

Federated Storage Infrastructures for HL-LHC and ErUM Communities

An ErUM-Data Consortium

Kilian Schwarz

Analysis Facility Workshop, June 18 2024

Federated Storage Infrastructures for HL-LHC and ErUM-Data

Table of contents

01 Motivation

- Central Storage Cloud
- Computing at NHR Centres
- Efficient Data Access

02 Consortium

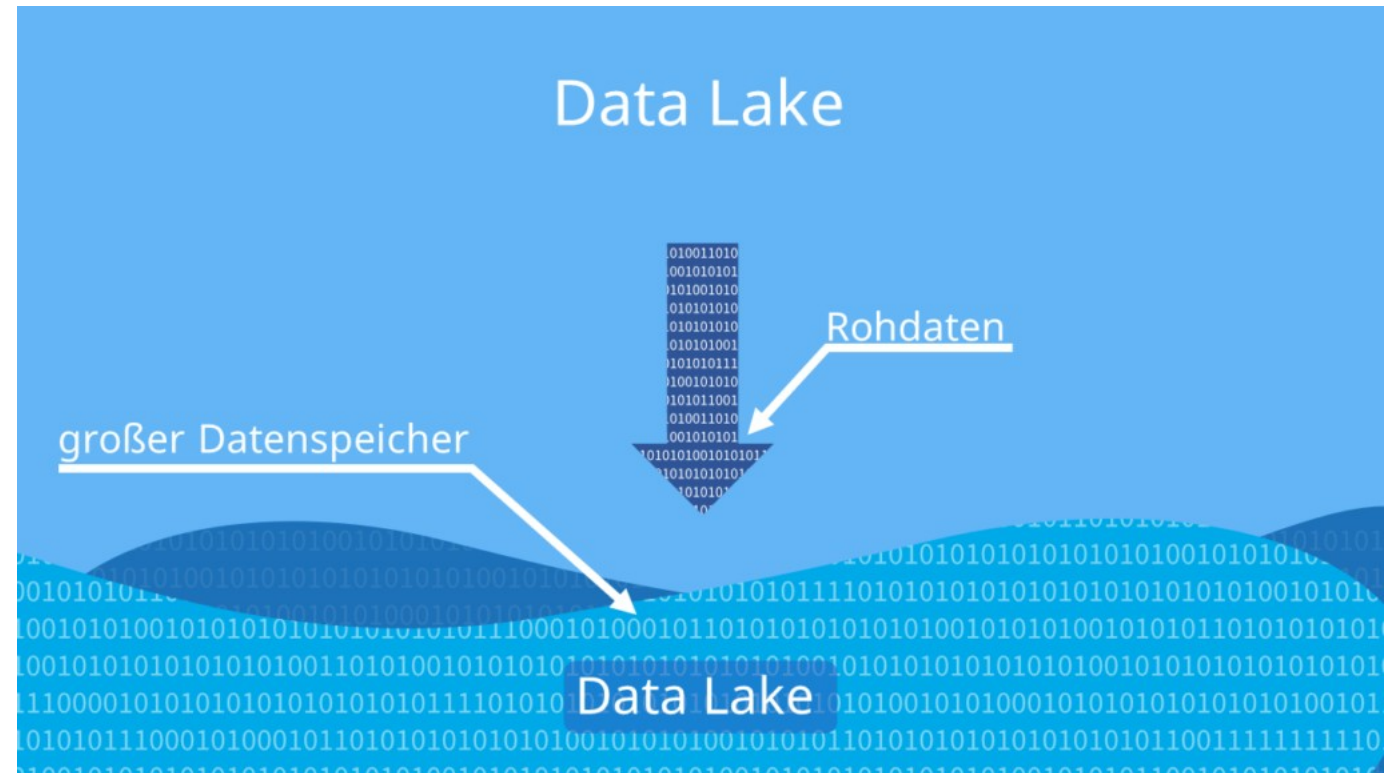
- Current members

03 Planned Work items

- Federated dCache
- Dynamic data cache

04 interaction with Analysis Facilities

- Efficient data access



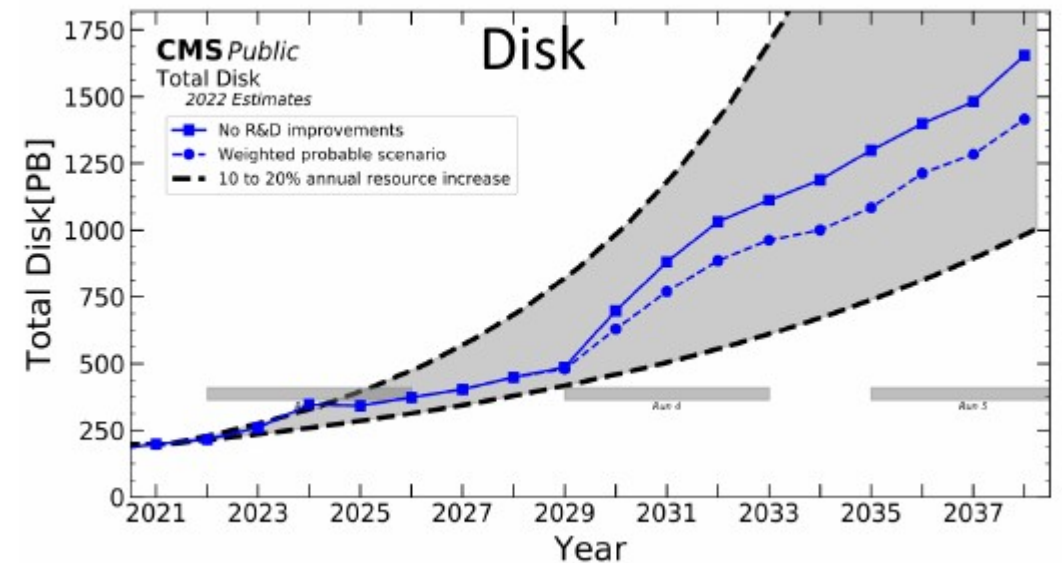
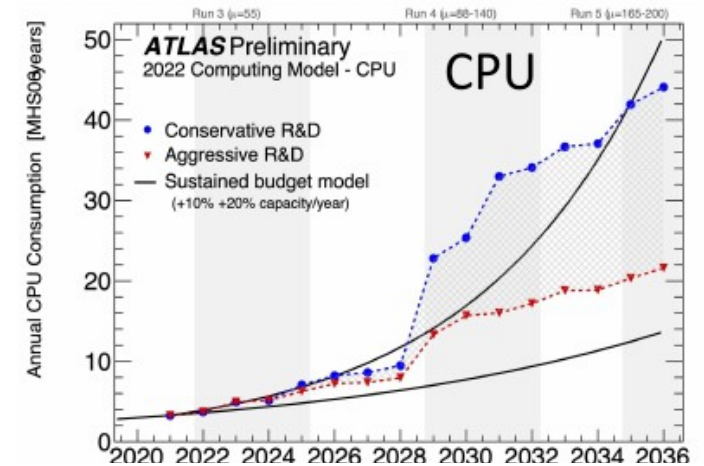
Motivation

Federated Storage Infrastructures for HL-LHC and ErUM-Data

Contact persons: K.Schwarz/DESY, C.Voss/DESY, T. Harenberg/U Wuppertal

Motivation and Background

- High storage rate of more and complex data
- Increased demand for resources
- This requires change in computing model
 - Concentration of mass storage on few larger centres (HGF) ==> central and federated storage cloud
 - Computing will be transferred to large German NHR HPC centres
 - Efficient data access needs to be provided from NHR HPC centres and opportunistic resources
 - More efficient and less personnel intensive operation



Consortium

Federated Storage Infrastructures for HL-LHC and ErUM-Data

Contact persons: K.Schwarz/DESY, C.Voss/DESY, T. Harenberg/U Wuppertal

Current Consortium Members

- DESY
- Wuppertal University
- Göttingen University
- LMU Munich
- Mainz University

