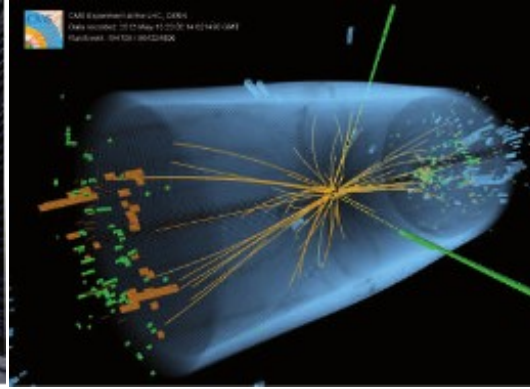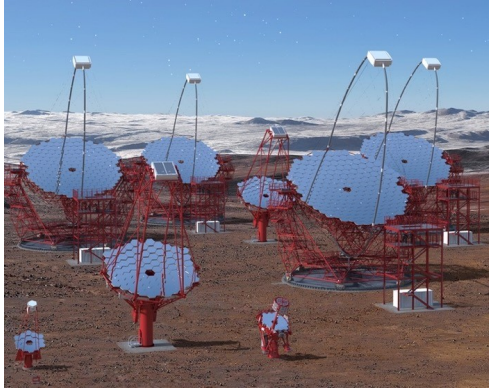# PUNCH 2.0:
Pillar on Federated Infrastructure
- Summary of Prep-Group Discussions -

Christoph Wissing (DESY)

Preparation Meeting at DESY – March 2025

Disclaimer:

This is an attempt to summarize more than three hours of intense discussions plus a several pages long shared document. Quite some choices are personal taste.

I do want to thank all contributors to the discussion sessions (on Friday afternoons) and the document.

Achievements from ongoing PUNCH4NFDI

Non-comprehensive list with focus on Federated Infrastructure topics

- Prototype of federated compute resources with (limited) AAI access
    - Mainly HTC optimized
- Prototype of storage instances accessible via very limited AAI
    - Groups should be possible, not tested yet
- Integration of HTCondor backend and S4P access from REANA
- Flexible Metadata catalog service
- JupyterLab interface within REANA
- Examples of scientific demonstrator workflows to exercise the infrastructure (incomplete!)
    - 'The Higgs OpenData' : various versions
    - 'LOFAR workflow' : connect C4P to a large archive
    - 'Cosmo simu post processing' : how to use a REANA workflow

Things still in reach during PUNCH "1.0"

Again focus on Federated Infrastructure topics (for sure incomplete!)

- Improvements to overcome some limitations of the AAI
    - At least a minimal set of group based AAI authorization schemata

- Improvements of container registry

- Progress on JupyterHub

- More demonstrator workflows

- Involving more groups which can make use of the available infrastructure

- Address some (certainly not all!) of the hurdles to use the infrastructure

Hurdles to use existing PUNCH Infrastructure

PUNCH infrastructure and tools not yet used in the broader communities

- Client tools appear way to complex for less advanced users

- Working with containers is not standard procedure throughout community

- Many analyses are still done in interactive mode

- Missing possibility to provide fine grained access control prevents wider usage e.g. in lattice, astro etc

- Connection between data and computing still not easy (getting code, data and CPU together)

- Documentation/setup to create workflows still complex

Wide agreement that PUNCH2.0 needs to be **more user focused**, e.g.

- Working group to educate users and understand their needs

- Community specific onboarding teams

Federated Infrastructure – Community View: HEP

Federated infrastructure means (slightly?) different things for each sub-community

Globally organized experiments at colliders

- Rather advanced federation of compute & storage at global scale

- Established global trust X509, tokens via Indigo IAM in progress

- Quite some coordination via WLCG

- Agreed set of protocols and methods for storage and job distribution

- Rucio is prominent for data management

- Metadata services experiment specific

- Several solutions for workflow execution: DIRAC (several VOs), Panda, …

- NHR CPU (perhaps other HPC-like) resources to be included in the next few years

- Analysis facilities as emerging topic in the community

- Little to gain from PUNCH2.0, international entities decide, perhaps some profit from tools for end-user analysis

Federated Infrastructure – Community View: HEP

Lattice-QCD

- Uses quite some data management components from WLCG
- Computing typically at HPC centers via individual applications
- Own metadata service (renewed in PUNCH consortium)
- Setup likely applicable for mid-size HEP experiments (and more data intense theory/pheno efforts)
- No use of the PUNCH/NFDI AAI yet (Indigo IAM instead) due to missing fine grained access control features

Small or mid-size experiments or phenomenology/theory, e.g. ALPS II, LUXE, Baby-IAXO

- As of now very little federation, but could be beneficial, because collaborations are international
- Quite some of the WLCG or Lattice developed could be employed

DESY Testbeam

- Use pattern is more like photon science (user groups come for measurement and typically take data home)
- Centrally provided metadata service could be of interest
- Data service to easily move data home of potential interest

Federated Infrastructure –
Community View: Nuclear & Hadron Physics

For huge collaborations like ALICE HEP/WLCG approach (mostly) applies

Other nuclear & hadron physics

- Huge spread in size, from 5 scientist to several hundred scientists (e.g. CBM at FAIR facility)
    - Huge spread in computing skills
    - Mohammad: "*Their Linux expert is the guy who knows to open a terminal to type the ls command*"

- Little federation as of now but looking for more federation because of the international nature of collaborations

- Interested in a model that ensures data accessibility across different sites while maintaining security and compliance (e.g., GDPR, FAIR Data Principles)

- Looking for common metadata services

- Interested in adopting containerization and workflow orchestration tools (REANA) for efficient cross-site deployment

Federated Infrastructure –
Community View: Astro (Particle) Physics & Astronomy

Combination of observations play a much bigger role than combinations of HEP results, multi-messenger astronomy

Usually observations become publically available after a (short) proprietary time

Large astro particle physics collaborations (e.g. CTA)
- Similar to HEP/WLCG, also similar tools like Rucio for data management
- International structures decide on tools, methods, protocols

Large (radio) astronomy collaborations (e.g. SKA, LOFAR, LSST/Rubin)
- WLCG tools & methods partly being used
- Data stored in archives often publically accessible (after registration),
- BUT close-by compute-resources often only available to members

Diverse spectrum of sky surveys, satellite missions, observations in different wavelengths..
- Access to data often provided via Virtual Observatory
- Very little federation

Theoretical astro physics and cosmology
- HPC usage via individual grants, local clusters typically not federated

PUNCH2.0 possible Topical View

Community onboarding

- Training and documentation
- Community specific example workflows
- Use of workflow languages and related tools

User-interface and client tools

- Easy to use clients or web services

Sustainability aspects

- Green IT
  (always asked for, almost luxury problem)
- Long term support

AAI

- Fine grained access rights propagated across federated resources
- Concepts and tools for community driven access policy distribution
- External policy engine

Storage

- More convenient data import and transport, e.g. employing FTS
- Real federation of storage instances, file catalog, replication rules
- Shared home directories

Compute

- Further integration of resources, also more HPC-like
- Monitoring and accounting
- Data location aware scheduling

The PUNCH2.0 Federated Infrastructure Plate