The PUNCH4NFDI Consortium

Particles, Universe, NuClei and Hadrons for the NFDI

Kilian Schwarz

DESY

PUNCH GM Meeting – TA6, 07.07.2025

Work Package 1 (M. Zacharias & K. Schwarz)

D-TA6-WP1-1 (Noticeboard) (01/22)

- Being installed at DESY to go public
- To be connected to helpdesk

D-TA6-WP1-2 (physics.tools) (01/23)

- Exchange platform for information on archives, data management software, hubs)
- See specific slides WP4

D-TA6-WP1-3 (communication with NFDI) (continuous)

- Synergies on cross-cutting topics to be discussed with NFDI initiatives
- Important point: https://intra.punch4nfdi.de/?md=/docs/NFDI/sections.md
  Fill this page, when you take part in NFDI section (working groups, etc) meetings

Work Package 1

D-TA6-WP1-3 (communication with NFDI) (continuous)

- Important point: https://intra.punch4nfdi.de/?md=/docs/NFDI/sections.md
  Fill this page, when you take part in NFDI section (working groups, etc) meetings

Common Infrastructures

List of Working Groups and members

Working Group	Email	General topics	PUNCH members	Meeting minutes	Hot Topic	Next meeting
Overall Architecture	Sign-up					
Data Integration	Sign-up		A. Montanari, J. Künsemöller			
Data Management Planning	Sign-up		H. Enke			23.02.24 12:00
Data Science & AI	Sign-up		A. Mohamed, N. Suvvi Neelakantaiah			
Data Transformation	Sign-up		C. Wissing, N. Suvvi Neelakantaiah			
Electronic Lab Notebooks	Sign-up					
Identity & Access Management	Sign-up		K.Schwarz			
Infrastructure and Data Security	Sign-up					
Long-term Archiving	Sign-up		J. Chodak			
Multi-Cloud	Sign-up	concept	K. Schwarz, C. Wissing, H. Enke, Neelakantaiah, J. Künsemöller			
Persistent Identifier	Sign-up					
Research Software Engineering	Sign-up	Marketplaces, JupyterHubs,etc	H. Enke, M. Zacharias, M. Blank			

Metadata

List of Working Groups and members

Working Group	Email	General topics	PUNCH members	Meeting minutes	Hot Topic	Next meeting
Cookbooks, Guidance and Best Practices	Sign-up					
Knowledge Graphs	Sign-up				B4N project approved	
Ontology Harmonization	Sign-up					
Research software metadata	Sign-up		C. Schmidt, O. Kaczmarek			
Search and Harvesting	Sign-up		M. Castro Neves			
Terminology Services	Sign-up					
TF Metadata			V. Tokareva			

Work Package 1

For the next 1+ years in PUNCH 1.0:

D-TA6-WP1-4 (services and tools) (continous)

- Further development of physics.tools and noticeboard
- PUNCH connection to Base service **nfdi.software**
- Services and tools will be made available to the NFDI

D-TA6-WP1-5 (continuation of D-3) (continous)

- Deepen ties with other NFDI consortia/sections/Base services
- Continue advertising PUNCH4NFDI at meetings (DPG, AG, etc.)
- start showing actual products and services
- PUNCH Lunches

Work Package 2 (K. Schwarz & D. Mallmann)

- D-TA6-WP2-1
  - Prototype PUNCH AAI (31 Mar 2022): done
  - Basic PUNCH AAI (31 Dec 2023): Unity dev requests #1, #2, #3
  - Extended PUNCH AAI (30 Sep 2026)
- D-TA6-WP2-2
  - Coordination with NFDI, national, international stake holders
    - Draft design (31 Dec 2022): done, published
    - Complete design (31 Dec 2024): done, published
- D-TA6-WP2-3
  - PUNCH AAI group management
    - Prototype (31 Dec 2022): done
    - Revised version (31 Dec 2024): in progress
    - Full group management (30 Sep 2026)

Work Package 2

- AAI requirements described in
  - https://results.punch4nfdi.de/?md=/docs/Documents/AAI/aai-requirements.md
  - These requirements are being implemented in the PUNCH AAI

- Status of Development requests to Unity
  - Request 1: „Group information in tokens“
    - Implemented by Unity even before we made an official request
    - Detailed testing in PUNCH still in progress
  - Request 2: „Claim filtering“ (only selected groups in tokens)
    - Implemented and tested
    - Detailed testing of workflows in PUNCH still in progress
  - Request 3: „Granular permissions“
    - In progress
    - In extended AAI meeting on June 04 decided for:
      - External policy service “behind“ AAI with door for “token exchange with trusted client“ to be kept open
      - Current options: Unity performs call directly or Unity launches external script being discussed also with FZJ
      - This request is completely supported by HIFIS
  - Additional: Indico access rights can be controlled via groups in PUNCH AAI
    - Currently being tested in the context of HIFIS

