Particles, Universe, NuClei and Hadrons for NFDI

Update on current technical implementations and work in progress

TA2 – Data management

- Storage4PUNCH: prototype with three instances
  - Two implementations: **dCache**, **XROOTD**
  - **Token** based authentication
  - File catalog candidates, e.g. **Rucio**, are being evaluated
- Compute4PUNCH: dynamic integration (**Overlay Batch System**) ot two sites
  - Single point of entry: traditional **login node** (available), **JupyterHubs** (in development)
  - Container registry available(Docker): CI pipeline to build and push in parallel
- Metadata Catalog: **International Lattice Data Grig** (ILDG)
  - Web services implementing a distributed “database”
  - Development to handle multiple metadata schemas

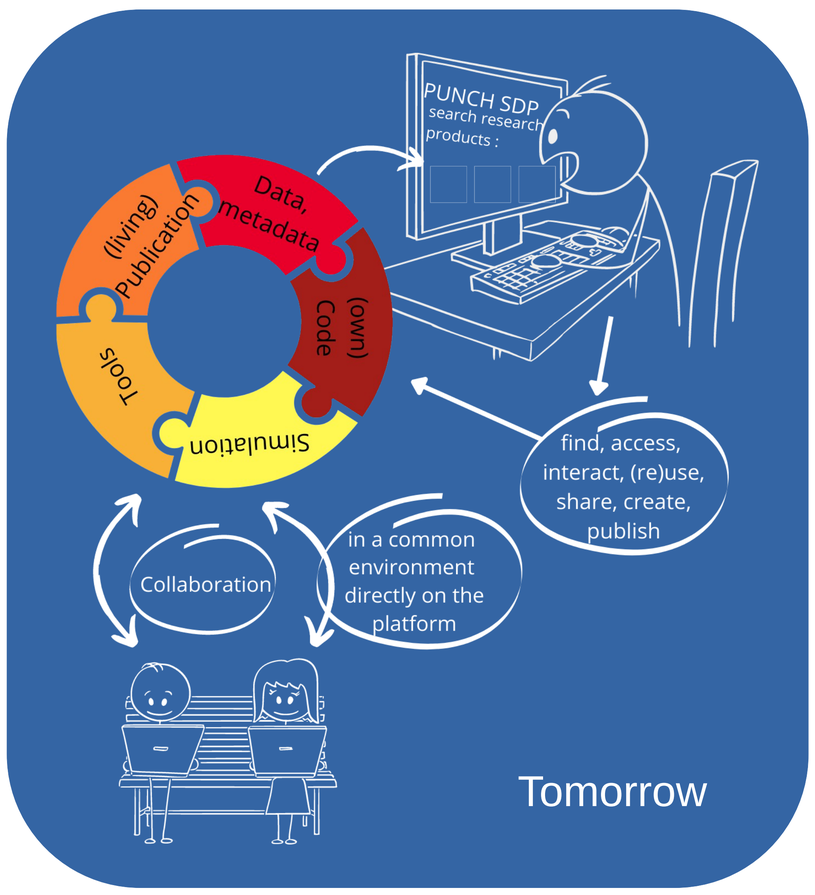
TA3 – Data transformations

- Development of tools for **statistical analysis**
  - Contribution to **BAT.jl**, e.g. on **neural spline sampling** and the python interface **BATty**
  - **GraphNeT**: the general neutrino telescope software
- Simulation codes:
  - **Benchmark testing** for several commonly used codes
  - Setup of codes and provisioning of makefiles at Jülich Supercomputing Centre
- Development of **automated ML** solutions
- Exploitation of ML algorithms
  - **Generative Networks** for radioastro surveys and **anomaly detection** for astro images
- First **implementation of containerized workflows** on the Compute4PUNCH infrastructure

