

The PUNCH4NFDI Consortium

# Particles, Universe, NuClei and Hadrons for the NFDI

Thomas Schörner (DESY)

DPG SKM 2021 Meeting, 28 September 2021





Gemeinsame  
Wissenschaftskonferenz  
GWK

**DFG**

Deutsche  
Forschungsgemeinschaft

## Nationale Forschungsdaten- infrastruktur (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

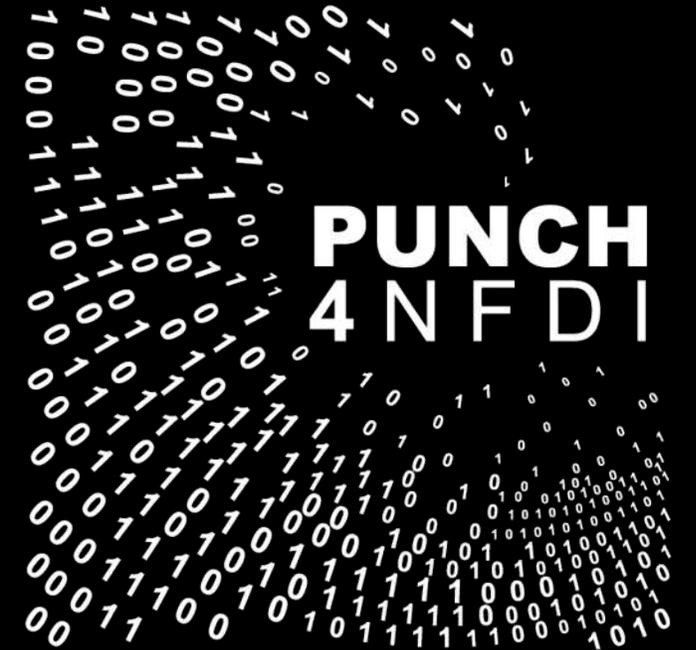
See also [DFG.de/nfdi](https://www.dfg.de/nfdi) and [nfdi.de](https://www.nfdi.de)



Gemeinsame  
Wissenschaftskonferenz  
GWK

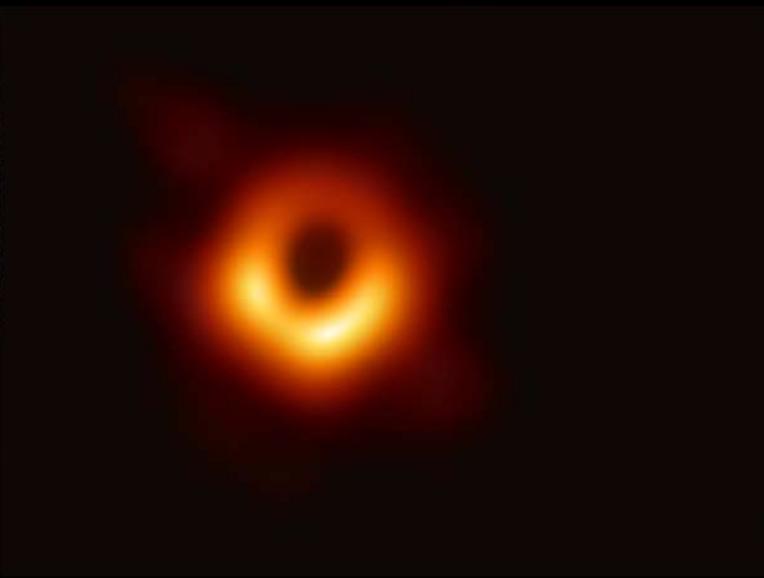
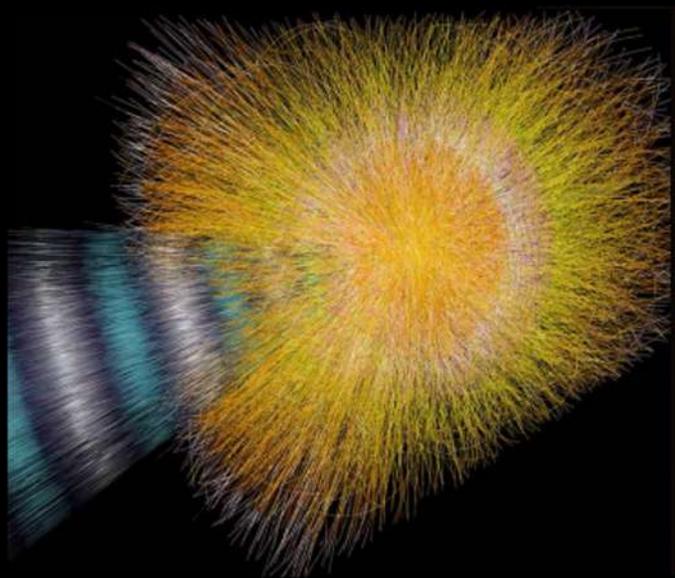
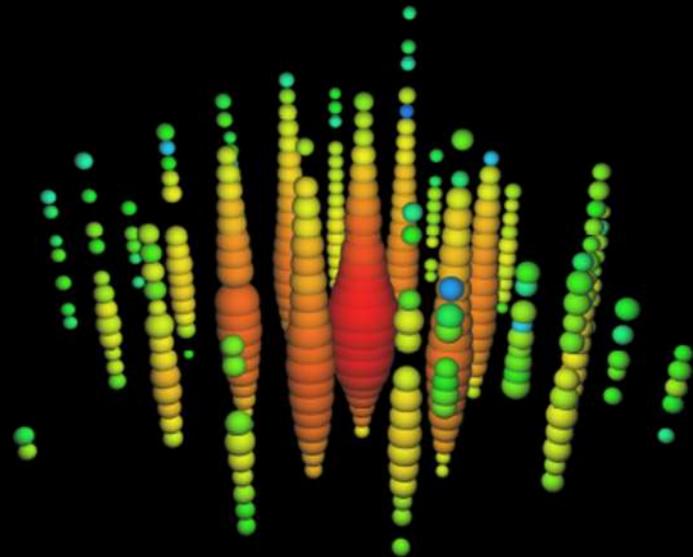
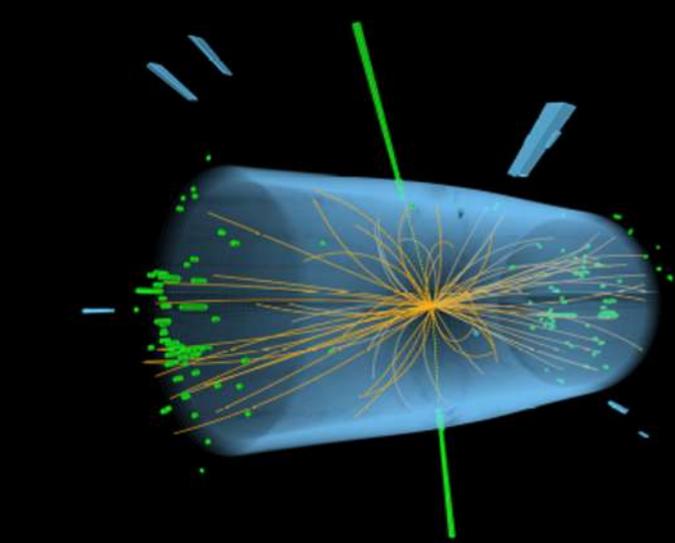
**DFG**

