

The PUNCH4NFDI Consortium

Newsletter Number 3

April 2022



Contents

- 1. News and Highlights***
- 2. NFDI and related topics***
- 3. Formal, legal and financial topics***
- 4. Status of the consortium***
- 5. Communication and collaborative tools***
- 6. Upcoming events and excitements***
- 7. Recent talks, results, and publications***

1. News and Highlights

“Open data” paper: Our first PUNCH4NFDI publication titled “Survey of Open Data Concepts Within Fundamental Physics” is out – the summary of the open data workshop held in early 2021. It has appeared in Computing and Software for Big Science and can also be found via Zenodo (just search for PUNCH4NFDI)¹.

High-level milestones: Following the publication of the detailed work programme with deliverables and timelines for all task areas and work packages, we have finally concluded on our high-level milestones for our work in the next few years². The milestones serve the dual purpose of reminding us of the main thrust of our work and of allowing for easy communication of the purpose of the consortium, also in view of future evaluations. In very short form, the milestones are:

- Milestone 1: An integrated prototype package of dynamic digital research products, the science data platform, and the Compute4PUNCH compute and storage resources (Storage4PUNCH) coupled with single sign on (AAI).
- Milestone 2: Data irreversibility solutions – a new kind of science in the age of too-large-to-be-stored data streams.
- Milestone 3: The PUNCH market place including notice board and communication tools as a place for efficient exchange of data management solutions, methods, services, standards, tools, protocols, and software is an offer to the NFDI as part of a federated system of discipline-specific market places.
- Milestone 4: Rolling out the PUNCH4NFDI outreach and education programme.

More information can be found in the intranet. We are now preparing a public document on our milestones.

Workshop on synergies, commonalities, tools etc.: On 5 April, a very successful workshop³ on commonalities and synergies across the physics-related and other NFDI consortia like DAPHNE4NFDI, FAIRMat, MaRDI, PUNCH4NFDI etc. was held together with representatives from the ErUM-Data initiative in order to discuss future directions in the collaboration, with more than 50 participants. All slides can be found in the INDICO agenda. The next step will be a series of more focused technical discussions of specific aspects raised during the workshop.

The **3rd general PUNCH4NFDI meeting** was held on 12 April, with reports from all task areas, the introduction of new people and some highlights. Have a look at the slides in INDICO⁴!

From 20-22 April, the DPG held a **workshop on “Forschungsdaten im Physikstudium“** („Research data in physics curricula“) in Bad Honnef⁵. PUNCH4NFDI was represented at the workshop – see below. The next step is a white book on ideas for data management in university studies.

¹ H. Enke et al., Computing and Software for Big Science (2022) 6,6; <https://doi.org/10.1007/s41781-022-00081-7>

² <https://intra.punch4nfdi.de/files/home/PUNCH-HighLevelGoals.pdf>

³ <https://indico.desy.de/event/33410/>

⁴ <https://indico.desy.de/event/33451/>

⁵ <https://www.dpg-physik.de/veranstaltungen/2022/workshop-forschungsdaten-im-physikstudium>

In case you have news that you want to share with PUNCH4NFDI, don't hesitate to send a mail to info@punch4nfdi.de. You can access all newsletters on our web site⁶.

2. NFDI and related topics

Base4NFDI: The Base4NFDI initiative of all already funded NFDI consortia has handed in a letter of intent for the DFG call for base services⁷. PUNCH4NFDI was involved in the Lol process, and Thomas Schörner was selected as one of the co-spokespersons of the Base4NFDI consortium initiative, with a responsibility for the task area 3 "Service Coherence Processes and Monitoring".

The next milestone for Base4NFDI is the full proposal – to be handed in to the DFG by 29 April. A decision on the funding of the consortium is expected by November 2022.

NFDI sections: So far, four NFDI sections have been established⁸, and a number of NFDI task forces have started working. PUNCH colleagues are engaged in all sections and most of their subfields as well as in the NFDI task forces.

A workshop to be held in May will decide on additional sections to be created within the NFDI association, and PUNCH4NFDI will need to decide on its participation in potential new endeavours.

Some information on the existing sections can be found in slides presented in the recent PUNCH4NFDI general meeting. In particular, all sections have recently elected new chairpersons, for terms of 2 years.

