HELMHOLTZ

RESEARCH FOR
GRAND CHALLENGES

HIFIS | HELMHOLTZ FEDERATED IT SERVICES

Integrated dataset placement service for scientists

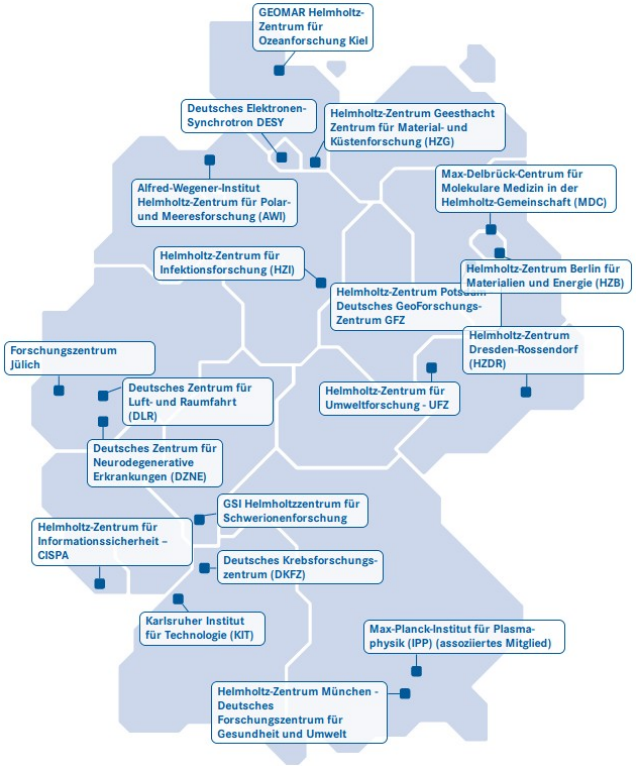
Tim Wetzel, Michael Schuh, Patrick Fuhrmann,
Thomas Beermann, Paul Millar, Johannes Reppin,
Muhammad Aleem Sarwar

@ Deutsches Elektronen-Synchrotron DESY on behalf of HIFIS

IBERGRID2022, Oct 12th, 2022

Helmholtz Association

- Founded in **1995** to formalise relationships between research centres
- **Members**: 19 autonomous research centres in Germany
- **Mission**: Contributions to grand challenges facing society, science and industry



- Increasing importance of cloud access to **common data treasure** and -services
- Rapidly growing **data exchange** from research instruments requires excellent data networking
- Growing connections between Helmholtz, **EOSC** and **FAIR**

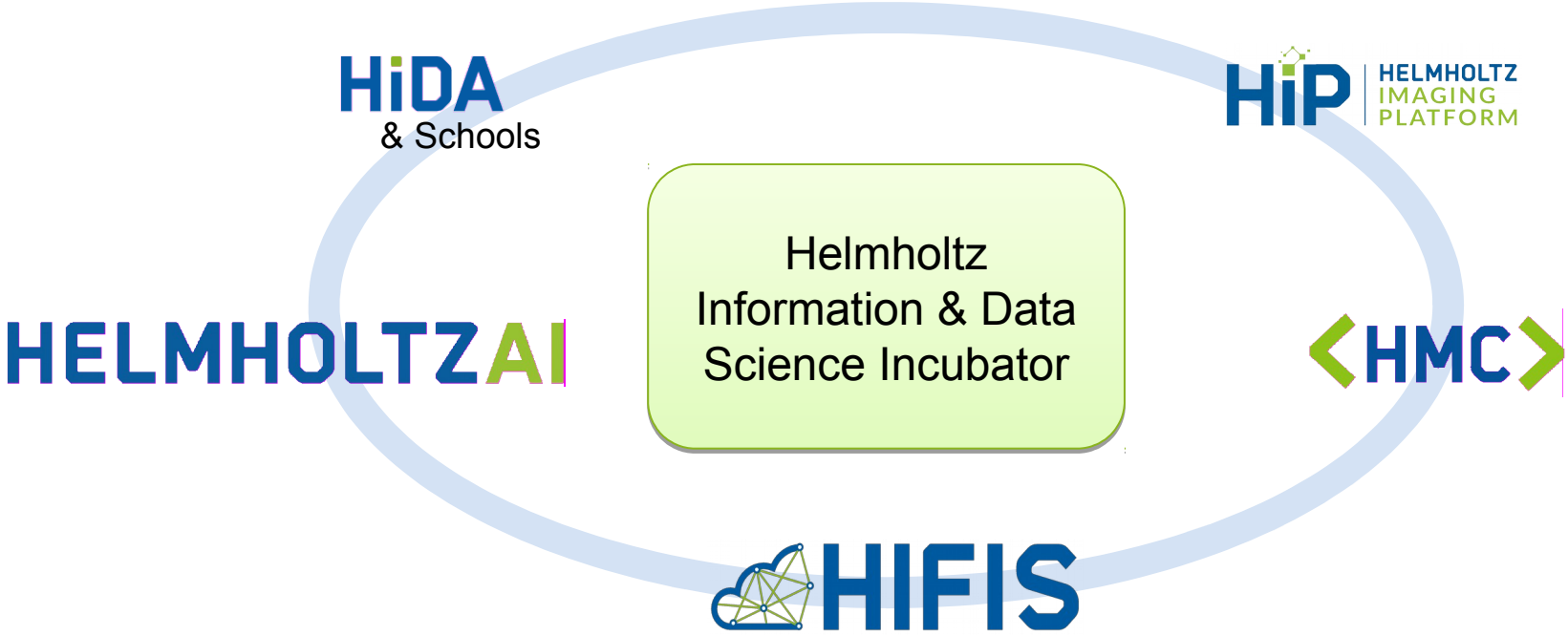
Helmholtz Association

- Founded in **1995** to formalise relationships between research centres
- **Members**: 19 autonomous research centres in Germany
- **Mission**: Contributions to grand challenges facing society, science and industry

