Automating scientific dataset management and processing using Onedata

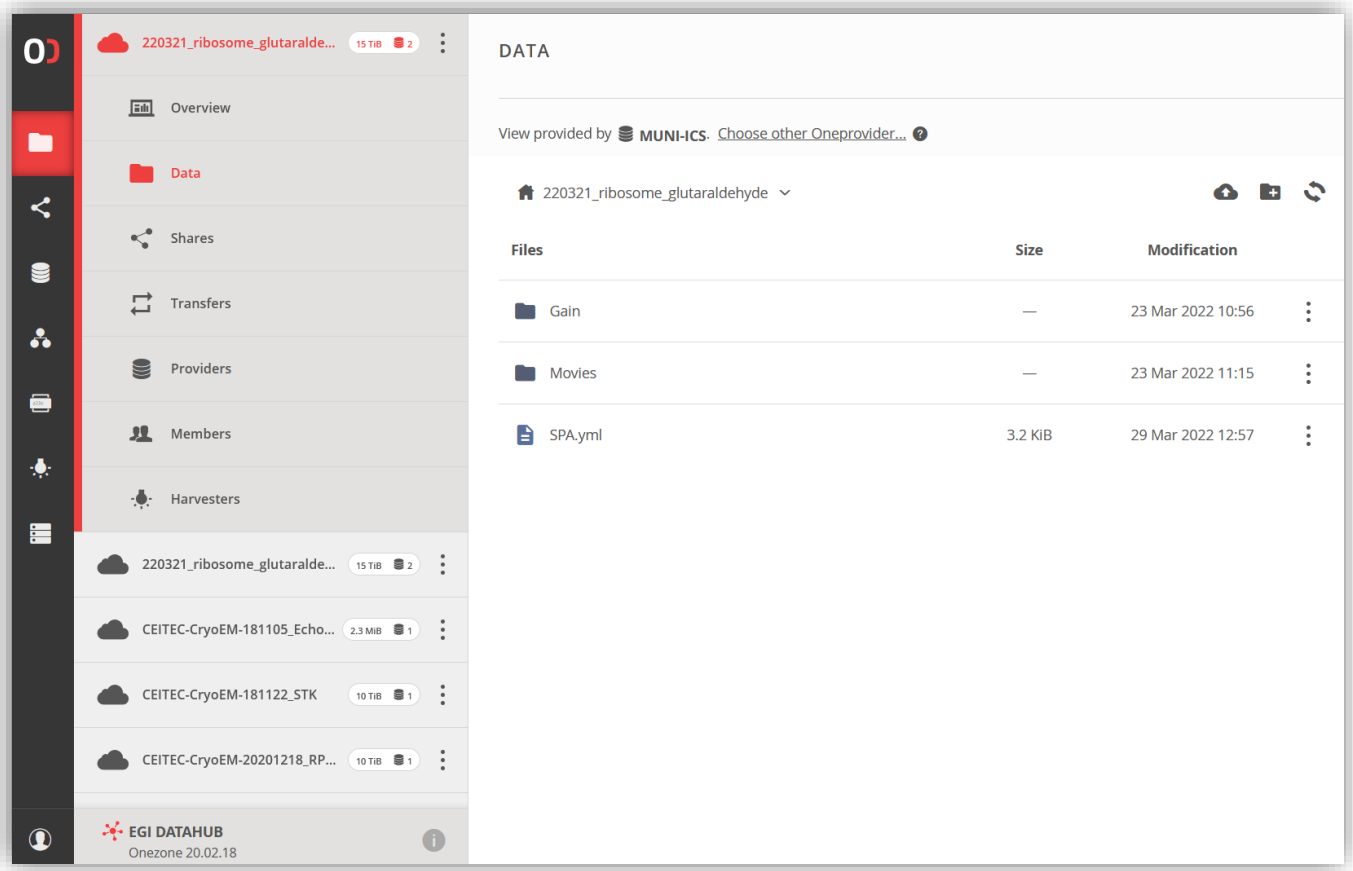
Tomáš Svoboda, Aleš Křenek, Josef Handl

CESNET & Masaryk University (Czechia)

