



Vlaanderen  
is supercomputing

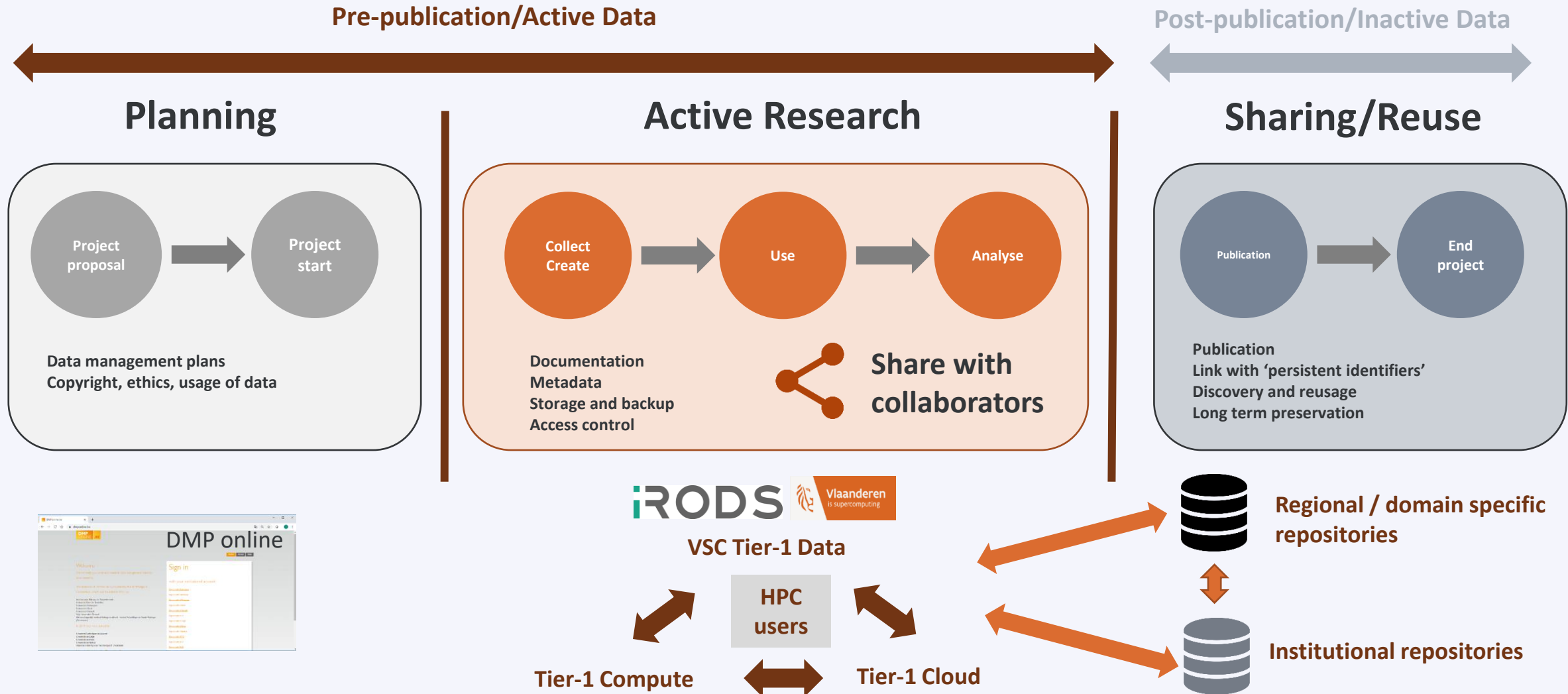
# iRODS User Training Introduction

# What is iRODS?

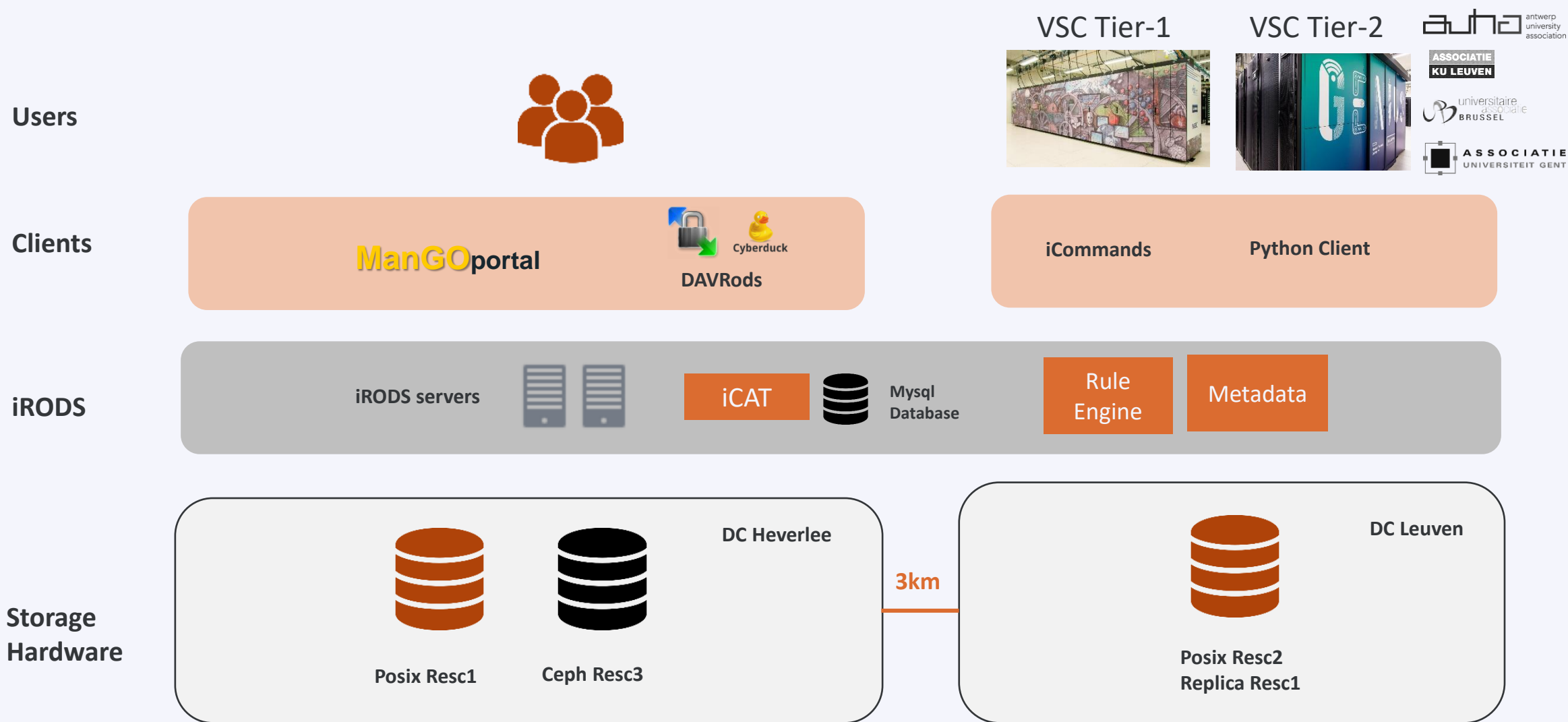
- iRODS (integrated Rule-Oriented Data System)
- Open Source distributed data and storage management system
- Configurable data management policies and workflows
- Scalable
- KU Leuven is part of the iRODS consortium