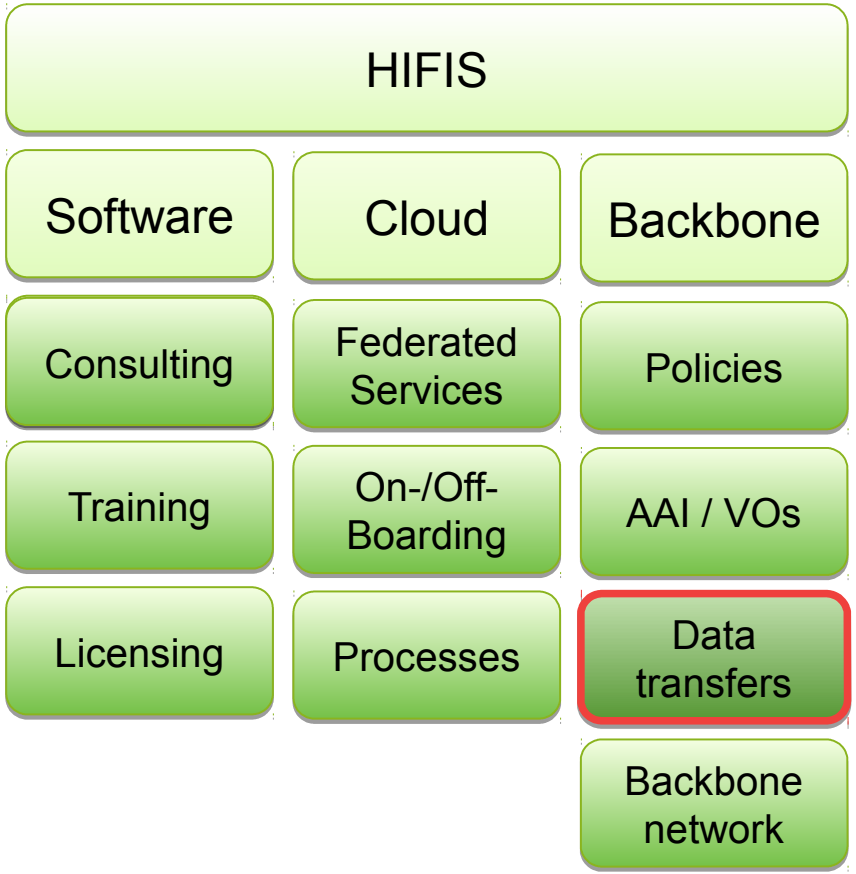
- Increasing importance of cloud access to **common data treasure** and -services
- Rapidly growing **data exchange** from research instruments requires excellent data networking
- Growing connections between Helmholtz, **EOSC** and **FAIR**

Helmholtz Incubator

Helmholtz aims for **joint research & information environment** for all Research Fields



HIFIS – Helmholtz Federated IT Services



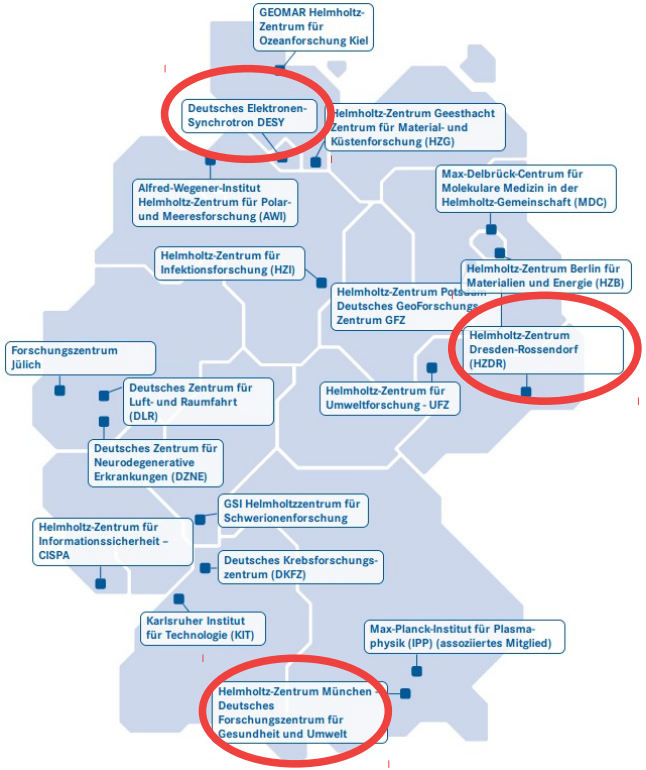
- Joint use of IT services provided by Helmholtz centres

- Exploitation of **synergy effects** by reducing the service provisioning efforts for each centre
- simplifying **collaboration processes** for scientists
- Team of **~120 people** from 11 Helmholtz centres
- **support and coordination platform** for federation in Helmholtz
- Close connections with other **incubator platforms** and all centres

Why do we need data transfers?

