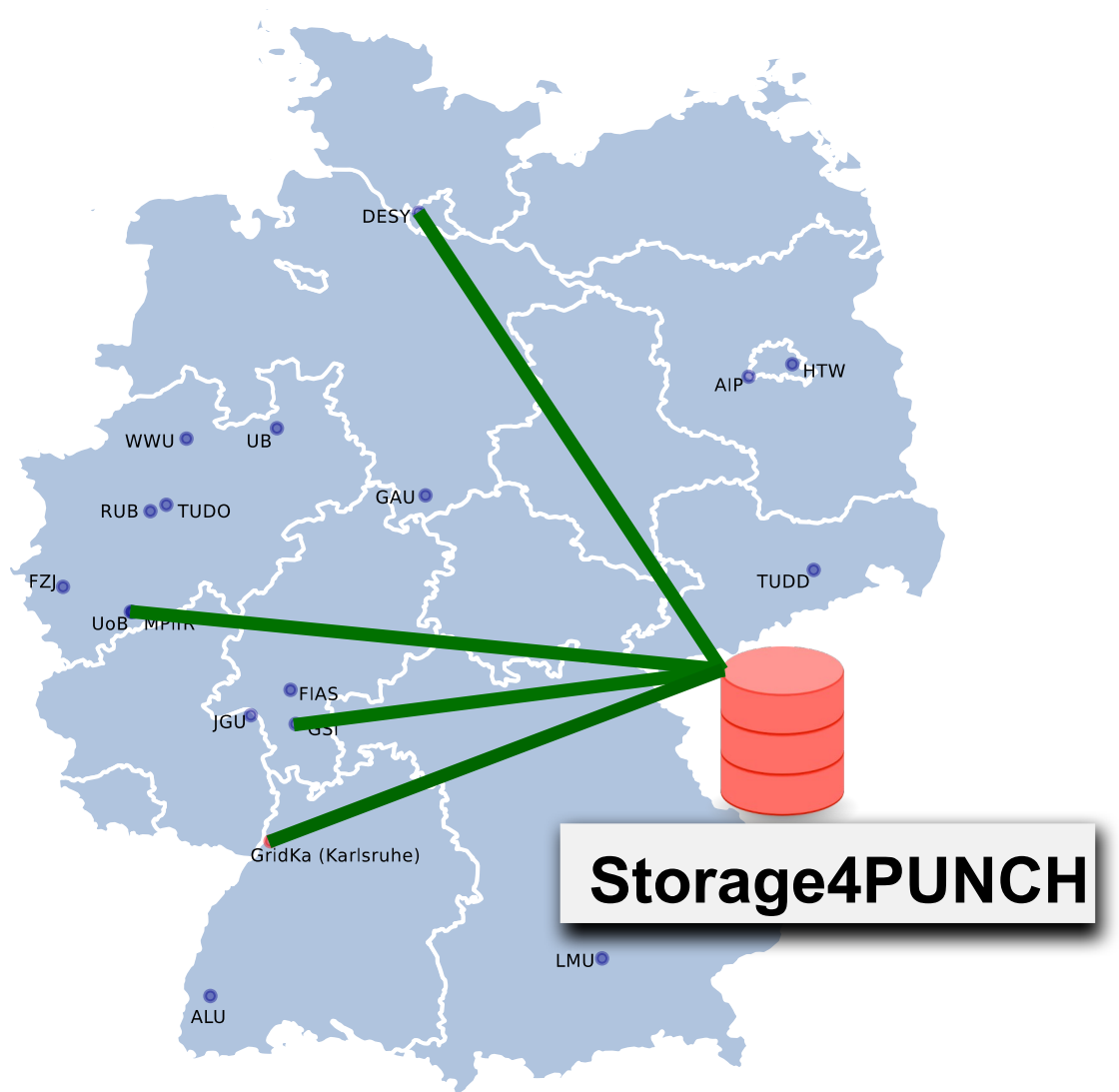


# TA2 WP1 – Brief Status Overview

# Storage4PUNCH - Overview

- Based on two storage technologies
  - dCache (Instances at DESY and KIT)
  - XrootD (Test instances at U Bonn & GSI)
- Token based access using PUNCH AAI
- Supported protocols: WebDAV & XrootD
- AAI still limited
- 



# Storage Instances

## DESY

- Based on dCache
- Production system serving primarily Belle-2, ILC research and local HEP experiments
- Tens of TBs for PUNCH available, can be expanded, if needed
- Endpoints & root path
  - WebDAV: `davs://dcache-desy-webdav.desy.de:2880/pnfs/desy.de/punch`
  - XRootD: `root(s)://dcache-desy-xrootd.desy.de/pnfs.desy.de/punch`

