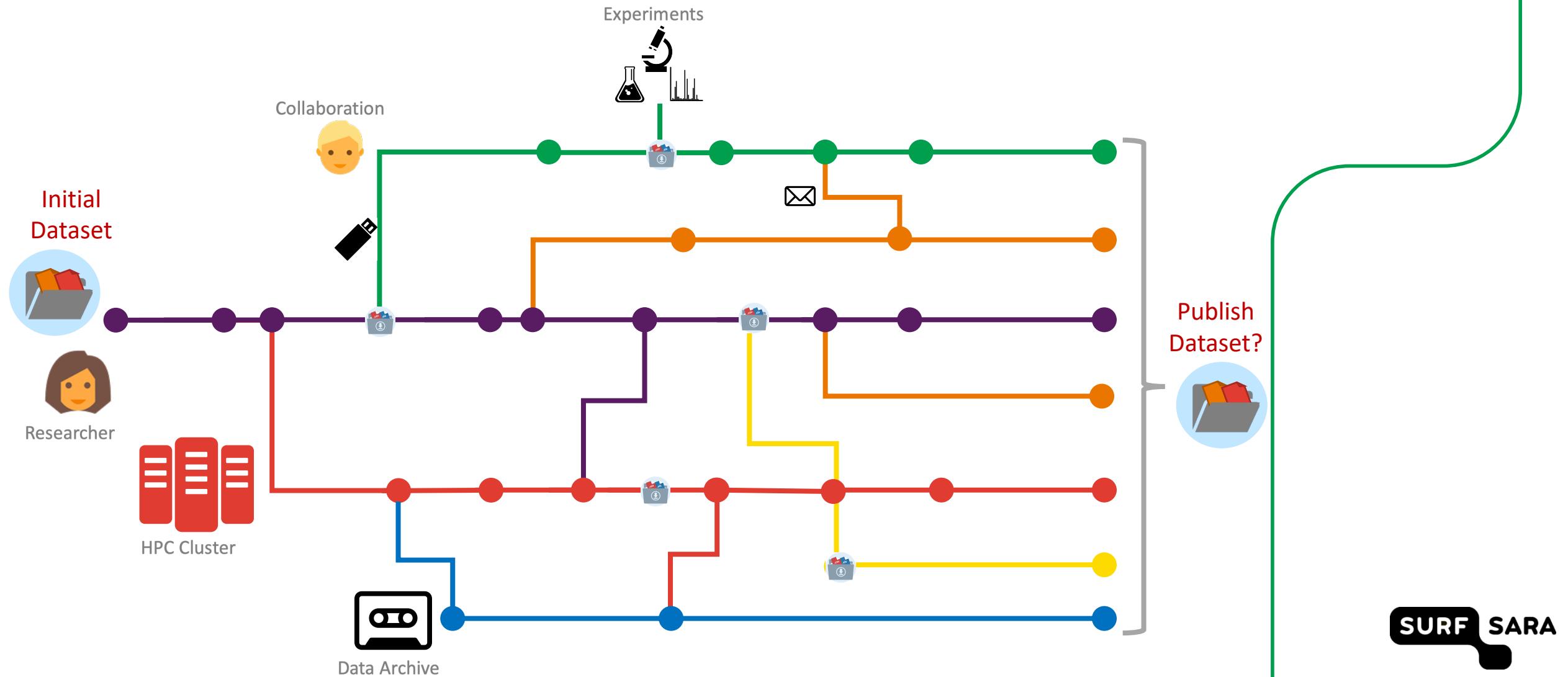