- Large data sets in **collaborative** research projects
- Data analysis often sensitive to **latency**
- **Data locality** is important!
- **HI** and **HelmholtzAI** projects use data storage at the computing sites and their own locations
- Collaborating centres **distributed** over Germany

➔ Reliable, comfortable and robust transfer endpoints and service needed

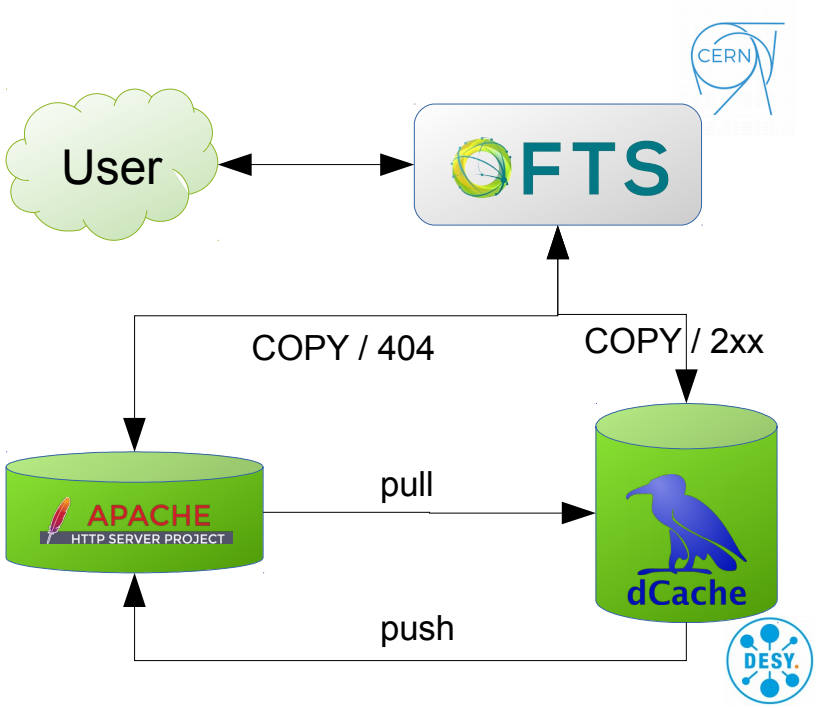


HIFIS Transfer Service (HTS)

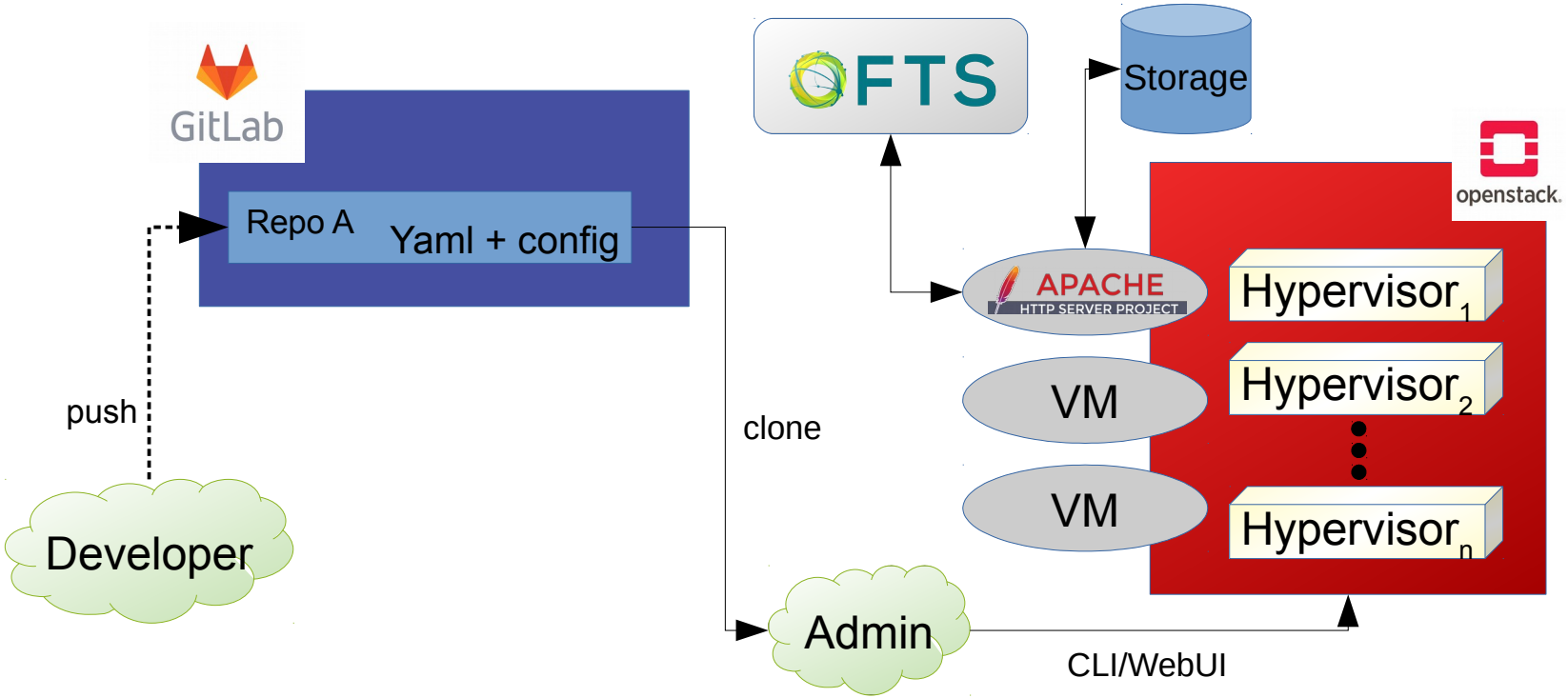
- Core service in **HIFIS** backbone
- CERN‘s **FTS3** as backend
- **webFTS** as comfortable WebUI
- **FTS3-REST** as CLI for scripted transfers
- Next: **Ru**RUCIO policy driven transfers

