PUNCH-2.0 – (preliminary) Toolbox working groups (WG)

WG 1: Software Development as a Service (SDaaS) - for existing workflows

- Continuous Integration / Continuous Development (CI/CD)
- Automated testing
- Container generation (in cooperation with "federated infrastructures", important: security)
- PUNCH Quality (PunQ) Award: gold, silver, bronze
 - Criteria: status of repository, containerization, use of infrastructures, documentation, ...

WG 2: ML as a Service (MLaaS)

- Optimization of existing AI-based workflows for use in HPC
 - GPU cluster, parallelization of training
- ML in real-time computing (FPGAs, re-training in real-time, ...)
- Analysis of huge data objects (memory-based computing)

WG 3: Curation

- Long-term support of workflows (Astroparticle / Future Collider / Radioastronomy / ...)
 - In cooperation with long-term support of archives (and experiments)
 - Container registry (in cooperation with "(meta) data management")

WG 4: User and Administration Tools for the medata/file catalog

- WGUI and command line tools for easy searching, accessing, uploading data and metadata
- Tools and Libraries for (automatic) generation, preparation and uploading of (meta-)data
- Administration and monitoring tools

WG 5: Training & Teaching

- Summer schools (on selected topics from WG 1-3)
- consulting service to facilitate the application of ML in astrophysics/particle physics research
- support for the use of large language models (LLMs)

PUNCH4NFDI Toolbox | Olaf Kaczmarek, Bielefeld University | March 04, 2025 | DESY Hamburg

Machine Learning in PUNCH 2.0 (Hermann Heßling)



Challenges for PUNCH (Gernot Meier) Punch4NFDI - Toolbox



- Data is only useful if it can be converted to physical quantities
- Two main challenges for PUNCH:
 - Intermediate data levels often in instrument specific data format: preserve / modernize interfaces
 - response functions required from (from MC simulations): preserve software & instrument models

DESY. | Presentation Title | Gernot Maier

Uni Hamburg:

- service for researchers unfamiliar with machine learning, services in ML data analysis.
- specialized consulting service to facilitate the application of ML in astrophysics/particle physics research.
- offer expertise in important areas such as data pre-processing, model selection and evaluation, algorithm implementation, and the interpretation of results. The service will operate as a collaborative hub.
- Workshops and training sessions will be conducted to enhance the computational skills of researchers, fostering a self-sustaining, knowledgeable community.
- offer support for the use of large language models (LLMs). We will test/tune/share models and act as consultants for the use of LLMs in teaching (e.g. specialised chatbots) and research (e.g. interface for data inference, statistics and plotting).

- has been organizing AI block courses and knowledge transfer events as part of EXC ORIGINS.
- has experience in large-scale data processing with e.g. dataframe based / pyhton tools, where we can put in more development work

DZA:

• will provide R&D on algorithms for selecting rare events in real-time out of huge astronomical data streams, for analysing huge data objects, for data visualization, and for Smart Green Computing.

FZJ:

• will contribute to **ML methods in general**, in particular in **optimizing the use of HPC computing** (GPU-computing and energy considerations).

FZJ, DESY, UBi and UR:

 will contribute the development of the user and administration toolbox for a modular distributed (meta-)data management, including GUI.

PUNCH4NFDI Toolbox | Olaf Kaczmarek, Bielefeld University | March 04, 2025 | DESY Hamburg

DESY:

software as a service for critical software infrastructures:

Software and workflows are critical parts of the research infrastructure and strongly intertwined with data.

framework for identifying and supporting and further **improve software tools**. Examples: Key4hep (future collider software; included tools like LCIO, Marlin, DD4hep) astropy (base package for astronomical community); CORSIKA (air-shower simulations);

A software layer is required to allow an **experiment-independent data acquisition and management framework** at DESY in the context of the local experiments, supporting seamless data storage, aggregation, and metadata handling for experimental detector and control systems. **Workflows using standardized workflow languages** (e.g., CWL) allow to share and connect research data,

software, and services.

contribute to **scientific workflow development** (languages and engines) and to provide a powerful platform for workflows (e.g., using the REANA workflow engine connect punch storage and compute resources). PUNCH benefit: Services (built on Punch4NFDI 1.0): workflow engine (e.g., REANA), version control (git), CI/CD pipelines (with access to Compute4Punch and Storage4Punch), package/container managers, documentation hosting

DESY:

improved build and test infrastructures for PUNCH software projects:

Analyze the existing build, test, and deploy infrastructure used by the dCache project;

Analyze centrally provided services by DESY for migration to the new infrastructure and identify required functionality; Harden DESY (or PUNCH) **Gitlab and Kubernetes services** for evergrowing demand on software development and testing infrastructure, etc.