



Vlaanderen
is supercomputing

Using iRODS to build a research data management service in Flanders

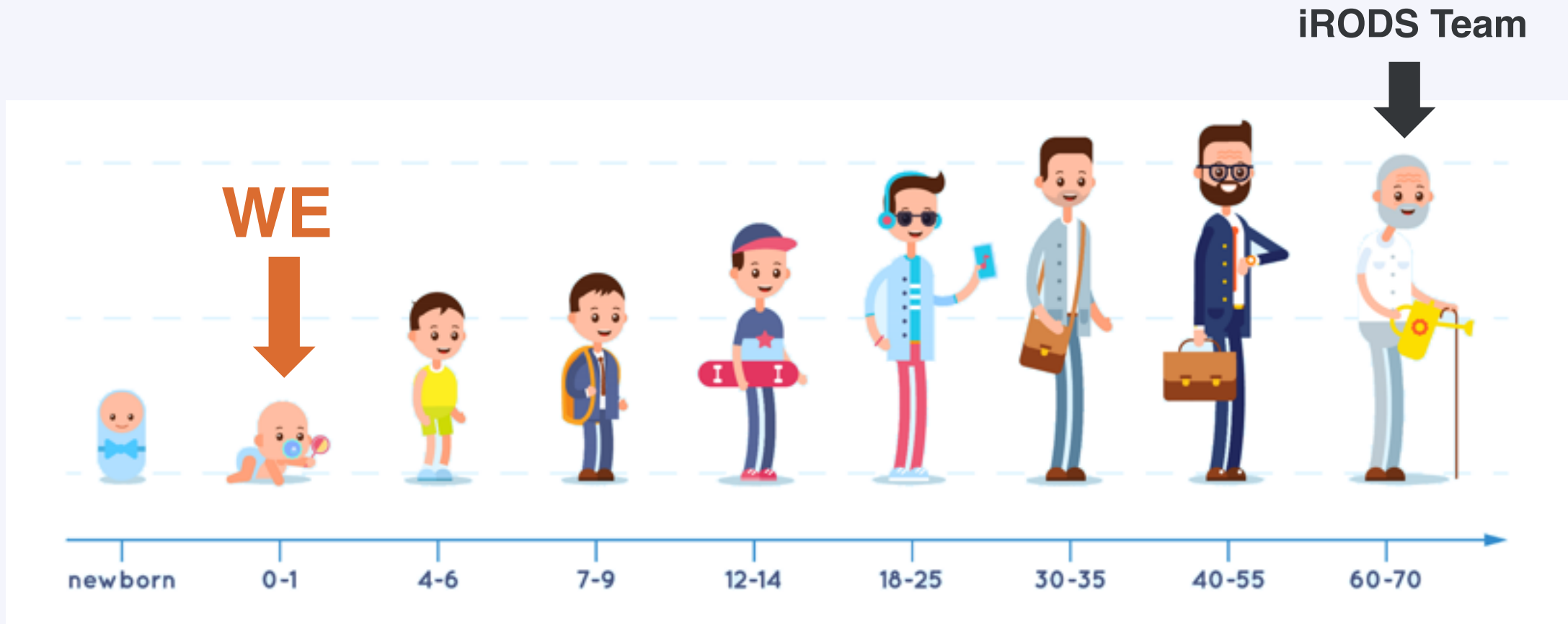
Ingrid Barcena, KU Leuven

iRODS User Group June, 10 2020

Agenda

- The Flemish Supercomputing Center (VSC)
- Tier-1 Data project
 - What
 - Why
 - How
- Pilot projects
- Future plans

iRODS user evolution



Months since first iRODS installation ...



The Flemish Supercomputing Center

VSC PARTNERSHIP



VSC HPC landscape



But ...