Work Package 3 (S Wagner & H. Hessling)

- D-TA6-WP3-1 (31 Jul 2022)
  - Reference guide on publishing data: living document
- D-TA6-WP3-2 (31 Dec 2022)
  - Reference guide for publishing software: living document
- D-TA6-WP3-3 (31 Dec 2023)
  - (Dynamic) Metadata frameworks for PUNCH-SDP: declared finished
  - Document (10 pages) by Tim Oelkers
  - To be followed up in PUNCH2.0 via DZA
- D-TA6-WP3-4 (31 Mar 2024)
  - Effelsberg data: data need to be converted, ongoing
  - Updated completion date: 31 Dec 2025
- D-TA6-WP3-5 (31 Dec 2024)
  - Converter for FITS/ROOT formats: done, being published
- D-TA6-WP3-6 (31 Dec 2025)
  - Metadata extensions: on track

Work Package 3

D-TA6-WP3-3

- Highlights
  - Tim Oelkers, "Overview of petabyte-scale metadata storage methods and frameworks“ (10 p, 2023)
  - Basis for future developments: SKAO project "Rucio-IVOA Metadata Integrations“
    (https://gitlab.com/ska-telescope/src/ska-rucio-ivoa-integration)

D-TA6-WP3-4

- Update
  - Conversion is ongoing. However, converted data need to be recalibrated, which is a demanding task and cannot be realized within PUNCH 1.0 (reduced funding)

D-TA6-WP3-5

- Highlight: a Root to FITS conversion is realized, which makes it possible to link astroparticle data with astronomical data.
- The conversion is a prototype solution. Optimization potential for further development into a plug & play solution.

D-TA6-WP3-6

- Highlight: concept developed. Implementation would, in particular, allow linking of multi-frequency data via „extended metadata".

Work Package 4 (G. Duckeck)

- D-TA6-WP4-1: survey of PUNCH tools
  - Initial overview (31 Jul 2022): done, being published
  - Final list (30 Jun 2026)
- D-TA6-WP4-2 (30 Jan 2023)
  - Reference repository with CI: done, being published
- D-TA6-WP4-4 (30 Jun 2026)
  - Data analysis examples
- D-TA6-WP4-5 (30 Jul 2023)
  - Software platform: done, merged with Physics tool

Work Package 4 & 1

New developments in **physics.tools**, a search engine for software referenced in publications.

Also part of the **Base4NFDI** proposal, in the **nfdi.software** group.

- The service is now on a server in Heidelberg University (here)

- The software database, containing software extracted from papers on arxiv
  - For now testing the deployed website with a subset of the paper database
  - Moved citations of issues/branches/discussions/etc. in GitHub repositories
    that were appearing in the old version of the website to the citation of the main GitHub repo.

- The search engine:
  - Polished output of the standard version: now ordered by GitHub repo star counts,
    showing the repo description and the main programming language
  - Working on the AI-assisted version (i.e., that understands keywords): will use a fuzzy-match algorithm
    with an embedding language model (w/ ollama and nomic-embed-text)

- Working on the feature *Add your software*
- Working on a Flask API to sync the database to the research software directory in nfdi.software

Alessandro, Michael, Stefan - Heidelberg University

Work Package 4 & 1

Future development **in physics.tools**, a search engine for software referenced in publications.

- Currently, only our team in Heidelberg University is working on this

- By the end of September 2025, we shall be able to:
  - Complete the database
  - Extend the engine output with other providers (i.e. Zenodo), but also private websites that are cited in the papers
  - Finalize the AI-assisted search engine
  - Make it possible for users to ask to add software
  - Finalize the Flask API for the synchronization with nfdi.software

- For the future after then:
  - Automatize the harvesting of arxiv papers to compile the database, and the index for the AI model
  - Define cronjobs for the Flask API and the synchronization
  - Possible new features, which partially depend on what is included in the Integration phase in nfdi.software

Alessandro, Michael, Stefan - Heidelberg University

Work Package 5 (S. Wozniewski & H. Enke)

- D-TA6-WP5-2 (30 Sep 2022)
  - Dynamic disk cache: done and published
- D-TA6-WP5-3 (30 Dec 2025)
  - Memory based computing
- D-TA6-WP5-4 (31 Dec 2025)
  - Interfaces to supercomputer, GPU, GoeGrid
- D-TA6-WP5-5 (31 Dec 2024)
  - COBalD/TARDIS: done, being published
- D-TA6-WP5-6 (30 Sep 2022)
  - MultiCloud resources: done, being published
- D-TA6-WP5-7 (30 Sep 2026)
  - Standard analysis software in JupyterHub
- D-TA6-WP5-8 (31 Dec 2023)
  - Services via API: ➔ changed to continuous
- D-TA6-WP5-10 (30 Jun 2024)
  - FTS and Rucio: in progress