## KIT

- Based on dCache
- System has roughly 2 PB usable for PUNCH4NFDI
- Endpoints & root path
  - `https://dcachewebdav-kit.gridka.de:2880/pnfs/gridka.de/punch4nfdi/`
  - `root(s)://dcachexrootd-kit.gridka.de:1094/pnfs/gridka.de/punch4nfdi/`

# Storage Instances

## Uni Bonn

- Based on XrootD
- Production system of the Bonn Tier-3, also used for ATLAS and other smaller
- Several TBs for PUNCH available
- Endpoints & root path
  - WebDAV: `https://xrootd.physik.uni-bonn.de:1094/cephfs/grid/helmholtz`
  - XrootD: `root(s)://xrootd.physik.uni-bonn.de:1094/cephfs/grid/helmholtz`

## GSI

- Based on XrootD
- Demo system, with very limited storage
- Endpoints & root path
  - WebDAV: `https://lxpunch1.gsi.de:1094/`
  - XrootD: `root(s)://lxpunch1.gsi.de:1094/`

# (Limited) AAI

Right now only one authentication mode

- Every PUNCH AAI can read and write in the punch name space

Cause was a limitation in XrootD, not supporting group claims in our AAI tokens

XrootD limitation regarding group claims has been fixed due to code contributions to XrootD

- Available in XrootD 5.7.0 (and higher)

Dcache ports groups in Unity AAI tokens

Configuration of AAI groups is required coherently on each storage instance

- Complex schemes could become challenging to manage

We will have some more AAI discussions during the workshop

# Unauthorized access for public data

Offering **public data** and derived data from public data

- Present token based access is complex (OIDC agent, registration in PUNCH)
- Access public would ease access

Configuration considerations

- Public READ access for parts of the directory tree, e.g. `<prefix>/<punch-prefix>/public`
  - Write access NOT for the general public!
- **Configuring existing entry points to allow such public read access seems difficult**
- Additional entry points (doors) for public access?
  - dCache and XrootD capabilities might be different
- Needs at least one additional AAI group to grant write access

# PUNCH Use Cases involving Storage4PUNCH

## LHC Open Data

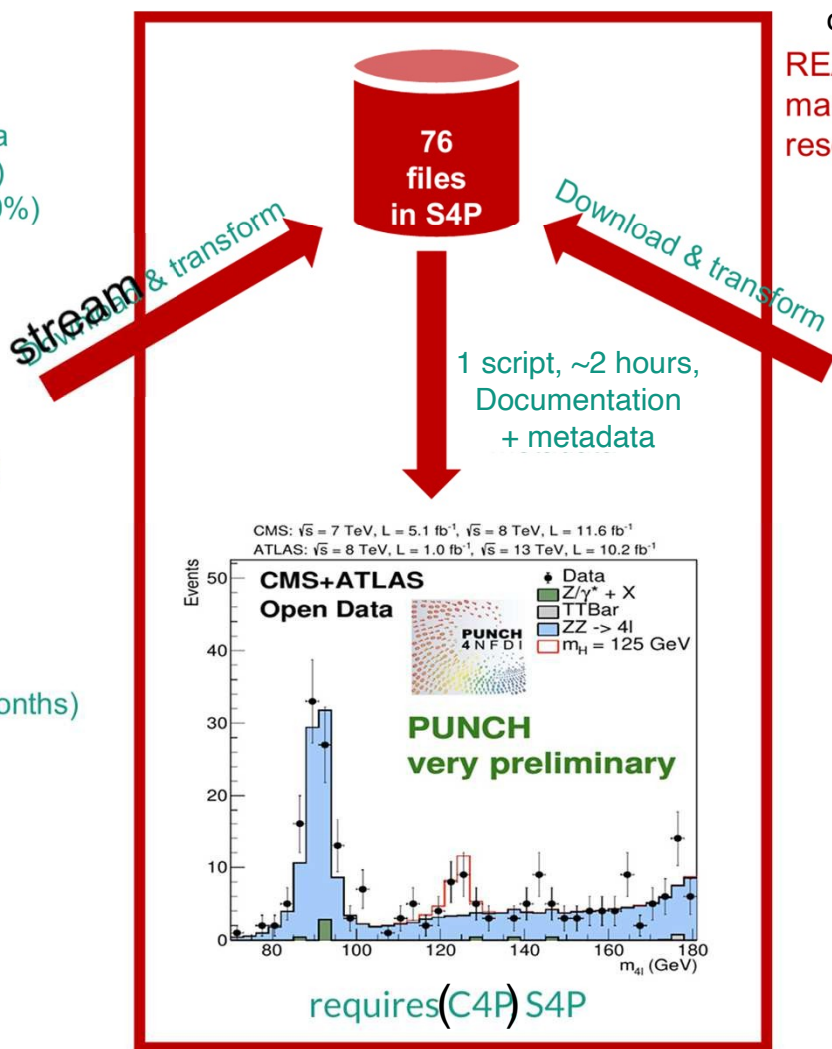
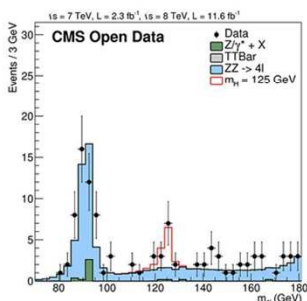
Original CMS legacy research data  
(2 PB via CERN Open Data portal)  
(2010 data 100%, 2011/12 data 70%)