Open science

Collaboration

Data management

Tier-1 Compute



SuSE

Windows

Interactive work

Workflow managers

Kubernetes

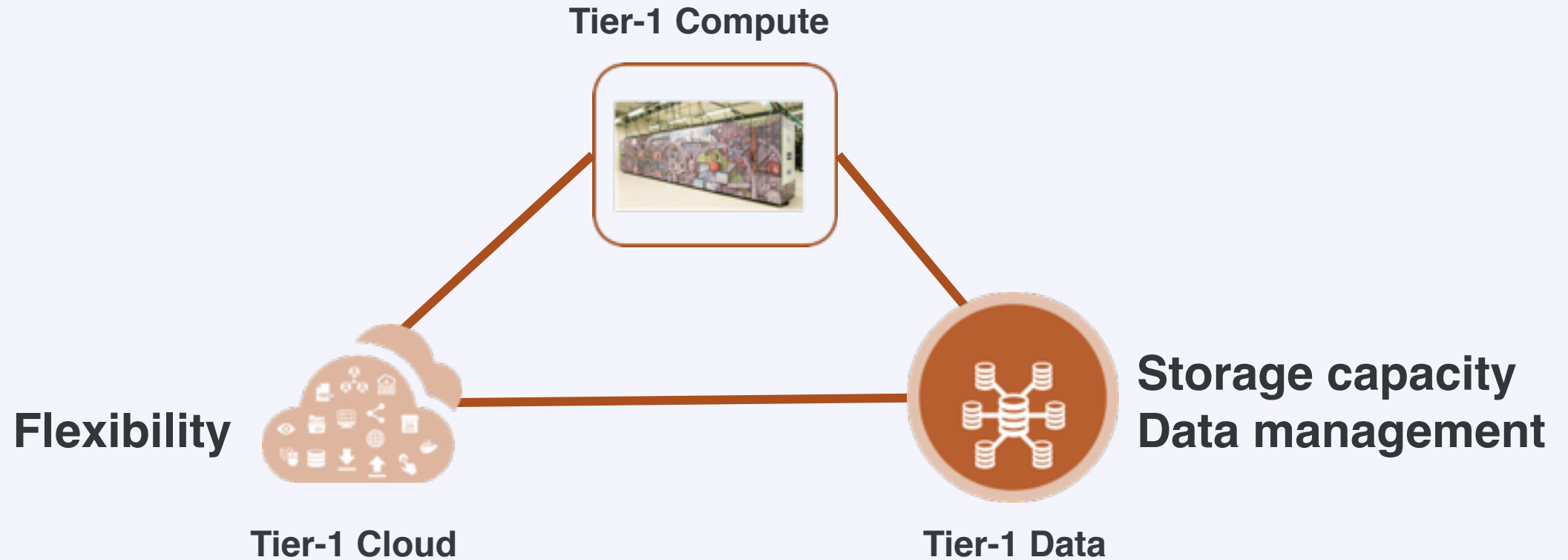
Containers

Portals

Databases

Ubuntu

Tier-1 Supercomputing as a Service





Tier-1 Data service

Tier-1 Data



Research Data repository with data management features

RESOURCES



WHAT



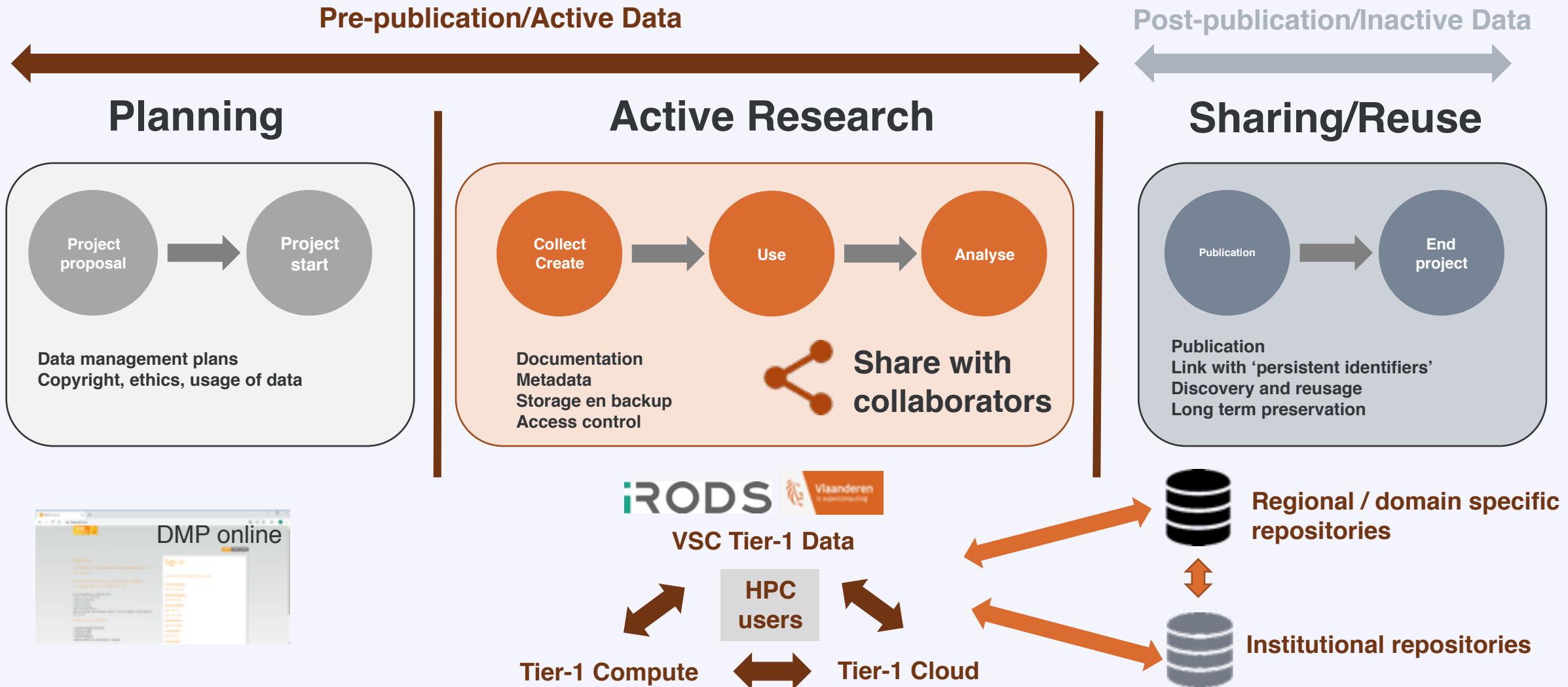
- ✓ ACTIVE DATA
- ✓ MANAGE DATA
- ✓ SHARING WITH OTHERS
- ✓ SEARCH AND DISCOVER

ADDED VALUE

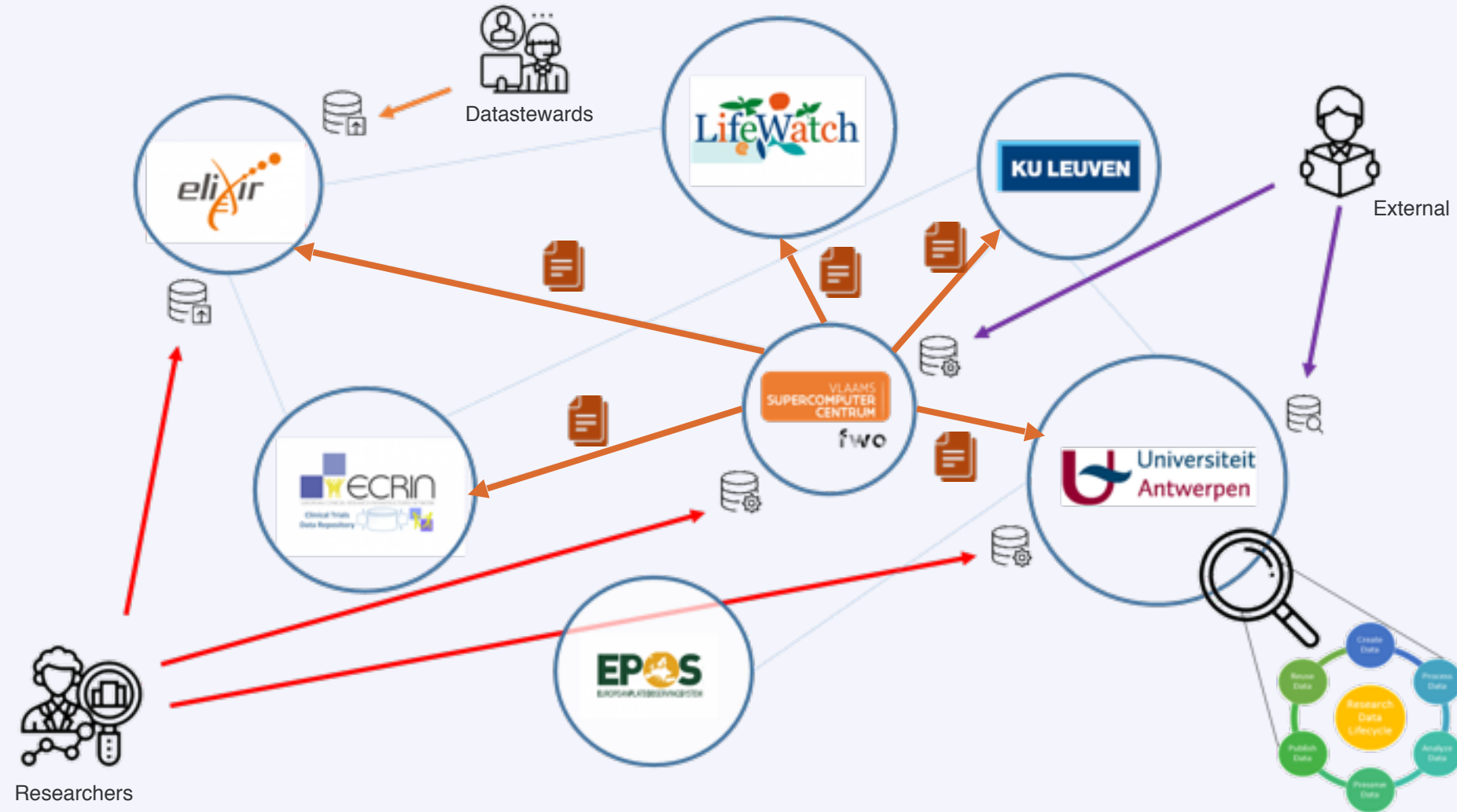