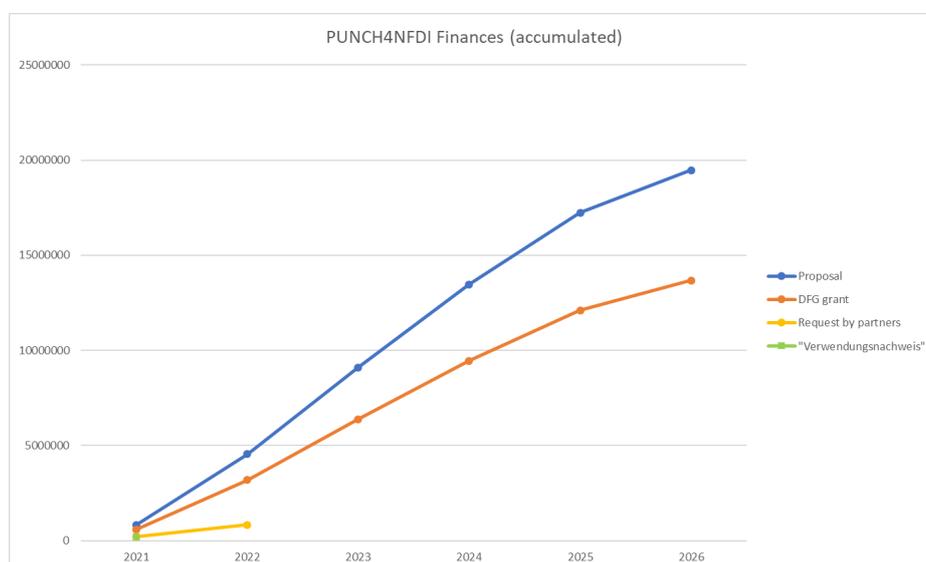


Figure 1: Spending profile of the PUNCH4NFDI fundw. Blue - proposal; orange - DFG grant; yellow - actual spending.

3. Formal, legal and financial topics

Figure 1 above shows the financial situation of the PUNCH4NFDI consortium as a function of time. Two effects become immediately apparent: the ~30% reduction of the

⁶ https://www.punch4nfdi.de/news_amp_events/newsletter/

⁷ https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen_2022/2022_base4_nfdi.pdf

⁸ <https://www.nfdi.de/sektionen/>

funding request in our proposal (blue) to the actual DFG grant (orange), and the serious underspending that we are experiencing (visible as the large difference between the orange and yellow curves).

All PUNCH4NFDI co-applicants are, therefore, working hard on filling the positions granted to them by the DFG, and on executing the tasks foreseen for them by the overall work programme. Funding requests for the second quarter of 2022 are due in the middle of May, and the financial contacts and principal investigators have been informed accordingly.

4. Status of the consortium

Reporting in PUNCH4NFDI

A first internal reporting cycle was started in late March, and the first quarterly report will be finalised in early May. Its aim is primarily to inform the SAB as our most important advisory body and to document our progress for later reporting and evaluation purposes. A template for the report – to be filled in by task area and work package leaders as well as by institute representatives – has been shown at the general meeting on 16 November 2021 in Thomas' presentation⁹.

New people

New people who contribute to the work programme of PUNCH4NFDI are our most valuable asset! In this category of the newsletter, we try to give short introductions of newcomers to the consortium – welcome to everybody!



Makarim Bouyahiaoui (MPIK, TA 3 / 6): I recently started as a new postdoc at MPIK in the Jim Hinton group. I will be involved in WP3.1, WP3.4, WP6.3 and WP6.4 for my postdoc duration of 2 years. My actual research field is in astroparticle physics, where I concentrate on cosmic-ray and gamma-ray astrophysics.
Contact: makarim.bouyahiaoui@mpi-hd.mpg.de

Christian Schmidt-Sonntag (Bielefeld, TA 3) received his Ph.D. in lattice QCD in 2003. After postdoc positions in Wuppertal, at Brookhaven National Laboratory and FIAS, he habilitated in Bielefeld in 2017. Christian is an expert in high-performance computing and code development and optimisation on GPUs. In PUNCH4NFDI, he will participate in porting and developing lattice QCD algorithms and conversation tools for data formats as foreseen in TA3. He will also participate in the development of appropriate meta data schemes for data products in WP4.
Contact: schmidt@physik.uni-bielefeld.de, [website](#)



⁹ <https://indico.desy.de/event/32298/contributions/113001/attachments/69833/88648/20211116.GeneralMeeting.pdf>

Maik Sowinski (FZJ, TA3/5): I am an astrophysicist (originally a mathematician by education), and I finished my MSc in Mathematical and Theoretical Physics at Bielefeld University in March 2022. Within PUNCH, I am now working in TA3/5. Recently, I started doing my Ph.D. at Forschungszentrum Jülich together with the University of Cologne. Here, I will be working on young star cluster classification at the Jülich Supercomputing Centre.

Contact: m.sowinski@fz-juelich.de



Caspar Schmitt (LMU, TA 3): I am developing software for high energy particle collider experiments. Currently, I am focusing particularly on how available software workflow management pipelines can be implemented in physics analyses. This serves to combine the many analysis steps into one well-documented and easily-reproducible workflow. This facilitates analysis reviews and preservation efforts considerably, among others.