CMS legacy software from public  
github via CERN Open Data portal)



Produce histograms (many CPU months)



or local CPU  
REANA for workflow  
management and  
resource organisation

## CERN Open Data H --> ll

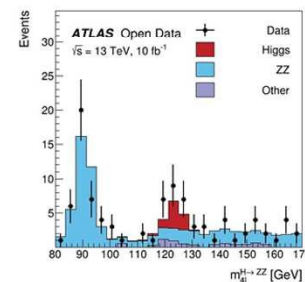
Non-public ATLAS legacy data



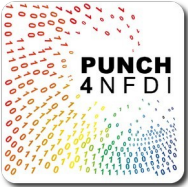
Simplified educational data sets 2012 /  
2016 via CERN Open Data portal or  
ATLAS Open Data portal



VM with dedicated software  
package or Jupyter notebook

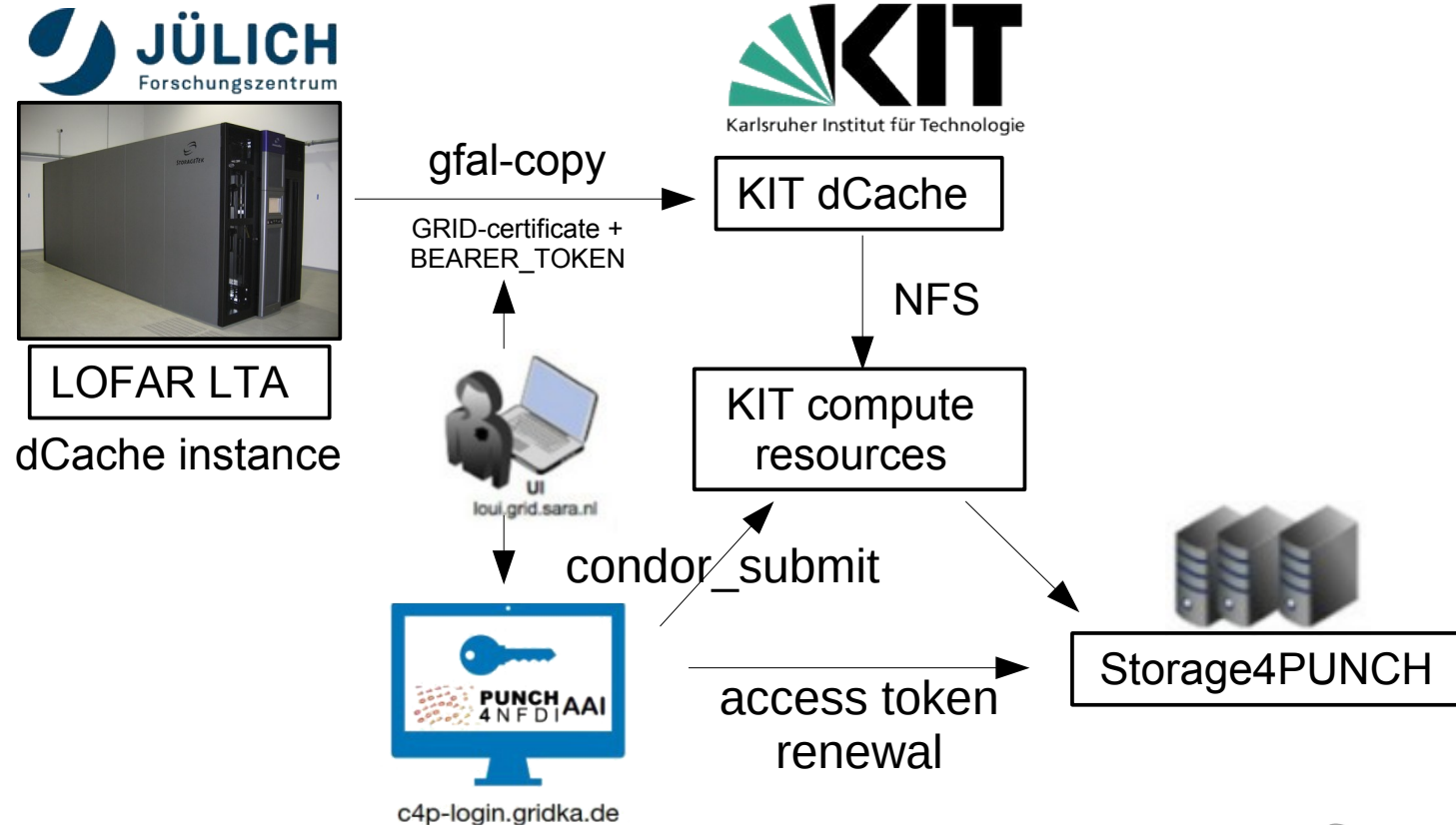


# PUNCH Use Cases involving Storage4PUNCH



Use case:

**Processing the LOFAR Two-Metry Sky Survey (LoTSS)**





# Rucio Instance for PUNCH4NFDI

---

- Small testbed instance currently being set up:
  - Running at DESY
  - Integrated with Helmholtz ID
  - FTS server running at CERN