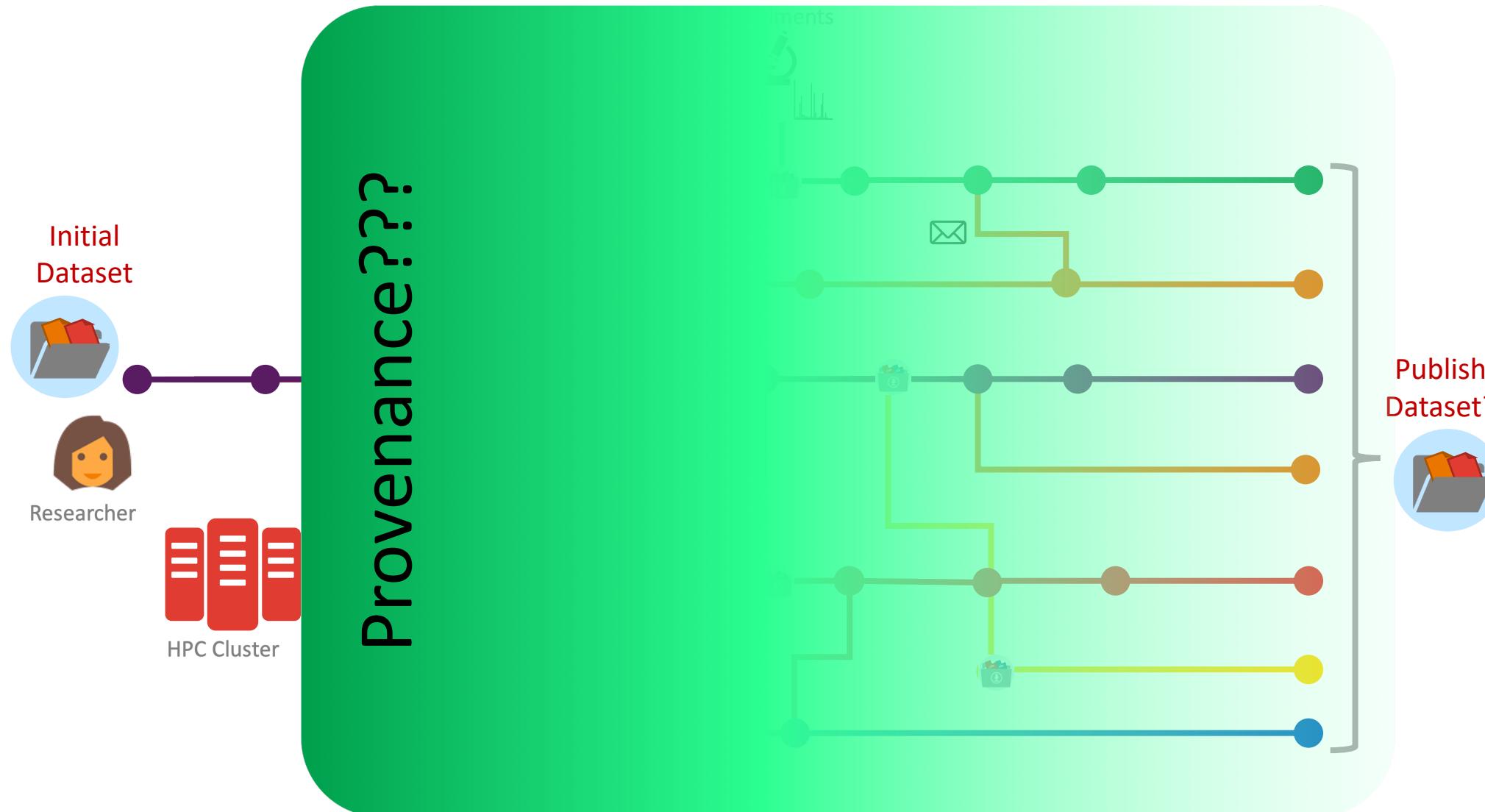


