iRODS User Training

Vlaanderen
is supercomputing

VLAAMS
SUPERCOMPUTER
CENTRUM

Innovative Computing
for A Smarter Flanders

vscentrum.be

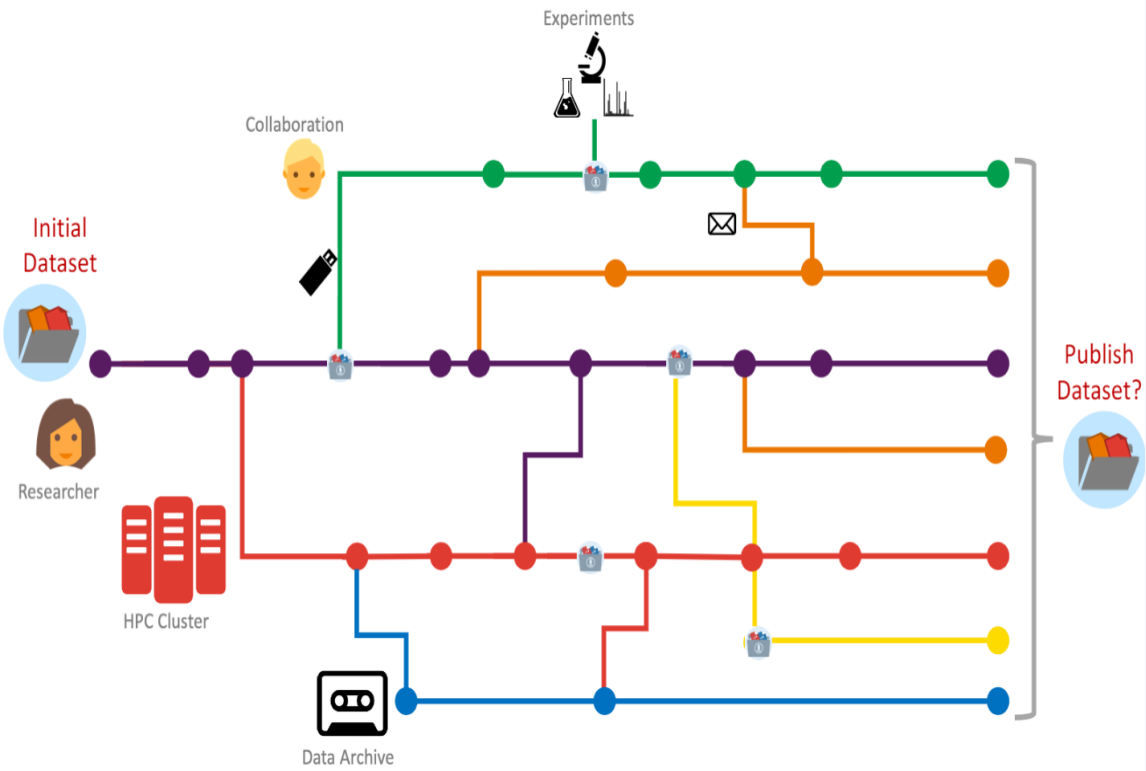
- Introduction
- Data and RDM
- Tier-1 Data platform Architecture
- What is iRODS?
- Functionalities