➔ Apache httpd as **passive** endpoint

➔ **Lightweight** solution for ad-hoc transfers

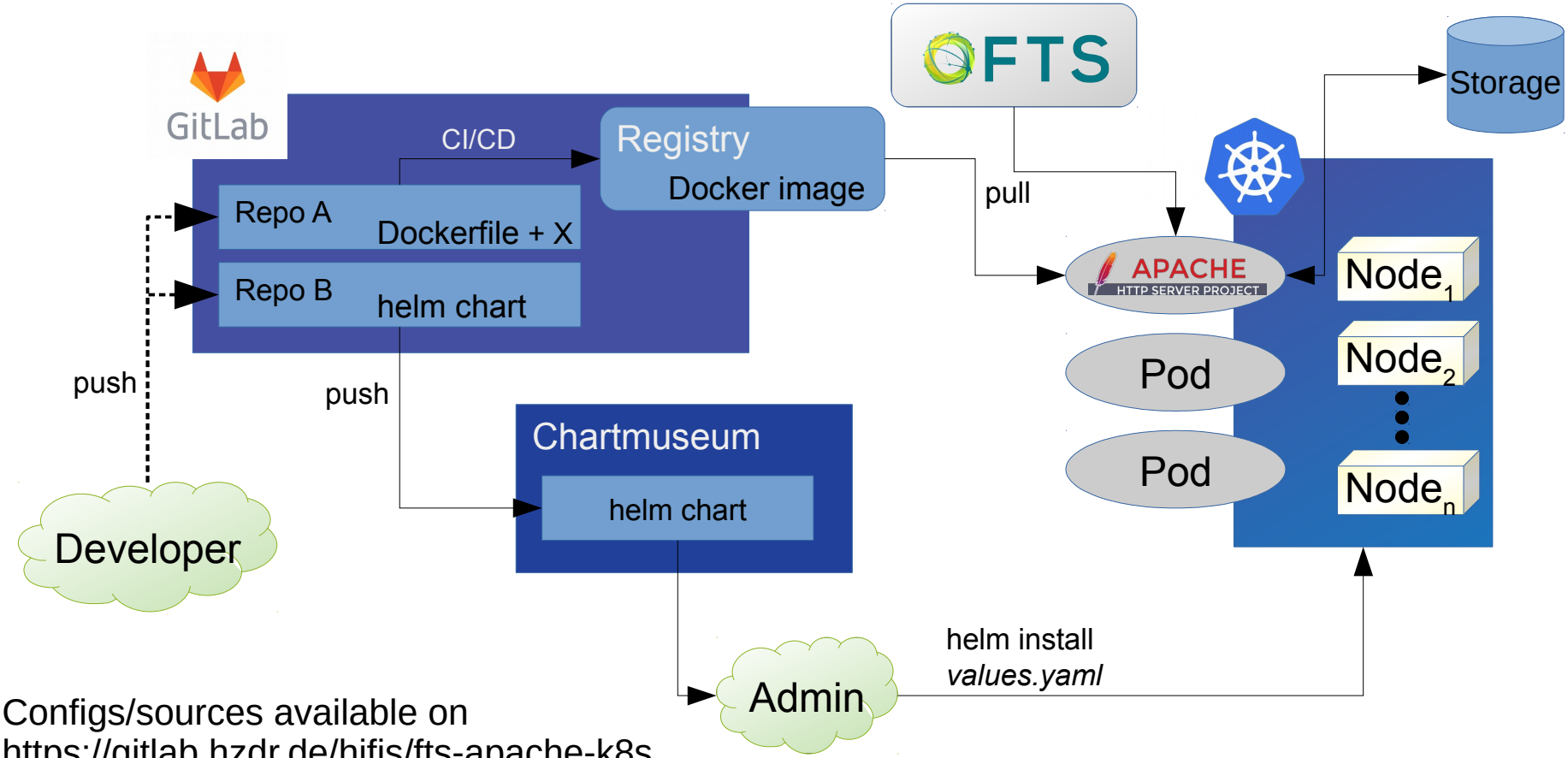


Rollout on Openstack VM



Configs/sources available on
https://gitlab.hzdr.de/hifis/hifis-transfer-service