Contact: caspar.schmitt@physik.uni-muenchen.de

Joeri Hermans (UHH, TA 3): My doctoral research at the University of Liège had a strong focus on machine learning for the physical sciences. In particular, my PhD focused on methodological advances in the field of simulation-based inference with applications in the field of astrophysics. Within PUNCH, I am responsible for the AutoML efforts in task area 3 as a Postdoctoral Research Associate.

Contact: joeri.hermans@desy.de



Marcel Trattner (HTW) grew up in southwest Germany. After backpacking through Australia, Asia and South America he moved to Berlin for his studies. He earned a bachelor's degree in International Media and Computing from the University of Applied Sciences (HTW) Berlin, and after that, he pursued a Masters in Applied Computer Science. In his major, he was involved in research projects that led him into the field of Big Data and he took the opportunity to join the PUNCH consortium.

Contact: Marcel.Trattner@Student.HTW-Berlin.de

Reports from task areas

The work in **task area 2** is centered around the setup of prototypes for storing and processing community data. Undoubtedly, there is tremendous experience in setting up and maintaining large scale data management systems within the PUNCH communities. However, we aim for the federation of existing infrastructures to provide the necessary resources for the PUNCH Science Data Platform and to enhance the

efficiency of the usage of the infrastructures. A first step in the prototype development is the consequent implementation of a token based AAI infrastructure that allows all PUNCH members to authenticate with the credentials of their home institution. Two sub instances have been setup in the prototype for distributed storage. One system located at DESY is a development instance based on dCache. The latter is also used in EU projects like ESCAPE and PaNOSC. The second system is based on the Xrootd technology and is located in Bonn. Data transfers between the two instances can be conveniently initiated via the HIFIS transfer service, which itself is based on WebFITS.

A demonstrator of the distributed Compute4PUNCH infrastructure is under construction. Initial computing resources at KIT and WWU Münster have been integrated using the COBaD/TARDIS resource manager. The integration of further resources is ongoing. The entry point is located at KIT and has been implemented as a server that allows ssh logins via AAI tokens. Containers are shipped as Singularity sandboxes via CVMFS. A closer integration of a container registry with the PUNCH Gitlab is ongoing.

All prototypes have first documentations on the internal PUNCH webpages and the first scientific applications are being ported to gain some experiences.

TA4 working groups have now established a regular meeting schedule. The four working groups have each made some progress with respect to their tasks:

- WP1 has added another use case, this time from astroparticles, to describe potential components of the DRP.
- WP2 has scrutinised various community Metadata schemas and compared these to the DataCite Metadata schema, evaluating their use with respect to the FAIR goals.
- WP3 had a common discussion round with WP2 and also discussed conversion between metadata formats (xml vs. json).
- WP4 worked out a first prototype format for the portal dashboard and collaborated with TA3 on a use case of an astrophysical simulation.

Additionally, various members of TA4 participated in discussions on the NFDI BaseService proposal and discussions in the NFDI e.V. sections.

The **first focus of TA5** is on concepts for defining metadata in order to capture dynamic workflows, enable validations and define interfaces. Such concepts for metadata in all workflows related to real-time data processing need to be established. In particular, a full description of underlying decisions which data to store has to be included in metadata.

For several upcoming experiments in the field of PUNCH sciences, a drastic increase of metadata volumes is anticipated. This will also require constant updates of “quality measures” of archived data and flexible data models (e. g. JSON data model). In order to evaluate performances related to the processing of dynamic metadata dedicated studies will be necessary.

In high energy physics hierarchical trigger schemes in hardware and software are established since years and extensions for more flexible workflows for online and offline processing are foreseen. One highly-relevant use case of anomaly-based triggering indeed requires extending existing sets of metadata: Detection of anomalous patterns beyond standard trigger schemes will be a particular focus of WP5

in TA5 with applications both in HEP and astrophysics. To this end, specific measures are foreseen to capture dynamic filtering/archiving workflows and in particular include relevant use cases of astrophysics modes of operation, e. g. time domain-based measurements, imaging, spectroscopy or polarisation.

In a previous study¹⁰ related to use cases of astronomical pipeline provenance, the extensive use of corresponding metadata has been identified as key to establish confidence within the detector systems, its final data products, and ultimately its scientific results, published in 13th International Workshop on Theory and Practice of Provenance (TaPP 2021).

As a starting point for further investigations, the cross-experiment data management system RUCIO is one option. Therefore, a dedicated PUNCHLunch with RUCIO experts is foreseen to discuss critical points.