# Tier-1 Data in the research Data Lifecycle



# Tier-1 Data architecture



# iRODS Core competencies



## Unified Storage Namespace

Data virtualization of distributed storage systems



## Automation

Rule Engine to enforce data policies



## Data Discovery

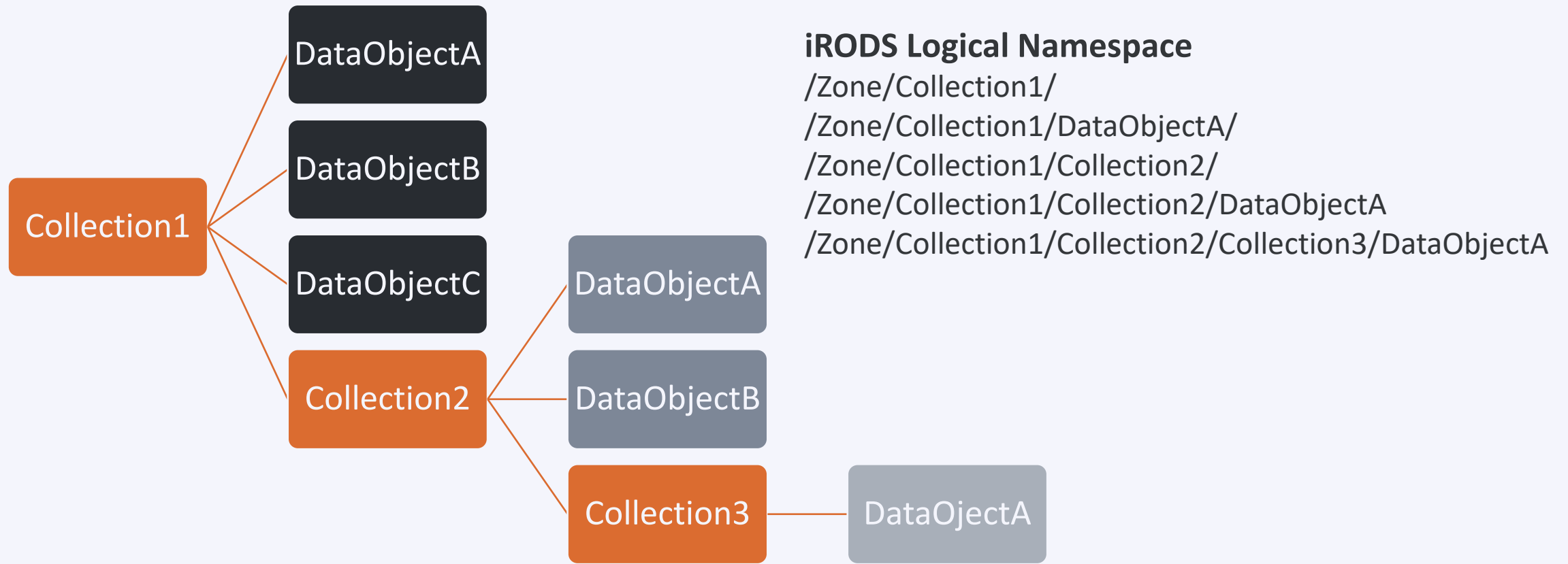
Rich Metadata for collections and data objects  
(System metadata and user-defined metadata)



## Secure collaboration

Three mechanisms: Permissions, Tickets and Federation.