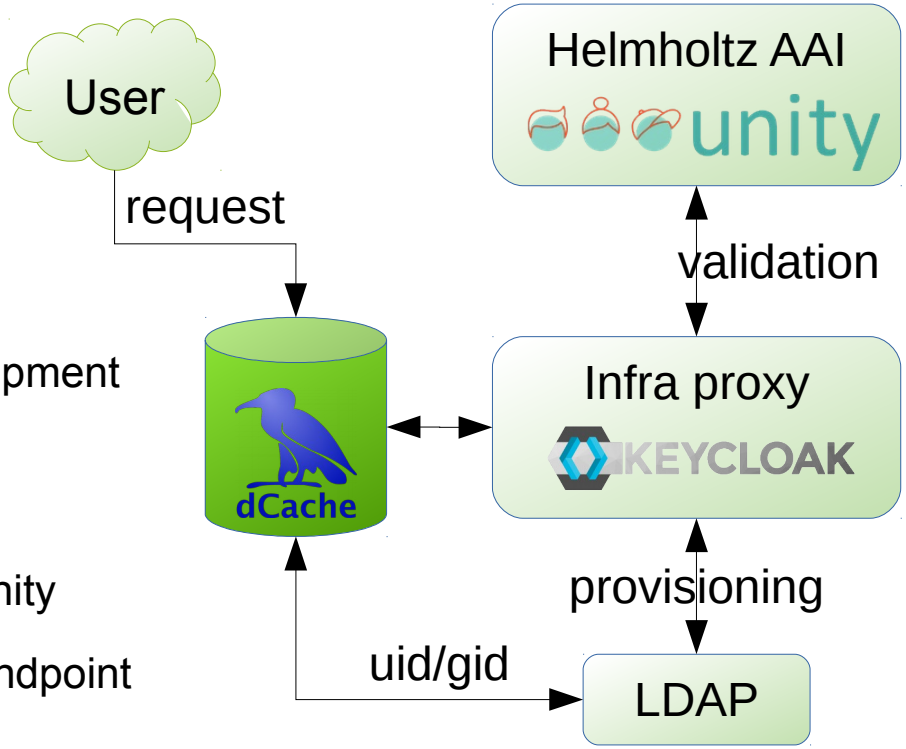
Rollout on Kubernetes via Helm charts



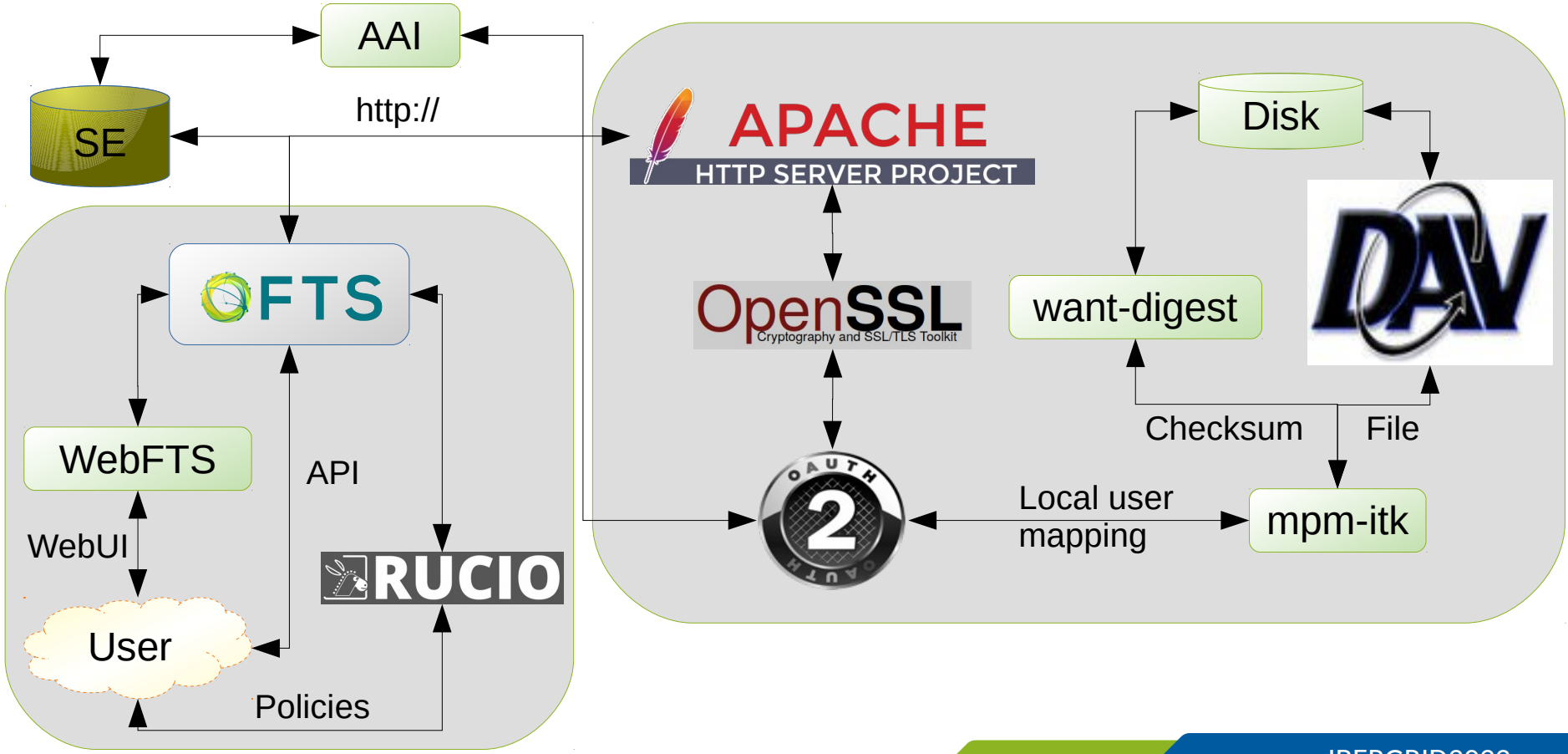
Configs/sources available on
https://gitlab.hzdr.de/hifis/fts-apache-k8s