Task area 6 targets cross-cutting activities that foster an exchange of concepts and developments among the PUNCH4NFDI community as well as with other consortia and the NFDI in general. The TA6 working groups have now established a regular meeting schedule and took up their work:

- WP1 has set up a marketplace prototype as an exchange platform for information on data management software, services, and cross cutting topics in general which is ready for testing. Moreover the deliverable D-TA6-WP1-1 (tools for communication) has been accomplished by having provided a set of collaborative tools¹¹.
- In WP2 a first milestone of the deliverable D-TA6-WP2-1 (Prototype of PUNCH-AAI) has been achieved by having set up the "Basic PUNCH AAI". Already many PUNCH users are registered with the PUNCH AAI and the PUNCH AAI is the standard method to access established PUNCH services as the PUNCH collaborative tools, the intranet and Storage4PUNCH. Moreover, first concepts of a Unity based group management have been implemented, allowing services to authorise users based on their group affiliations.
- WP3 has started to evaluate tools for dynamic metadata.
- WP4 (open source data analysis tools) has started their regular meeting schedule.
- WP5 has set up a list of initial service prototypes and identified responsible contact partners for each of these prototypes.

The TA has sent representatives to the NFDI task force "Tools" and to the NFDI section "Metadata".

5. Communication and collaborative tools

For more details on PUNCH4NFDI AAI and collaborative tools efforts, please refer also to the intranet and to the last newsletter.

Most of the facilities - still mostly intern to the consortium - are using the PUNCH-AAI for Authentication:

- Cloud storage: syncandshare.desy.de
- Intranet: intra.punch4nfdi.de

¹⁰ https://www.usenix.org/sites/default/files/tapp2021_johnson.pdf

¹¹ https://www.punch4nfdi.de/services/collaborative_tools

- gitlab: gitlab-p4n.aip.de
- chat channel: mattermost-p4n.aip.de
- Compute4PUNCH and
- Storage4PUNCH

With these services established, the other “A” in AAI – Authorisation – became an urgent topic. We now have established a concept for groups (and subgroups) that can be used by the resource providers (e.g. gitlab) to enable a more finely grained control for access and usage. The PUNCH intranet contains a more extensive description of this¹².

Additionally, Compute4PUNCH uses the gitlab registry and continues integration (CI) facilities for building and sharing the Singularity containers to external sites with compute facilities; in particular they they are accessible by the TARDIS system. For members of PUNCH4NFDI who develop workflows or software, the gitlab environment provides not only a CI facility, but also a S3-capable cloud storage in order to keep the Docker containers in local registry and share them to collaborators. Those are the first steps for adding elements to the PUNCH science data platform.

6. Upcoming events and excitements

- A complete list of PUNCHLunches can be found in the intranet and in INDICO: <https://indico.desy.de/category/743/>). If you have suggestions for the seminar series – let us know at info@punch4nfdi.de.
- For a complete list of TA and other working meetings, see the INDICO category <https://indico.desy.de/category/741/>
- NFDI InfraTalk on 2 May 2022: “Data management in the PUNCH sciences” (details coming soon)
- NFDI ToolTalk on 11 May 2022: AAI and collaborative tools in PUNCH4NFDI efforts so far (Harry Enke and Kilian Schwarz, more information coming soon)
- PUNCH4NFDI Annual (in person) Meeting: 28/29 September – either Hamburg or Göttingen.

7. Recent talks, results, and publications

To be found at least partly on the web page and in ZENODO (<https://zenodo.org> – just search for “PUNCH4NFDI” or – if you are interested in the broader scope – “NFDI”).

- DPG spring meeting Heidelberg – Andreas Haungs: The PUNCH4NFDI Consortium¹³
- DPG spring meeting Mainz – Kilian Schwarz: The PUNCH4NFDI Consortium
- H. Enke et al., Survey of Open Data Concepts in Fundamental Physics, Computing and Software in Big Science, (2022) 6,6; <https://doi.org/10.1007/s41781-022-00081-7>

¹² <https://intra.punch4nfdi.de/?md=/docs/Consortium/AAI-Doc/AAI-Group-Policies.md>

¹³ <https://zenodo.org/record/6383637#.YIPgdyORr3U>

- Workshop on “Forschungsdaten im Physikstudium“ („Research data in physics curricula“), 20-22 April in Bad Honnef¹⁴: The goal of the workshop was to develop strategies for incorporating data and IT skills into physics curricula, as these skills are becoming increasingly important in science and in the job market. The workshop was attended by relevant stakeholders, including representatives from the KFP (Conference of Physics Departments), the (Young) German Physical Society, physics-related NFDI consortia, libraries, and physics education practitioners. The discussions were divided into plenary and topical sessions and revolved around the desired data and IT skills, priority areas in education, formal aspects of the various curricula, and best practices. Among the many aspects addressed, programming and research data management skills, the role of electronic lab books, and statistics, simulations, and numerical methods were discussed in depth. The outcome of the workshop is a concrete plan to write a whitebook on IT and data skills in physics curricula to guide physics departments in developing their bachelor's and master's degree programs.

¹⁴ <https://www.dpg-physik.de/veranstaltungen/2022/workshop-forschungsdaten-im-physikstudium>