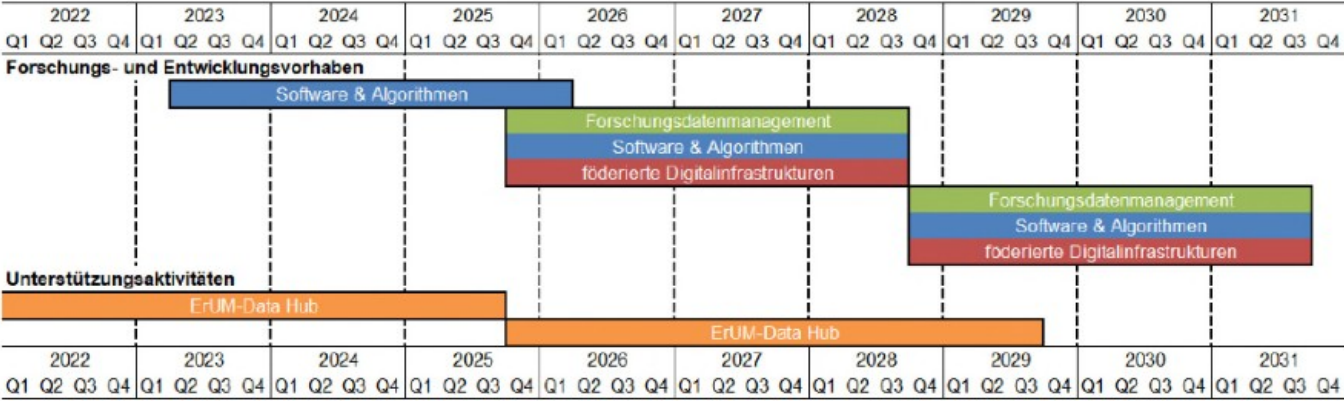
Associated Members

- KIT
- Freiburg University

Supported by

- PUNCH4NFDI

ErUM-Data: Timeframe and implementation



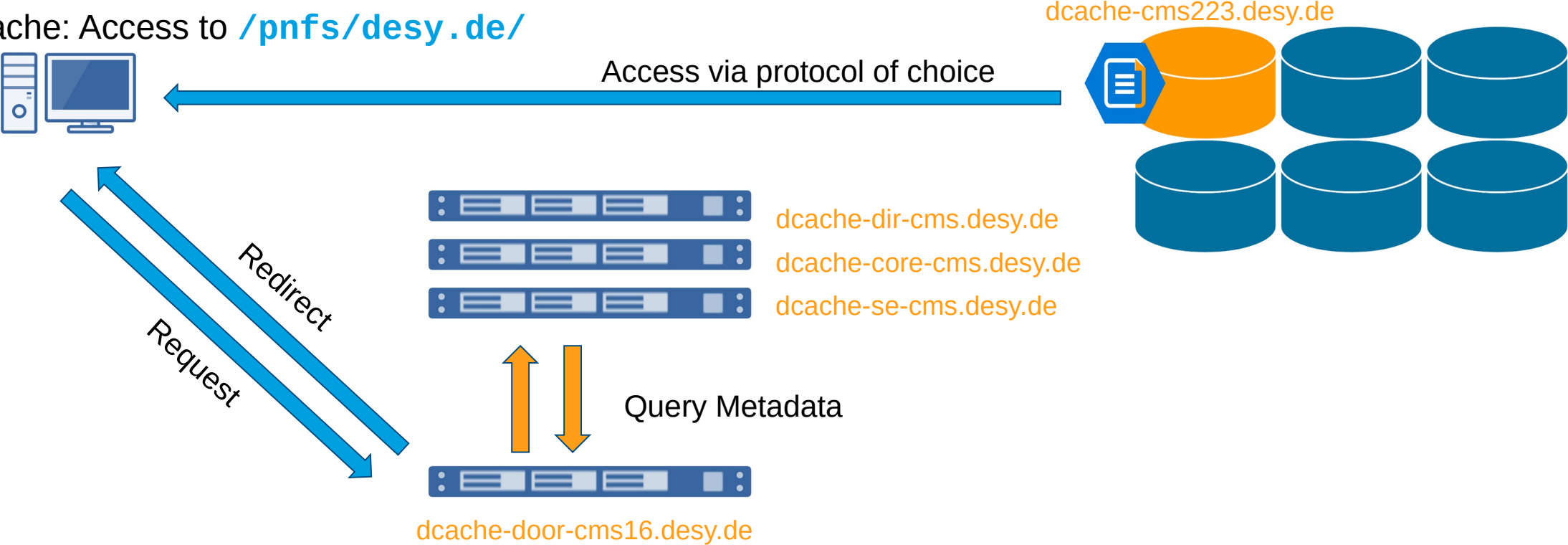
- Prisma strategy meeting on January 23rd/24th, 2024 in Hamburg