Helmholtz Imaging storage

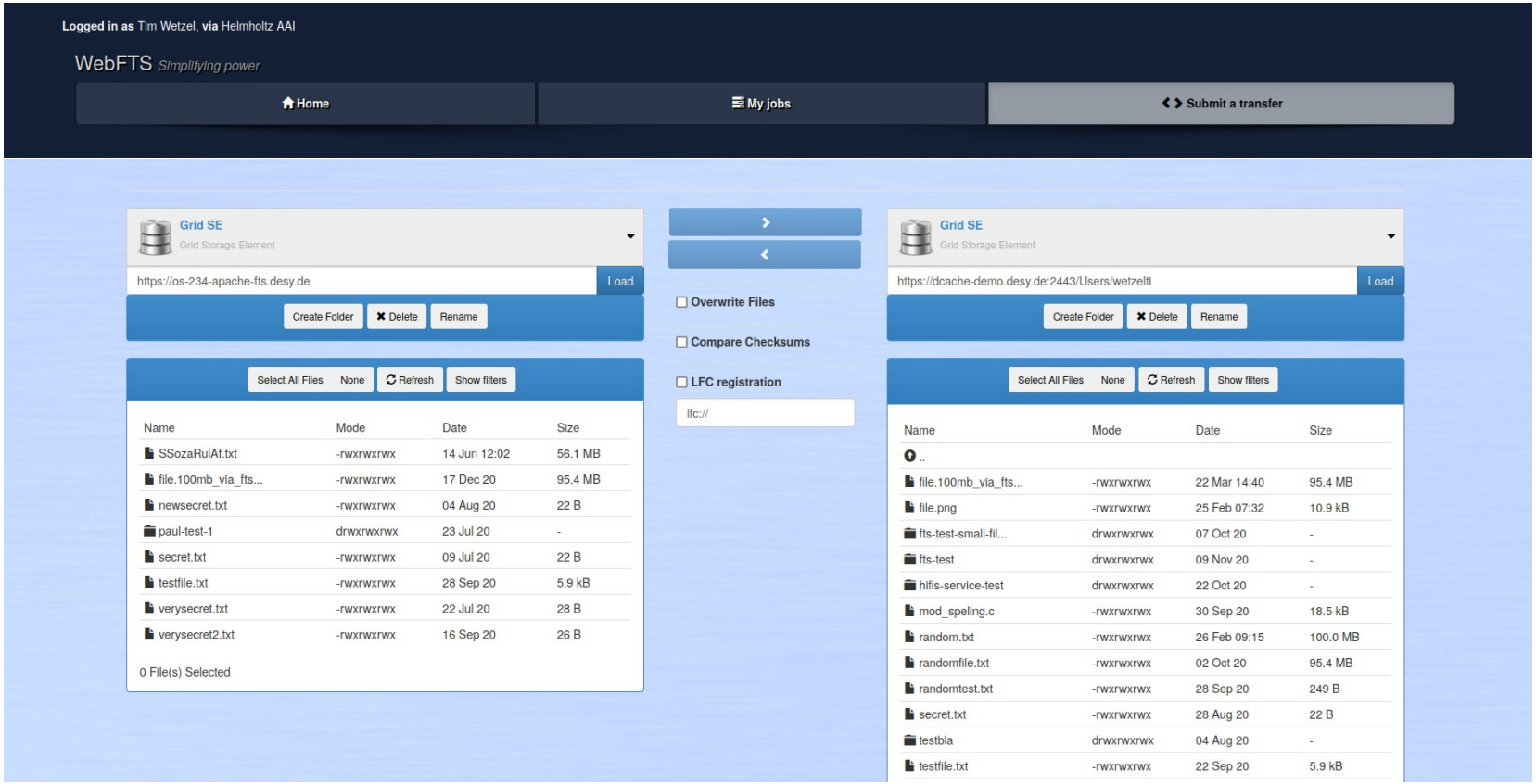
- Central data repository
  - Requested by HI community
  - → dCache @ DESY
  - AuthN/Z via OIDC
  - Automated user provisioning in development
    - → Connection between Keycloak, HelmholtzAAI and DESY-LDAP
- Setup shows how to connect a community
- → Process to be integrated into Apache endpoint



Detail view on interaction



WebUI WebFTS



REST-API

CLI bindings:

```
fts-rest-whoami -s <ENDPOINT> --access-token $TOKEN
fts-rest-transfer-submit -s <ENDPOINT> --access-token $TOKEN <SRC> <DEST>
fts-rest-transfer-status -s <ENDPOINT> --access-token $TOKEN <JOB_ID>
```

Direct API calls with cURL:

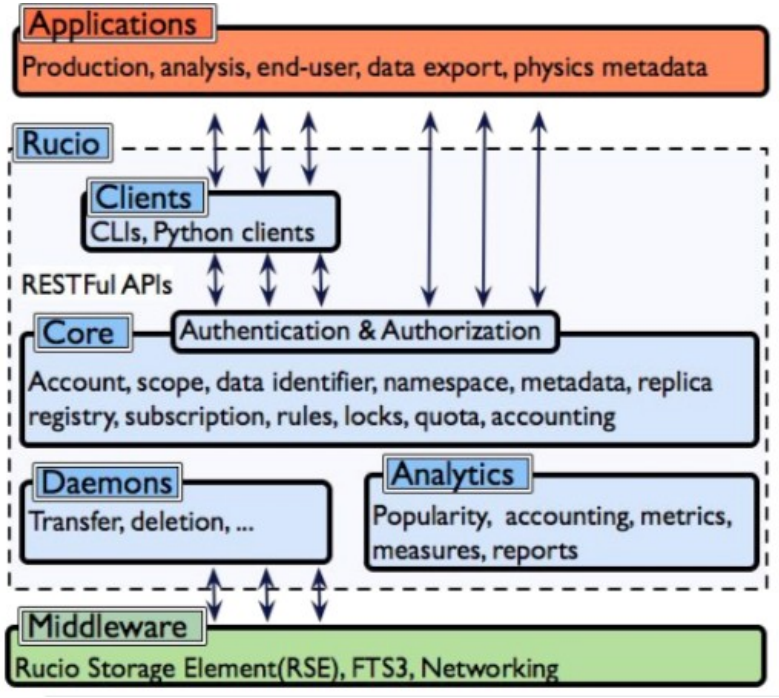
```
curl -s --capath /etc/grid-security/certificates/ -X POST -H "Content-Type:
application/json" -H "Authorization: Bearer ${TOKEN}" --data @submission.json
https://fts-public-002.cern.ch:8446/jobs
```

