Tier-1 Data

Contents

VLAAMS
SUPERCOMPUTER
CENTRUM

Tier-1 Data in the VSC workflow

Tier-1 Data

Storage capacity

Data management

Flexibility

Compute power

Tier-1 Cloud

Tier-1 Compute

ASSOCIATIE
KU LEUVEN

ASSOCIATIE
UNIVERSITEIT GENT

ASSOCIATIE
UNIVERSITEIT GENT

VLAAMS
SUPERCOMPUTER
CENTRUM
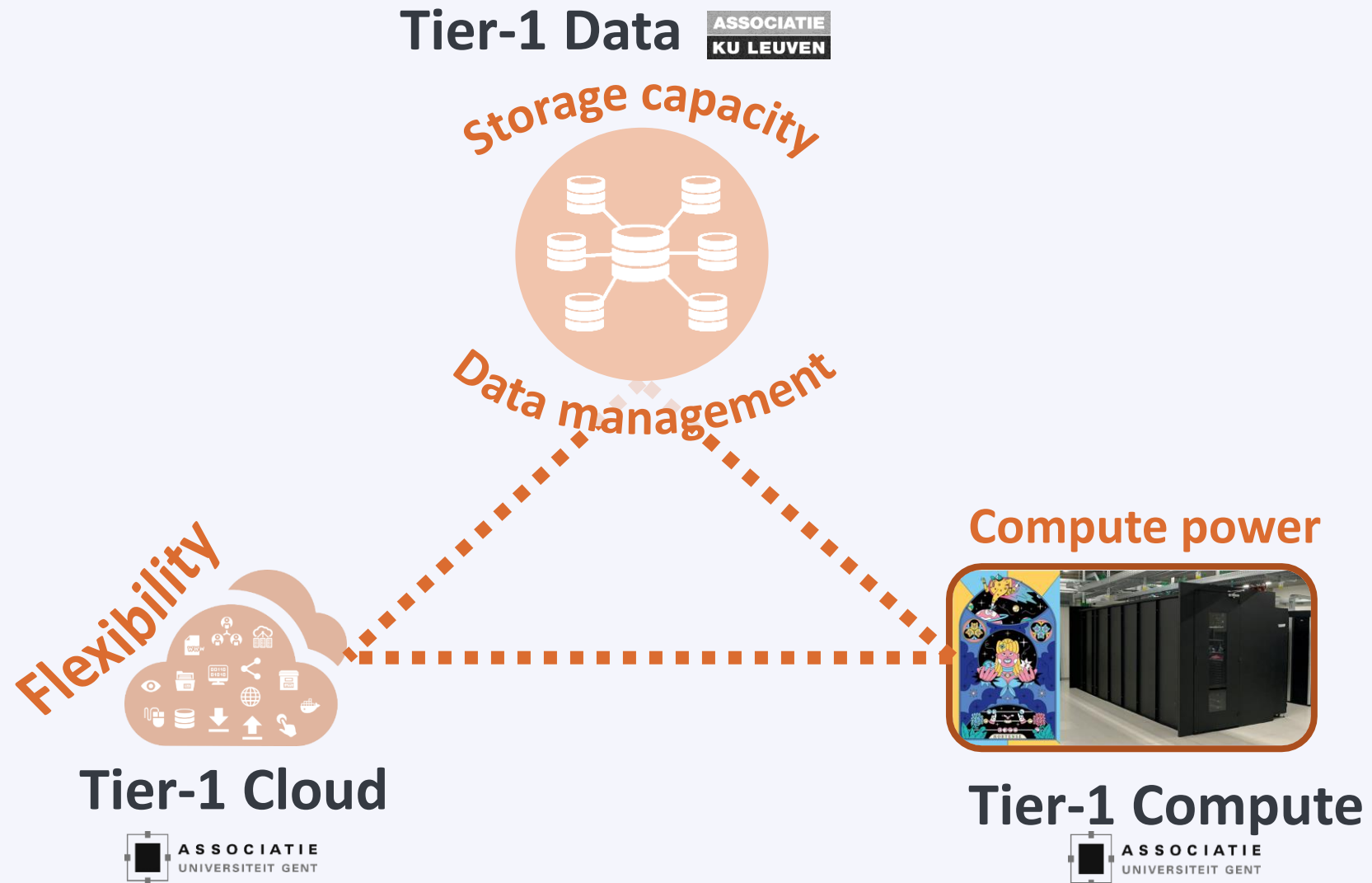
Tier-1 Data in the research data lifecycle

