iRODS User Training Introduction

- Introduction

- Data and RDM

- Tier-1 Data platform Architecture

- What is iRODS?

- Functionalities

VLAAMS
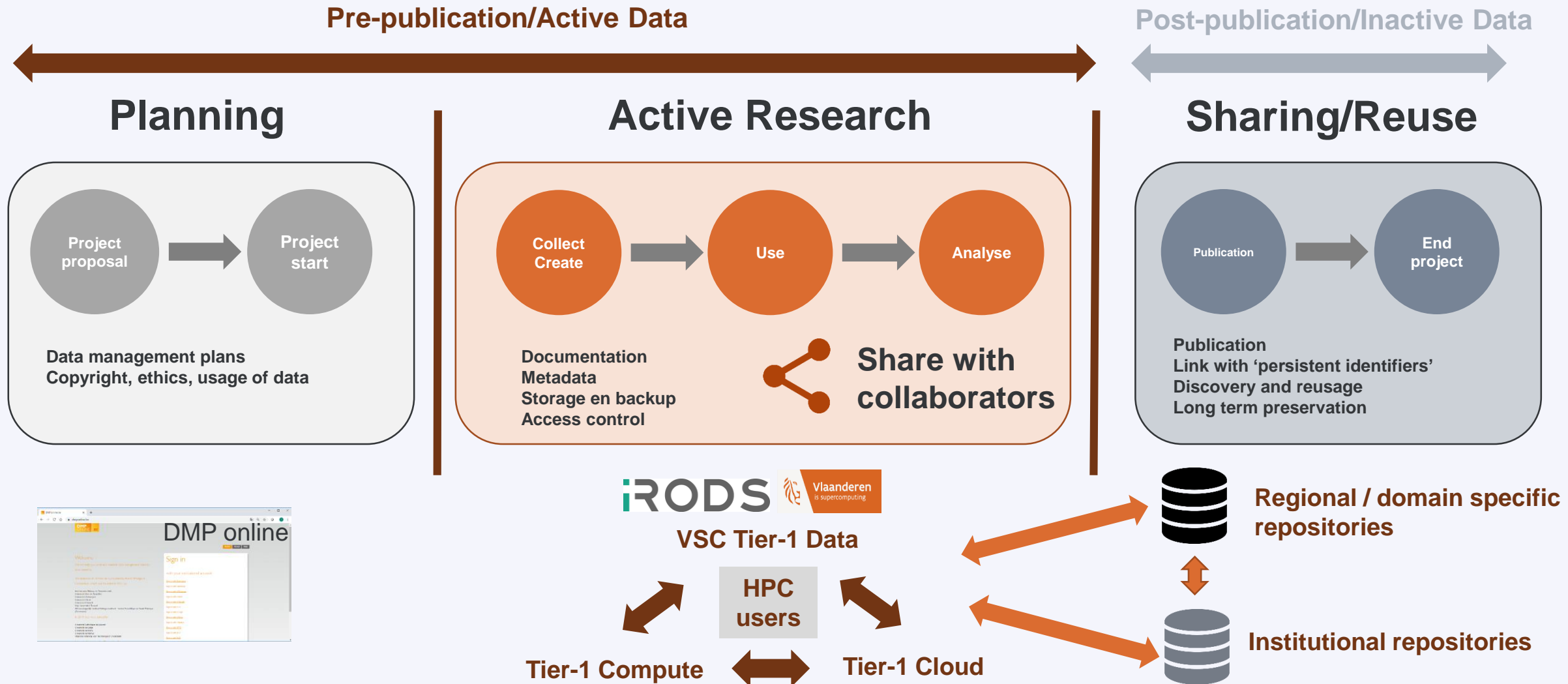SUPERCOMPUTER
CENTRUM

Introduction

- The aim of the training is to explain basics of Tier-1 Data Service.

- iRODS training consists of general introduction, iCommands, VSC-PRC, basic irules and portal client (yoda, metalnx).

- This training is planned for VSC users.

- It includes hands-on sessions.

- Whole training may take around 3 hours, the introduction will take 15 minutes.

- Any questions, feedbacks will help us improve the quality of the training.

