

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

RON TROMPERT



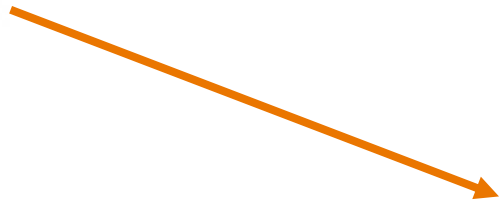
SURF

Research Drive Recap

- Once upon a time there was.....



FLEXIBLE RESEARCH DATA SHARING



PERSONAL SYNC-AND-SHARE



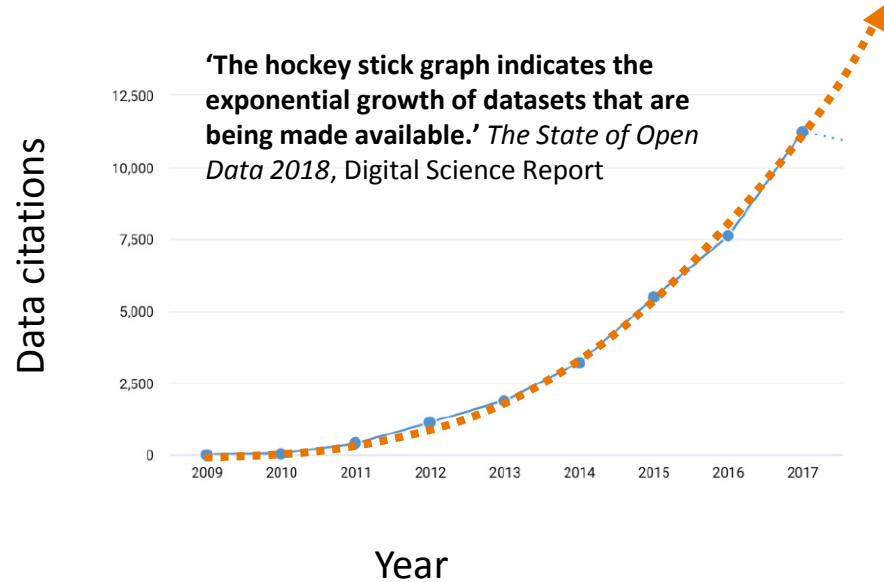
Research Drive Recap

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



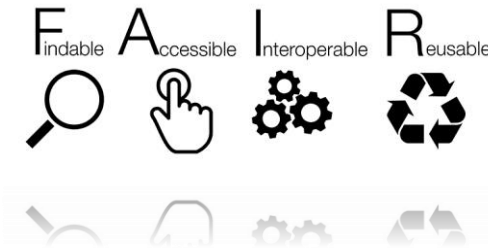
Research Drive Recap

Lots of data



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

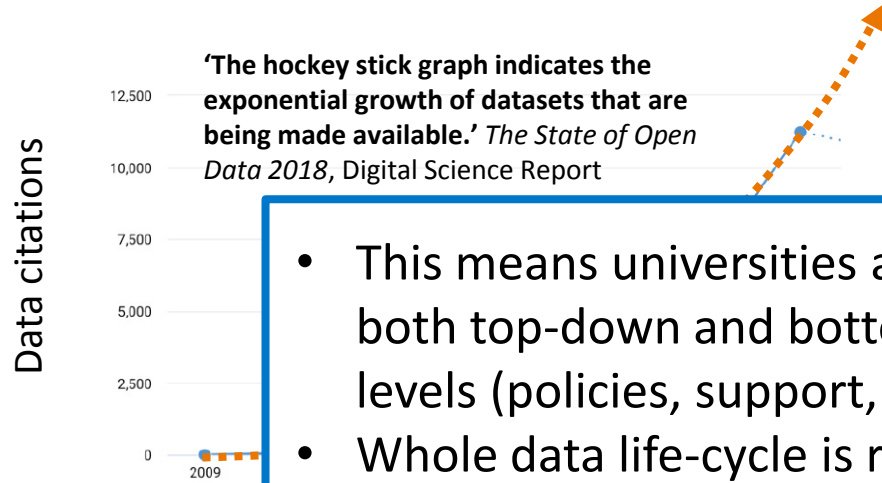


The FAIR Principles and Open Science are on the agenda of university boards, funders and the government.