# **Introduction RDM in HPC – iRODS and data processing**

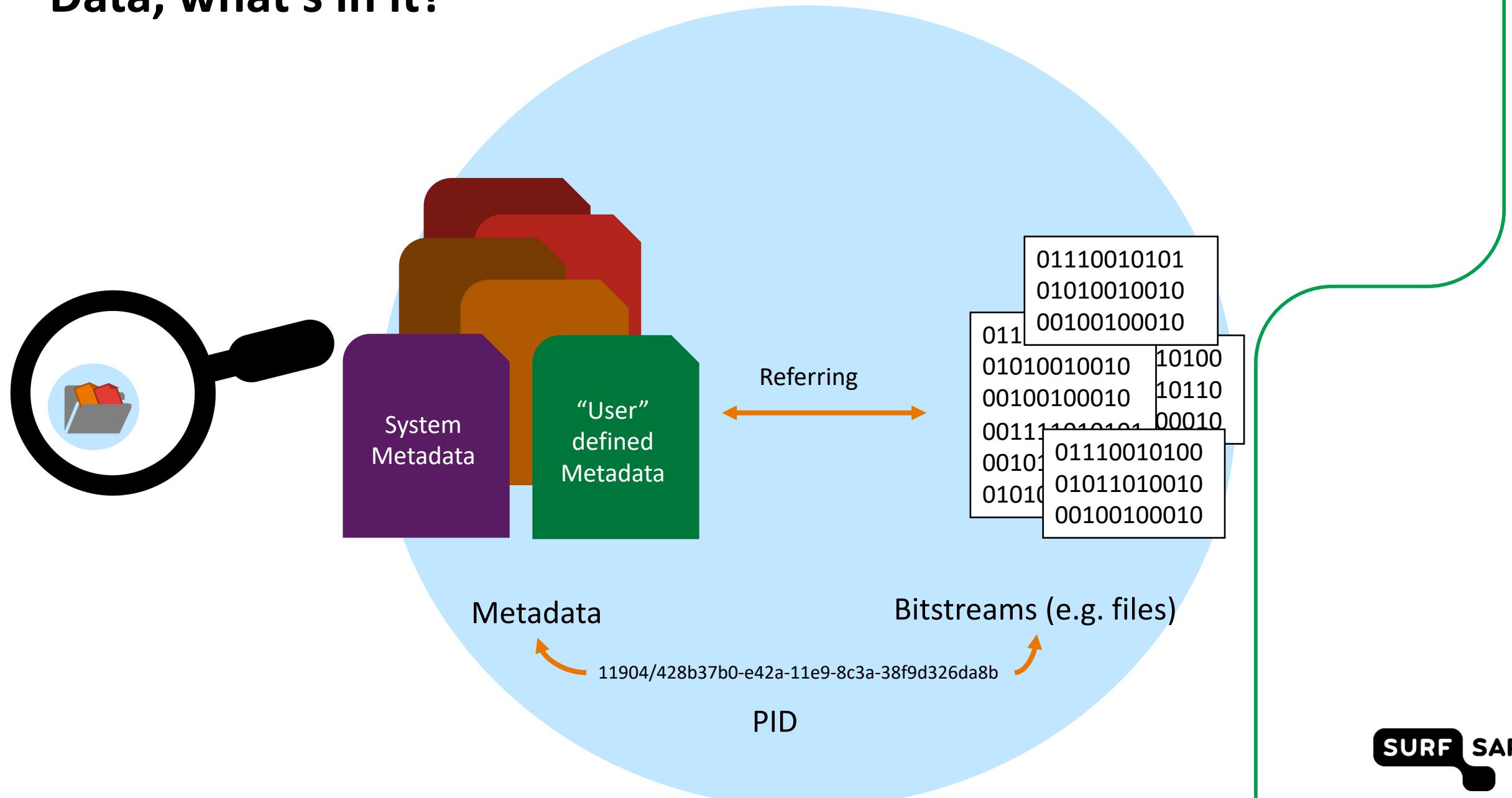
# Data, what's the problem?



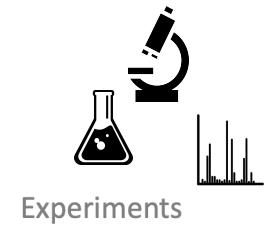
# Data, what's the problem?