VLAAMS
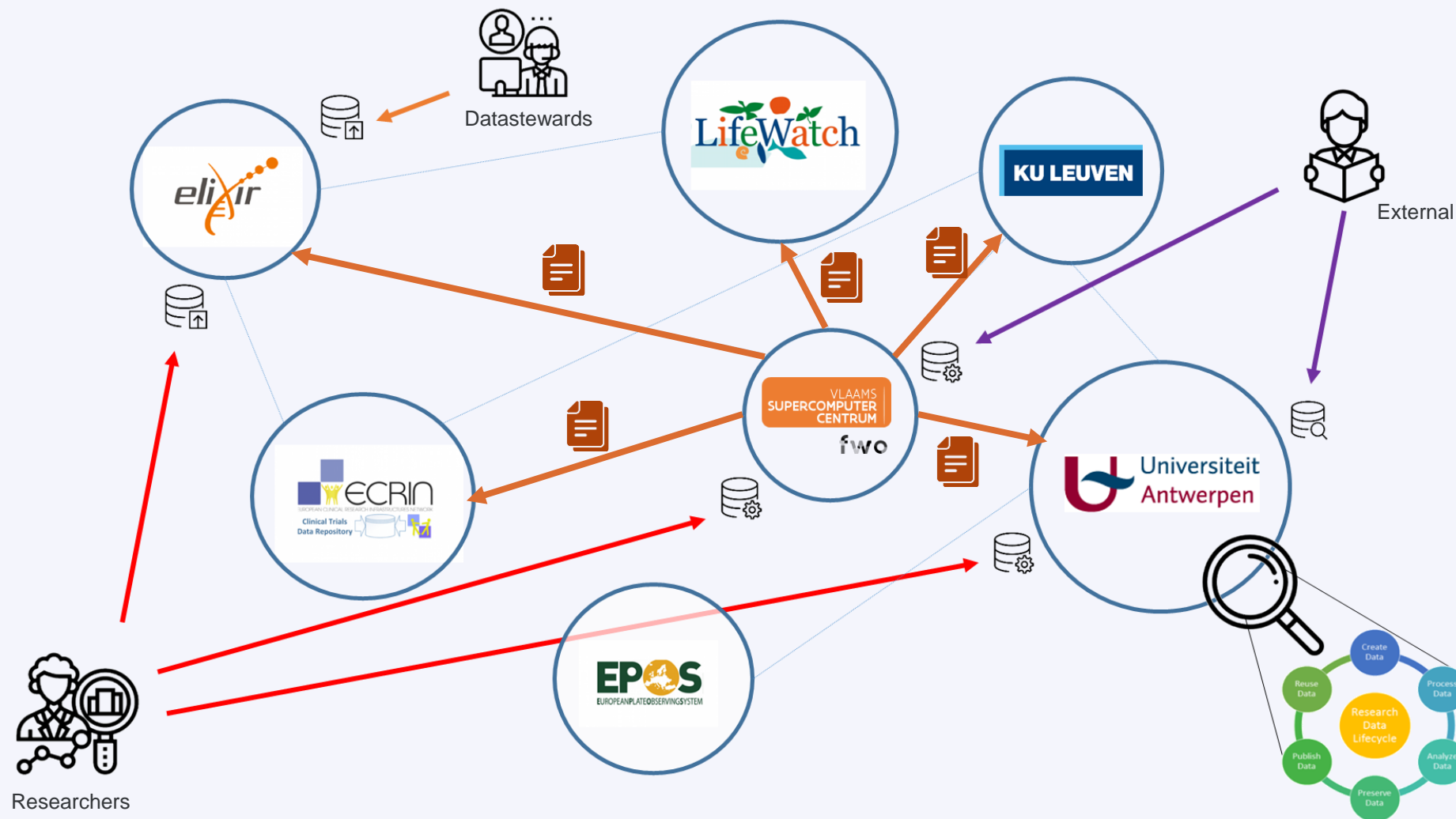SUPERCOMPUTER
CENTRUM

Data and RDM

Data; what is in it? What is the problem with it?