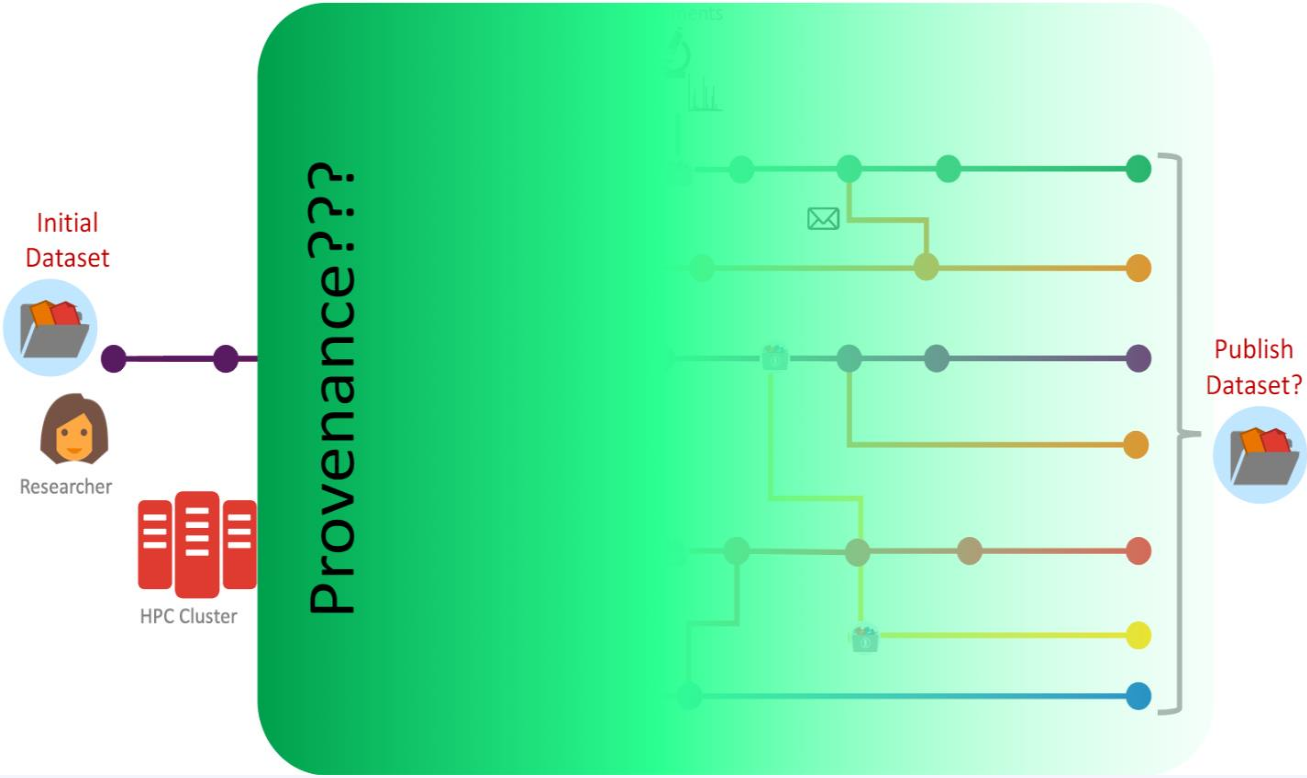
Introduction

- 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.

Data and RDM

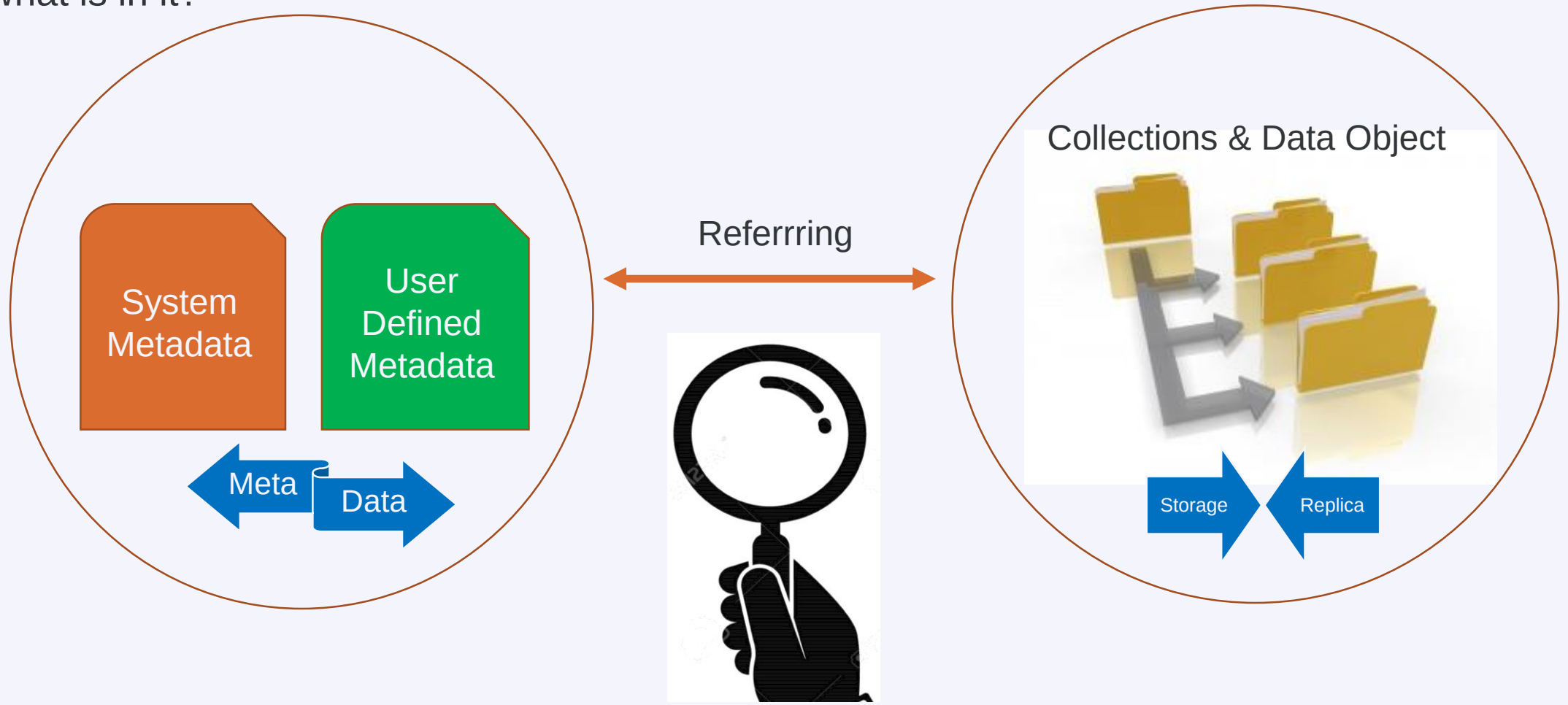
Data, what is the problem with it?