# Data organization in iRODS

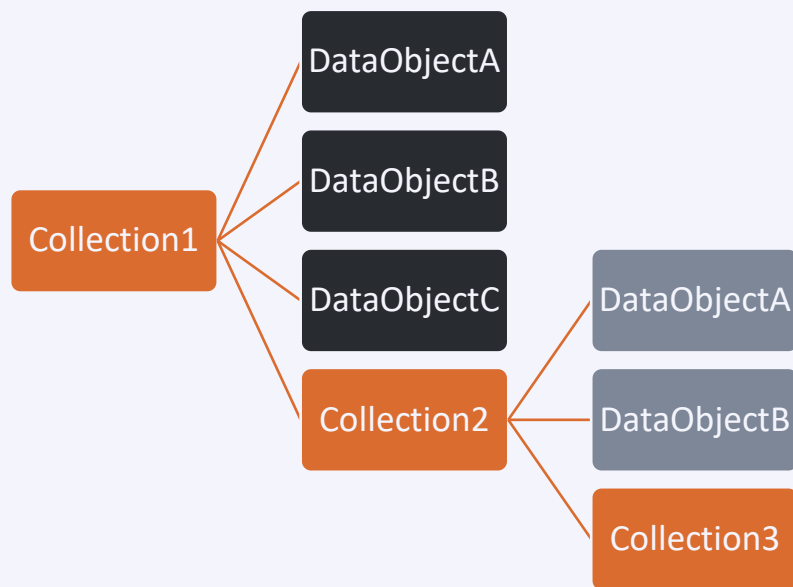


Collections ~ Directories

DataObjects ~ Files

# Data virtualization in iRODS

## Logical representation



## iRODS Logical Namespace

/Zone/Collection1/

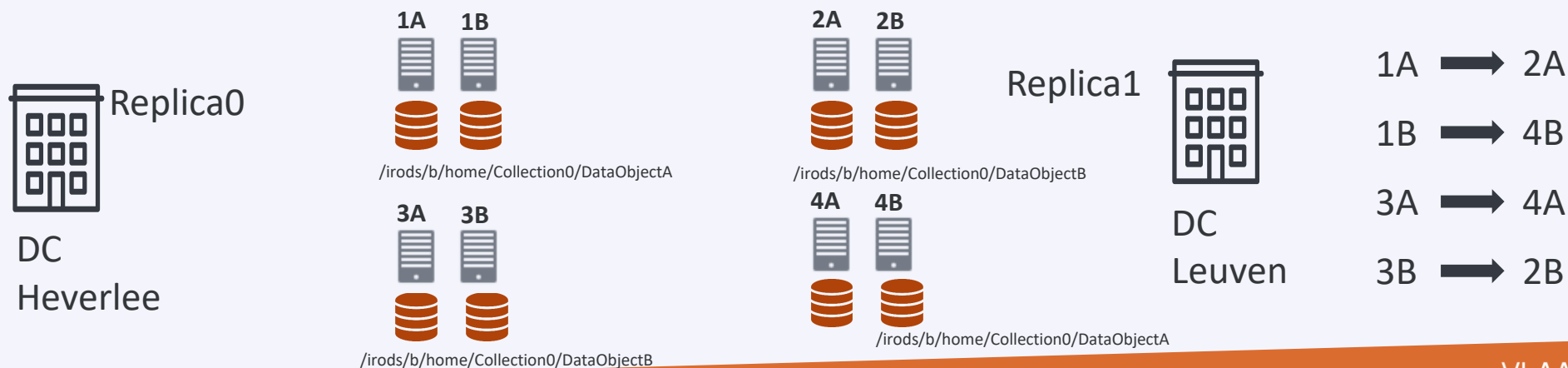
/Zone/Collection1/DataObjectA/