System Metadata

User Defined Metadata

Meta

Data

Referrring

Collections & Data Object

Storage

Replica

VLAAMS
SUPERCOMPUTER
CENTRUM
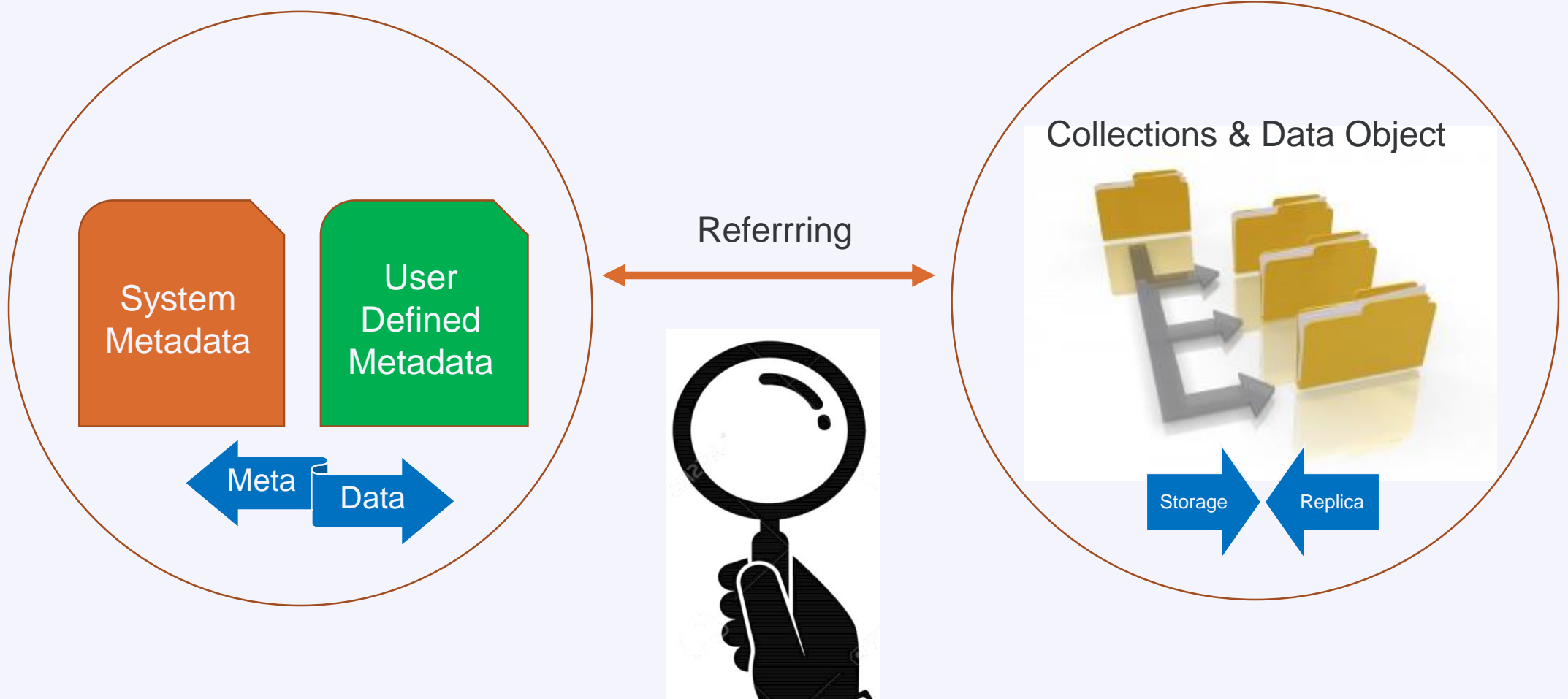
Tier-1 Data in the research Data Lifecycle

Pre-publication/Active Data

Planning

Active Research

Sharing/Reuse

Project proposal → Project start

Data management plans
Copyright, ethics, usage of data

Collect Create → Use → Analyse

Documentation
Metadata
Storage en backup
Access control

Share with collaborators

Publication → End project

Publication
Link with 'persistent identifiers'
Discovery and reusage
Long term preservation

DMP online

iRODS Vlaanderen is supercomputing

VSC Tier-1 Data

Regional / domain specific repositories

HPC users

Tier-1 Compute ↔ **Tier-1 Cloud**

Institutional repositories

The RDM landscape

Tier-1 Data architecture

VSC Tier-1

VSC Tier-2

User Clients

YODA   **metalnx**   **DAVRods**   **iCommands**

iCommands   **PRC-Python API**

VSC iRODS 1 zone

HA

Rule Engine

iCAT

MySQL cluster

Posix Resc1

Ceph Resc3

3km

Posix Resc2 Replica Resc1
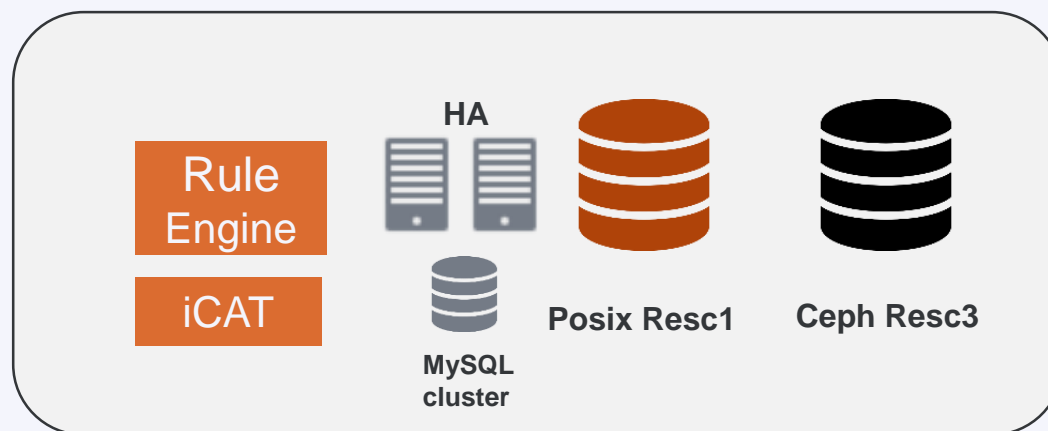
DC Heverlee

DC Leuven

What is iRODS?