Data and RDM

Data, what is in it?



Data Life Cycle

Descriptive

sensitive

Provenance

Experiments

CREATING

PROCESSING

ANALYSING

PRESERVING

SHARING

RE-USING

Operational Metadata

System Metadata

HPC Cluster

Access rights

Data Archive

VLAAMS SUPERCOMPUTER CENTRUM

Tier-1 Data in the research Data Lifecycle

Pre-publication/Active Data

Post-publication/Inactive Data

Planning

Project proposal → Project start

Data management plans
Copyright, ethics, usage of data

Active Research

Collect Create → Use → Analyse

Documentation
Metadata
Storage en backup
Access control

Share with collaborators

Sharing/Reuse

Publication → End project

Publication
Link with ‘persistent identifiers’
Discovery and reusage
Long term preservation

iRODS Vlaanderen is supercomputing

VSC Tier-1 Data

HPC users

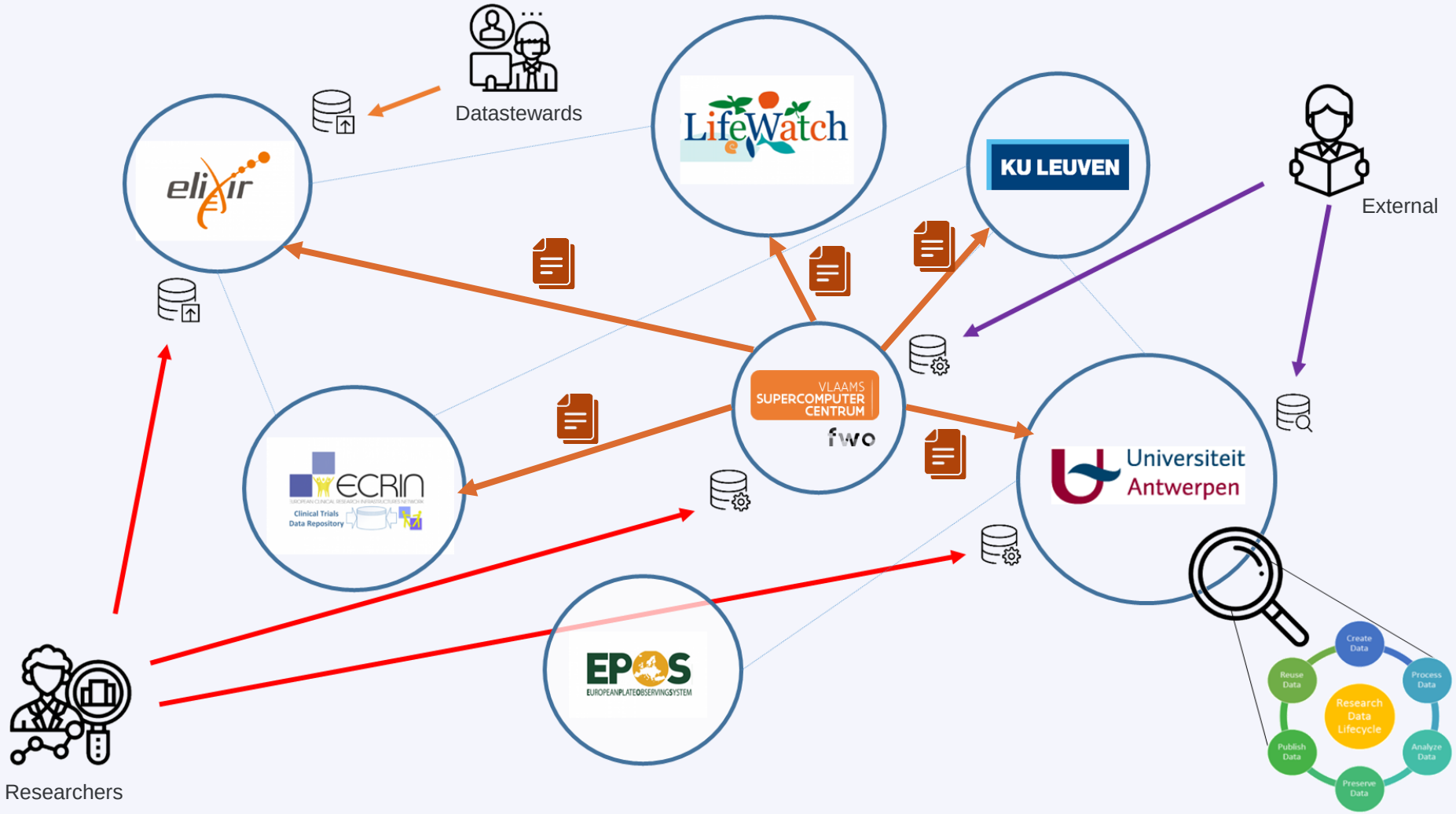
Tier-1 Compute

Tier-1 Cloud

Regional / domain specific repositories

Institutional repositories

The RDM landscape



Tier-1 Data architecture

VSC Tier-1

VSC Tier-2

User Clients

YODA

metalnx

Cyberduck

DAVRods

globus online

iCommands

iCommands PRC-Python API

globus online

VSC iRODS 1 zone

Rule Engine

iCAT

HA

MySQL cluster

Posix Resc1

Ceph Resc3

3km

Posix Resc2
Replica Resc1

DC Heverlee

DC Leuven

VLAAMS SUPERCOMPUTER CENTRUM

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

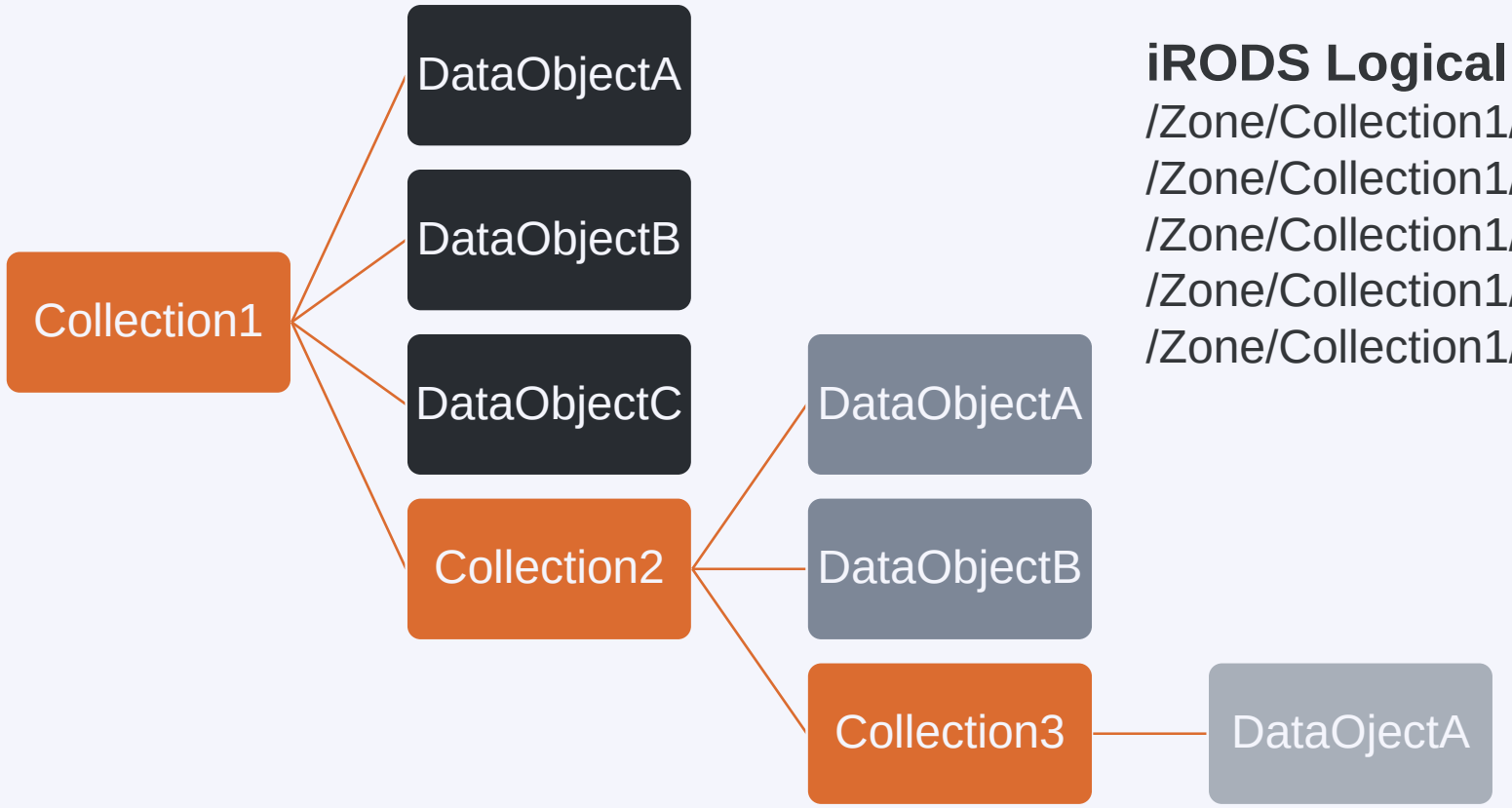
Data Discovery

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

Secure collaboration

Three mechanisms: Permissions, Tickets and Federation.