IBERGRID 2022, Faro

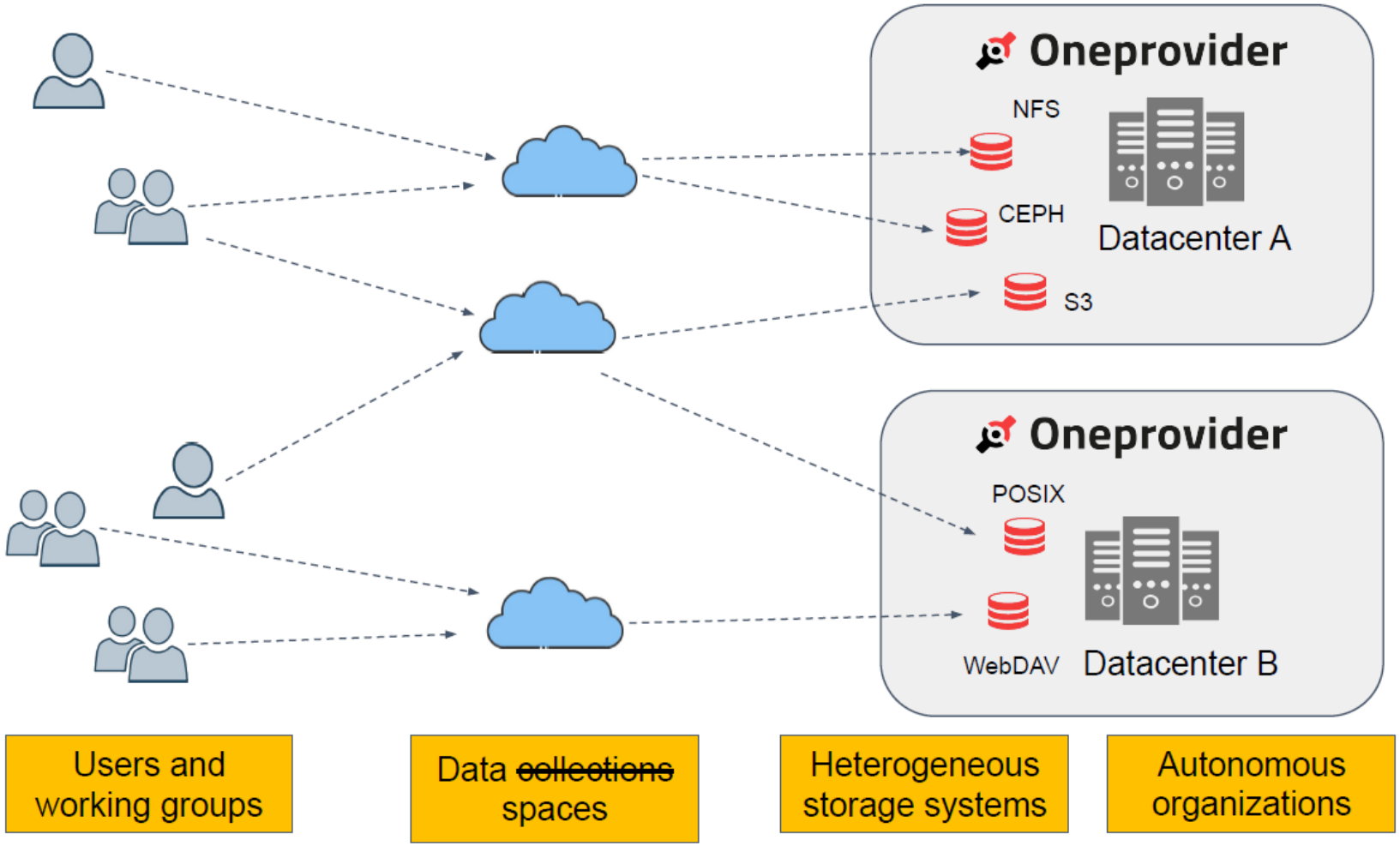
Onedata

- Global data access management system
  - storing, sharing and publishing data
- Provide access to data in a similar way which users know from widespread cloud solutions (Dropbox, …)
- Adapted for use in a scientific environment:
  - Compliant with FAIR recommendations
  - Ready to HPC
- Data accessible by several ways:
  - Web interface
  - Client application
  - REST API
- Multiple storage backends
  - POSIX, S3, Ceph, …