- ✓ STANDARDS
- ✓ BEST PRACTICES
- ✓ AUTOMATION

Tier-1 Data in the research Data Lifecycle



The RDM landscape



The challenge

2019

200 users
70 Projects
300,000 jobs



2020

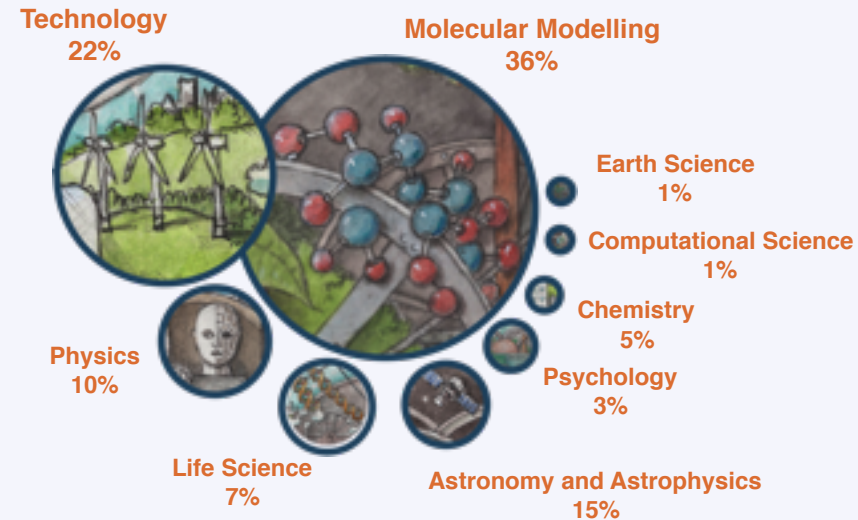


+



Leuven

Gent



Diversity!!!



Scientific domains



Data types



Data sizes



Scientific domains

The challenge

2019

Diversity!!!

"Data platforms that work for every use case, exists only in marketing brochure"
Nirav Merchant



Leuven

+



Gent



Scientific domains

Divide and conquer!