- iRODS (integrated Rule-Oriented Data System)

- Open Source distributed data and storage management system

- Configurable data management policies and workflows

- Scalable

- iRODS consortium ensures sustainability by:

  - Guiding further development of the software;

  - Growing the user and developer communities; and

  - Facilitating iRODS support, education, and collaboration opportunities.
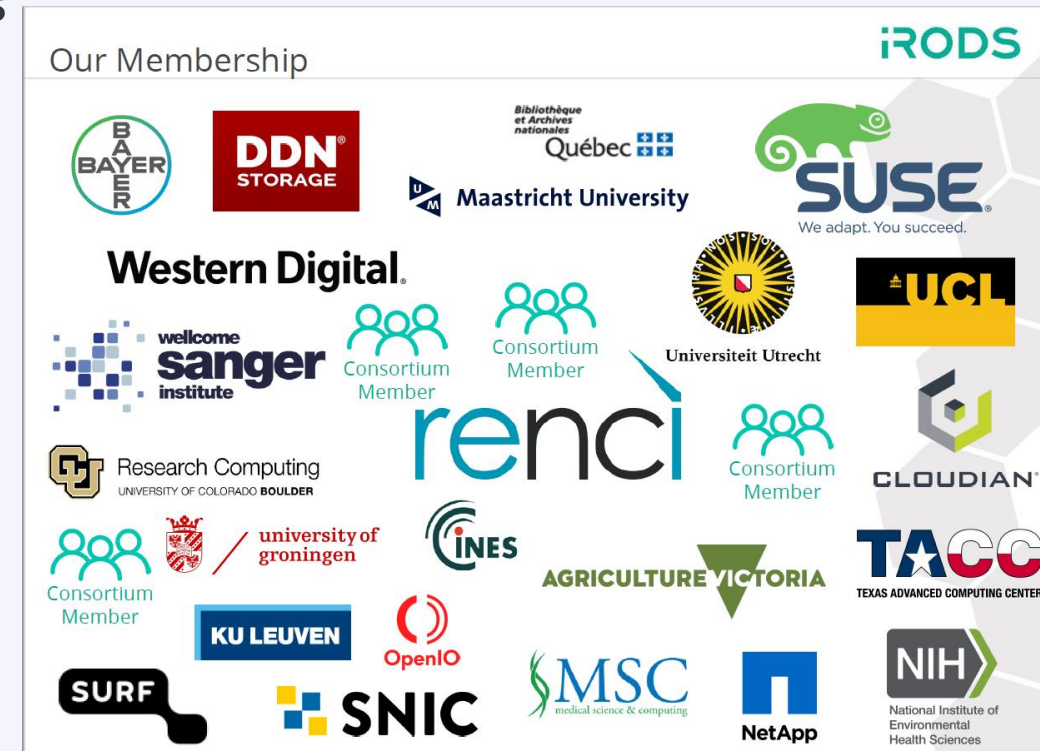
iRODS Core competencies

Unified Storage Namespace

Data virtualization of distributed storage systems

Automation

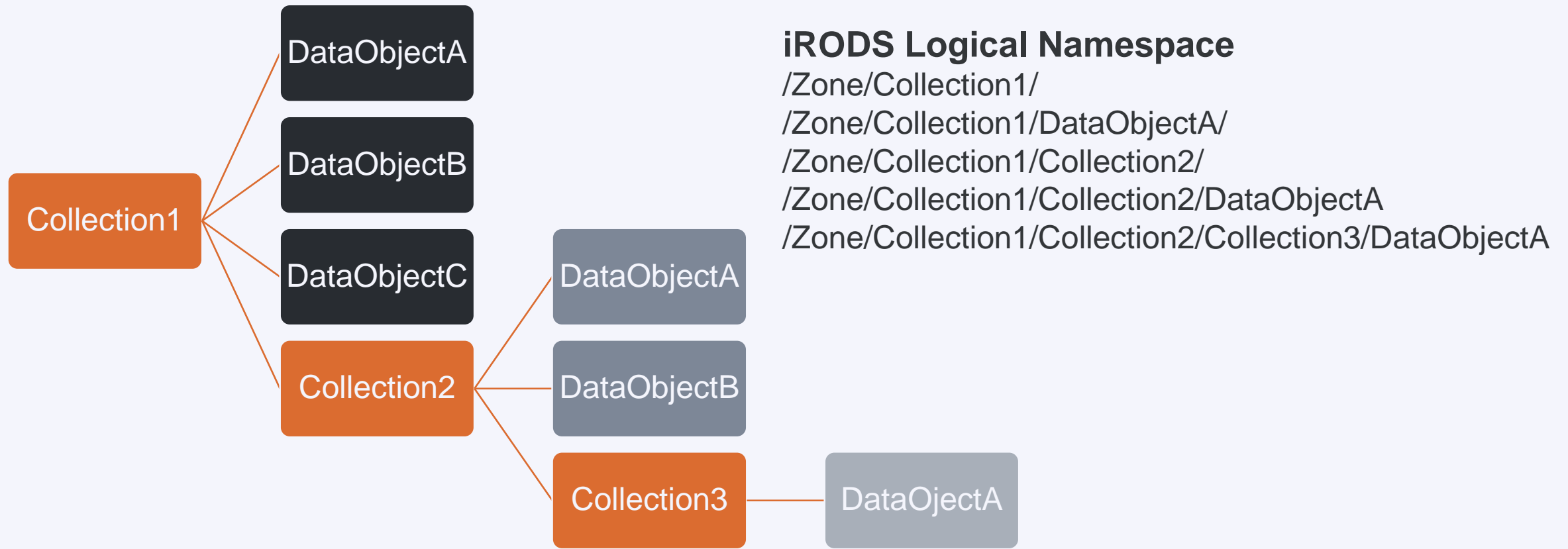Rule Engine to enforce data polices

Data Discovery

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

Secure collaboration

Three mechanisms: Permissions, Tickets and Federation.

VLAAMS
**SUPERCOMPUTER
CENTRUM**

Data organization in iRODS

iRODS Logical Namespace
/Zone/Collection1/
/Zone/Collection1/DataObjectA/
/Zone/Collection1/Collection2/
/Zone/Collection1/Collection2/DataObjectA
/Zone/Collection1/Collection2/Collection3/DataObjectA

Collections ~ Directories
DataObjects ~ Files

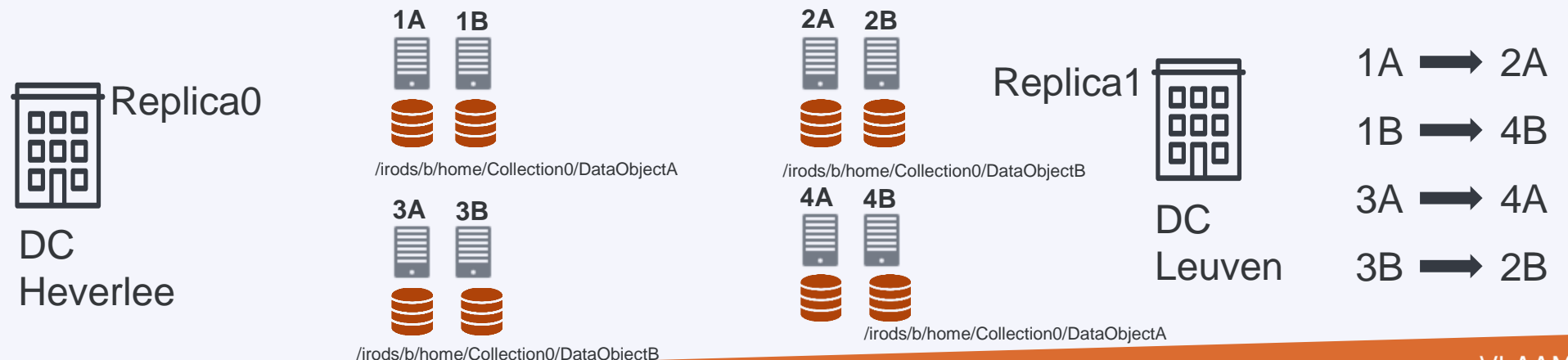# Data virtualization in iRODS
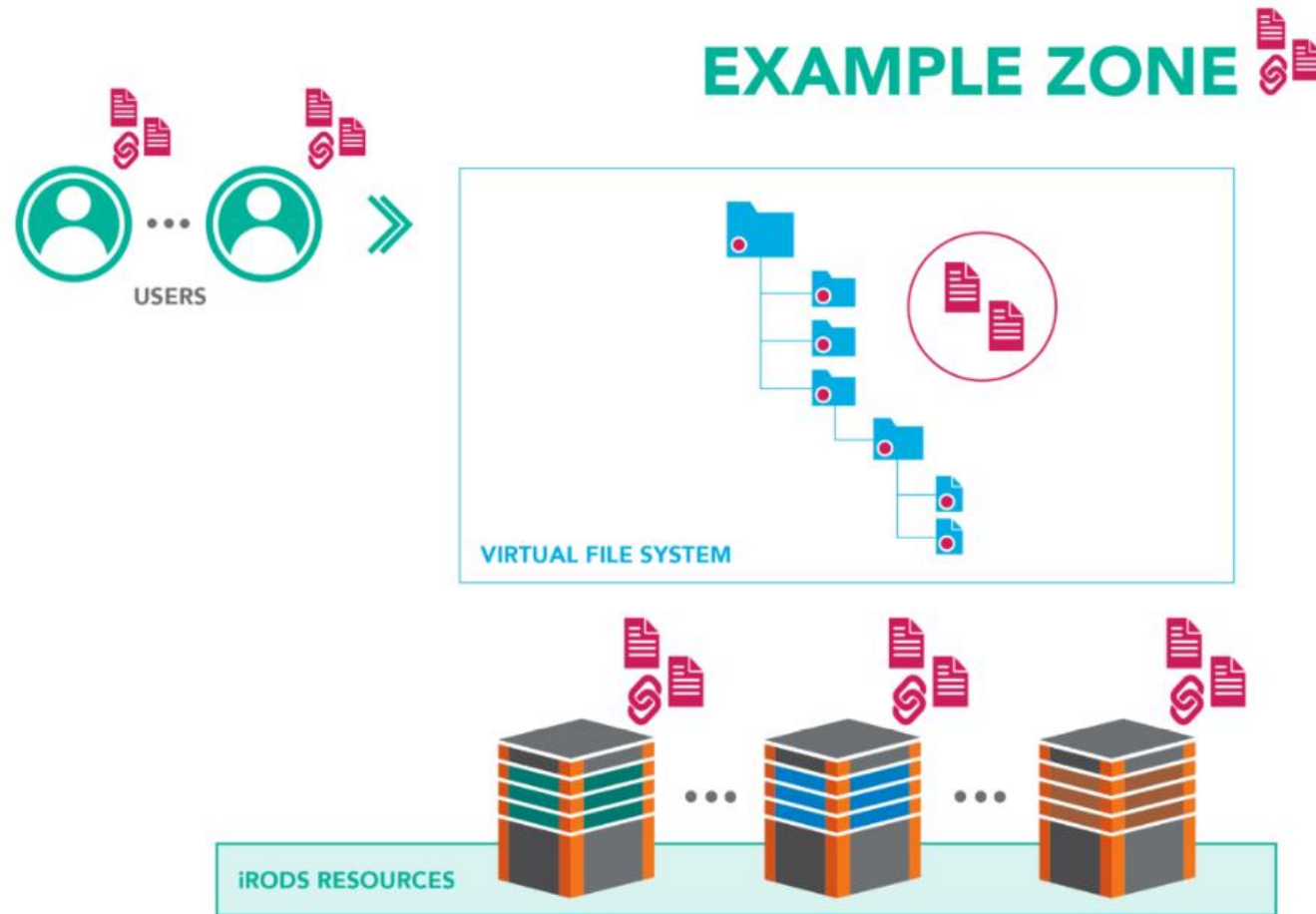
Logical representation
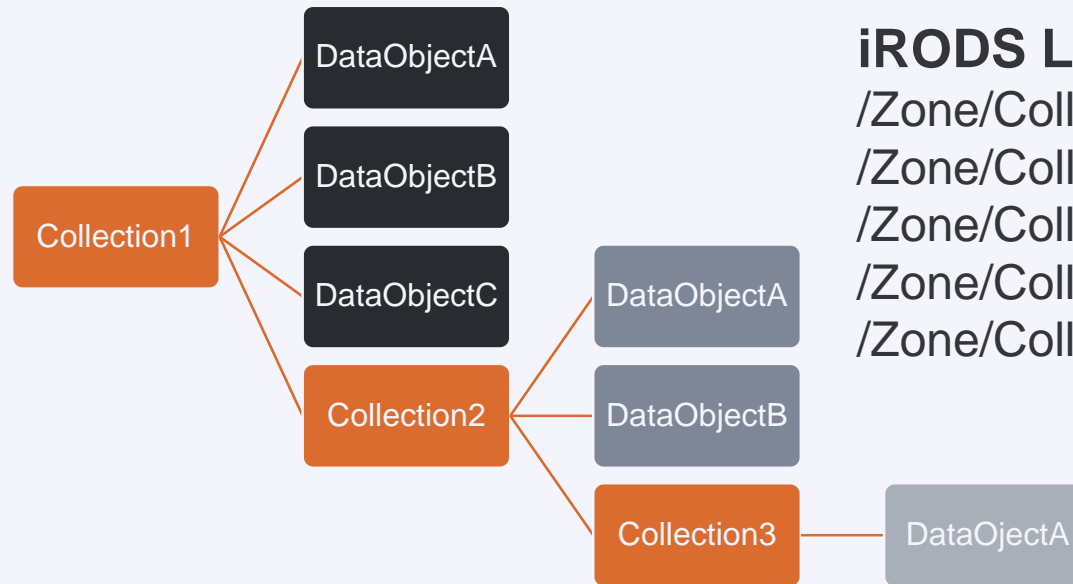
iRODS Logical Namespace
/Zone/Collection1/
/Zone/Collection1/DataObjectA/
/Zone/Collection1/Collection2/
/Zone/Collection1/Collection2/DataObjectA
/Zone/Collection1/Collection2/Collection3/DataObjectA