Onedata

Unifies access to geographically and organizationally distributed data



Source:
Opiola Lukasz, Onedata 101

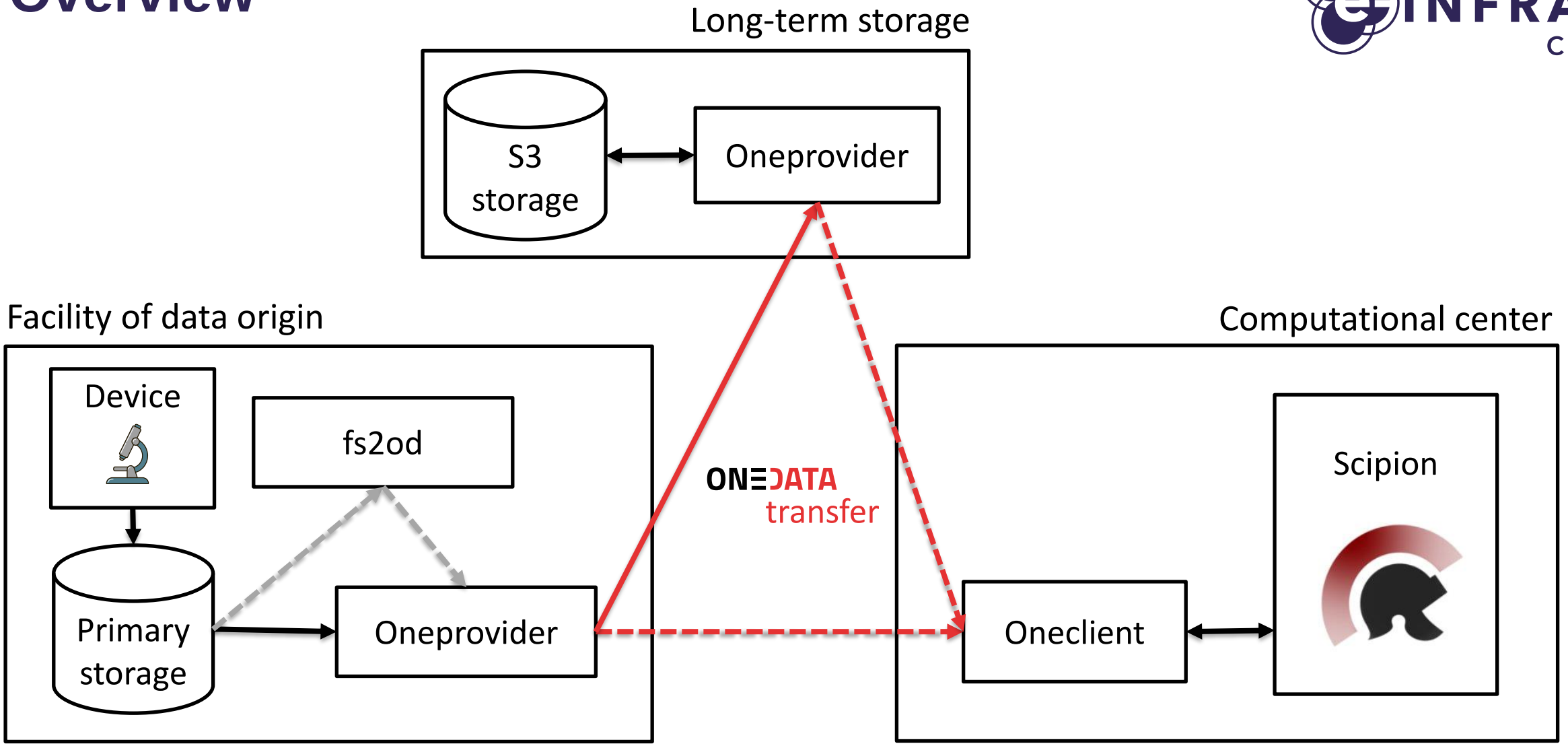
Introduction

fs2od

- Supporting software for automating managing scientific datasets
  - According to defined policies
- Developed on Cryo-EM use-case
- Consist of:
  - Data acquisition from specialized devices
  - Sharing datasets to users
  - Efficient access to computational processing
  - Replicating to long-term storages
  - TODO: Exporting to existing repositories

- Documentation - fs2od.readthedocs.io

Overview


Long-term storage
S3 storage
Oneprovider
Facility of data origin
Computational center
Device
fs2od
ONEDATA
transfer
Scipion
Primary storage
Oneprovider
Oneclient

Demonstration