Research Drive Recap

Lots of data

Lots of attention, lots of ambition



F indable A ccessible I nteroperable R eusable

- 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.

Open Science
data management
open access
best practice
communication
privacy protection
copyright
archival
experiment

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

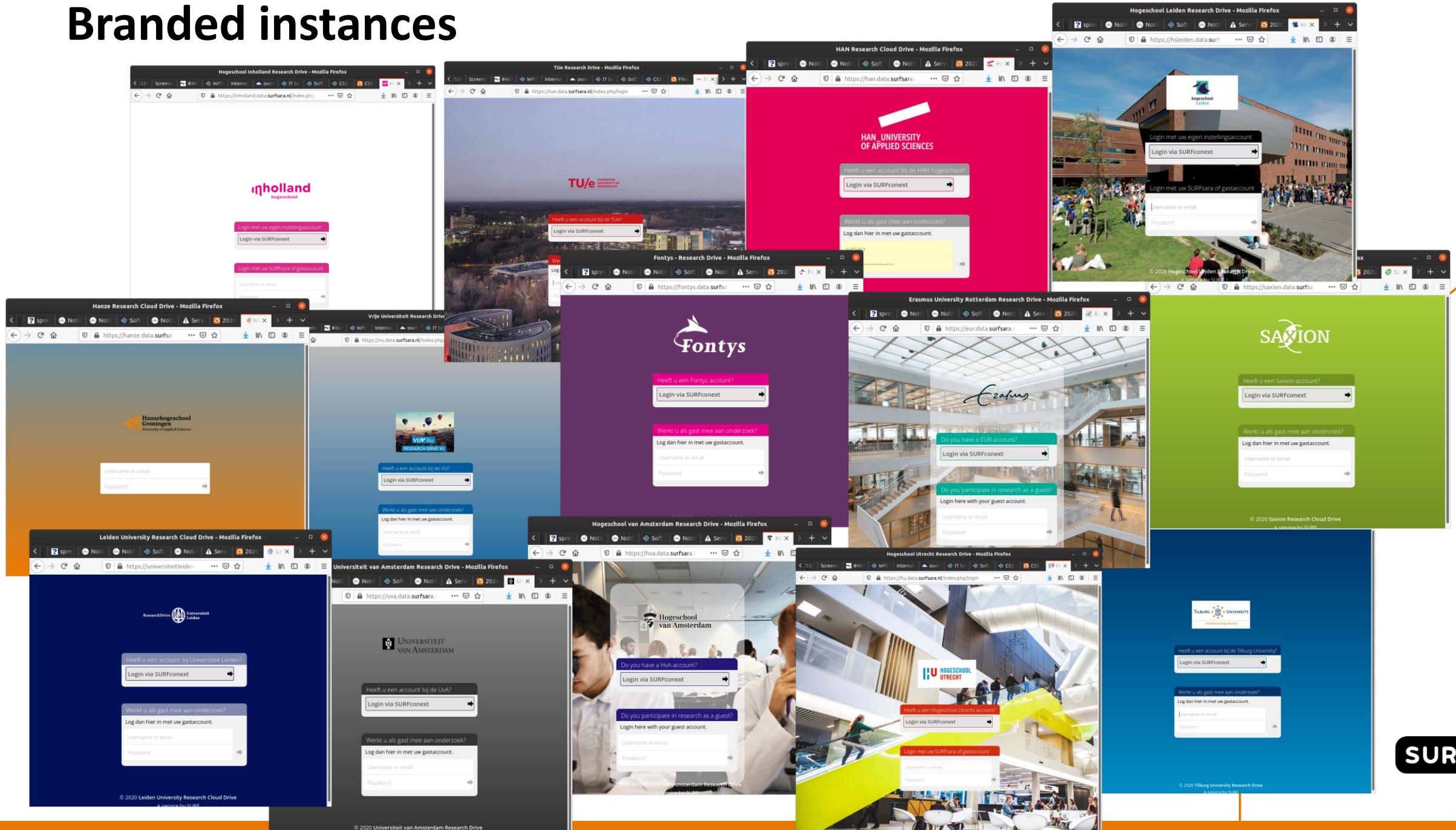
- Sync-and-Share Service with focus on the needs of the research community
 - Owncloud-based
 - Features
 - Your own branding
 - Sharing outside of the organisation.
 - Multiple AUTHN methods
 - Integration with external storage systems
 - Functional accounts supporting data stewardship
 - Quota as large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