/Zone/Collection1/Collection2/

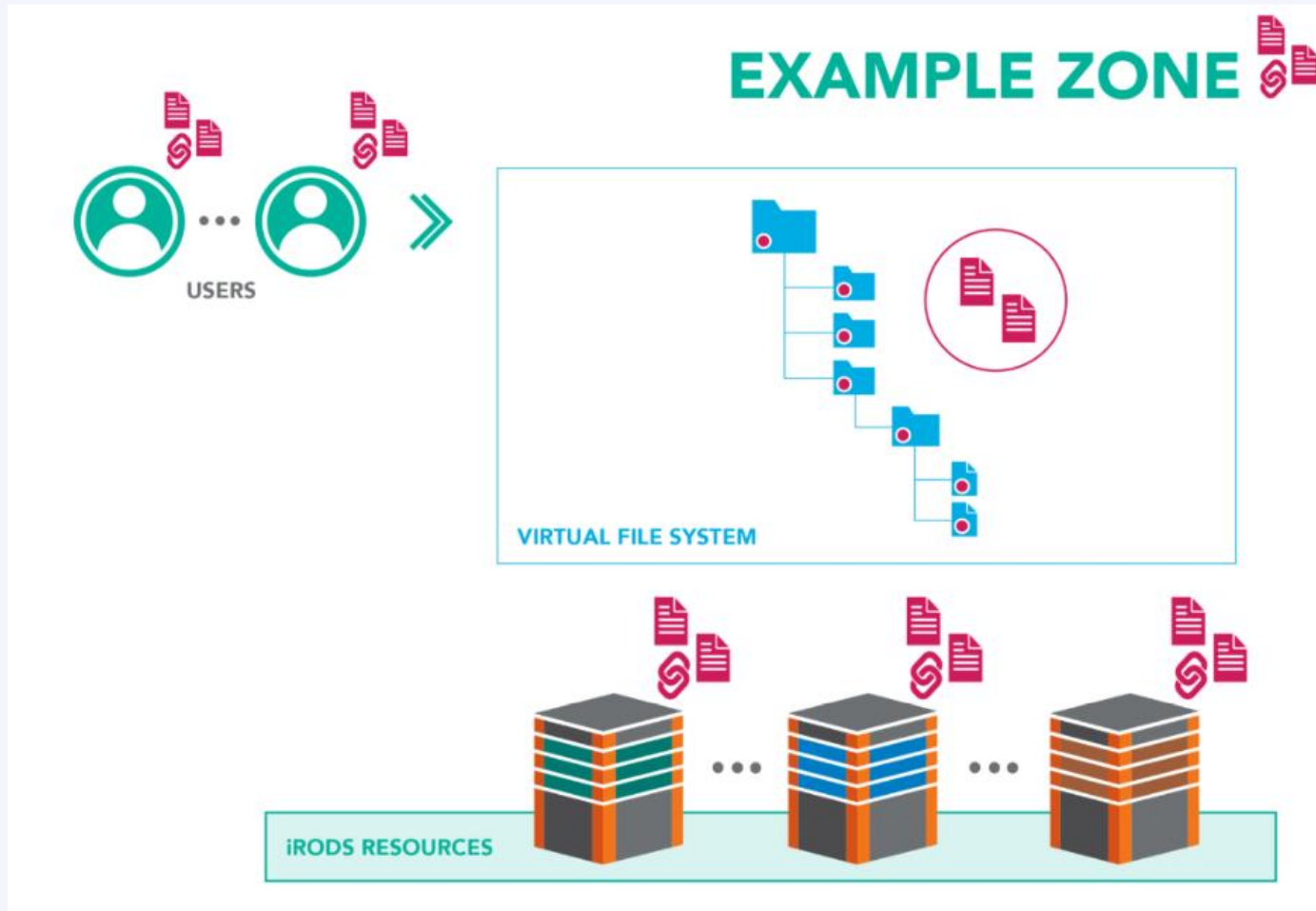
/Zone/Collection1/Collection2/DataObjectA

/Zone/Collection1/Collection2/Collection3/DataObjectA

## Physical representation



# Metadata in iRODS



## System Metadata:

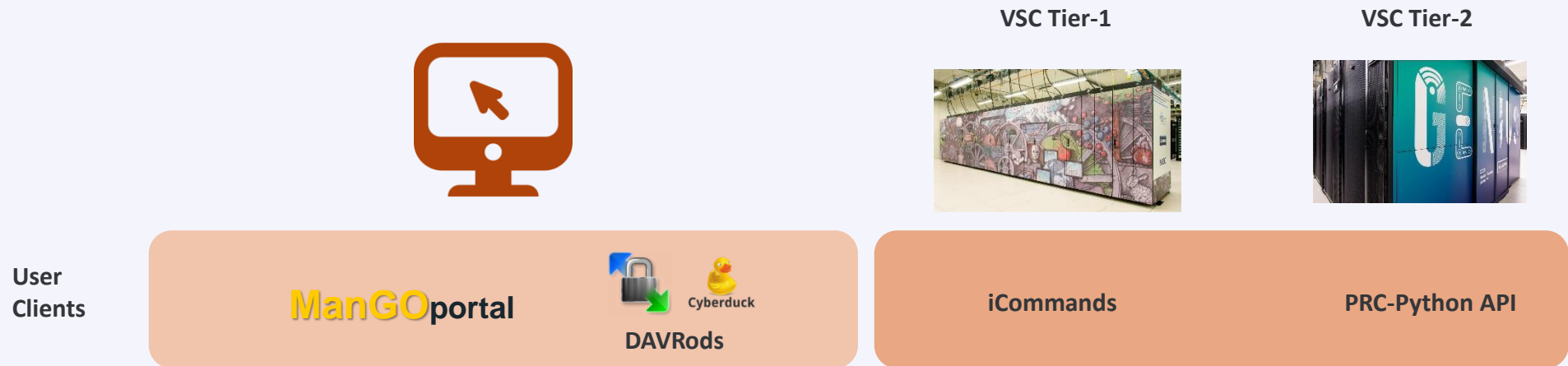
- filename, file size, creation date ...

