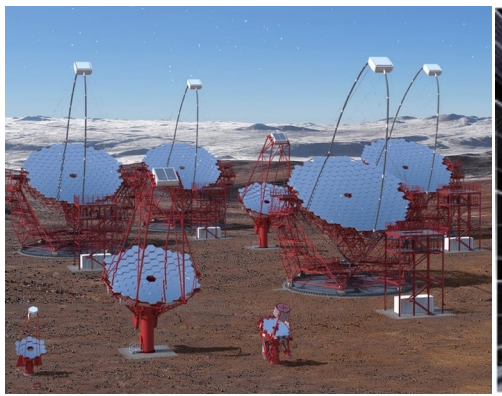
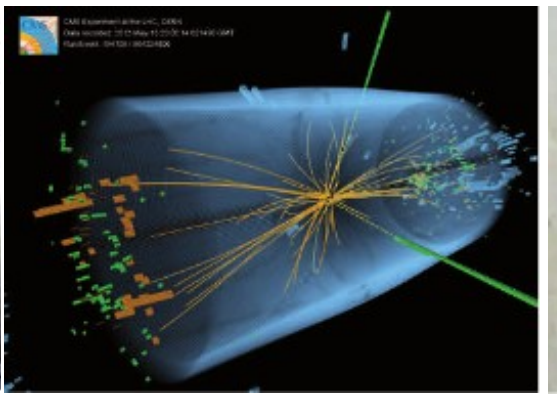
PUNCH 2.0:
Storage + Catalogs Component Box

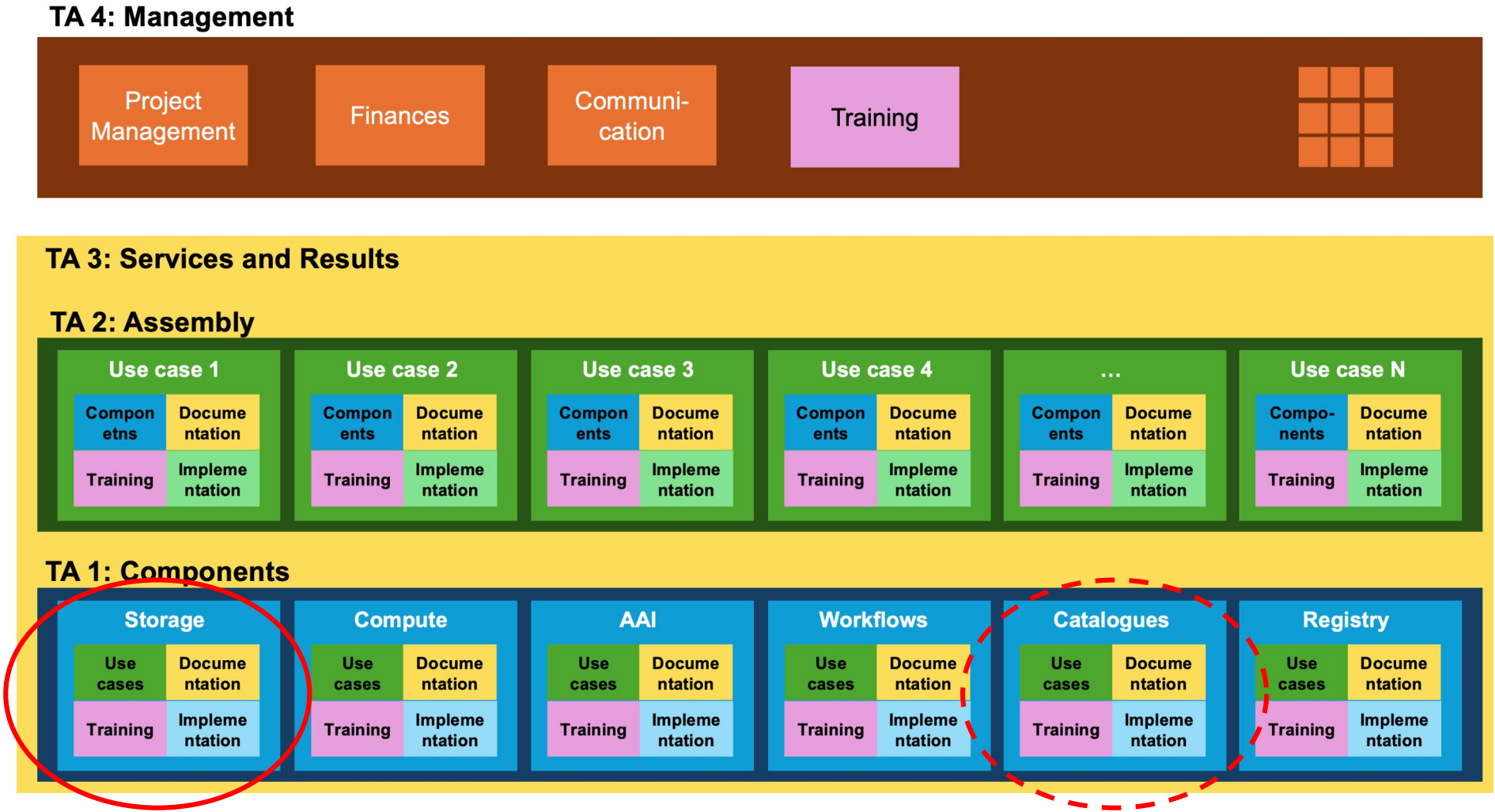





Philip Bechtle (Uni Bonn), Oliver Freyermuth (Uni Bonn), Michael Hübner (Uni Bonn),
Hubert Simma (DESY), Christoph Wissing (DESY)

Preparation Meeting – 14th April 2025

Home works were given along this structure (which is still in flux!)



What is in the Box(es)

Storage and catalogs (for files and metadata) should not be separated

Technology view

- Established storage middleware from WLCG, also used in PUNCH1.0: dCache & xrootd
- Teapot: Tool to make POSIX storage work with Tokens and FTS file transfers
  - Presently developed in interTwin EU project
- S3: Well established cloud storage, however tricky to integrate with our AAI
- PUNCH1 Metadata catalog: Flexible schema, e.g. for LatticeQCD, plain file-catalog, ...

Interfaces and Integration

- FTS (File Transfer Service)
- Rucio (optional tools for distributed data management)
- AAI – this will be key and remains challenging!

Main Deliverables / Milestones

Storage for smaller research groups

- Provide storage for research efforts that cannot do it on their own
- Has to come with long term commitments

Reference installations

- Reference and prototypes for “bigger” research efforts to copy it for own infrastructure
- Documentation and installation/configuration support
- Consulting and best practices

Integration

- Very wide topic!
- Interoperability with
  - AAI
  - Data management components, e.g. Rucio
  - File and meta data catalogs
  - Data import / export services, e.g. at HPC sites
- Selection of integration efforts is use-case driven

Resource Requirements

Very hard to estimate, if work sharing between Use Cases and Components is not resolved

If all integration and support work comes from the Storage and Catalogs box, this would be sizable: about 5 FTEs over the full time

Type of personnel

- Code maintenance and contributions to (upstream) software stacks will be required
- Need software developers and/or domain scientist with strong computing and integration skills
- Substantial amount of guidance by experienced (senior) staff expected (likely in-kind)

Some in-kind hardware

- Prototypes
- Infrastructure for “smaller” groups

Institutions

Some extract from a preparation document for the Federated Infrastructure Pillar indication of interest for storage/catalog related PUNCH2.0 contributions

DESY: dCache, teapot, PUNCH1 catalog like to HIFIS

KIT: dCache in S4P

FZJ: dCache archive for LOFAR, ILDG

JGU Mainz: S4P integration

AIP: S3 storage

GSI: XrootD

Uni Bonn: XrootD

DZA: Storage & archives for various (radio) astronomy research (incl. SKA)

Community Coverage

Main targeted sub-communities

- Mid-size HEP experiments: e.g. Axion program at DESY
- Lattice-QCD and potentially other theory
- Radio astronomy (perhaps not to SKA scale)
- Small to mid-size hadron and nuclear physics experiments

Not really targeted or likely not such a great match for

- Global experiments or observatories (like LHC experiments, Belle-2, SKA)
  - Experience and tools will certainly be used, actually required
  - Collaboration will not really care about NFDI
- Astronomy research centered around the Virtual Observatory

Gap Analysis

Relevant components are addressed

However coordination is required to ensure coherent developments tailored towards the needs of use-cases


PUNCH2.0