The image shows a screenshot of a web browser window displaying the login page for Miskatonic University. The browser's address bar shows the URL `https://miskatonic.data.surfsara.nl/index.php/login`. The page features the university's logo and two login options: "Heeft u een Miskatonic-account?" with a "Login via SURFconext" button, and "Werkt u als gast mee aan onderzoek?" with a "Log dan hier in met uw gastaccount." form. Below the login forms, a Jupyter Notebook interface is visible, titled "jupyter guust (unsaved changes)". The notebook's code cell contains the following Python code:

```
In [4]: import matplotlib.pyplot as plt
import matplotlib.image as mpimg
img=mpimg.imread('guust.jpg')
imgplot = plt.imshow(img)
plt.show()
```

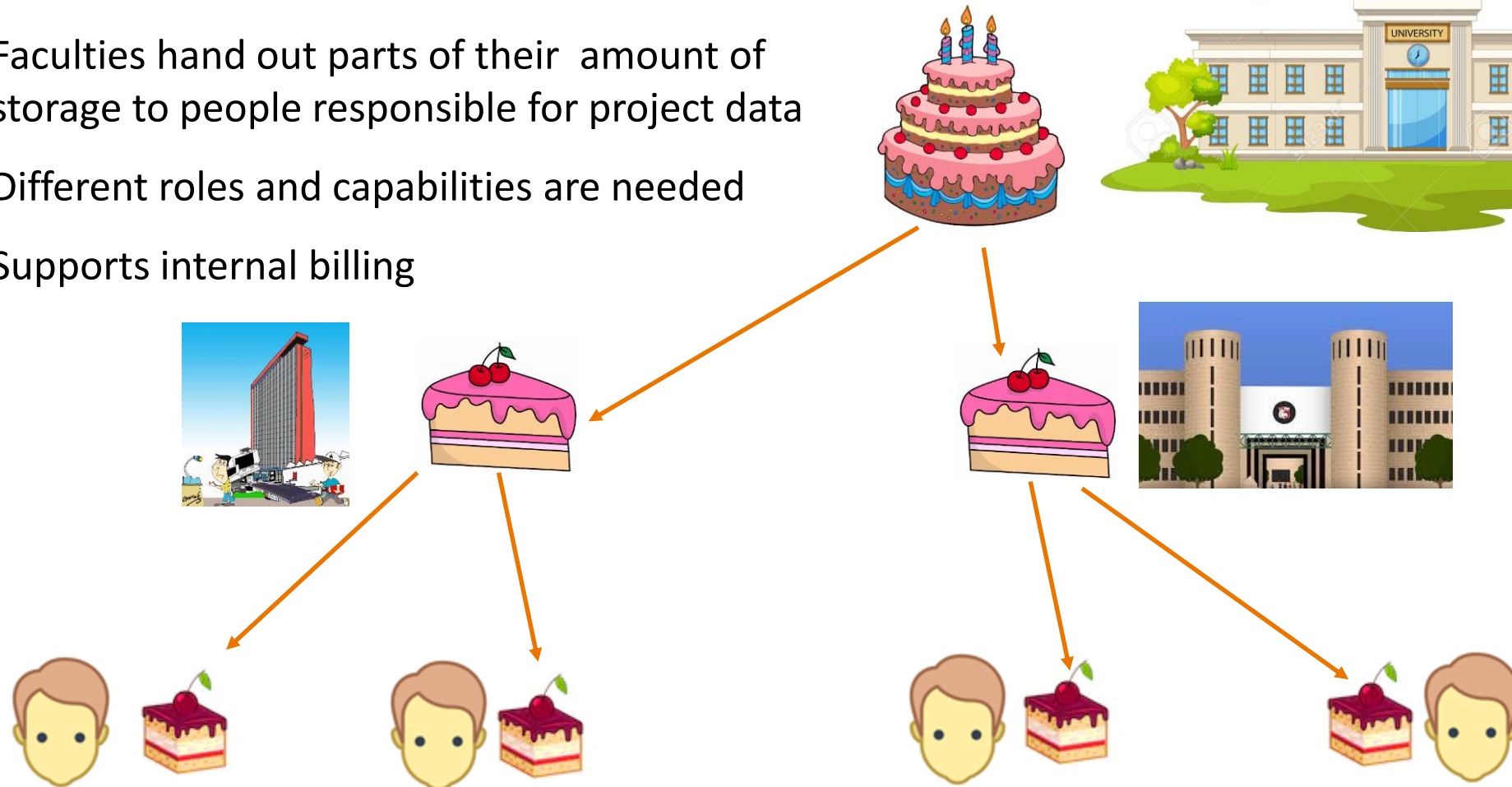
The output of the code is a plot showing a cartoon character (Guust) in a hammock, with a y-axis ranging from 0 to 700 and an x-axis ranging from 0 to 1000.

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

Menu  Dashboard



belfort@harvard-example.edu



Dashboard

User belfort@harvard-example.edu
Name belfort@harvard-example.edu
Email

Projectfolders

Gebruikersaccounts

Contracten

Gebruiker Uitnodigen

Zoeken

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

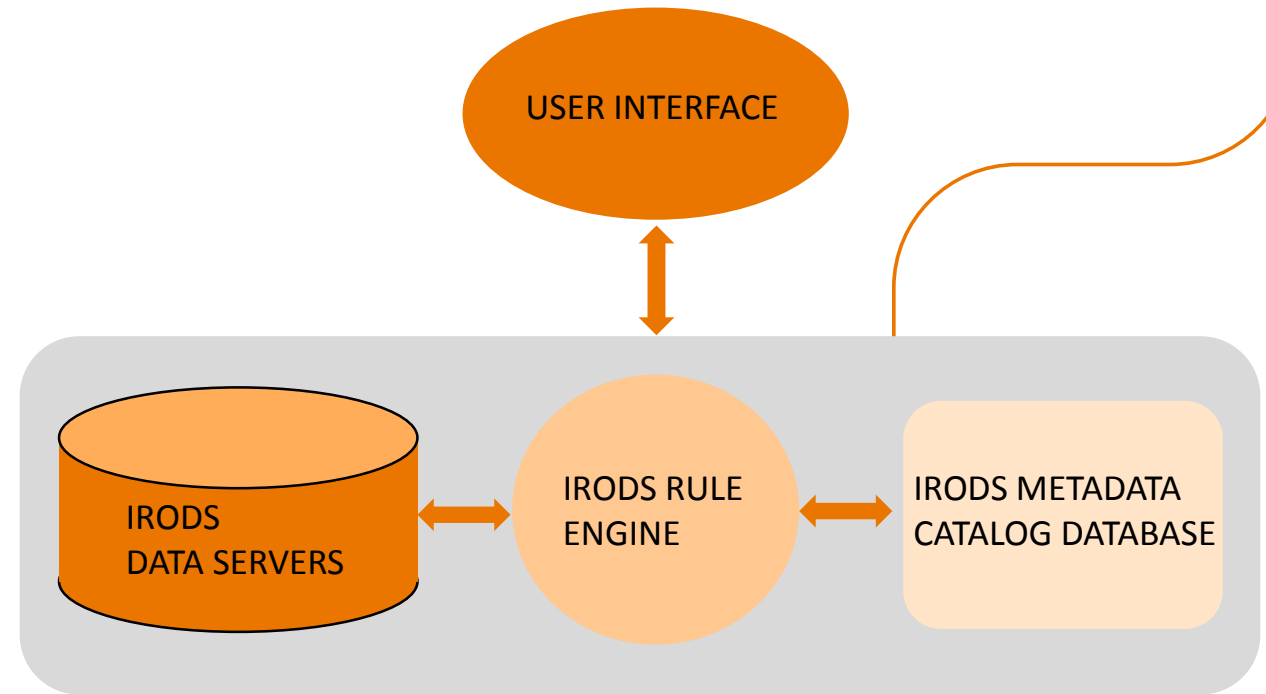
Moving up higher in the stack

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

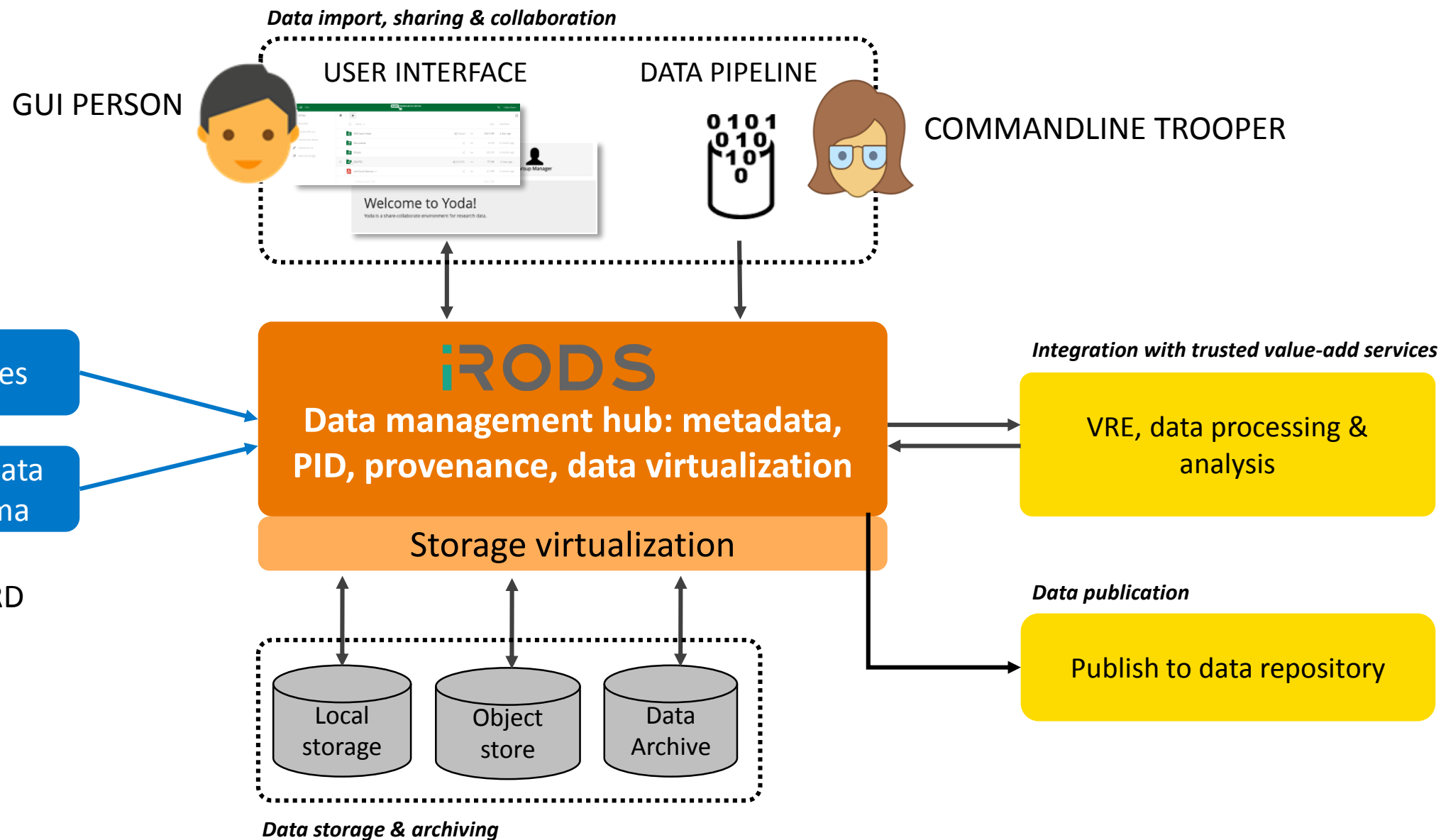


iRODS

- 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

The screenshot shows the SURF RESEARCH DRIVE interface. At the top, there is a green header with 'Menu', 'Files', 'SURF RESEARCH DRIVE', a search icon, and the name 'Mara'. Below the header is a list of items:

Item	Size	Time
Documents	35 KB	8 minutes ago
iRODS	Pending	5 minutes ago
Photos	663 KB	8 minutes ago
ThermalSensation	268 KB	seconds ago

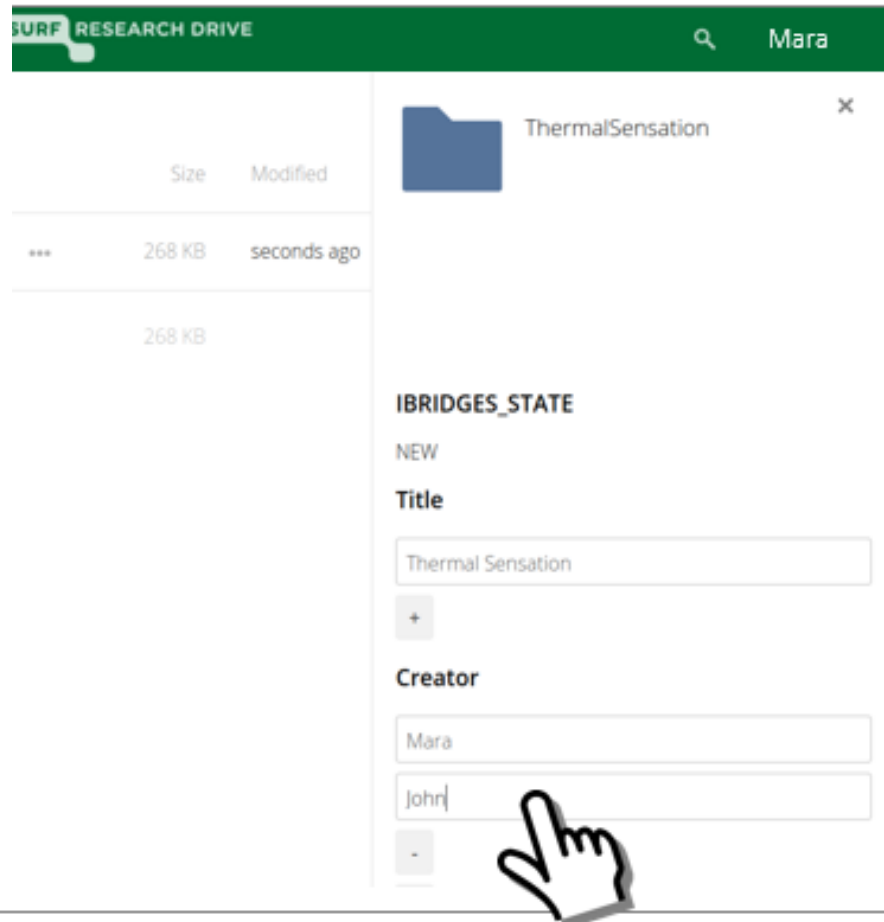
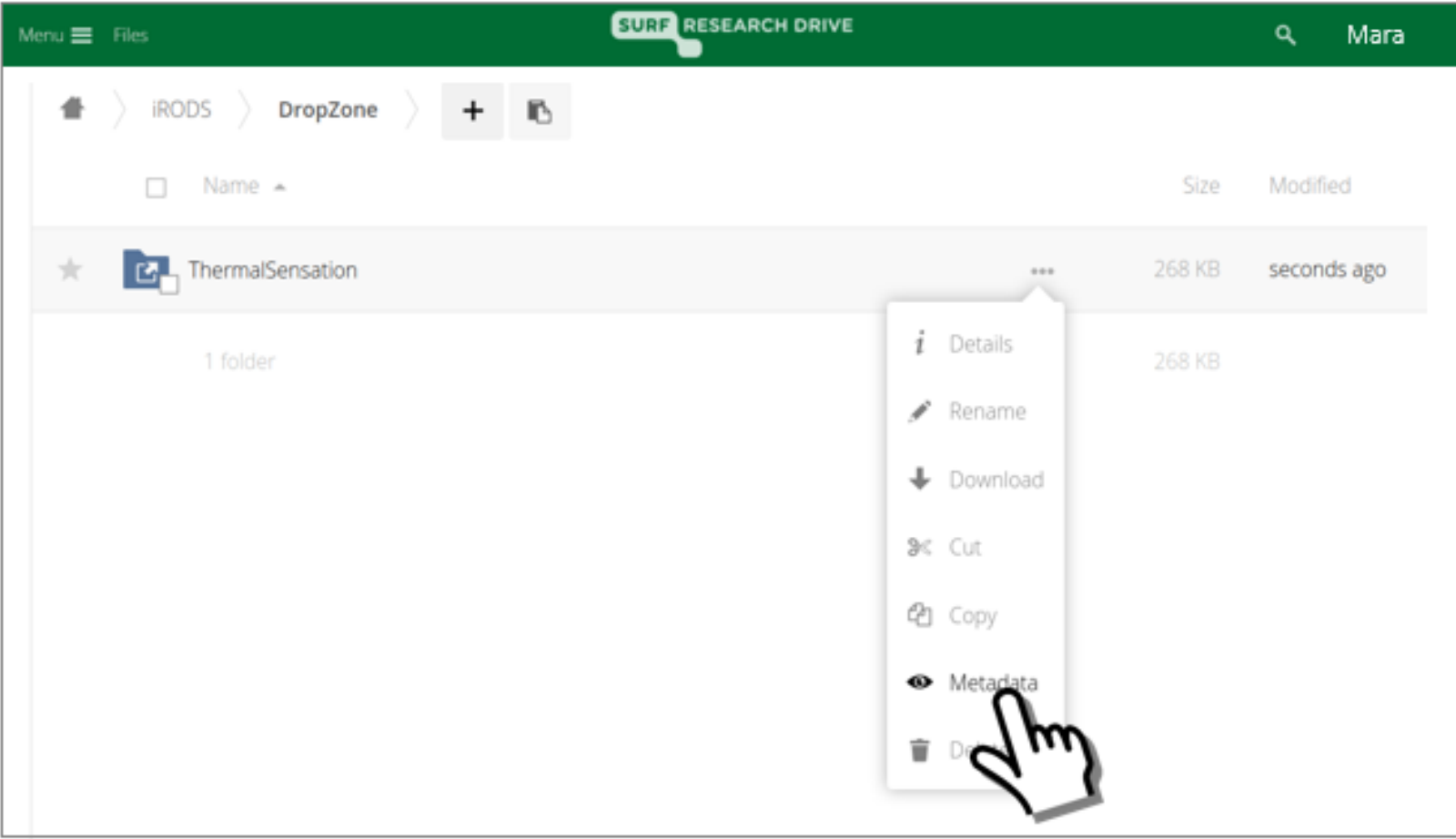
ownCloud Manual.pdf

4 folders and 1 file

- i Details
- ✎ Rename
- ↓ Download
- ✂ Cut
- 📄 Copy
- 🗑 Delete

The screenshot shows the SURF RESEARCH DRIVE interface for the 'iRODS DropZone'. The breadcrumb path is 'iRODS > DropZone'. There is a '+', a document icon, and a 'Paste 1 item' tooltip. A hand cursor is pointing at the document icon. Below the icons is a folder icon and the text 'No files in here' and 'Upload some content or sync with your devices!'.

Researcher selects folder for submission and proceeds to add metadata



Researcher submits collection to Archive

The screenshot shows the SURF RESEARCH DRIVE interface. The top navigation bar is green with the SURF logo and 'RESEARCH DRIVE' text. Below the navigation bar, the breadcrumb path is 'iRODS > DropZone'. A table lists the contents of the 'DropZone' folder:

Name	Size	Modified
ThermalSensation	268 KB	seconds ago

Below the table, it indicates '1 folder' with a size of '268 KB'. On the right side, a detailed view of the 'ThermalSensation' folder is shown. It includes a 'Title' field with the value 'Thermal Sensation', a 'Creator' field with the value 'Mara', and a 'Description' field. At the bottom of this panel, there are 'Save' and 'Submit' buttons. A hand cursor is pointing at the 'Submit' button.

Data steward selects submitted collection

Menu Files SURF RESEARCH DRIVE Ayoub

Name	Size	Modified
★ mara - ThermalSensation	Pending	6 minutes ago
1 folder	Pending	



- Details
- Download
- Cut
- Copy
- Metadata

DRIVE Ayoub

Size	IBRIDGES_STATE
Pending	SUBMITTED
Pending	Title Thermal Sensation
	Creator Mara John
	Reject
	Revise
	Approve

Researcher checks that the data collection is now in the Archive

The screenshot shows the SURF Research Drive interface. At the top, there is a green header with the SURF logo and 'RESEARCH DRIVE' text. Below the header, a breadcrumb trail shows the path: Home > iRODS > Archive > ThermalSensation. A message below the breadcrumb states: 'You don't have permission to upload or create files here'. A table lists the files in the archive:

<input type="checkbox"/>	Name		Size	Modified
<input type="checkbox"/>	 Thermal_Sensation_Vote.jpg	...	257 KB	seconds ago
<input type="checkbox"/>	 Correlations.xlsx	...	11 KB	seconds ago
2 files			268 KB	

A hand cursor icon is positioned over the 'Name' column header.

- All files
- Favorites
- Shared with you
- Shared with others
- Shared by link
- Tags
- External storage
- Deleted files
- Settings

Home > 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*

Status*

Title*

PID*

Authors*

Date Of Creation*

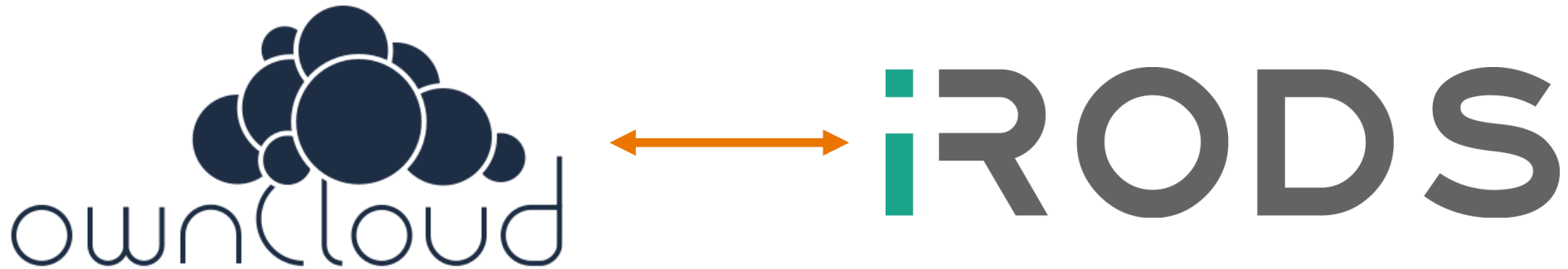
Ownership

Use Rights*

Release Date

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

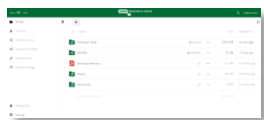
The screenshot displays the MUMC DataHub ResearchDrive interface. The top navigation bar is green and contains the text "MUMC DataHub ResearchDrive" and a search icon with the name "Hylke Koers". On the left, there is a sidebar with navigation options: "All files", "Favorites", "Shared with you", "Shared with others", "Shared by link", and "External storage". The main content area shows a list of files and folders. A red dashed box highlights three files: "data_set1.xlsx", "data_set2.xlsx", and ".data_set3.pdf". A blue callout box with the text "Researcher drags files into archiving folder" has a red arrow pointing from the highlighted files to a yellow folder named "collection_alldata_week3". Below this folder is another folder named "ARCHIVE".

Name	Size	Modified
data_set1.xlsx	57.8 MB	2 days ago
data_set2.xlsx	9 KB	a month ago
.data_set3.pdf	9 KB	a month ago
collection_alldata_week3		
ARCHIVE	17.6 MB	a month ago

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.
 - User 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 ♡