ACTIVE

Data collection

Data storage

Data processing

Metadata attribution

Tier-1 Data

PUBLISHED

Data publishing

Metadata publishing

LONG TERM

Data archiving

Data preservation

VLAAMS
SUPERCOMPUTER
CENTRUM
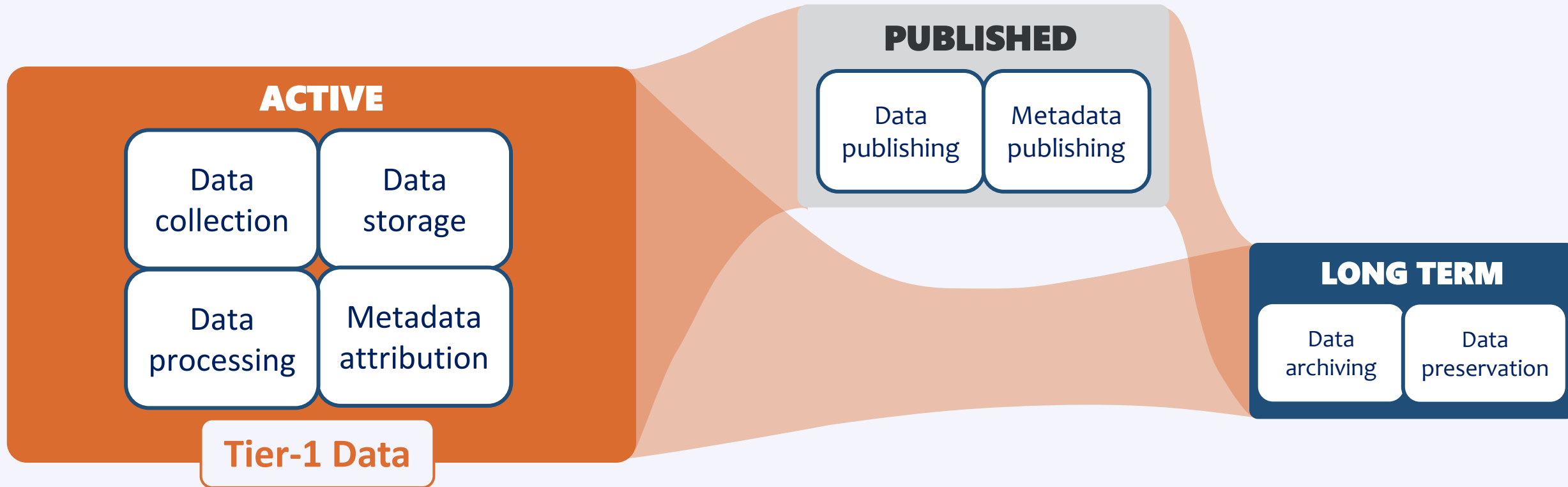
What is Tier-1 Data?

Management of **ACTIVE** research data

Based on open-source software **iRODS**

More than storage (metadata, automation, sharing)

Accessible via various clients

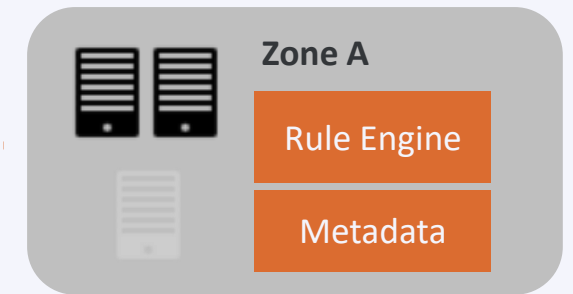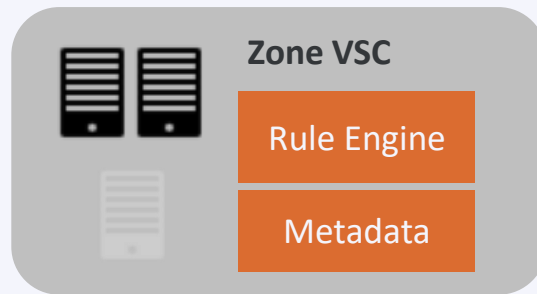To help researchers **manage** their data according to best practices