## User Metadata:

- Manual introduction
- Metadata templates
- Automation (rules/microservices)



# Clients



## Interaction with iRODS

# Functionalities

## ■ icommands:

`iput - iget- irsync -imeta...`

- uploading/downloading data
- adding metadata to data objects/collections
- querying based on metadata
- deleting data objects/collections
- synchronization of data
- ACLs to data objects/collections

```
vsc33731@login1 ~  
$ imkdir research  
  
vsc33731@login1 ~  
$ icd research  
  
vsc33731@login1 ~  
$ iput dataset1  
  
vsc33731@login1 ~  
$ iput dataset2  
  
vsc33731@login1 ~  
$ ils  
/kuleuven_tier1_pilot/home/vsc33731/research:  
dataset1  
dataset2
```

# Functionalities

## ■ VSC-PRC:

Python3, python-irodsclient

- working with data objects/collections
- adding metadata to data objects/collections
- querying based on metadata
- deleting data objects/collections
- listing the disk usage
- ACLs to data objects/collections

```
In [1]: from vsc_irods.session import VSCiRODSSession
In [2]: session = VSCiRODSSession(txt='-')
In [3]: irods_path = session.path.get_irods_home() + "/research"
In [4]: session.path.imkdir('research')
In [5]: session.path.ichdir('research')
In [6]: session.bulk.put("./dataset*", irods_path)
In [7]: for item in session.search.find(irods_path, types='f'):
...:     print(item)
...:
/kuleuven_tier1_pilot/home/vsc33731/research/dataset1
/kuleuven_tier1_pilot/home/vsc33731/research/dataset2
```

# Functionalities

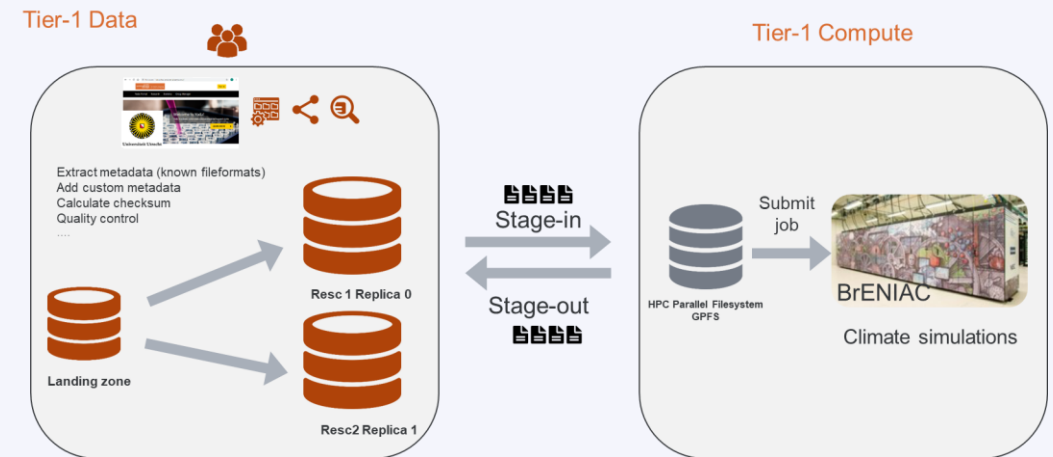
- HPC\_to\_Data:

- icommands:

```
input - iget- irsync -ibun
```

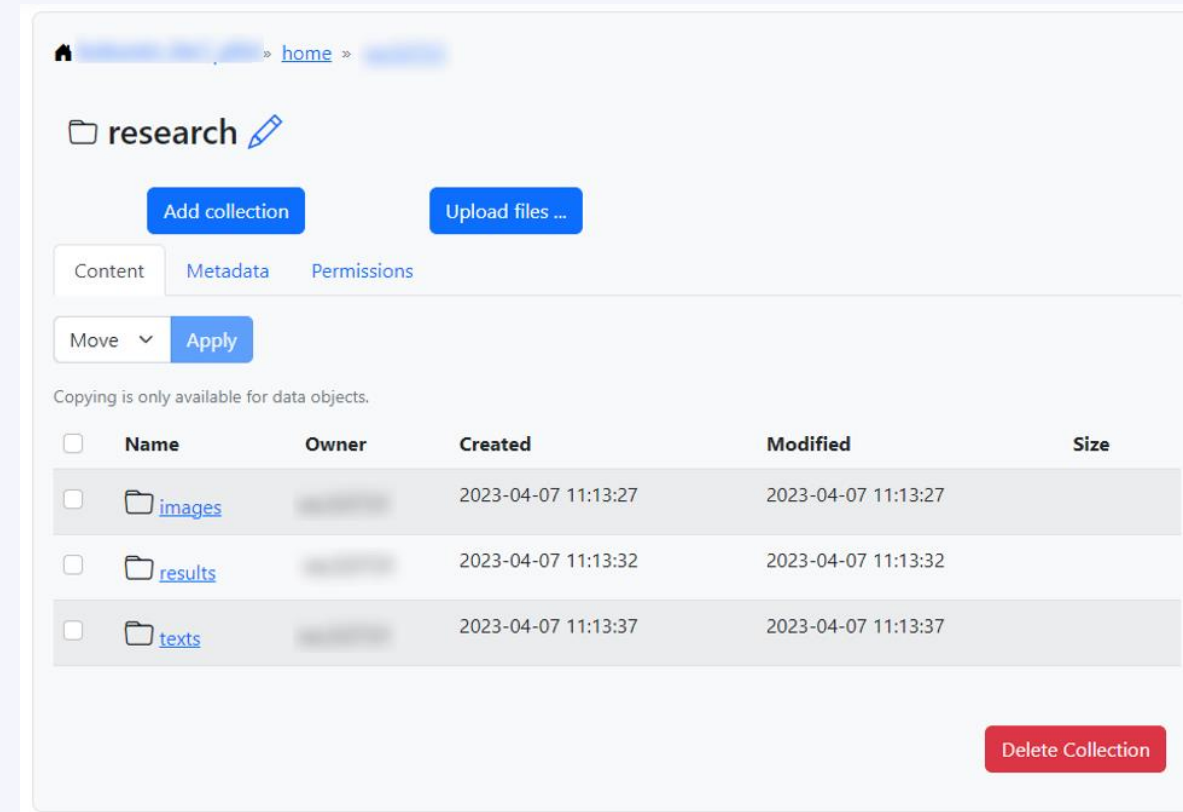
- VSC Python client:

```
vsc-prc-iget- vsc-prc-input
```



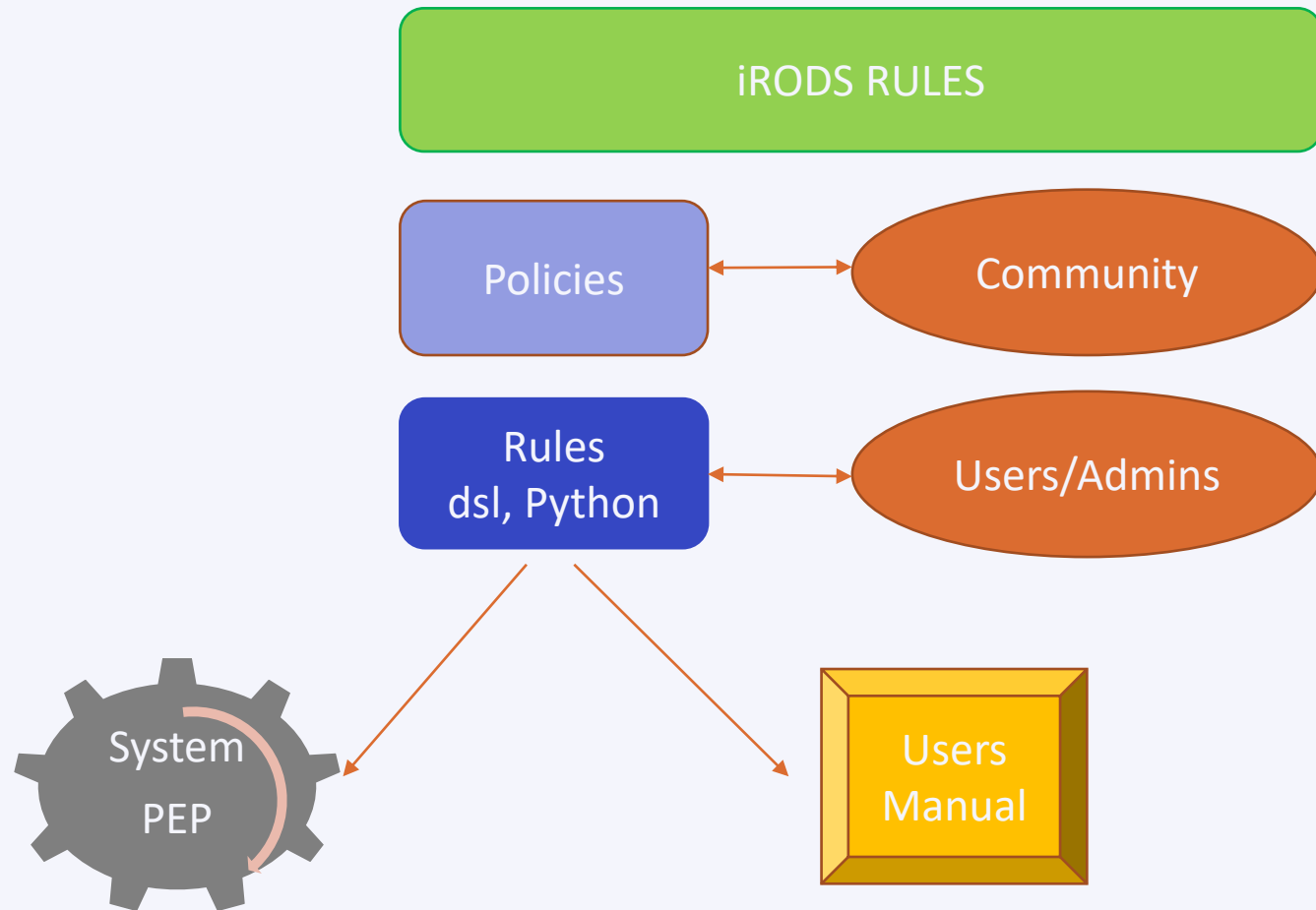
# Functionalities

- ManGO portal:
  - Graphical user interface
  - Full data management capabilities
  - Focus on metadata
    - Metadata schemas
    - Automatic extraction from file headers



# Functionalities

- irule:
  - user level rules
  - written into a local file
  - execute it when you need



# Functionalities

- Other clients (not mentioned in video):
  - webDav
  - Cyberduck
  - WinSCP

# Documentation and support

- Documentation

[https://docs.vscentrum.be/en/data\\_m/](https://docs.vscentrum.be/en/data_m/)

- Support

`data@vscentrum.be`



*Stay Connected  
to VSC*

**Linked**  <sup>®</sup>