Work Package 5

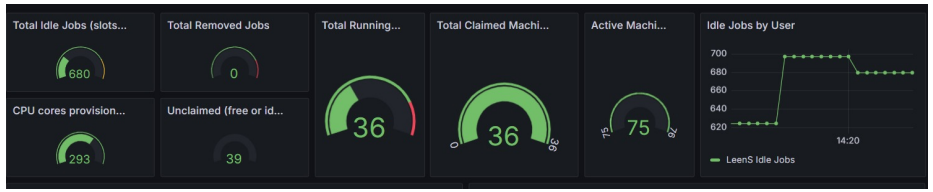
5 deliverables left to be completed till end of PUNCH (1):

- D-TA6-WP5-3 (due Dec 2025):
  Gen-Z consortium was developing an open standard for memory-based computing, being explored and offered in this WP as a solution, in particular for astronomy use cases in the CASA framework.
  Gen-Z was replaced by Intel CXL (Compute Express Link), which requires some porting of CASA (ongoing).
  Worked on by: HTW Berlin

- D-TA6-WP5-4 (due Dec 2025):
  PUNCH4NFDI EXPLORE service offering compute resources (hosted at Uni Göttingen) for LHC open data analysis – *service is active*
  Currently offered via own registration website and login – sorting out obstacles with PUNCH AAI ongoing
  Worked on by: Baida A.

- ...

Public Access to CERN Open Data via GoeGrid — EXPLORE



- As of now, a total of **3,965 HTCondor clusters** have been submitted for ATLAS Open Datasets processing on **EXPLORE**.
- Each cluster contained between **9 and 29 individual jobs**, depending on the performed analysis.
- This results in an estimated total of **35,685 to 114,985 jobs** executed.
- These jobs were designed to process discrete samples in parallel, enabling scalable and efficient use of available compute resources

- **Aim:**
  Provide public access to GoeGrid resources for CERN Open Data analysis, available since Q4 2024 **(TA6-WP5-4).**
- **Service Status:**
  Operational since November 2024, fully optimized after alpha/beta testing with 12 users and HEP Masterclass students.
- **Target Users:**
  - Students (school and university level)
  - Citizen scientists and hobbyists
  - Independent and external theorists
- **Access & Authentication:**
  - Email-based registration with SSH key login; no third-party authentication required.
  - Under/over-18 flag and parental consent required for minors.
  - Consent to Terms of Use mandatory.
  - Account lifetime is 3 months, after which the user data is deleted.
- **Resources & Support:**
  - Dedicated compute resources via University of Göttingen
  - Tutorials for HZZ, TTbar, and Hyy ATLAS Data analyses, including job execution and data visualization. Additional tutorials planned!
  - Materials hosted publicly on PUNCH GitLab repository

EXPLORE User Outreach & Growth

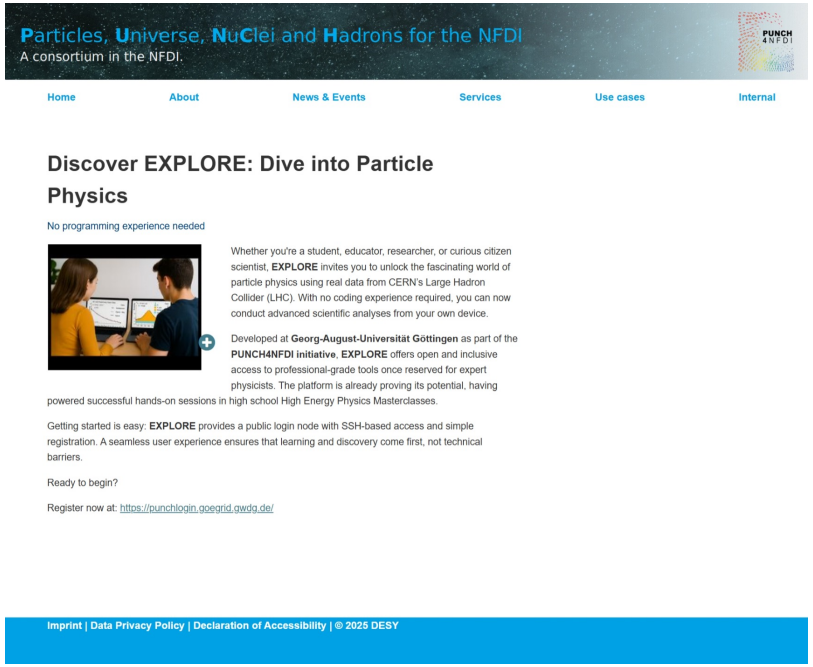
• **Promoted on the PUNCH website** and featured in the **NFDI newsletter (May 2025).**

• Scheduled **presentation at** [CoRDI 2025 in Aachen 26-28 August:]
"EXPLORE: A Scalable Infrastructure for LHC Open Data Analysis and FAIR Data Provisioning" (Track: RDM Infrastructures).

