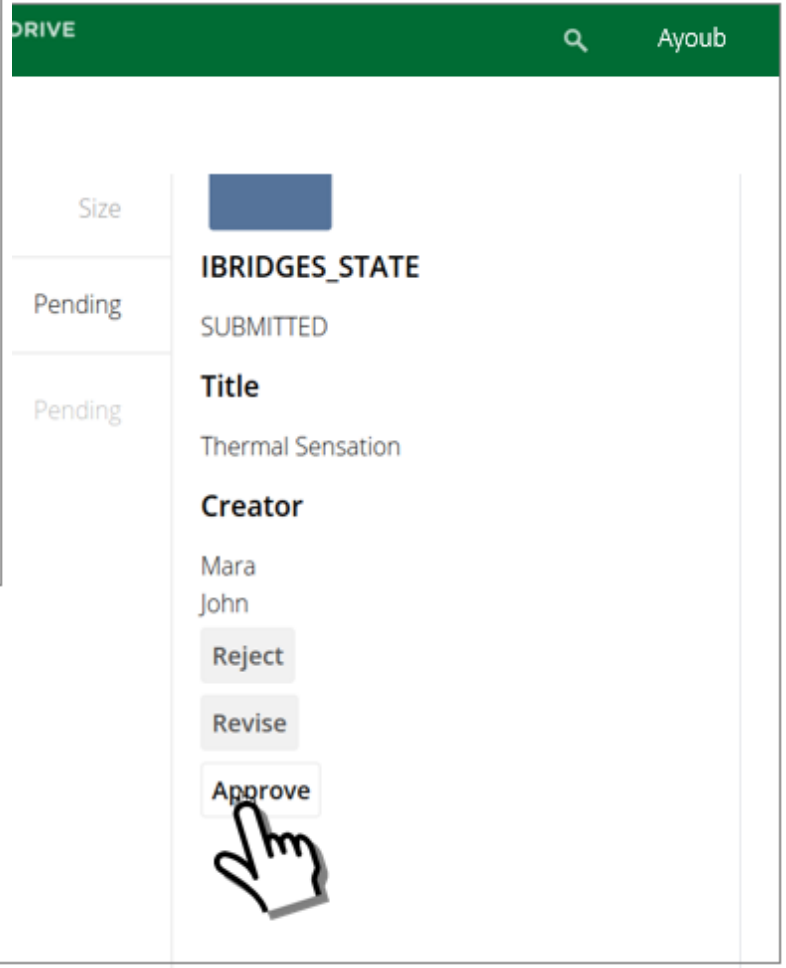



# Researcher submits collection to Archive

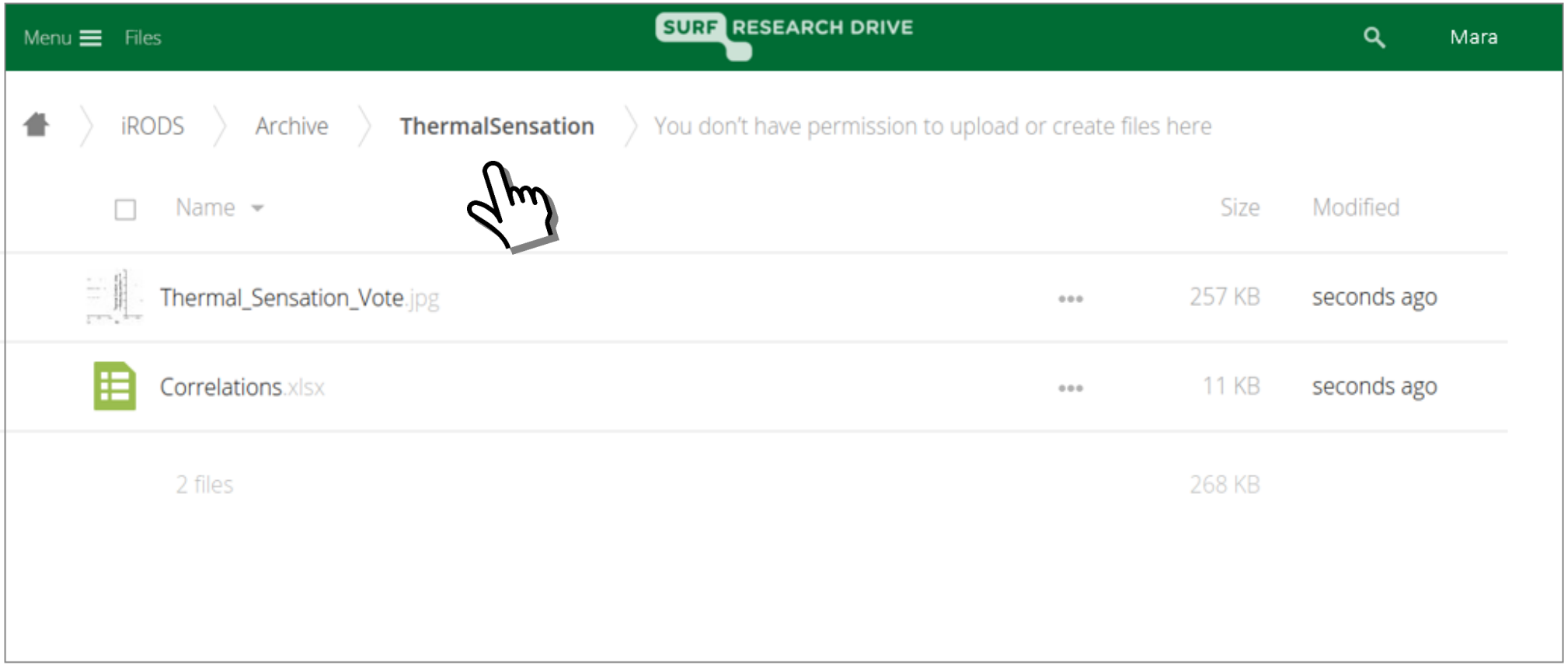


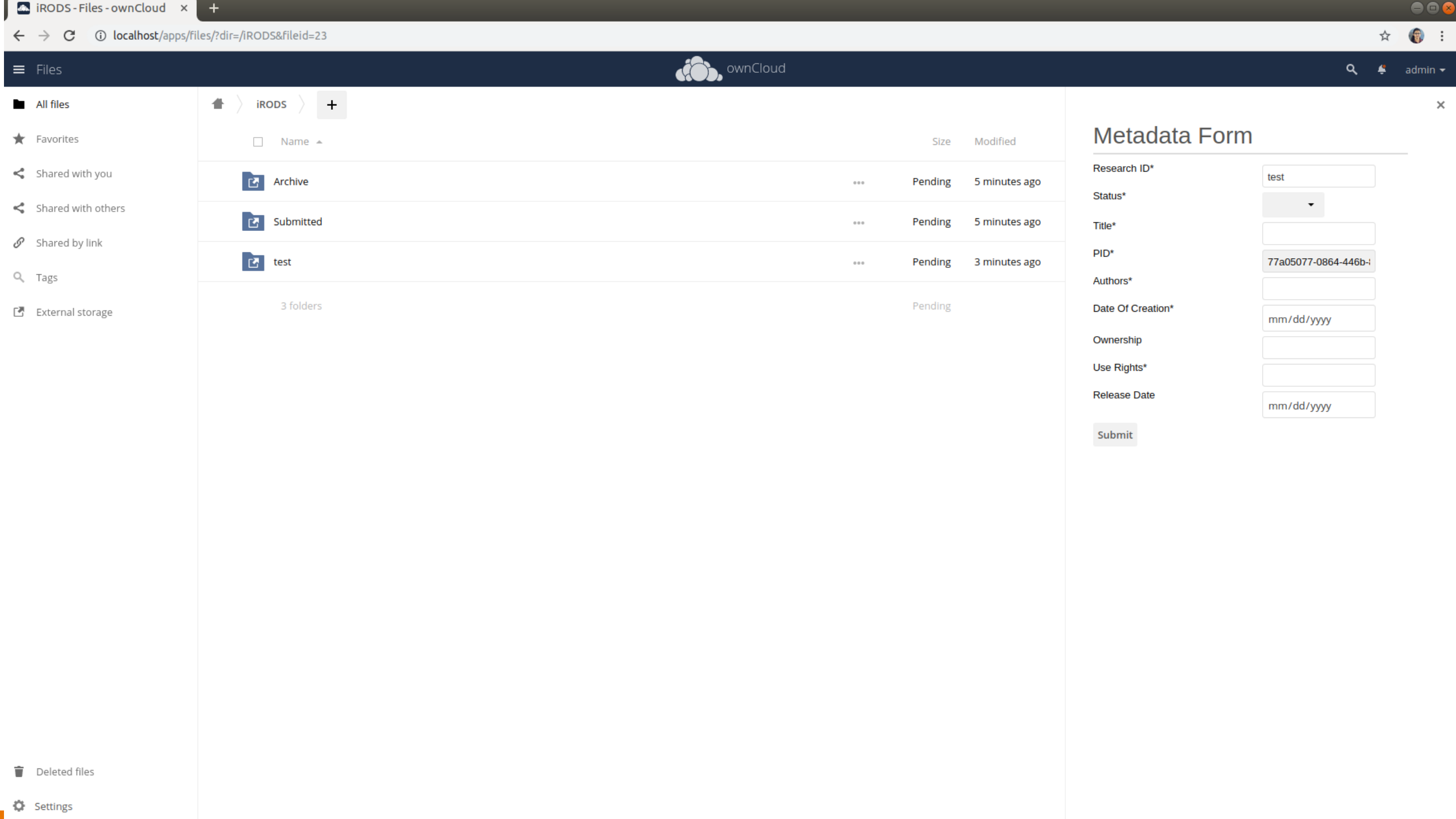
# Data steward selects submitted collection



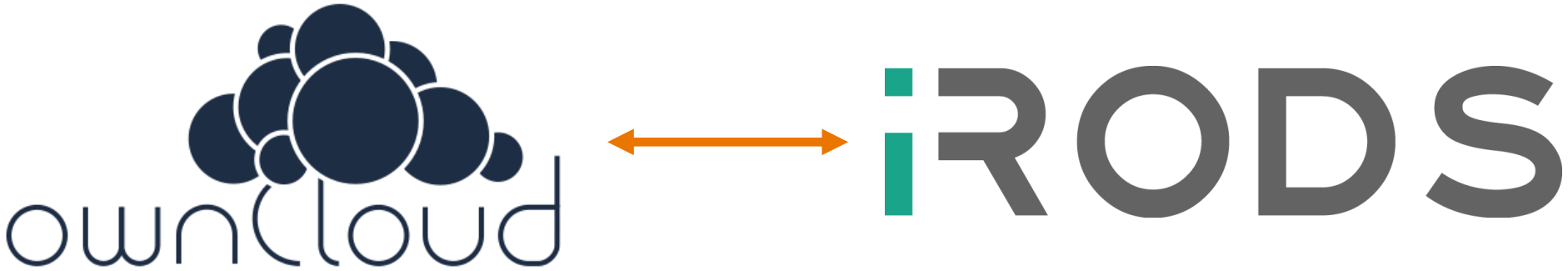
# Researcher checks that the data collection in now in the Archive




iRODS - Files - ownCloud
localhost/apps/files/?dir=/iRODS&fileid=23
Files
ownCloud
admin
All files
Favorites
Shared with you
Shared with others
Shared by link
Tags
External storage
Deleted files
Settings
iRODS
Name
Size
Modified
Archive
Pending
5 minutes ago
Submitted
Pending
5 minutes ago
test
Pending
3 minutes ago
3 folders
Pending
Metadata Form
Research ID*
test
Status*
Title*
PID*
77a05077-0864-446b-
Authors*
Date Of Creation*
mm/dd/yyyy
Ownership
Use Rights*
Release Date
mm/dd/yyyy
Submit

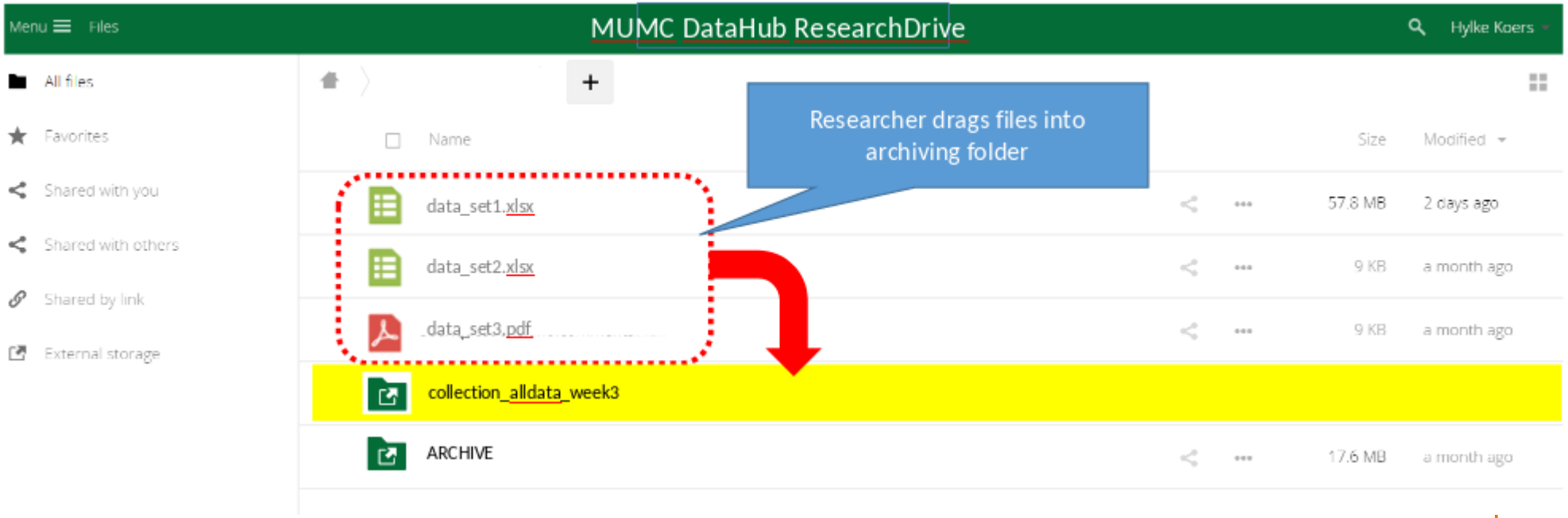
# iRODS and Owncloud are integrated

- Files and metadata inserted into iRODS through Owncloud is visible and accessible through iRODS CLI and API
- Files and metadata inserted into iRODS directly is visible and accessible in Owncloud
- Metadata is configurable per customer



# iRODs

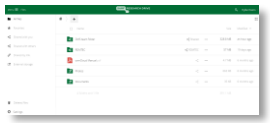
- Pilot with Saxion and Maastricht UMC
  - Integration with local archiving solutions



# SURF Research Drive & iRODS – combining the best of both

So, we set out to extend Research Drive by integration with iRODS:

- User Experience:
  - User can **add metadata** from within the Research Drive environment.
  - Use can '**archive**' or '**publish**' from the Research Drive environment.
- Behind the scenes, Research Drive is integrated with iRODS
  - iRODS maintains the **'source of truth' metadata** records
  - iRODS serves as point of integration to ensure **consistent user experiences** between Research Drive users , iRODS command-line users , and institutional data steward
- 'Archival' and 'Publication' workflows codified in Apache Airflow, working in unison with iRODS rule engine



& iRODS = HAPPY ( )

# Acknowledgements


- Stefan
- Tom
- Wybrand
- Jean-Marie
- Thirsa
- Tasneem

Happily
ever after
Starts here