Documentation:https://fts3-docs.web.cern.ch/fts3-docs/fts-rest/docs/api.html

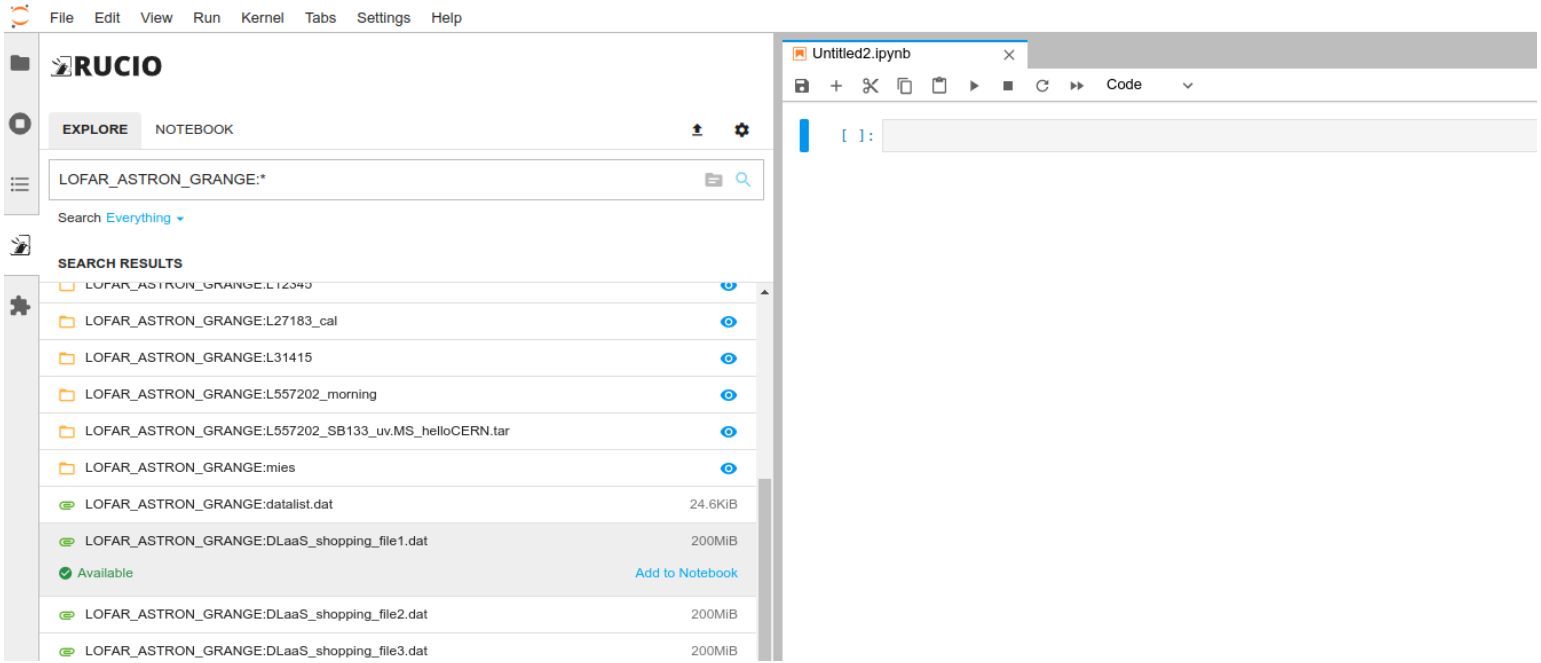
(Rough) Introduction to Rucio

- Orchestration layer on top of FTS3
- Policy- & rule-based organisation of data transfers
- Multi-VO capable
- Token-based AuthN/Z
- Access via WebUI or CLI

Integration tests with JupyterLab-plugin have been shown by colleagues in the context of the ESCAPE datalake