General use

Base configuration
Tools and libraries
General Policies
Metadata standards
Performance

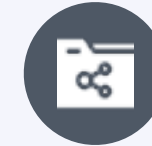
Domain Specific

Domain Policies
User Metadata
Tools and libraries

Pilot selection



Collaboration



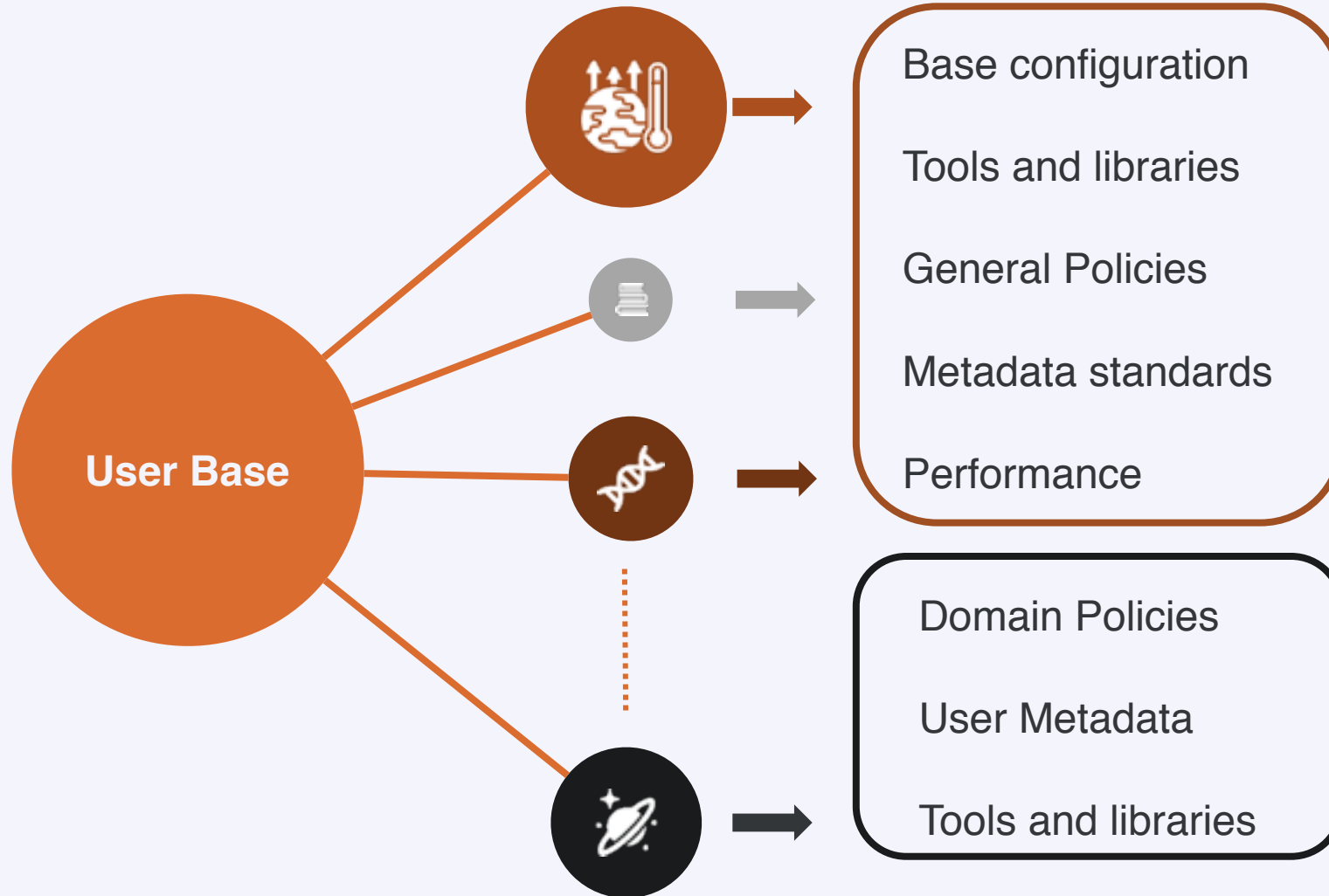
Shared datasets



Shared workflows



Community impact



Goal of the pilots



Test iRODS configuration and capabilities in a real environment



Gain experience on the development of iRODS rules for a real use case

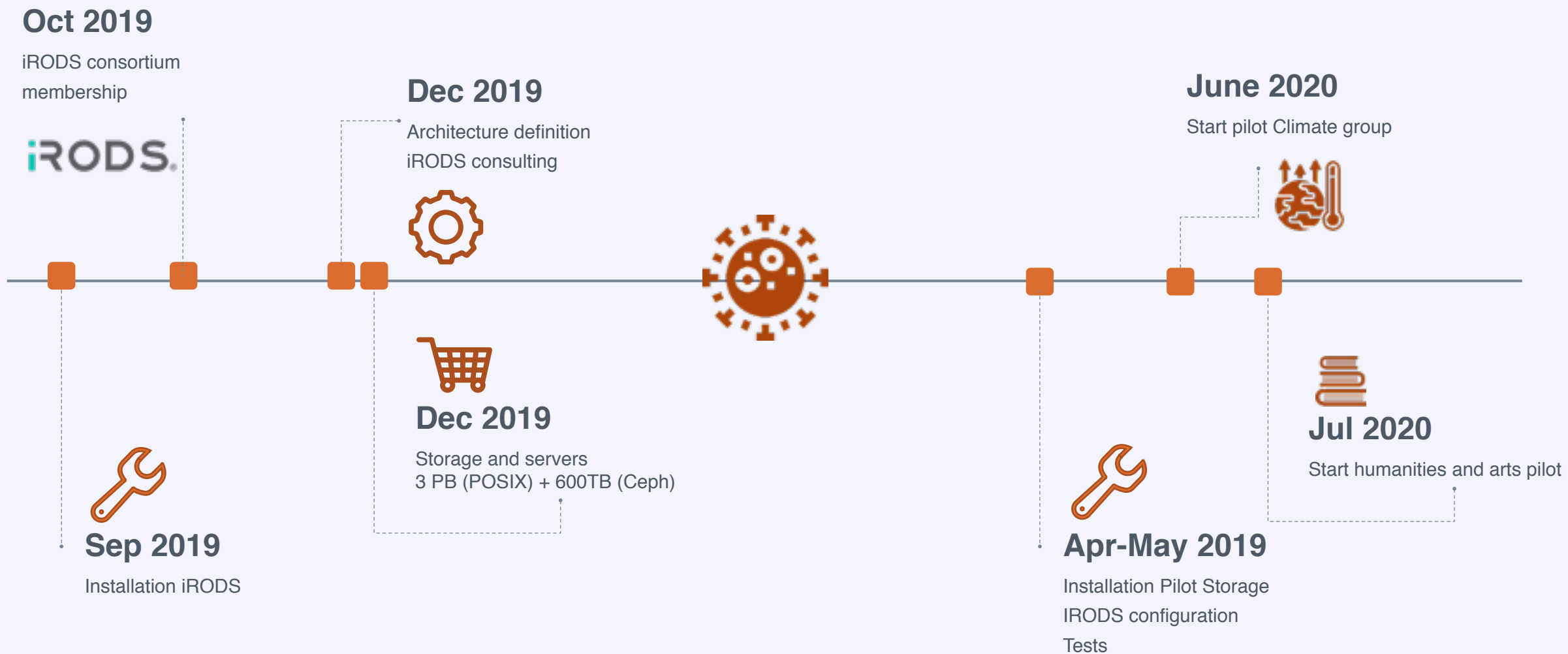


Implement and experiment data management procedures



Test interoperability with other repositories and services