Data organization in iRODS



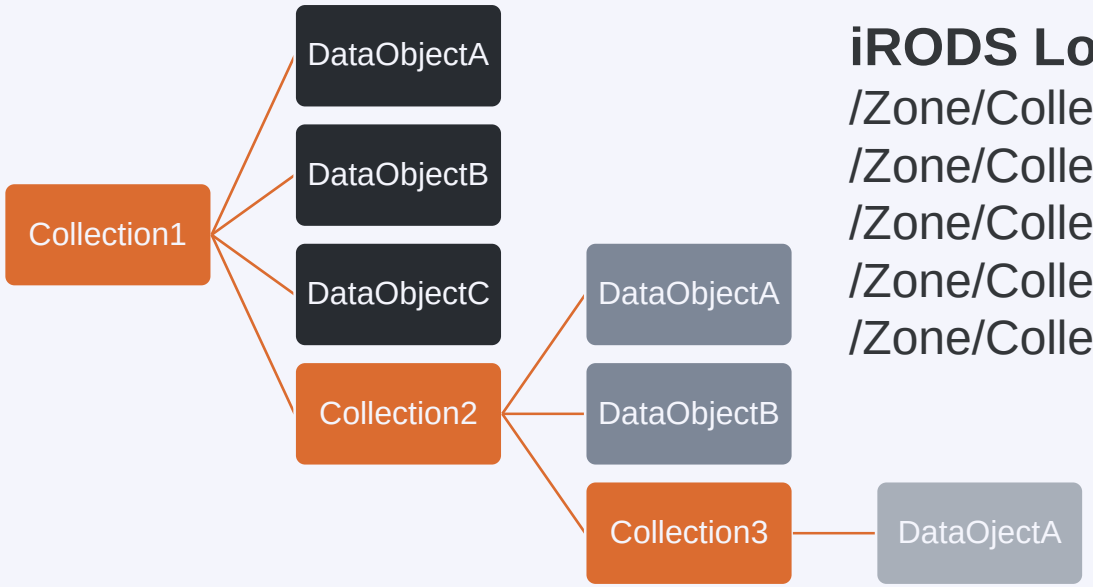
iRODS Logical Namespace

/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

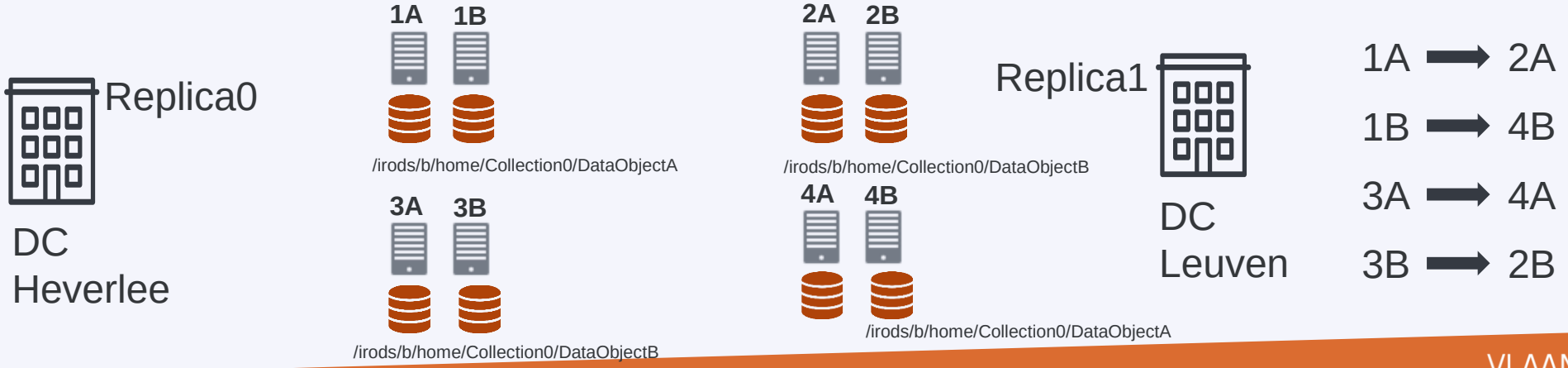
Logical representation



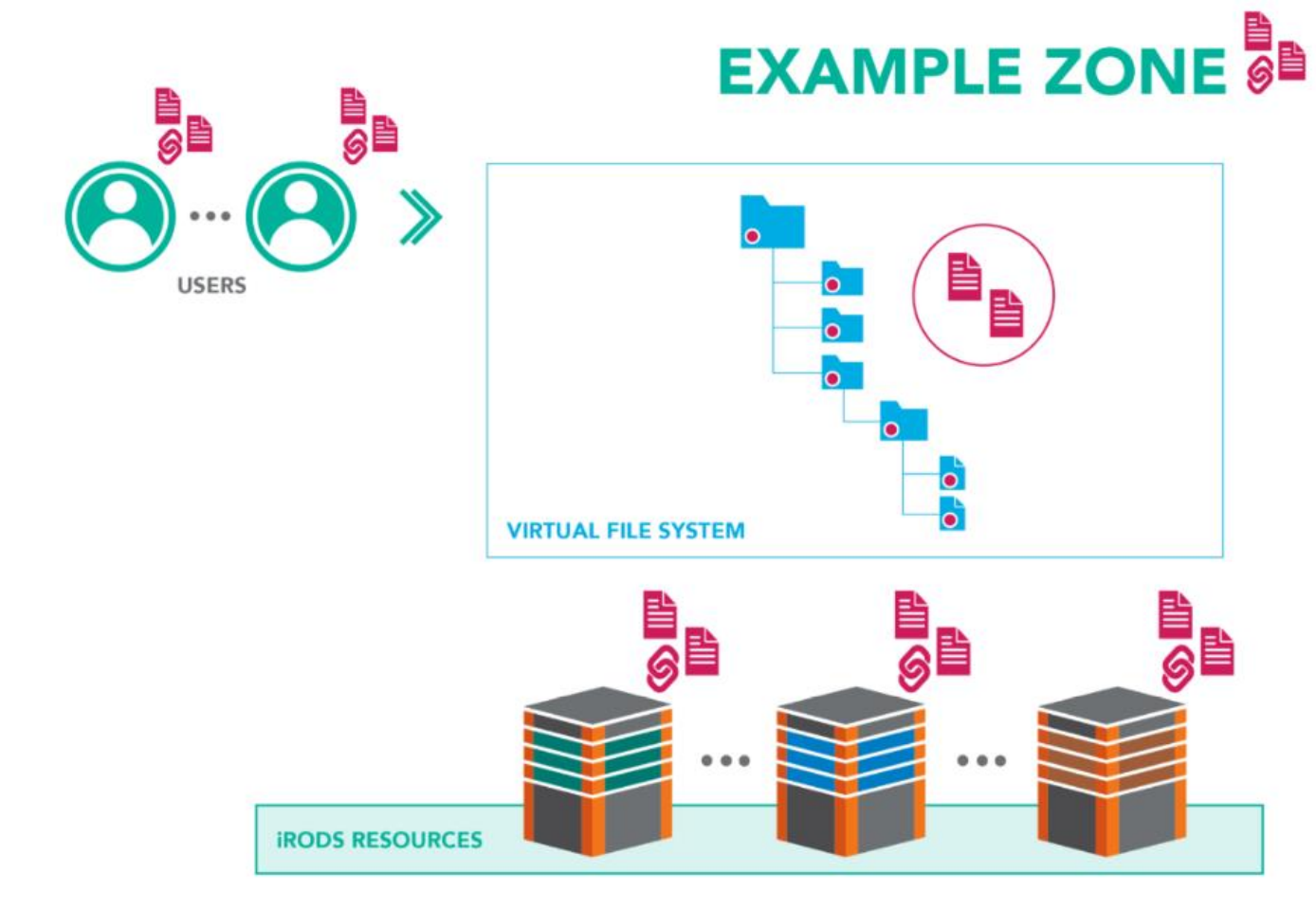
iRODS Logical Namespace

/Zone/Collection1/
/Zone/Collection1/DataObjectA/
/Zone/Collection1/Collection2/
/Zone/Collection1/Collection2/DataObjectA
/Zone/Collection1/Collection2/Collection3/DataObjectA