• Considering integrating AAI for EXPLORE to boost user uptake, **but modifications to the AAI are needed before use**; growing the base of regular users remains a key challenge.

• We encourage community members to help advertise the service and invite potential users to register and explore the platform at: EXPLORE Registration.

•PUNCHLunch planned in September 2025, reserve the date



EXPLORE

@ Georg-August Universität Göttingen

Access and Analyze Real High-Energy Physics Data from the LHC Experiments at CERN.
Gain insights into particle physics, learn from tutorials, and contribute to scientific discoveries.

FREE OF CHARGE SERVICE TO SUPPORT FAIR SCIENCE!

PUNCH 4NFDI

Provided by PUNCH4NFDI

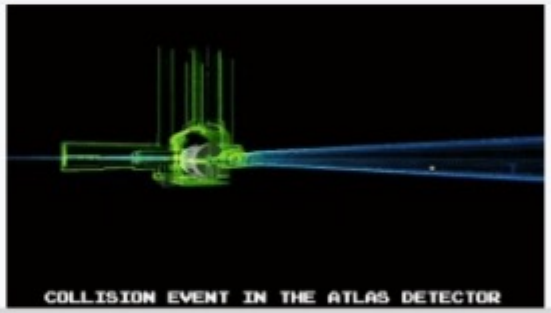
PUNCH4NFDI, the Nationale Forschungs-Daten Infrastruktur (NFDI) Consortium of Particle, Astro-, Astroparticle, Hadron and Nuclear physics, is dedicated to making research data FAIR: Findable, Accessible, Interoperable, and Reusable.

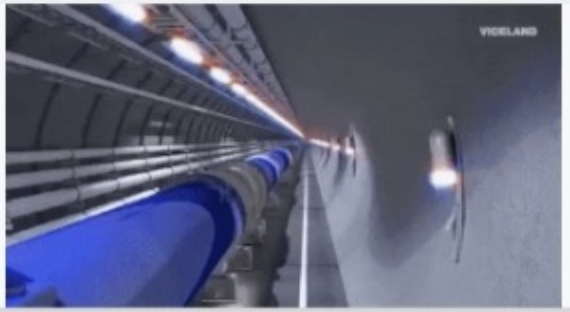
About the Service | **Tutorials** | **Give a Feedback**

Are You a Student, Teacher, or High-Energy Physics Enthusiast, with no Affiliation to an LHC-CERN Experiment or Institution?

⏩⏩ Register for our service and start exploring High-Energy Physics! 😀

Register Now





Privacy Impressum Contact Us

Work Package 5

- ...

- D-TA6-WP5-7 (due Sep 2026):

  Integration of standard analysis software and newly developed tools into the JupyterHub platform at University of Münster via use case specific notebook images. Currently upgrade of the JupyterHub-Cluster.
  Service and images have already been used for classes of the PUNCH Young Academy, in particular.

  Worked on by: Michael W.

- D-TA6-WP5-8 (due Sep 2026 / continous):

  Discussion of AUPs and related policies and identification of corresponding texts (for services accessed centrally via the PUNCH-AAI)

  Worked on by: WP5/interested people, Harry E. & Sebastian W. as WP-Leads, resource providers

- D-TA6-WP5-10 (due June 2024)
  - Rucio and FTS instances in place
  - Configuration to a large extent in place
  - Extensive testing still do be done
  - Worked on by: TA6/TA2: K. Schwarz, C. Wissing

TA6 related delivery reports

Deliverable reports

- D-TA2-WP2-1, D-TA2-WP2-4: Demonstrator for a federated compute infrastructure
- D-TA5-WP1-1: Report on impact of on-line filtering on discovery potential
- D-TA5-WP2-1: Curation & metadata schemes for Dynamic Filtering
- D-TA5-WP2-2: Strategy concept for identifying highly complex (multi-parametric) signals in huge data streams
- D-TA5-WP3-1: Specifying the concept of a Dynamic Archive
- D-TA6-WP1-2 and D-TA6-WP4-5: physics.tools
- D-TA6-WP1-3: Synergies on cross-cutting topics within NFDI
- D-TA6-WP5-2: Dynamic disk caching for opportunistic resources
- D-TA7-WP1-2: Establishment of the PUNCH4NFDI Young Academy
- D-TA7-WP1-4: Critical Review of the Measures Developed and Development of a Feedback System to Guide Education and Training

Acknowledgements

This work was [in part] supported by DFG fund „NFDI 39/1“
for the PUNCH4NFDI consortium.

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation