The PUNCH4NFDI Consortium

Particles, Universe, NuClei and Hadrons for the NFDI

Thomas Schörner (DESY)
DPG SKM 2021 Meeting, 28 September 2021
Nationale Forschungsdateninfrastruktur (NFDI)

- Sustainable utilisation of research data
- Establishment of FAIR data management
- Connection to European and international efforts (like EOSC)
- Bottom-up approach of independent consortia

See also DFG.de/nfdi and nfdi.de
See also DFG.de/nfdi and nfdi.de
Particles, Universe, NuClei and Hadrons for the NFDI
PUNCH4NFDI Genesis

- 2018
  - First contact
  - Ext. abstract
  - Letter of intent

- 2019
  - Full proposal
  - Evaluation
  - Protocol
  - Expert vote

- 2020
  - Letter of intent
  - Full proposal
  - Evaluation
  - Protocol
  - Expert vote

- 2021
  - Final vote GWK

- 2022
  - Start of funding

Astronomy

Astro@NFDI

Particles, Universe, NuClei and Hadrons for the NFDI

Particles
Hadrons & nuclei
Astroparticles
Our Partners

Universities, Helmholtz, Max Planck, Leibniz
PUNCH data are diverse
- in size and rate
- in complexity and purpose
- in abstraction level

PUNCH4NFDI expertise
- Big data and open data
- Data irreversibility and reduction
- Harnessing heterogenous resources
- Highly collaborative globally distributed data management
Challenges
- FAIRification: data & workflows
- “big data” and “open data”
- Irreversibility challenge and data loss
- heterogeneous data & infrastructures
- transfer of knowledge
Task Areas

TA 1: Management and governance

TA 2: Data management

TA 3: Data transformations

TA 4: Data portal

TA 5: Data irreversibility

TA 6: Synergies & services

TA 7: Education, training, outreach, citizen science
Flexibility, efficiency, scalability:
data volumes, number of users and analyses, heterogeneous resources; data combinations
TA 2: Infrastructure Evolution
Access to data, federated computing, automation, data lake prototype

Today
- Grid site CPU+Storage
- Data archival Storage
- Grid site CPU+Storage
- HPC

Future
- Compute site
- Storage
- CPU + storage
- HPC
- Data Lake
- TA 4 Data portal

Now: very heterogeneous; different approaches for communities
HEP: >170 HTC-based grid centres very community-specific
Astro: local, isolated data archives

PUNCH: Generic solutions with standardised protocols for archive / compute sites, suited for “all” communities
Globally distributed data lake with large storage and compute resources and portal access
Opportunistic resources in federated science cloud
Digital Research Product

- Access via portal
- Interaction with other RPs
- Interfaced to tools and infrastructures

Collaboration data repositories

Compute resources, Dockers, clouds, storage

Astropy, ASCL, DebianAstro, FAIROOT, ...

arXiv, journals, ADS, open access, ...

Github, gitlab(s), ROOT, CASA, ...

DOI, IVOA registry, data&resource provider,

AAI, interactive environments, collaborations

Public data repositories

RP

RP

RP
Task Area 3: Data Transformations
Integration of common tools into a data infrastructure based on code-to-data principle
Provision of tools for parallel processing of huge data sets on heterogeneous resources

User
Selection of tools / workflows via portal

Data and metadata

Scientific result

Tools common to many science fields

Image processing  Particle selection  Data modelling

Statistical methods  Numerical methods + simulations  Machine learning  Across-dataset methods

Data and metadata

Statistical methods
Numerical methods + simulations
Machine learning
Across-dataset methods
PUNCH-SDP

The science data platform for RPs

Research product contains executable workflow

Metadata description of workflow

Code / tools

Computing / storage

Many options for the user

Execute existing workflow

Learn, (ex)change, add, create

Link & use ext. sources

Data set(s)
Simulation
Tools
Code
Metadata
Publication

Portal

AAI

Search & find
Access
Modify & add
Use & compute
Create & Store

RP 1
...
RP N
New RP
MaRDI: micro-metadata schemes, statistical procedures, use cases

NFDIxCS: easier access to and use of HPC centres

NFDI4Ing: archetypes, terminology services, work on basic services, AAI, ...