Physical representation



Metadata in iRODS



System Metadata:

- filename, file size, creation date …

User Metadata:

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

Functionalities

VSC Tier-1

VSC Tier-2

User Clients

YODA

metalnx

Cyberduck

DAVRods

globus online

iCommands

iCommands PRC-Python API

globus online

Interaction with iRODS

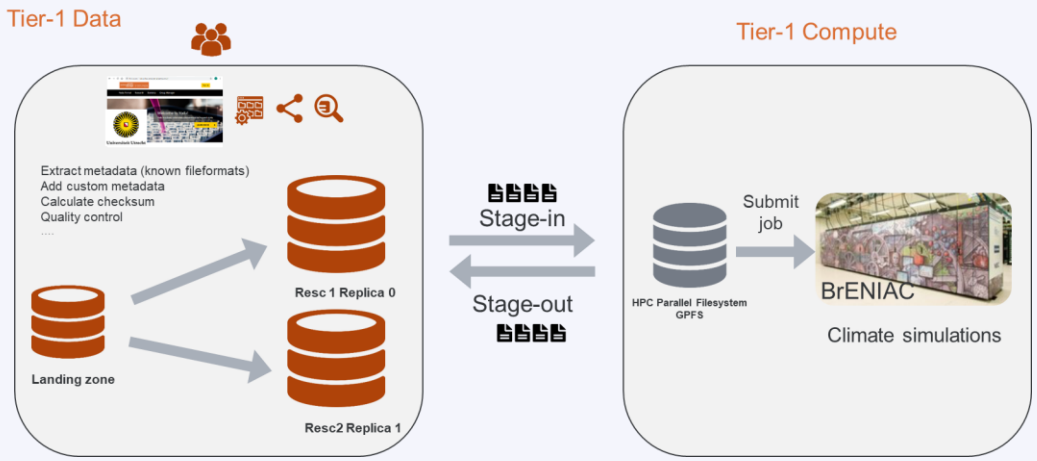
Functionalities

- icommands:

  `iput – iget- irsync -ibun`

- VSC Python client:

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



Documentation and support

- Documentation

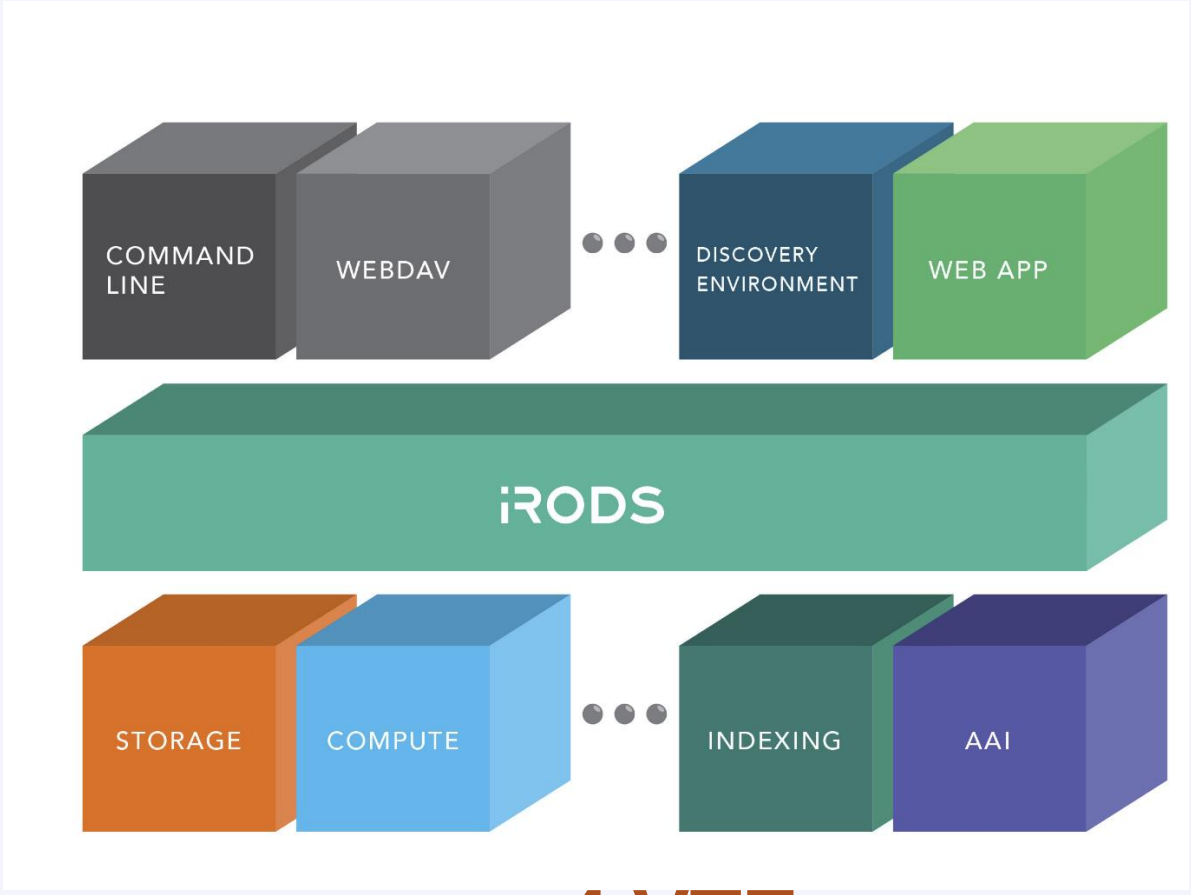
  https://vlaams-supercomputing-centrum-vscdocumentation.readthedocs-hosted.com/en/data_m/data/tier1_data_main_index.html