Planned Work Items

Basic Setup

Standard Single Site Setup

- Use dCache: Access to [pnfs/desy.de/](https://pnfs.desy.de/)



This is currently the state at almost all centres
Which means there is need for local storage admins

Layout of Federated dCache

Simplest, most Centralised Layout



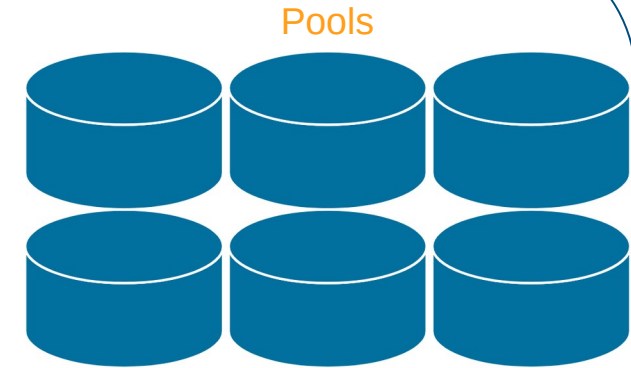
dcache-dir-cms.desy.de

dcache-core-cms.desy.de

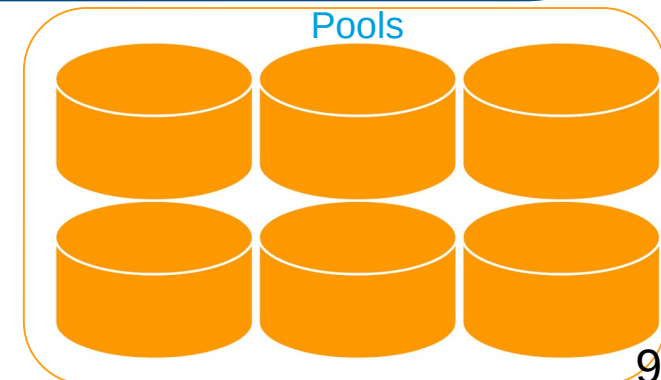
dcache-se-cms.desy.de



dcache-door-cms16.desy.de



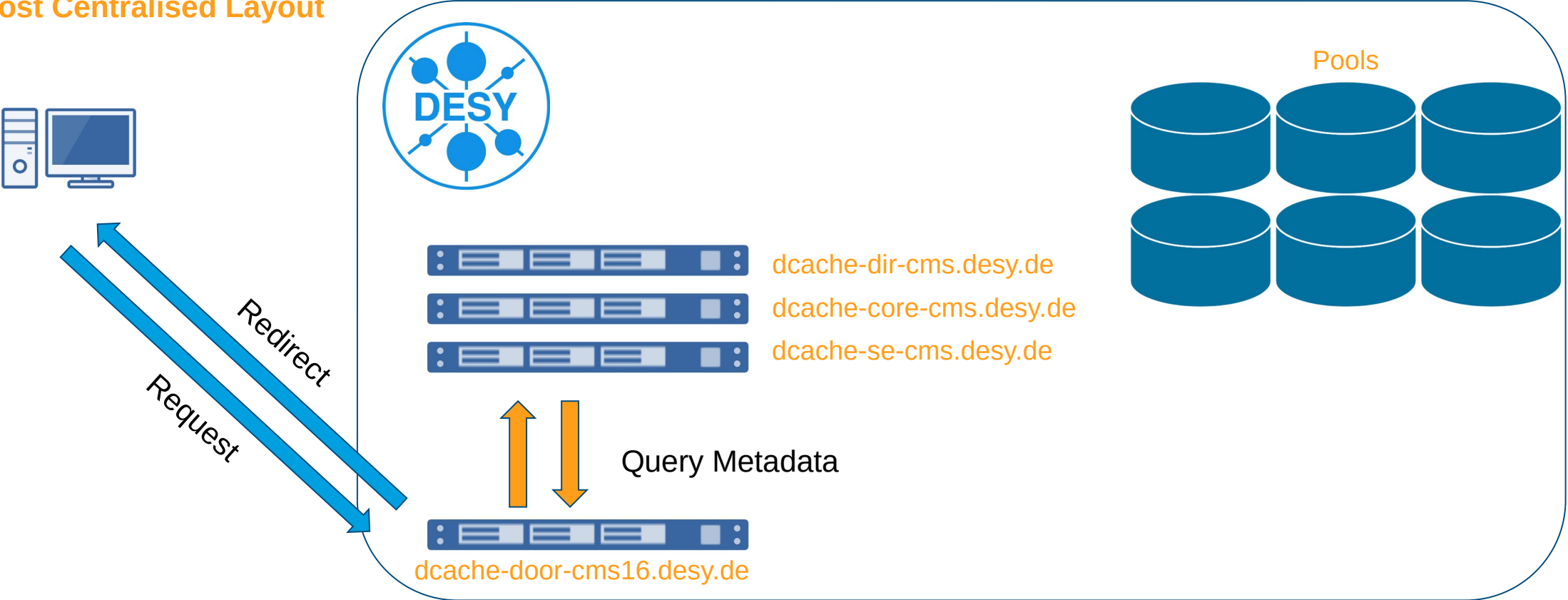
- At remote site only pools are deployed
- All management services located at central site
- Central accesspoint
- Centralised AAI-interface and namespace
- This model is already supported by dCache



Remote Site

Layout of Federated dCache

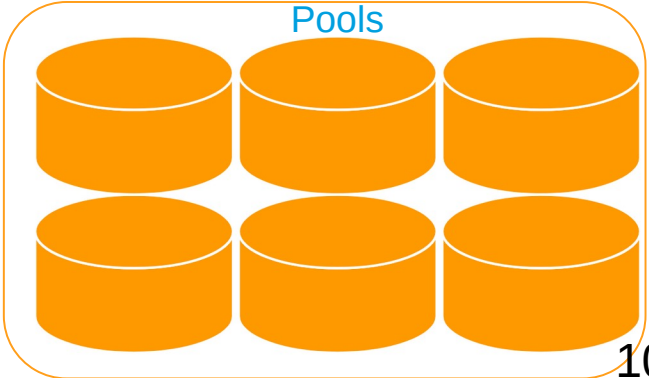
Simplest, most Centralised Layout



- Use dCache: Access to [pnfs/remote-site.de/](https://pnfs.remote-site.de/)



Access via protocol of choice



Remote Site

Federated Storage Infrastructures for HL-LHC and ErUM-Data

Contact persons: K.Schwarz/DESY, C.Voss/DESY, T. Harenberg/U Wuppertal

Federated dCache

- Central or local administration instance
- Single or local name space
- Easy deployment
- Data locality at NHR sites
- Connecting HPC, T3, and opportunistic resources
- Required development works
- Example installations
- POSIX interface

Dynamic data Caches

- Dcache as dynamic data cache
- Cache aware data management
- Cache deployment in heterogeneous environments
- Integration of parallel cluster file systems
- Node local storage access

Other topics

- (Real time) Monitoring and accounting
- Combination with experiment workflows
- Token based (fine granular) data access
- Integration of external endpoints
- Tape workflows
- Automatic data replication

Interaction with Analysis Facilities

Federated Storage Infrastructures for HL-LHC and ErUM-Data

Interaction with Analysis Facilities

- Efficient and consistent data access from AFs to central storage cloud (e.g. via dynamic data caches)
- Data staging from central storage cloud
- AF triggered data replication
- Data analysis on FAIR/Open Data



Thank you

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Kilian Schwarz

IT/Scientific Computing

kilian.schwarz@desy.de

040 8998 2596