# Data, what's in it?



# Data Life Cycle



Access rights



Provenance

RE-USING

SHARING

PRESERVING

CREATING

ANALYSING

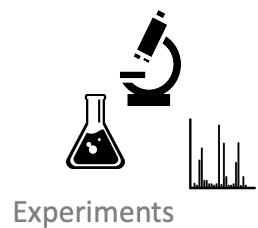
Descriptive  
sensitive

Operational  
Metadata  
System  
Metadata



Data Archive

# Data Life Cycle

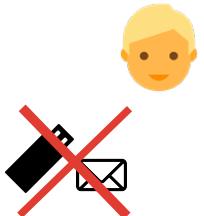


Access rights

Provenance

RE-USING

SHARING



PRESERVING



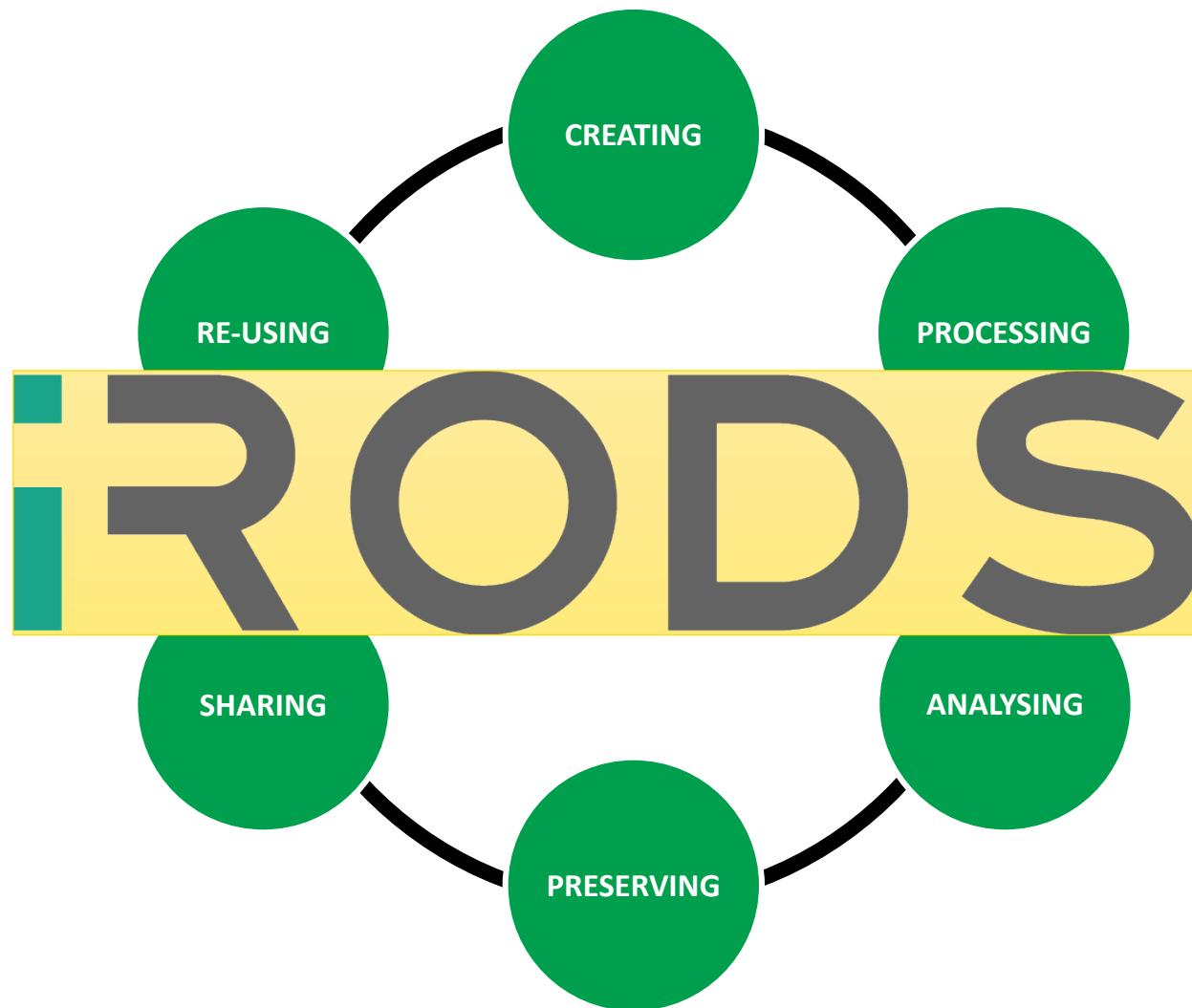
Data Archive

Descriptive  
Sensitive

Operational Metadata  
System Metadata

Findable   
Accessible   
Interoperable   
Reusable

# Data Life Cycle

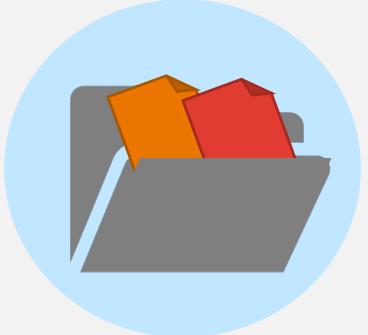


# iRODS: integrated Rule-Oriented Data System

- Open source data and storage management system
- Governed by iRODS consortium:  
SURF and several universities are members
- Scalable storage system for multipetabyte scale
- Metadata is a first class citizen