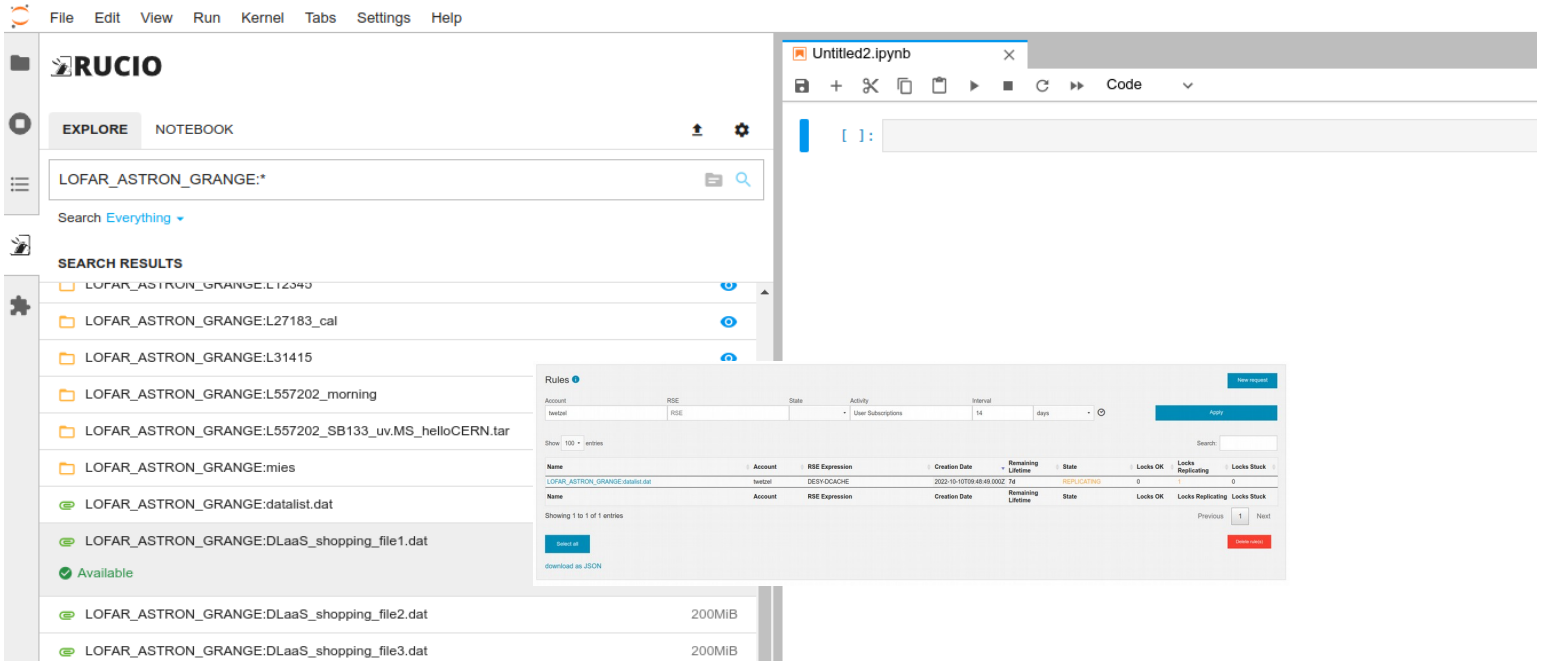


Rucio integration with Jupyter



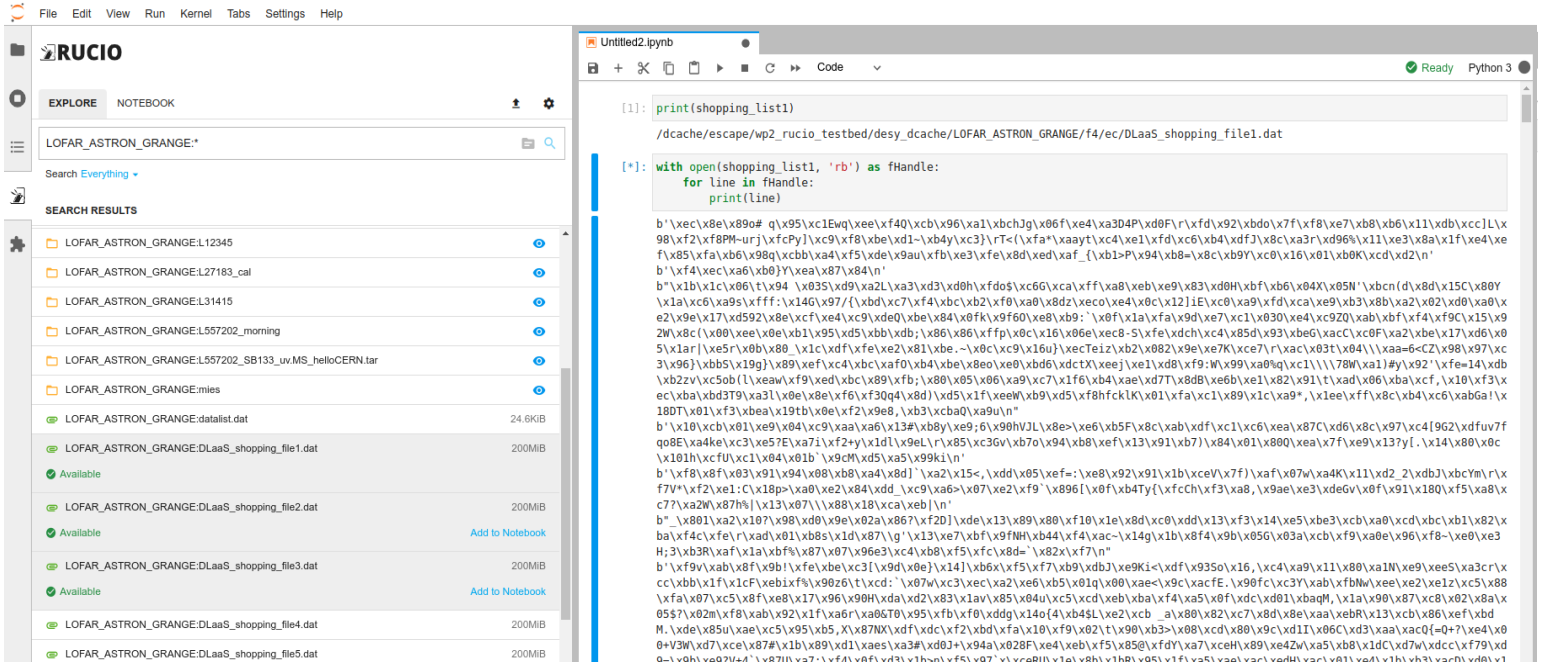
List existing datasets in scope → make available → use file path from mounted storage element in notebook

Rucio integration with Jupyter



List existing datasets in scope → make available → use file path from mounted storage element in notebook

Rucio integration with Jupyter



List existing datasets in scope → make available → use file path from mounted storage element in notebook

Summary & Outlook

- HIFIS Transfer Service
  - Starting to be established in Helmholtz
  - Endpoint installation to be simplified
  - Outreach is increasing
- RUCIO
  - Testing instance set up
  - AuthN/Z token integration next
  - Endpoint integration afterwards

HIFIS

HELMHOLTZ
FEDERATED
IT SERVICES

Thank you! Questions?