  - 3 RSEs at DESY, KIT, Uni Bonn

Presented by Thomas Beermann in November in a TA2 meeting

# Meta Data Catalog - Overview

## Handling of "scientific" MD

- several collections of MD (XML)
- each described by free schema (XSD)
- supports complex search queries (xpath)

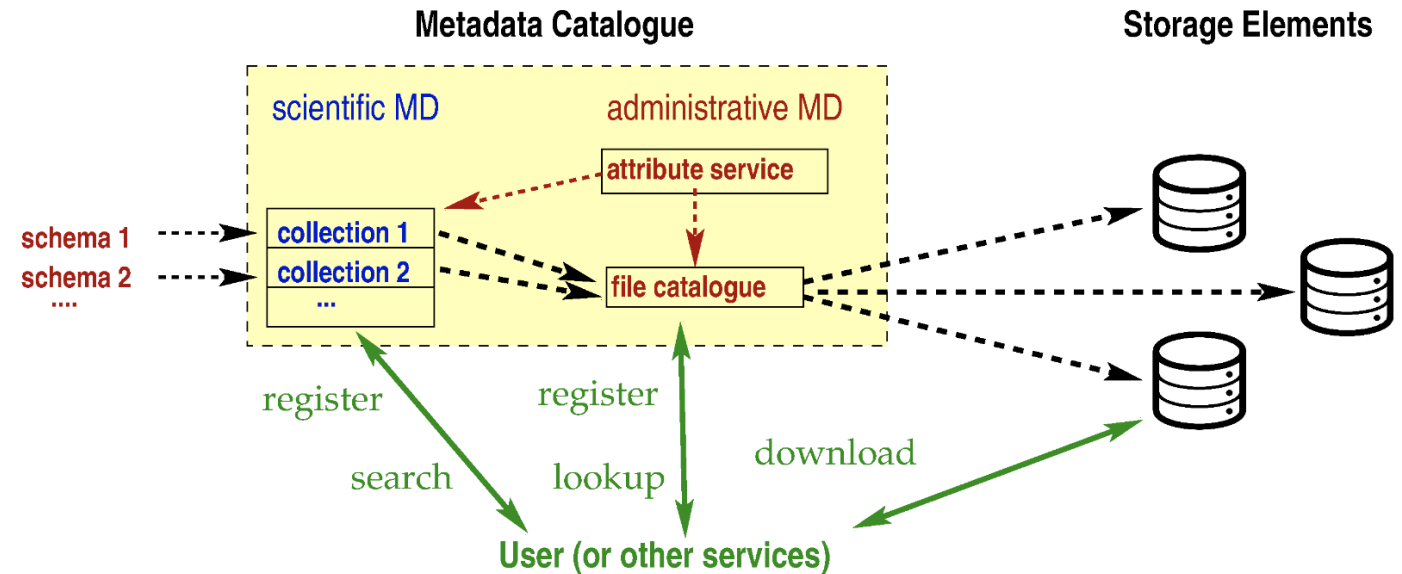
## Handling of "administrative" MD

- file catalogue: ID → storage URL
- attribute service to allow fine-grained R/W access control to metadata, file catalogue, and storage – presently focused in IndigoAM

## First production ready instances for LQCD

- implemented as webservice with REST API
- deployment as docker containers

More detailed presentation by Hubert



Proceedings of the Lattice22 conference

<https://pos.sissa.it/430/244>

## Optimization

- Various directions
- Cannot have all at once