VLAAMS
**SUPERCOMPUTER
CENTRUM**
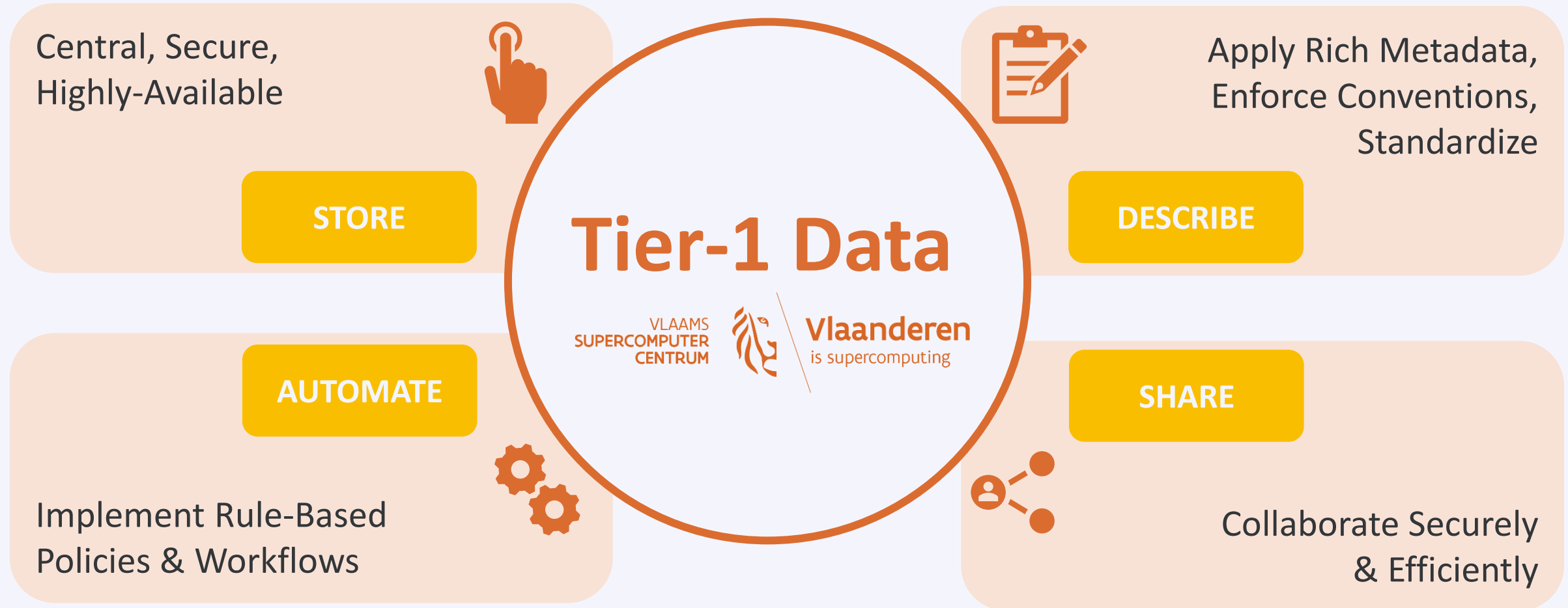
Architecture