# What is iRODS? - Core competencies

Unified storage



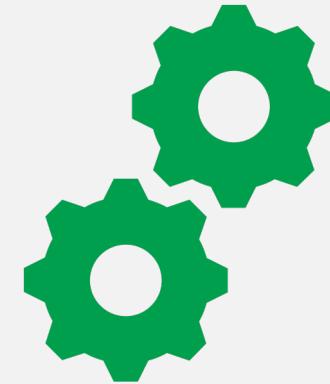
Data Virtualization  
of disk and tape

Data Discovery



Metadata of data  
objects / collections

Automation



Rule Engine to  
enforce policies

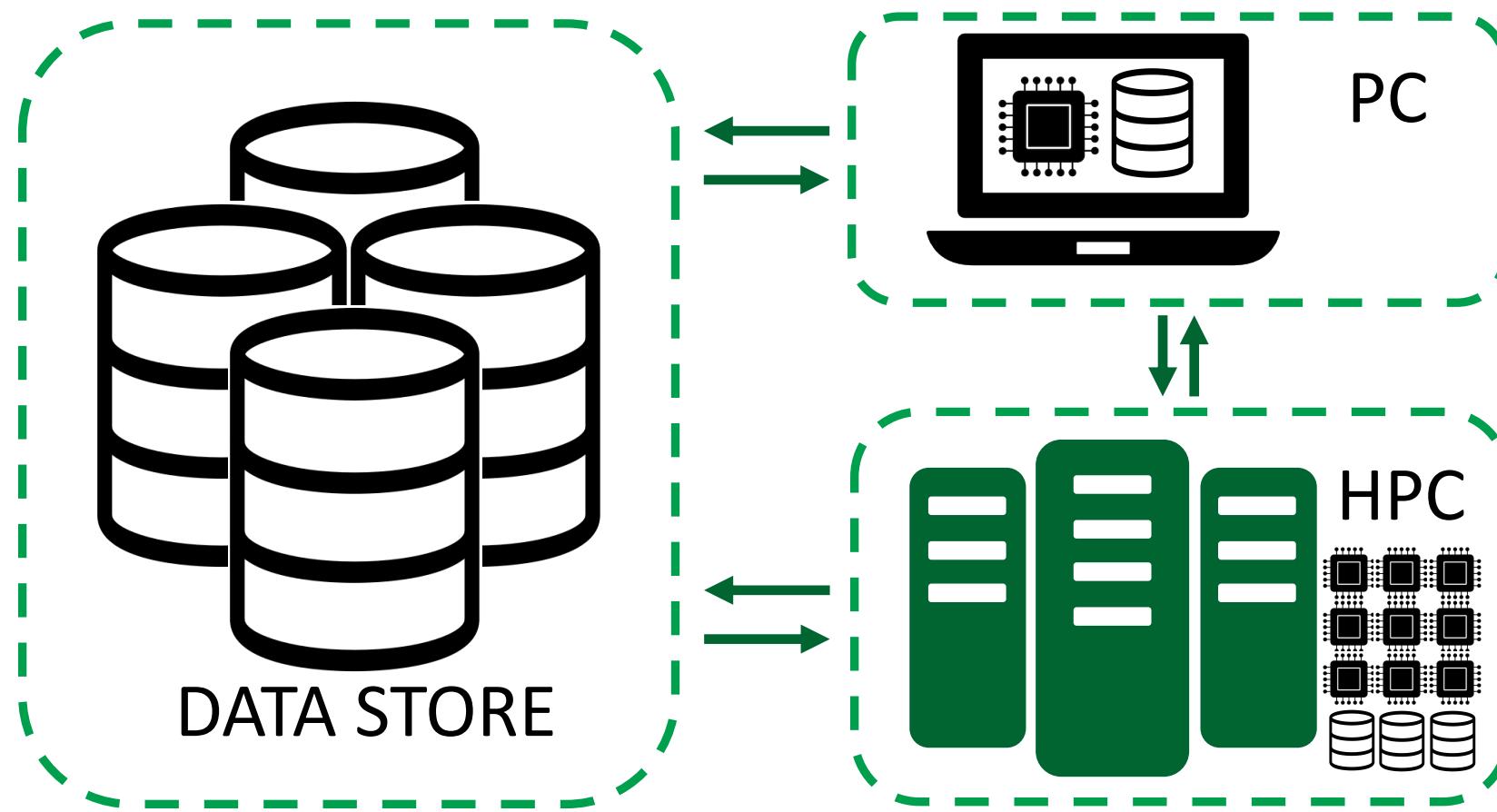
Secure Collaboration



Federation between  
different instances

# Data storage from PC to HPC

- Typically on your PC, the compute and storage system are on the same machine
- Research using HPC systems; compute and storage need to be separated.  
iRODS can be used as central data store.



# Why iRODS for RDM at HPC?

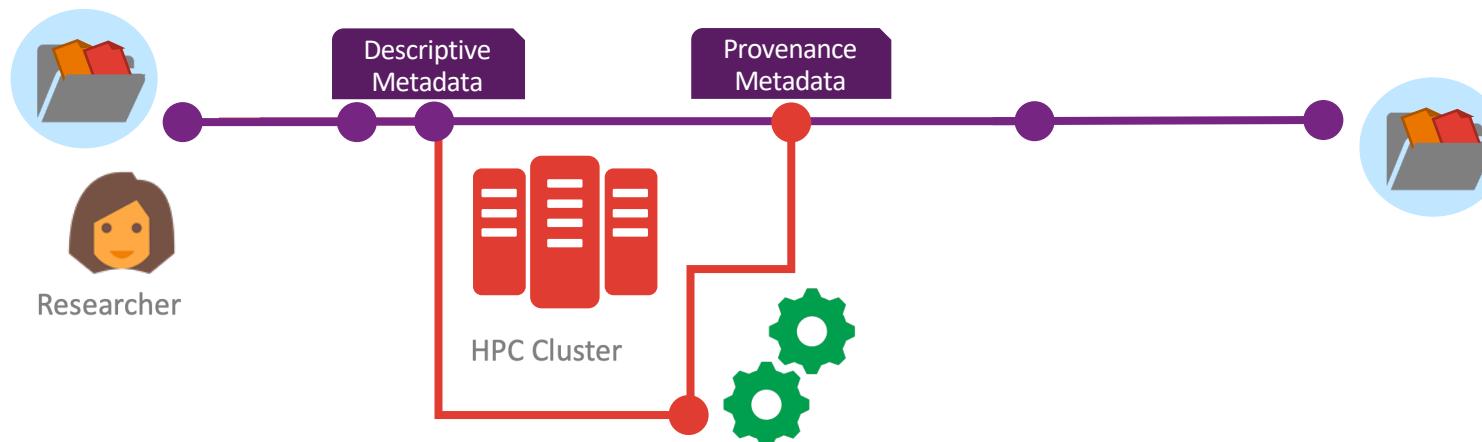
- High performant data transfer protocol
- Data accessible programmatically through icommands, C API, Python API
- Scalable for very large datasets
- Integrates data with metadata
- Open source

# iRODS at SURF

- RDM storage scale-out
- iRODS hosting service
- (Future service) iCE: iRODS Community Edition

# Hands on for data processing with iRODS

- Learn how to handle data via iRODS icommands and use in HPC jobscripts
- Learn about iRODS concepts: data objects, collections, metadata handling querying
- Learn how to find data based on metadata (not on some knowledge of hand made directory trees) and use it in an HPC system



# Some questions for the audience

- Who got a sdemo (or your own) account for Lisa?
- Who changed their password and was able to login to Lisa via a ssh client?
- Who knows the basics of Bash?
- Who knows iRODS already before this course?
- Who is doing data intensive compute jobs already on Lisa/Cartesius/somewhere else?
- Who is ready for the course!?

**<https://edu.nl/ba9e4>**