Physical representation

Replica0
DC Heverlee

/irods/b/home/Collection0/DataObjectA

/irods/b/home/Collection0/DataObjectB

Replica1
DC Leuven

/irods/b/home/Collection0/DataObjectB

/irods/b/home/Collection0/DataObjectA
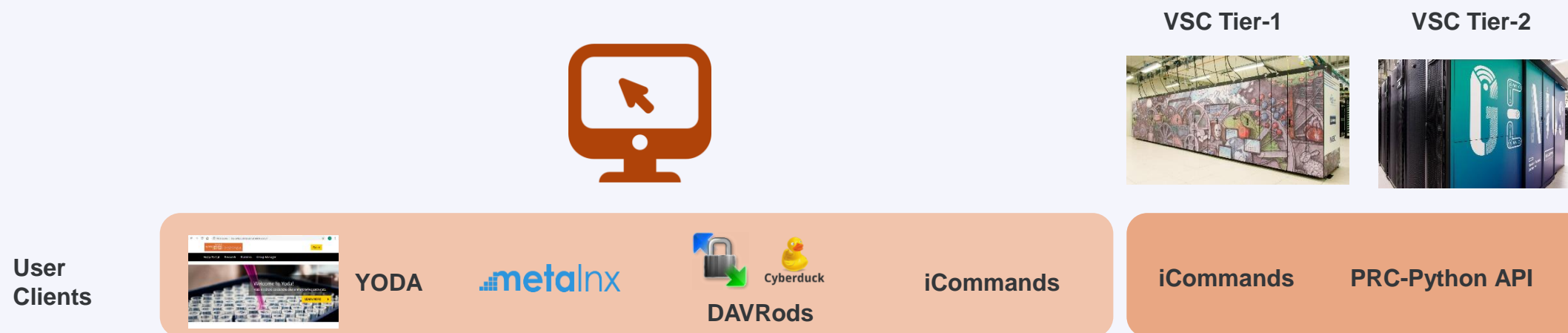
1A ➡ 2A
1B ➡ 4B
3A ➡ 4A
3B ➡ 2B

Metadata in iRODS

System Metadata:
- filename, file size, creation date …

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

VLAAMS
SUPERCOMPUTER
CENTRUM

Functionalities

VSC Tier-1   VSC Tier-2

User
Clients   YODA   metalnx   Cyberduck   iCommands   iCommands   PRC-Python API

DAVRods

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

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

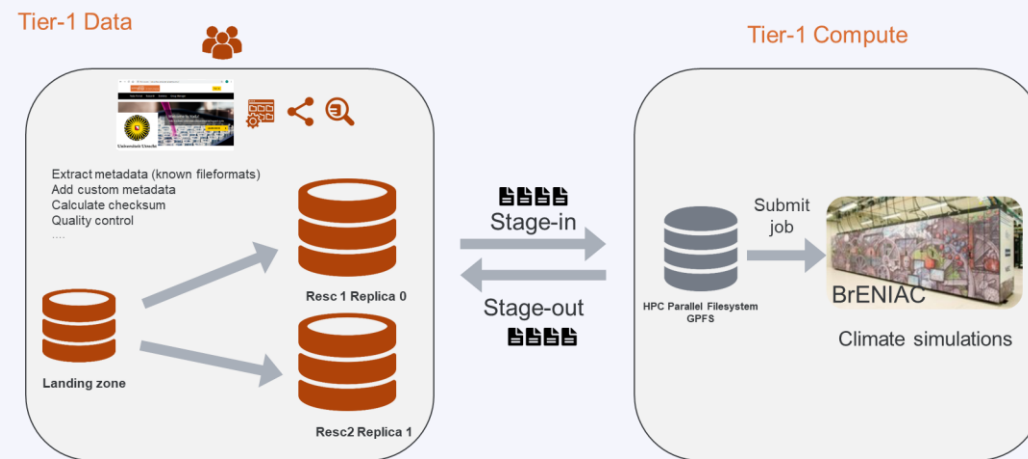- ACLs to data objects/collections

Functionalities

■ HPC_to_Data:

■ icommands:

```
iput – iget- irsync -ibun
```

■ VSC Python client:

```
vsc-prc-iget- vsc-prc-iput
```

Functionalities

■ Yoda:

- graphical user interface easiness
- working with data objects/collections
- adding metadata to data collections
- querying visually
- downloading data objects
- group management
- RDM workflow - UU

■ MetaInx:

- graphical user interface easiness
- working with data objects/collections
- adding metadata to data objects/collections
- downloading data objects
- permission
- iRODS design

VLAAMS
SUPERCOMPUTER
CENTRUM

Functionalities

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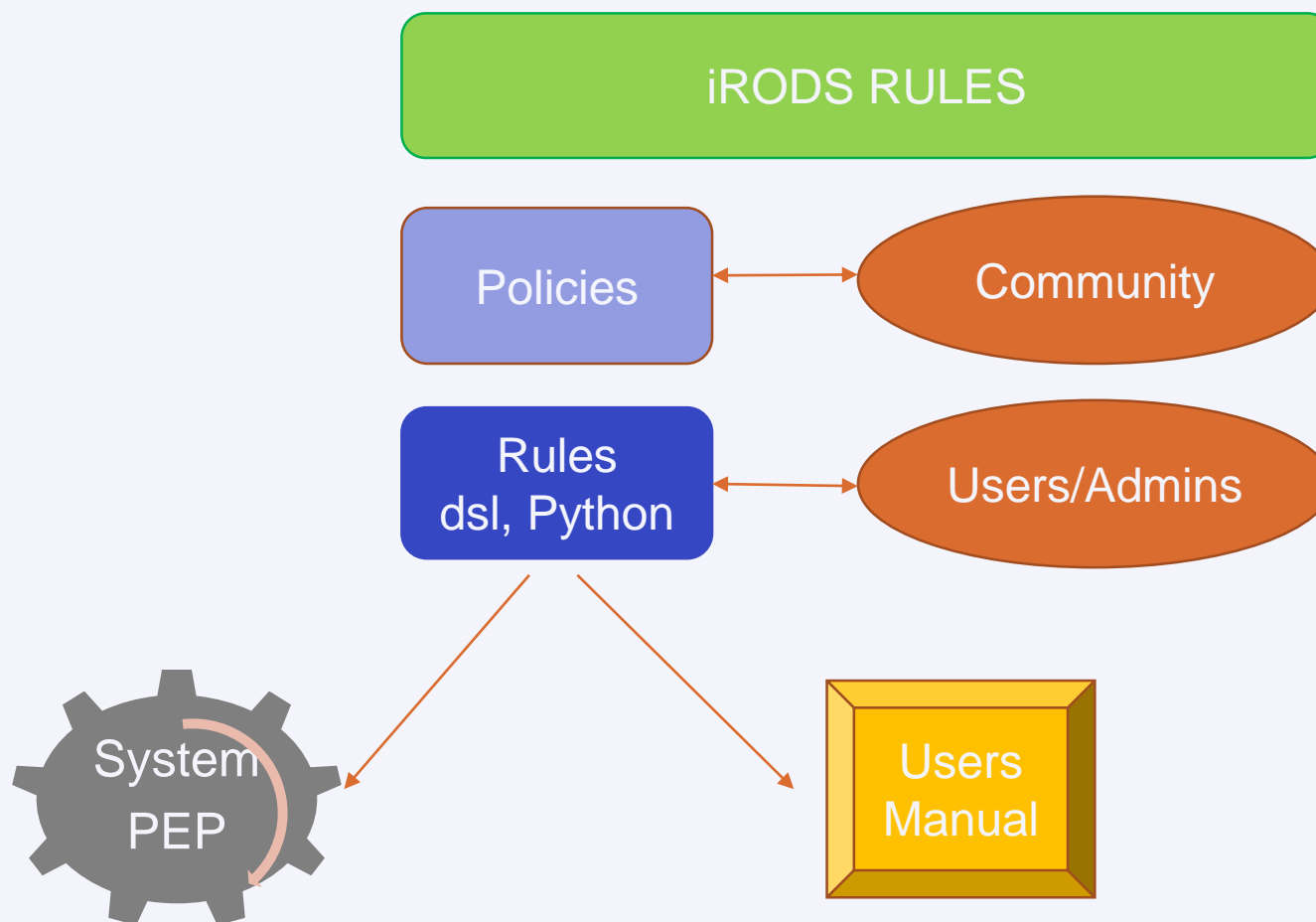
Documentation and support

- Documentation

https://vlaams-supercomputing-centrum-vscdocumentation.readthedocs-hosted.com/_/sharing/cxplsgyxzaizmf4xg7wl5jj8

NOTE: After you click the special link above, you will reach the latest version of the VSC Documentation which doesn't include "Tier-1 Data Service". Hence you should click the version arrow on the right below side of the incoming page and chose the "data_M" version.

- Support

    data@vscentrum.be

Questions