- Support

  data@vscentrum.be

Questions

Tier-1 Data: the team



~ 4 VTE

Paul Borgermans

Mustafa Dikmen

Maxime Van den Bossche

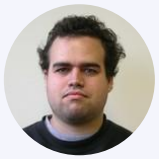
Jef Scheepers

User Support
User tools
Policies
Documentation

Peter Verraedt

iRODS middleware
iRODS servers

Tom Vanmierlo



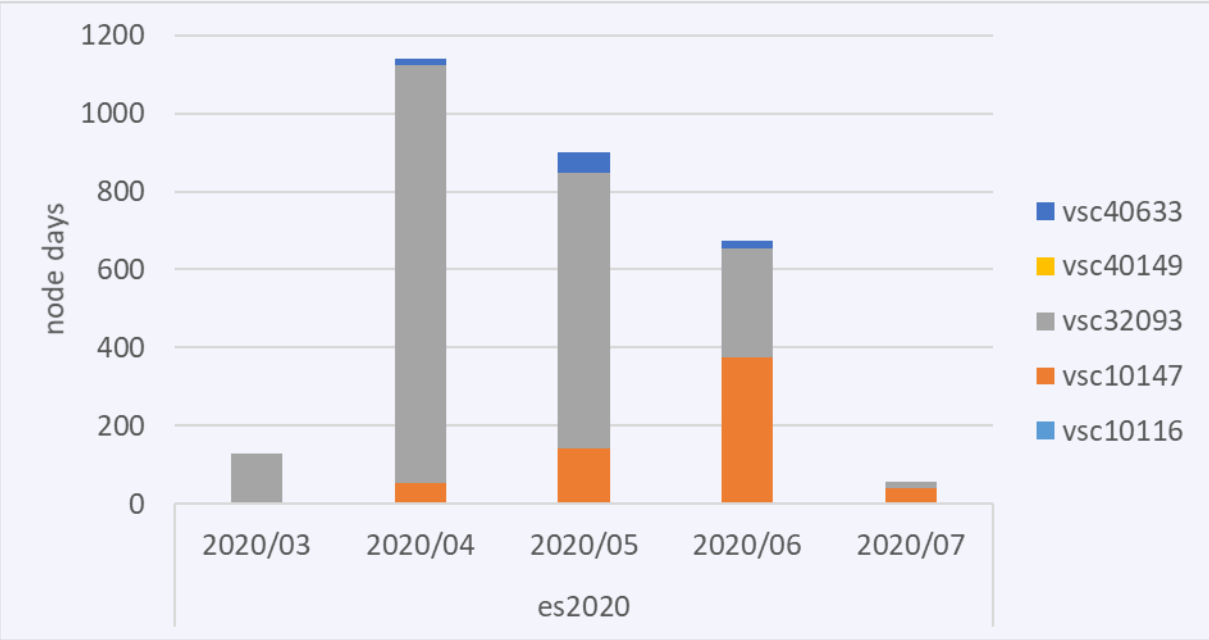
Kristoff Van Buggenhout

Bart Vanneste

Storage

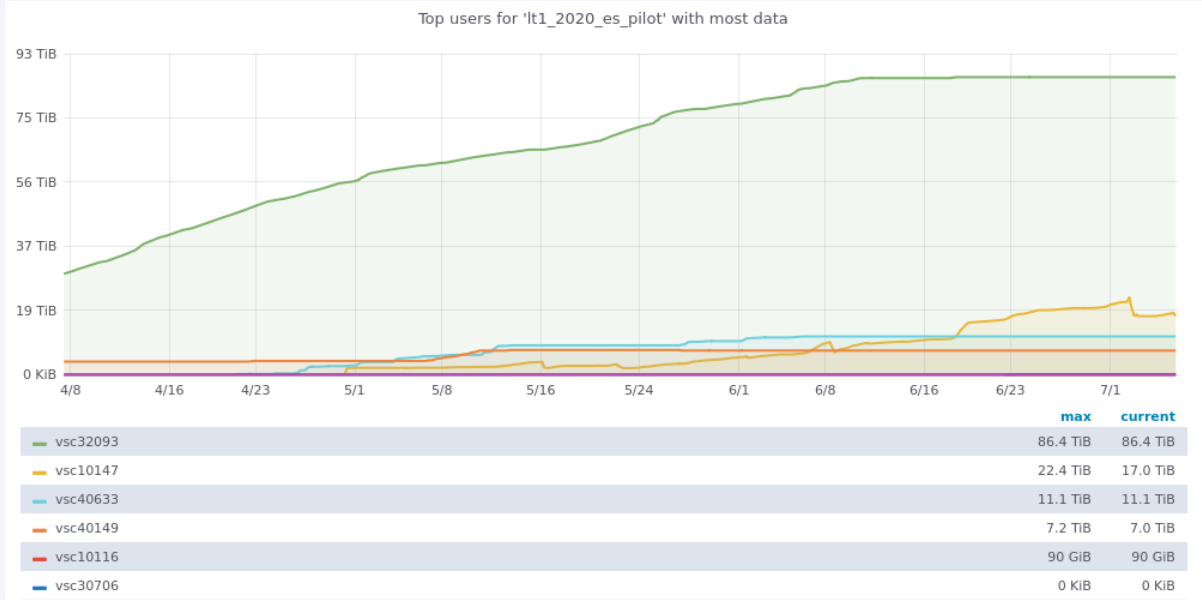
Tier-1 Compute: usage

Compute time



Total used: 2,899
Total available: 17,100

Scratch Storage



Total used: 127 TiB
Total available: 300 TiB

Working from the HPC systems

Start an iRODS session

From Tier-1

```
$> ssh irods.hpc.kuleuven.be | bash
You are now authenticated to irods. Your session is valid for 168 hours.
Connection to irods.hpc.kuleuven.be closed.
```

From Tier-2

```
$> ssh irods.tier1.leuven.vsc | bash
You are now authenticated to irods. Your session is valid for 168 hours.
Connection to irods.tier1.leuven.vsc closed.
```

VSC Belnet connection will be used to transfer data from Tier-2 to the iRODS Resource Server

Working from other systems: authentication

- Get a temporary password:

  https://vsc-passwd.icts.kuleuven.be
- Login with your institutional account
- Click on Authorize
- Copy the password
- Password expire after some time