NFDI4Microbiota: use cases, AAI and cloud standards

NFDI4Culture: legal aspects of AAI and FAIR data management

NFDI4Chem: common work on metadata & data resusability

DAPHNE4NFDI & FAIRMat & DPG: address physics-specific aspects

Marketplace, PUNCH-SDP, data portal

Knowledge fabric, digital research products, metadata services

AAI infrastructure, dynamic disk caching

Big data management & data storage services

Machine learning services and real-time applications

IT resources via Compute4PUNCH interactive analysis interface

Cloud-based testbed

Teaching and education

(Non-exclusive examples)
PUNCH4NFDI and Epidemiology

A first example for „outreach“ beyond our PUNCH sciences

PUNCH statistics software (BAT.jl and Geneva) being applied to established epidemiology studies (V. Priesemann et al.) – collaboration on further development of epidemiological models

PUNCH expertise planned to be applied for corresponding data management solutions

---

**The challenges of containing SARS-CoV-2 via test-trace-and-isolate**

Sebastian Contreras¹, ², X, Jonas Dehning¹, X, Matthias Loidolt¹, X, Johannes Zierenberg¹, F. Paul Spitzner¹, Jorge H. Urrea-Quintero¹, Sebastian B. Mohr¹, Michael Wilczek¹,³, Michael Wibrul¹, and Viola Priesemann¹,³

To calculate $N_{\text{cum}}^{0n}$ in terms of the model’s parameters, we insert equations ⁴–⁶ into equation ⁷:

$$N_{\text{cum}}^{0n} = \lambda_n H_n (1 + \eta R_0) \left( 1 - \frac{1}{1 + \nu R_0} \right).$$

---

Fit, EA

---

CSEIS

0 100

0 300 500

$N_{\text{cum}}^{0n}$, data

$N_{\text{cum}}^{0n}$, mode expectation

0 25 75 100

t

PUNCH4NFDI  |  DPK SKM Meeting  |  28 September 2021  |  TS
Training Experts

1. Training in the **application of the tools and practices** developed within PUNCH4NFDI
2. Coordination and organization of specific **training measures and events** on NDFI topics for physicists. Promotion of measures for **career advancement** in the data sciences. Emphasis on the promotion of women.
3. Critical review of the measures developed and development of a **feedback** system to guide education & training
4. Documentation and long-term **archiving** of training material, the service documentation and tool descriptions in coordination with the TIB Hannover

Educating Students

1. Market **survey** of available teaching concepts and material
2. Development of **standardized curriculum**
3. Compilation and development of **teaching material** for courses and independent learning.
4. Aggregation of **data resources** and **access to computing infrastructure**
5. Coordination and initiation of **educational events**, e.g. visiting seminars.

Outreach and Schools

1. Website, blog, twitter, communication of PUNCH products
2. Networking, PO training, study on effectiveness of outreach, development of evaluation criteria
3. Central **access** to data and software for education and entertainment
4. Provide tutorials and **material** for teachers and students, teaching and learning material for masterclasses
5. Study and test **pedagogical approaches** to promote data science for children
6. Pilot extracurricular activities **(AG/Projektkurs)** in selected schools, promote changes in school curricula
7. Organize/support **hackathons**, rent a scientist, **masterclasses** (Netzwerk-Teilchenwelt), **VO-days**, etc.

Citizen Science

1. Map out potential research applications in **Citizen Science**
2. Prepare data sets and providing soft- and hardware infrastructure
3. Run pilot projects for 6-12 months by engaging schools and universities and evaluate results
4. Prepare and launch of further projects in collaboration with the physics community
PUNCH4NFDI: AAI + GitLab + Intranet + Mattermost + CloudShare

Collaborative Tools connected by PUNCHAAI

DESY CloudShare
- Sign in with PUNCH4NFDI AAI
- Remember me

PUNCH AAI
- https://login.helmholtz.de/punch/

PUNCH4NF DI Cloud

ORCID
GitHub
Google

DFN AAI
Helmholtz AAI
CERN AAI

CERN
UBonn
UBonn
“PUNCH” Lines

PUNCH4NFDI – an open consortium by the community for the community - and beyond

A leader in data management with expertise in big data, open data, FAIRification, use of heterogeneous resources

Use the experience, expertise & work programme to take data management to the next level and solve the problems that also others will have tomorrow
Thank you!

The PUNCH4NFDI Consortium

Spokesperson:
PD Dr. Thomas Schörner (DESY, thomas.schoerner@desy.de)
DESY, Notkestr. 85, D-22607 Hamburg

Contact:
Mail:   punch4nfdi@desy.de
Web:    www.punch4nfdi.de
Twitter: @punch4nfdi