Where are we?



Tier-1 Data architecture



VSC Tier-1



VSC Tier-2



auha antwerp university association

ASSOCIATIE KU LEUVEN

universitaire associatie BRUSSEL

ASSOCIATIE UNIVERSITEIT GENT

User Clients



YODA



DAVRods

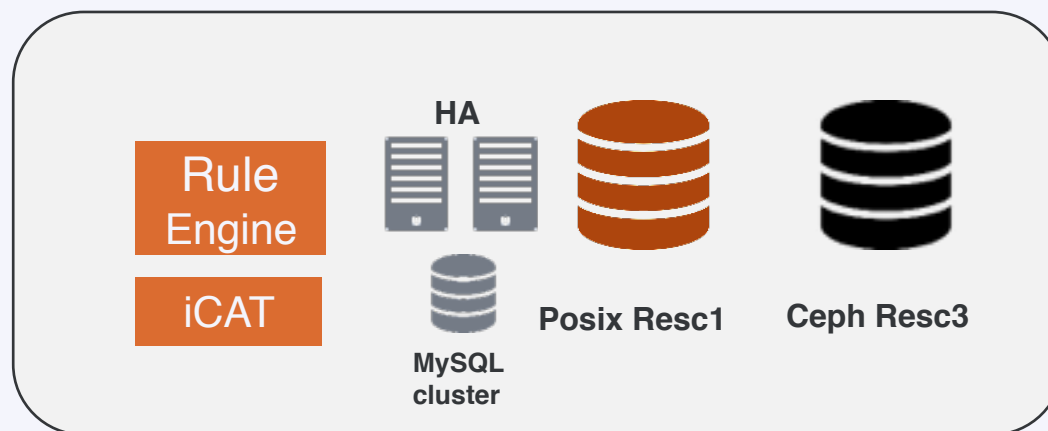


iCommands

iCommands PRC-Python API



VSC iRODS 1 zone



DC Heverlee

3km



DC Leuven

Use case: Earth Science pilot



Hydrology and Climate team (HCG)

Prof. Dr. Diego Miralles

VITO

Dr. Hendrik Wouters



Remote Sensing and Data Assimilation (RSDA)

Gabrielle De Lannoy

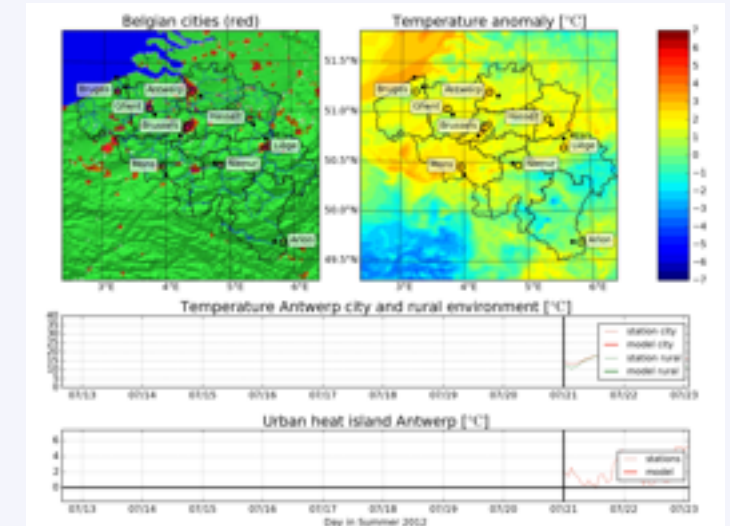
Regional Climate Studies (RCS)

Nicole van Lipzig



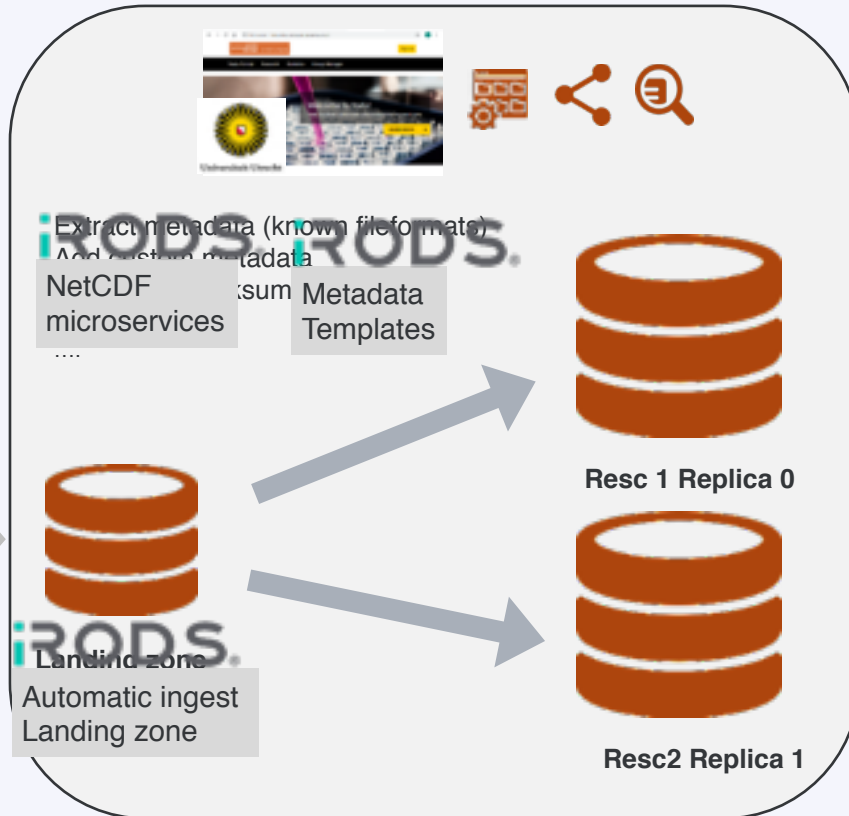
BCLIMATE Group

Prof. Dr. Wim Thiery



Earth Science pilot: workflow

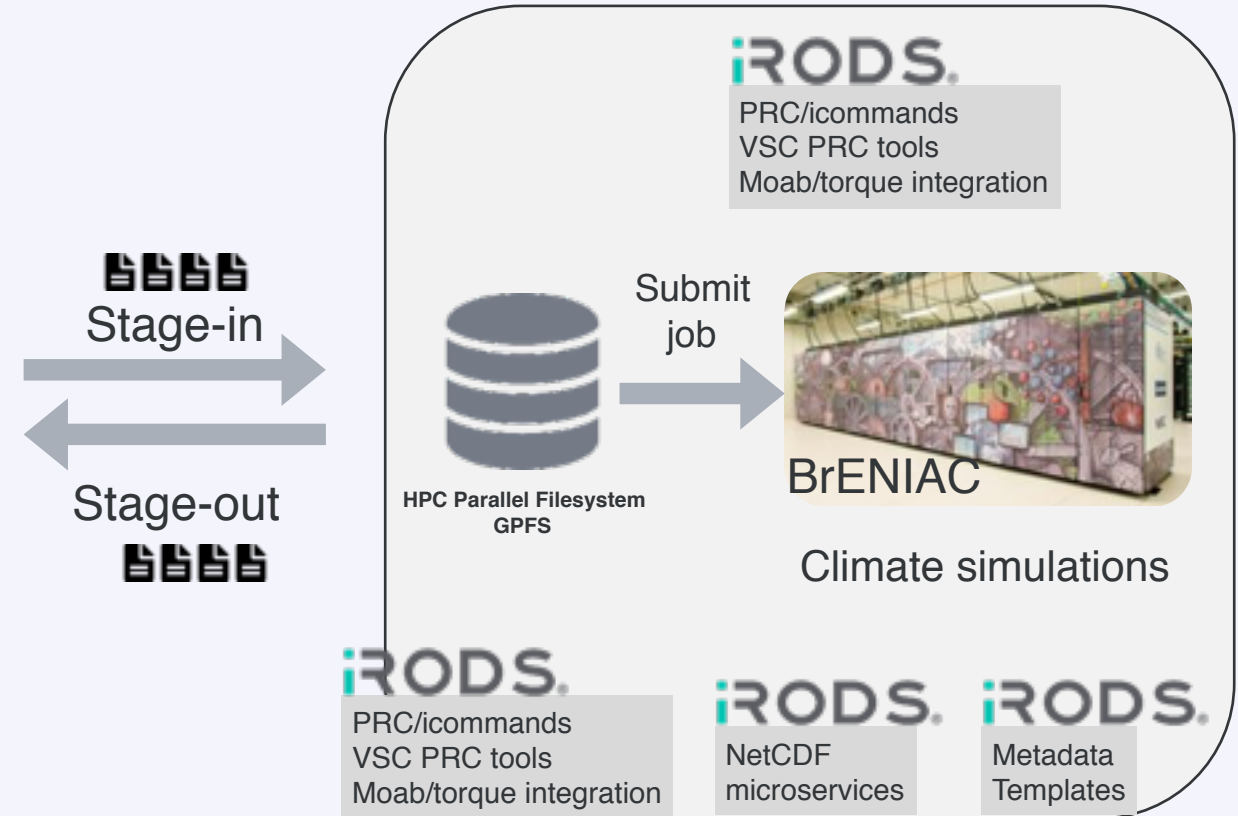
Tier-1 Data



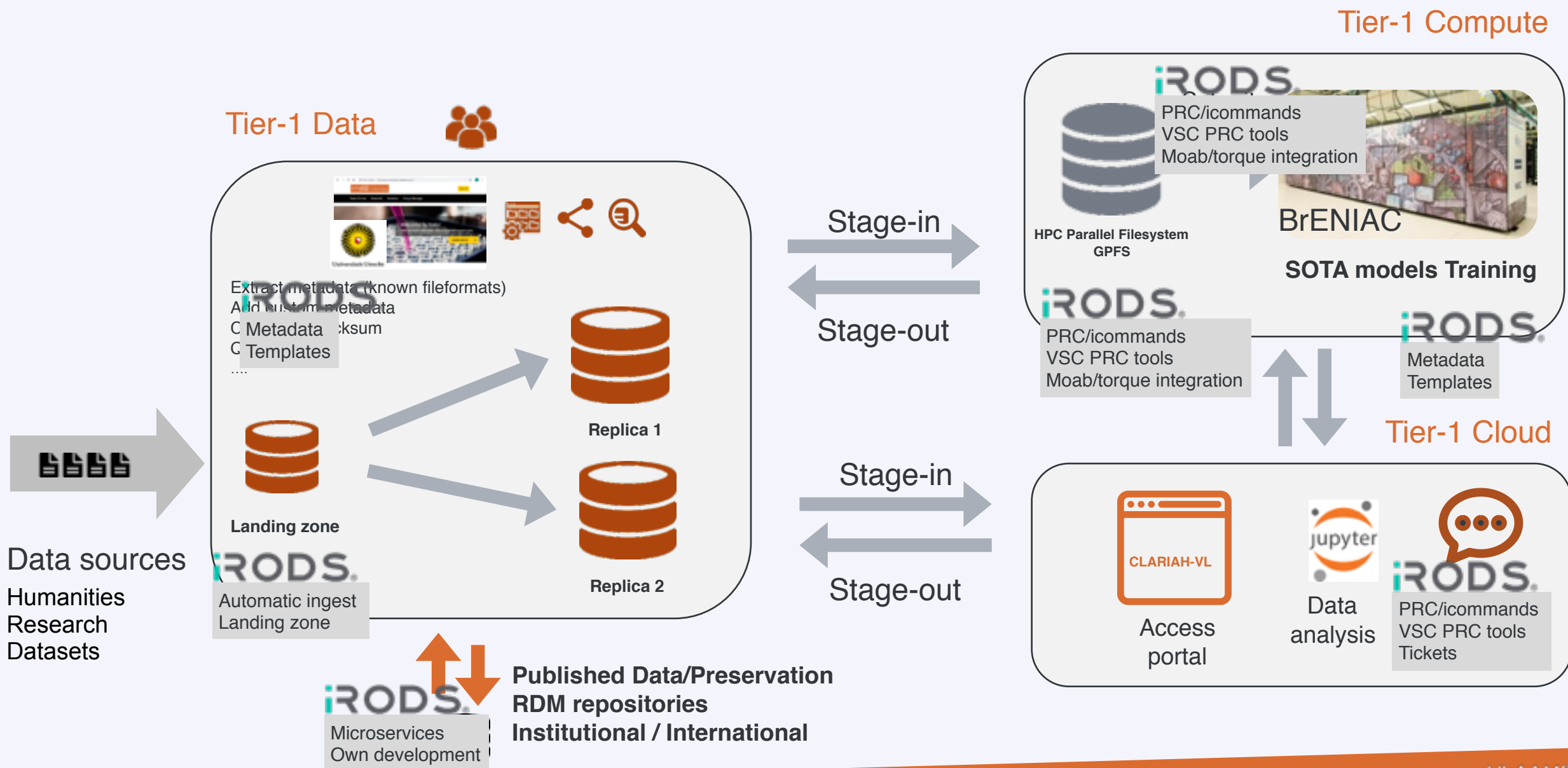
Data sources

- CMPI6
- CORDEX
- LUMIP
- ...

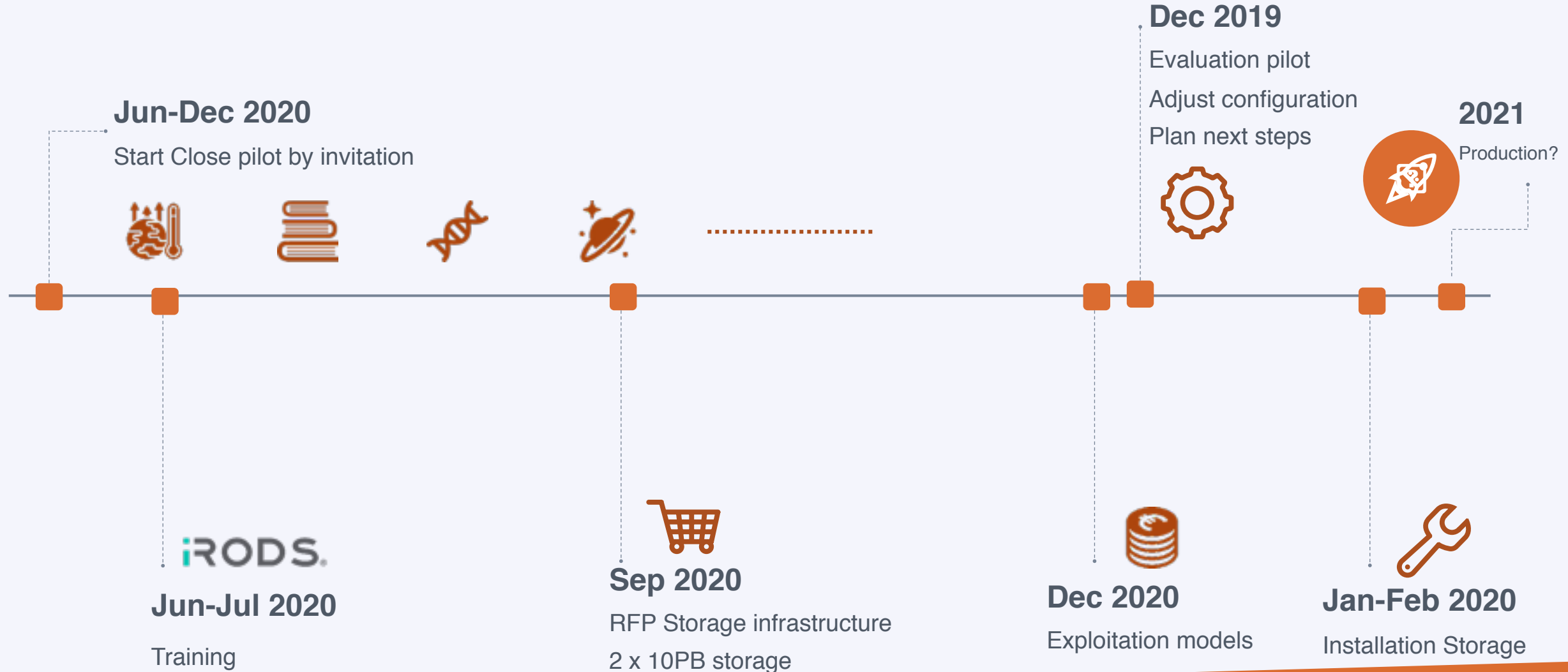
Tier-1 Compute



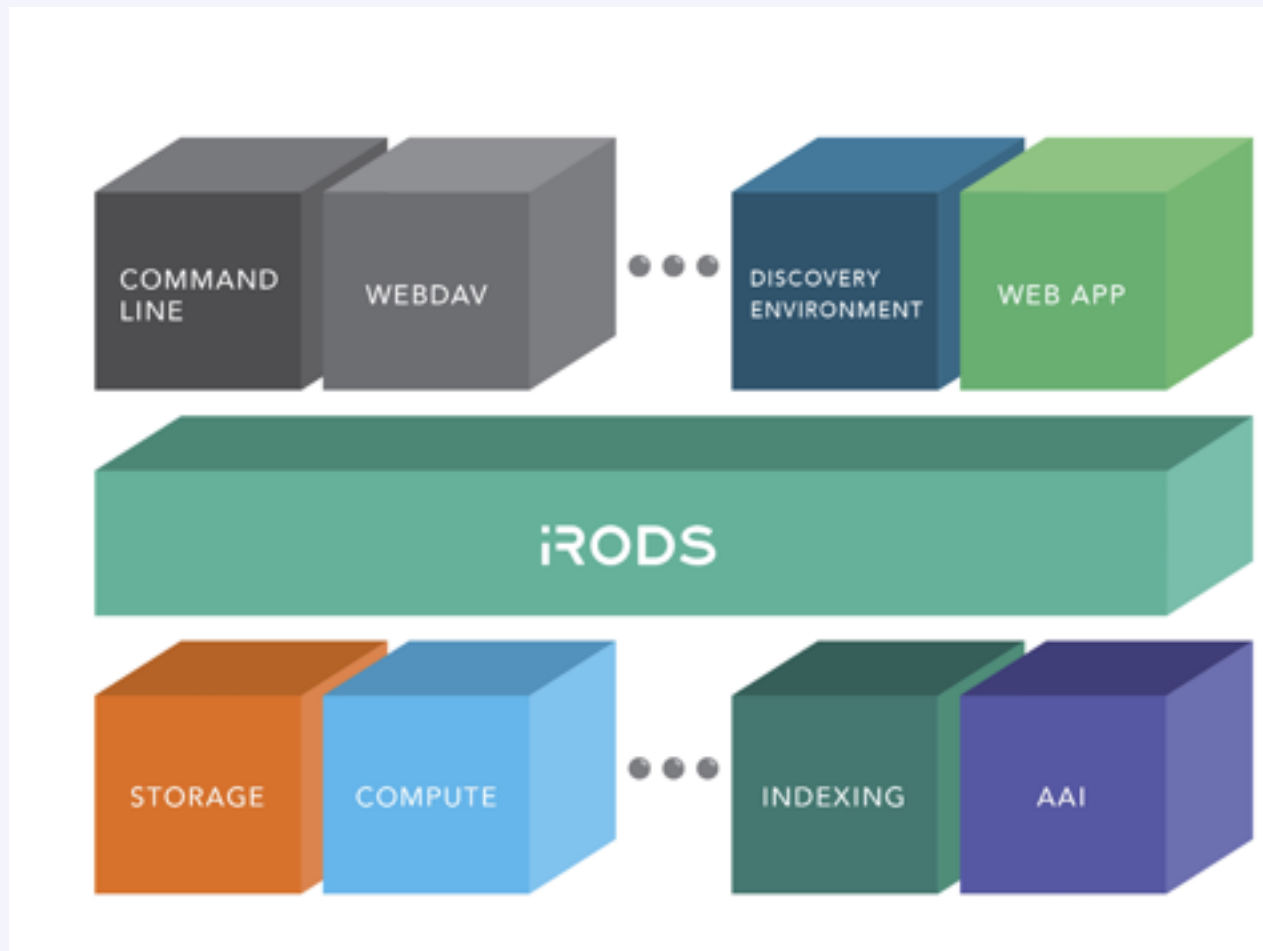
Use Case: Humanities, social sciences and arts (CLARIAH-VL)



Where we want to go?



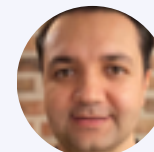
Tier-1 Data: the team



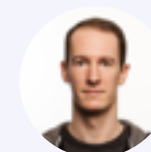
~ 4 VTE



Paul
Borgermans



Mustafa
Dikmen



Maxime
Van den Bossche

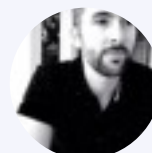


User Support
User tools
Policies
Documentation

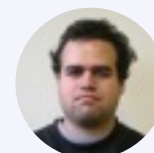
iRODS middleware



Peter
Verraedt



Tom
Vanmierlo



Kristoff
Van Buggenhout



Bart
Vanneste

Storage



Thank you!