TA4 – Data portal

- Prototype of **data portal** to the Science Data Platform (SDP): web interface
- Assembly of **elements of portal services**
  - **Reana**: workflow engine
  - **Docker** and **Kubernetes** infrastructures
  - Gitlab (CI and Registry for Docker images)
  - results.punch4nfdi.de (publication of documents, guides, preliminary results)
- Prototype design of **Dynamic Research Products** (DRPs): from machine-executable program to data analysis workflows
  - Metadata schemas
  - Persistent IDentifiers
  - Capturing workflows with **SciTrace**

    => Workshop at AIP
- Prototype design of the PUNCH **DRP database**
- Prototype metadata describing the **interaction of software and data**
- Development of **technical interfaces** to external resources



TA5 – Data irreversibility

Real-time workflows

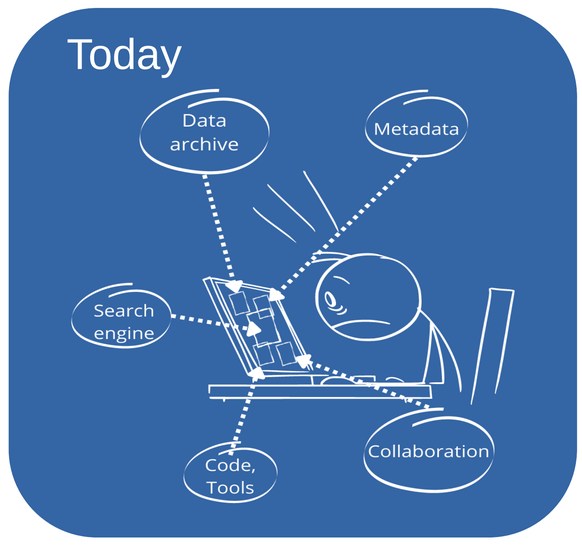
- Comparison of data (experimental data and simulations) with other data and theories, in the presence of **irreversible information loss**
- Concepts and environments for **identifying highly complex signals in huge data streams**
- Interconnection **between dynamic filtering** and **dynamic archiving**
  - Dynamic filter
    - time-dependent process deciding whether some data is kept or discarded
  - Dynamic archive
    - Product of a dynamic filter and/ or used to create dynamic filter
- **Neural networks on FPGAs** for high-throughput processing => *Workshop with EuXFEL at DESY*
- **Scalable workflows** with parallelizable code
  - Applications in simulations for pulsar searches
  - Pattern recognition in high energy physics

TA6 – Synergies & services

- Development of a **Software Marketplace** in close collaboartion with partners in the NFDI
- **Software products** used in PUNCH sciences
  - ArXiv study on commonly used software
  - Creation of a **repository** for commonly used open software products
- Further development of **AAI** in close collaboration with NFDI IAM group
  - **Gap analysis**: e.g. token delegation, file access control
  - Composition of a **requirements document**
- Evaluation of tools for a **dynamic metadata catalog**: Unicore, XTENS 2, MASi
- Organisation of seminar series: **PUNCH Lunch**

High-level goal

An integrated prototype package of Dynamic Research Products including metadata, a science data platform and compute and storage resources coupled with single-sign-on (AAI).



TA7– Education, Training, Outreach, Citizen Science

- **PUNCH Young Academy**: training for PUNCH PhD students & Postdocs
  - Organisation of soft skill workshops for career development in collaboration with **LHC ErUM-FSP office**
  - Support of female scientists
- White book with recommendations for curriculum design concerning **research data in physics studies** => *phone interviews*
- Support of **Machine Learning Masterclass** in collaboration with **Netzwerk Teilchenwelt**

  => *preparation of material & documentation*
- Presentation at **MS Wissenschaft**
- Conversion of **pulsar data** to make them accessible for **Citizen Science**

Contact

Are you interested in sharing expertise? **Get in touch!**

info@punch4nfdi.de – christiane.schneide@desy.de

nfdi Nationale Forschungsdaten Infrastruktur

DFG Deutsche Forschungsgemeinschaft

@punch4nfdi@nfdi.social

@punch4nfdi