TA3 Status Report

Joe Mohr, Thomas Kuhr
for TA3 Colleagues

TA3 Work Packages

1. Statistical methods
    a. HEP: Cornelius Grunwald (TU Dortmund)
2. Numerical methods and simulations
    a. Astro: Frank Wagner (JFZ)
    b. Lattice: Tilo Wettig (Uni Regensburg)
3. Machine learning
    a. HEP: Anna Hallin (Uni Hamburg)
4. Methods for analyses across datasets
    a. HEP: Thomas Kuhr (LMU)
    b. Astro: Aditya Singh (LMU)

WP 1: Statistical Methods

Lead: Cornelius Grunwald

WP 1 - Goal: develop & provide statistical tools needed for the analysis of large data sets and complex models

Deliverables:

D-TA3-WP1-1 (30 Sep 2026): Statistical inference in the limit of large datasets and highly parallel computing.

D-TA3-WP1-2 (30 Sep 2026): Integration of a broad set of statistical methods; further development of a subset of methods into a common set of cross-community tools.

- Plan to achieve deliverables by releasing a new major version (v.4.0) of the BAT.jl tool for Bayesian data analysis with updated functionalities for more parallel computing and new sampling algorithms for more complex models
- Plan to examine accessibility of publicly available statistical methods for PUNCH users (w/ TA6)
- People funded:
  - Cornelius Grunwald, 50% (until end of PUNCH 1.0)
  - Nicola Malavasi (~25% 1st year)

WP 2: Numerical Methods and Simulations (Astro)

Lead: Frank Wagner and Tilo Wettig

Deliverables:

D-TA3-WP2-1 (30 Sep 2026): Optimisation of performance-critical routines entering data analysis and simulation software on GPU systems, heterogeneous compute clusters and upcoming processor generations.

D-TA3-WP2-2 (30 Sep 2026): Provision of data-/compute-heavy algorithms with a focus on algorithmic/technical aspects of scientific reproducibility (resiliency, uncertainties).

- Objectives until end of PUNCH 1.0:
  - Extend the DESTINY database for Solar System dynamics by adding new simulation data and automatic data publishing tools
  - Port, benchmark, and optimize the plasma community code PEPC on the exascale computer JUPITER
  - Publish a report on FAIR astrophysical simulations & data
- People funded:
  - Until end of PUNCH 1.0: Frank Wagner, FZJ

WP 2: Numerical Methods and Simulations (Lattice)

Lead: Frank Wagner and Tilo Wettig

Deliverables:

D-TA3-WP2-1 (30 Sep 2026): Optimisation of performance-critical routines entering data analysis and simulation software on GPU systems, heterogeneous compute clusters and upcoming processor generations.

D-TA3-WP2-2 (30 Sep 2026): Provision of data-/compute-heavy algorithms with a focus on algorithmic/technical aspects of scientific reproducibility (resiliency, uncertainties).

- Objectives until end of PUNCH 1.0:
  - Extend functionality of qcd_ml and SIMULATeQCD libraries
  - Continue benchmarking activities and tuning of algorithms on new hardware architectures (e.g., H100)
  - All deliverables on track to be completed
- People funded:
  - Until end of Punch 1: Daniel Knüttel (Regensburg), Giovanni Pederiva (FZJ)
  - Earlier: Nils Meyer, Christian Schmidt (55%, 01/22-09/24), Simran Singh (04/23-09/24), David Clarke (01/22-12/22), Basavaraja Bheemalingappa Sagar (09/24-01/25)

WP 3: Machine Learning

Lead: Anna Hallin

- Deliverables
  - D-TA3-WP3-1 (30 Sep 2026): Framework for AutoML on scientific data based on the PUNCH domain **changed** to Foundation models for high energy physics [10.1088/2632-2153/ad66ad, https://doi.org/10.48550/arXiv.2412.10504, http://doi.org/10.25592/uhhfdm.16505, https://doi.org/10.48550/arXiv.2501.05534] (UHH-particle)
  - D-TA3-WP3-2 (30 Sep 2026): Tools and solutions for distributed learning using very large datasets: publication on self-supervised models to explore sky surveys plus the accompanying code is underway (UHH-astro); Star clustering algorithm (Jülich). This addresses the usage of large datasets, but will not address distributed learning.
- People funded
  - Nicolas Baron Perez, Universität Hamburg
  - Anna Hallin, Universität Hamburg
  - Maik Sowinski, FZJ, funding ended in 2025

WP 4: Methods for Analyses Across Datasets

Leads: Thomas Kuhr and Aditya Singh

- D-TA3-WP4-1 (31 Dec 2024): Framework for conversion/reading of data for combined analyses; implementation of selected conversion/reading methods on heterogeneous systems.
  - Plan to set up joint analysis after demonstrating code-to-data (with TA2)
- D-TA3-WP4-2 (30 Sep 2026): Tools to define, test, and execute scalable workflows; library of template workflows.
  - A Workflow Management System Guide (arXiv:2212.01422), REANA (TA4)
  - Employ DRP registry to save template workflows (with TA4)
    - DISPERSE and MeerKAT radio imaging DRPs, joint analysis DRP, others?
- D-TA3-WP4-3 (31 Dec 2024): Standard interface for the publication of likelihoods, including a catalogue for the definition of common parameters.
  - Constructing model-agnostic likelihoods, a method for the reinterpretation of particle physics results (EPJC 84 (2024) 7, 693)
  - Joint HEP Astro analysis under discussion
- People funded
  - Aditya Singh, LMU (through mid 2026)
  - Lorenz Gärtner, LMU (through mid 2026)