Users   VSC-accounts

Clients   iCommands   PRC   ManGO Portal   Globus   SFTP

iRODS

Zone VSC

| Rule Engine |
| Metadata |

Zone A

| Rule Engine |
| Metadata |

Storage

27 PB (double copy)

The four pillars

Central, Secure,
Highly-Available

STORE

Apply Rich Metadata,
Enforce Conventions,
Standardize

DESCRIBE

Tier-1 Data

VLAAMS
SUPERCOMPUTER
CENTRUM

Vlaanderen
is supercomputing

AUTOMATE

SHARE

Implement Rule-Based
Policies & Workflows

Collaborate Securely
& Efficiently

VLAAMS
SUPERCOMPUTER
CENTRUM

Centralization & Accessibility

ManGO Portal
iCommands
PRC
Globus
SFTP

Tier-1 Data

Active data

Inactive data

Regional / domain specific repositories

Institutional repositories

VLAAMS
SUPERCOMPUTER
CENTRUM
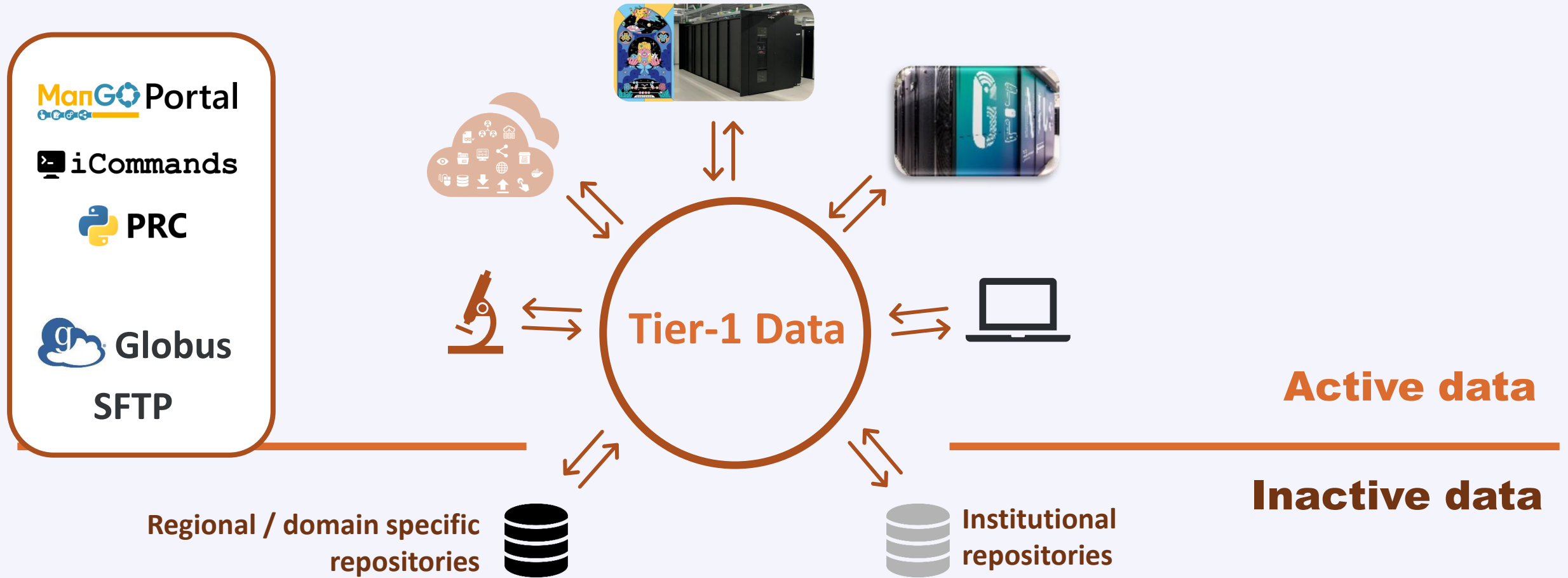
Sharing & Collaboration

Group-based

Access Levels:
1. Null
2. Read
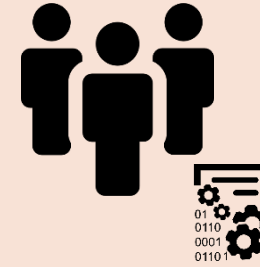3. Write
4. Own

@ Objects
@ Collections

Data creators

Create data
Upload data

📁 Raw Data

Data analysts

Search data
Analyze data
Upload results

📁 Raw Data

📁 Processed Data

Data consumers

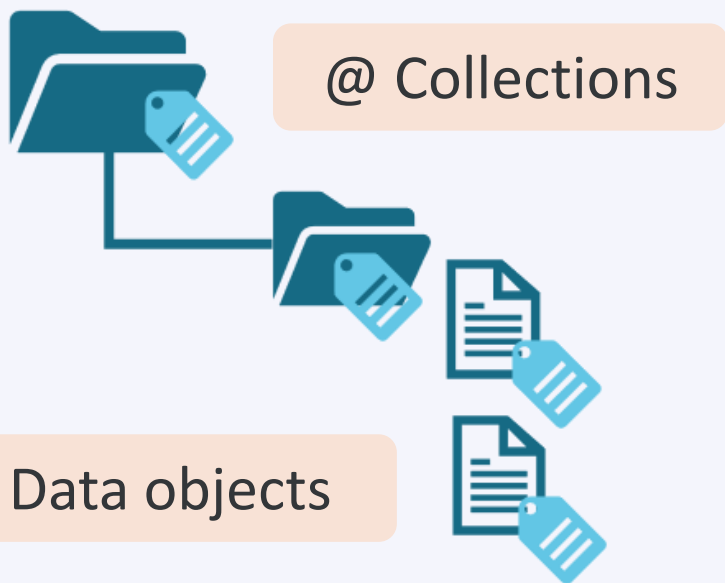Search data
Read data
Download data

📁 Processed Data

🟢 Write Access   🟣 Read Access

VLAAMS
SUPERCOMPUTER
CENTRUM

Metadata: Where & How

@ Collections

@ Data objects

MANUAL ⚠️

Add metadata

Name

Value

Unit

Cancel Add

MANUAL-GUIDED (METADATA SCHEMA'S)

Metadata schema: literature 2.0.0

Language*
Choose one of 7 options.
○ english
○ french
○ dutch
○ spanish
○ italian
○ japanese
○ chinese

Genre
Choose one of 3 options.
○ fiction
○ non-fiction
○ poetry

Save metadata

SEMI-AUTOMATIC (INSPECT & IMPORT)

Tika

AUTOMATIC (SYSTEM OR USER RULES)

Automation

	Client-side	**Server-side**
Language	iRODS Python Client	Integrated Policy Engine
How	Triggered by metadata Manual execution Cronjobs	In consultation with support
Examples	▪ Add metadata to a batch ▪ Check data status and take an action ▪ Pre-process data after acquisition from HPC	▪ Restrict access ▪ Log for audit and reporting ▪ Send notification ▪ Execute a process on the file

Tier-1 Data clients

iCommands

```
$ icd myPath
$ imkdir test
$ itouch test/myfile
```

```python
# this file is used to query which data objects have no attached metadata
import os, os.path
import json
from irods.session import iRODSSession
from irods.column import Criterion
from irods.models import Collection, DataObject, DataObjectMeta

# Get environment file
env_file = os.getenv(
    "IRODS_ENVIRONMENT_FILE",
    os.path.expanduser("~/.irods/irods_environment_python_vsc.json"),
)

# Provide path to iRODs collection
iRODsPath = "/vsc/home/                              /%"

# provide the metadata field attribute that we are looking for
MDname = "mgs.image_attributes.%"

# establish session
session = iRODSSession(irods_env_file=env_file)

# list which of these files do not have metadata attached from the specif
lMDtrue = []
qMDtrue = (
    session.query(Collection.name, DataObject.name, DataObjectMeta.name)
    .filter(Criterion("like", Collection.name, iRODsPath))
    .filter(Criterion("like", DataObject.name, "%.czi"))
    .filter(Criterion("like", DataObjectMeta.name, MDname))
)
for item in qMDtrue:
    lMDtrue.append(f"{item[Collection.name]}/{item[DataObject.name]}")
```
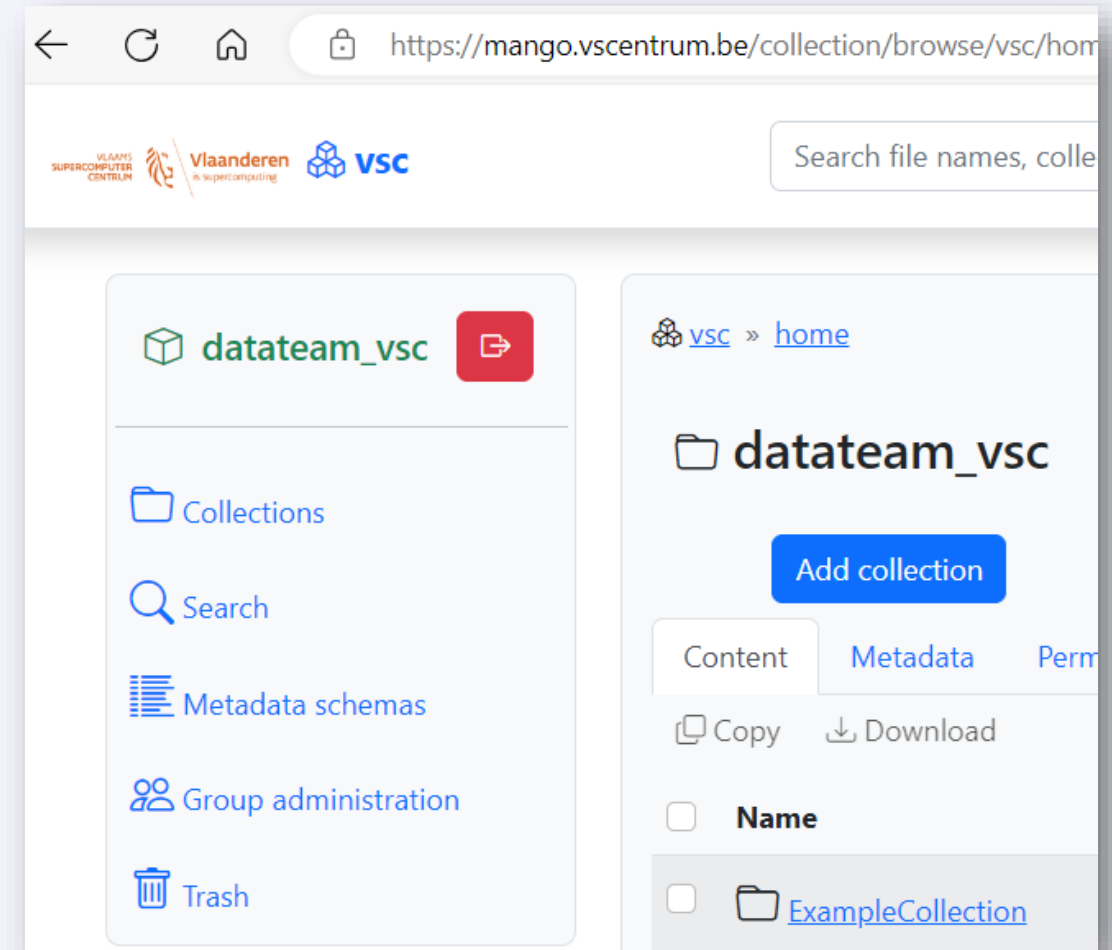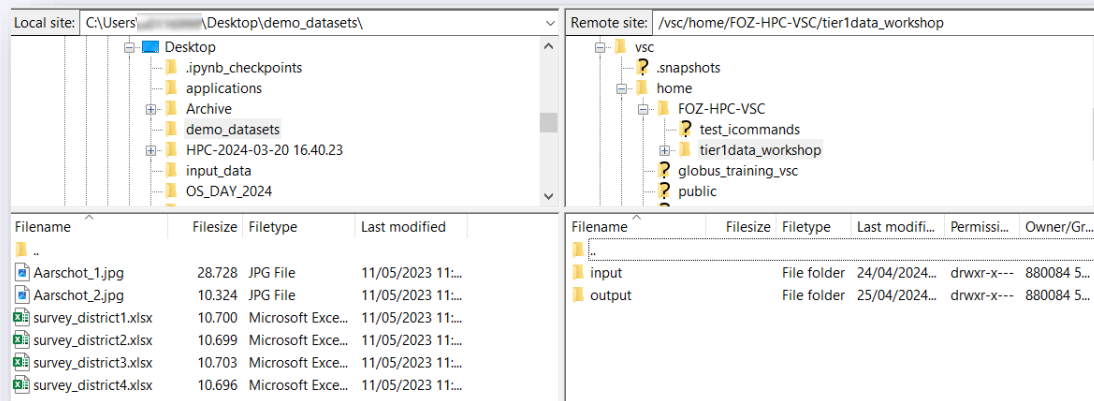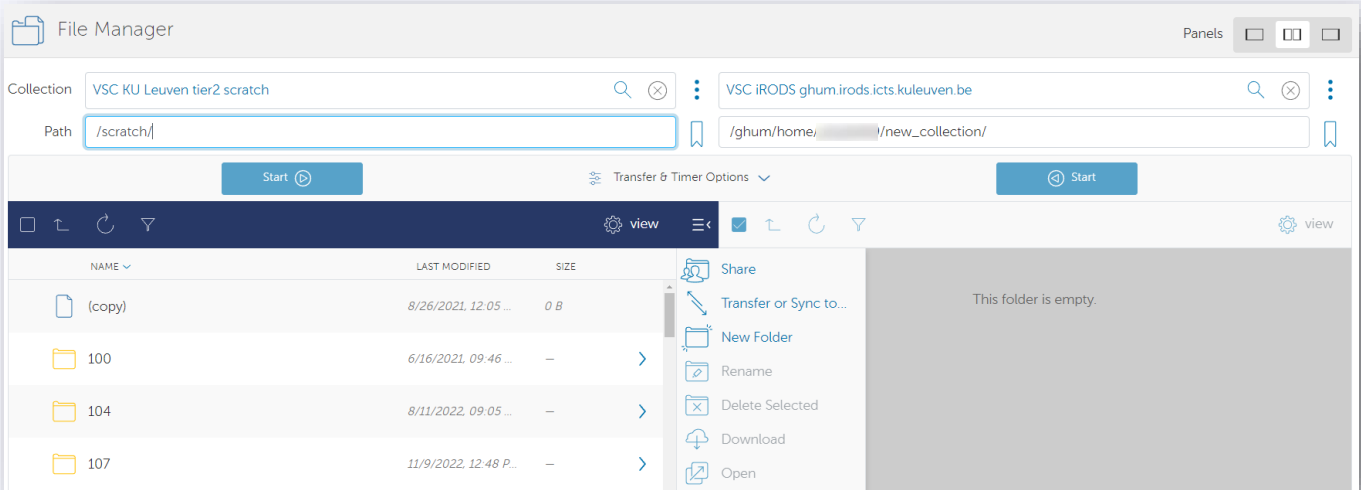
VLAAMS
SUPERCOMPUTER
CENTRUM

Tier-1 Data clients

SFTP

Globus

© Globus online

Authentication

INSTITUTIONAL ACCOUNT

PERSONAL ACCOUNT

VLAAMS
SUPERCOMPUTER
CENTRUM

Tier-1 Data overview

Data repository with data management feature for HPC related data

Resources

iRODS

globus online

What

- ACTIVE DATA
- MANAGE DATA
- SHARING WITH OTHERS
- SEARCH & DISCOVER

Added Value

- DATA LIFE CYCLE
- STANDARDS
- BEST PRACTICES
- AUTOMATION

VLAAMS
SUPERCOMPUTER CENTRUM
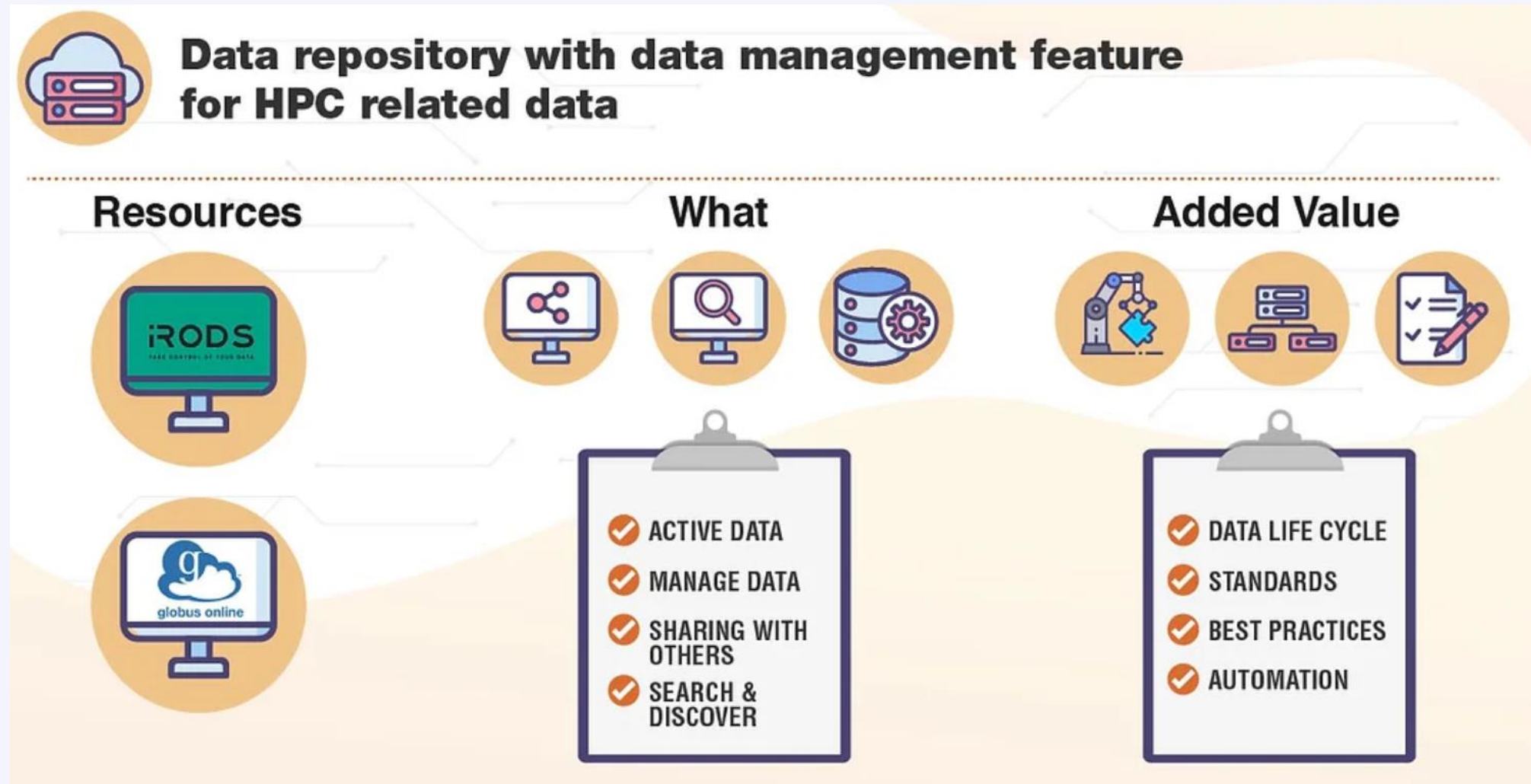
Realistic workflow example

Upload raw data

Organize
Describe

**Tier-1
Data**

Share

Publish

Query
Download

Describe

Upload

```
query data
download data
run some process
-------
-------###
SUCCESS
upload results
add metadata
rinse
repeat
```

Process
Analyze

The **Dataverse**®
Project

VLAAMS
SUPERCOMPUTER
CENTRUM
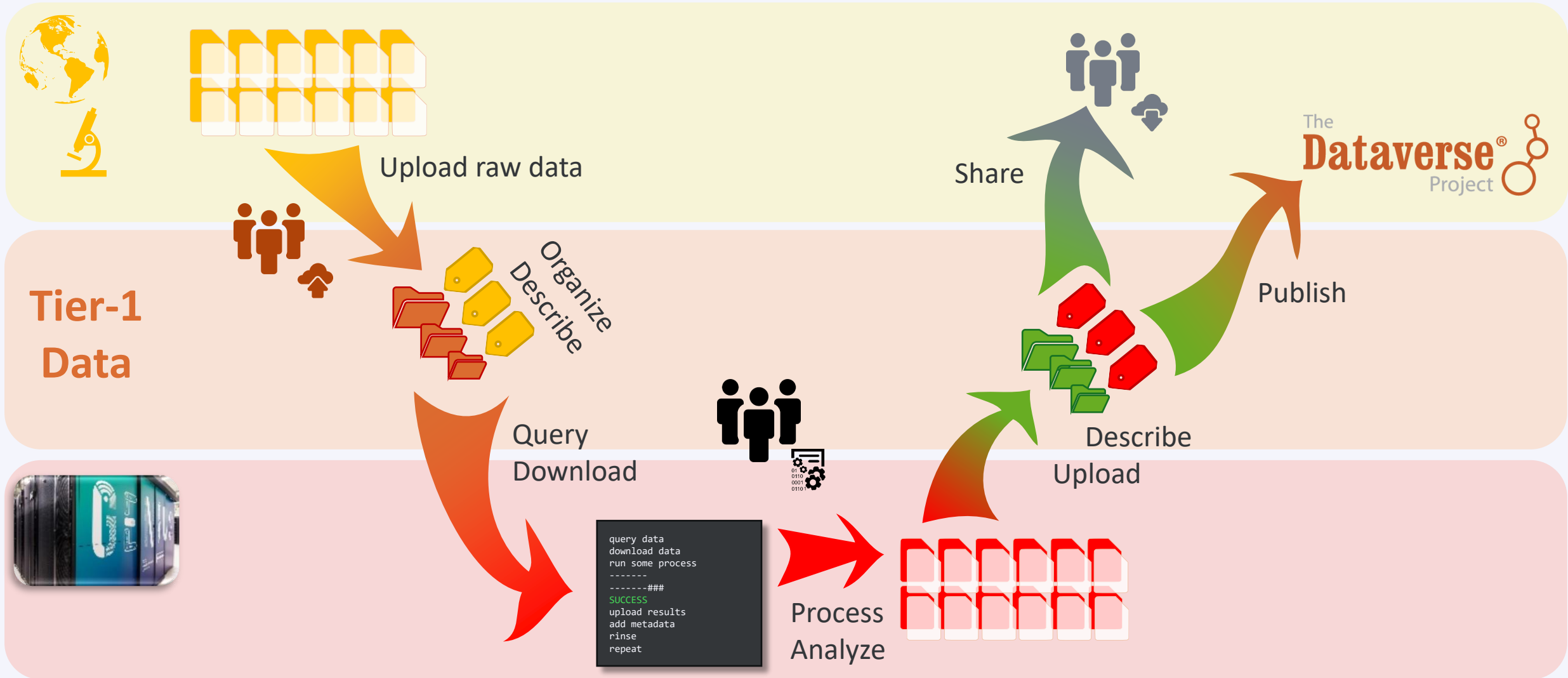
Practicalities: Get started & Get support

Requesting a Tier-1 Data project

- Project-based access
- Criteria:
    - Public Research Institution
    - Active data in the context of VSC: Tier-1/Tier-2 compute, Tier-1 Cloud
    - Judged on workflow feasibility and research data management aspects
- Types:
    - Starting grant: 5TB for 8 months
    - Full grant: 5TB-1PB
- Request at any time
- All info on https://www.vscentrum.be/data

User support
- Helpdesk: data@vscentrum.be
- Documentation: https://docs.vscentrum.be/data/tier1_data_service.html

VLAAMS
**SUPERCOMPUTER
CENTRUM**

Thank you!