Deutsche  
Forschungsgemeinschaft



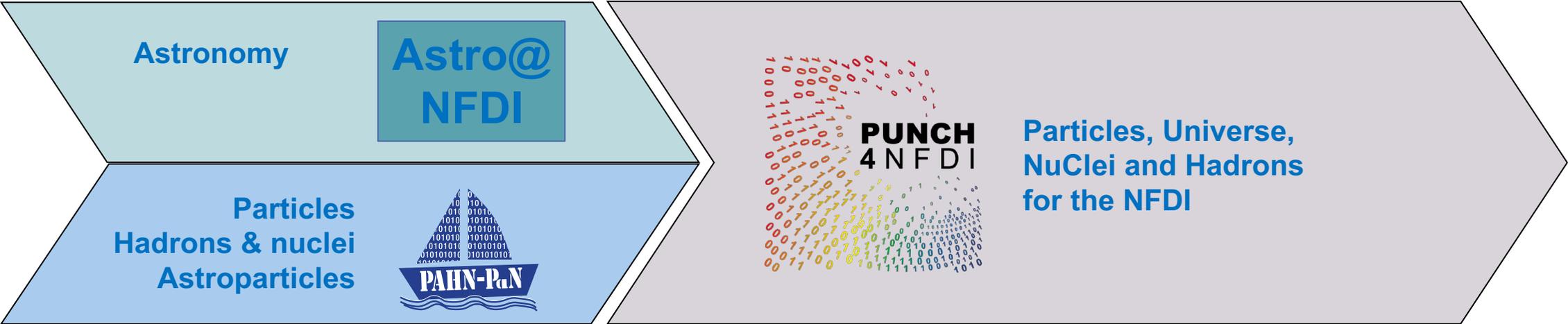
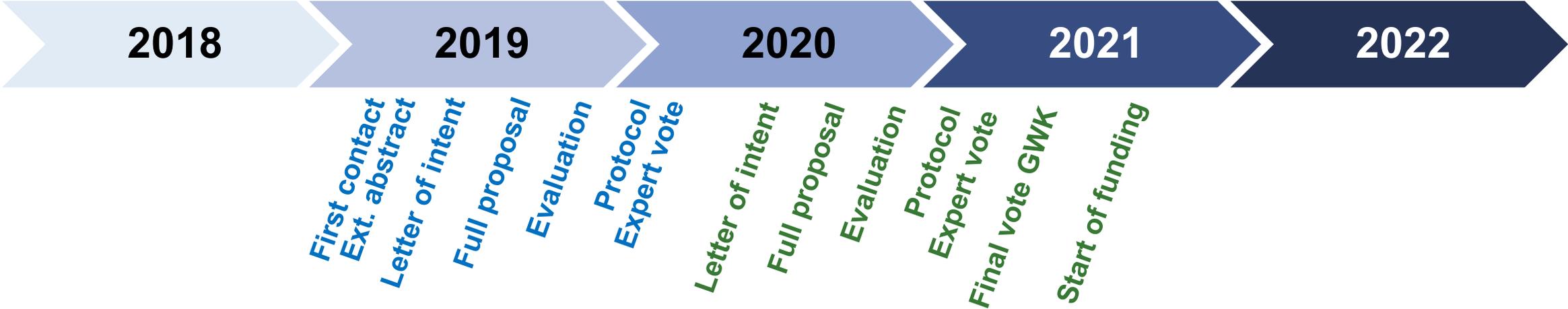
**Particles, Universe, NuClei  
and Hadrons for the NFDI**

See also [DFG.de/nfdi](http://DFG.de/nfdi) and [nfdi.de](http://nfdi.de)



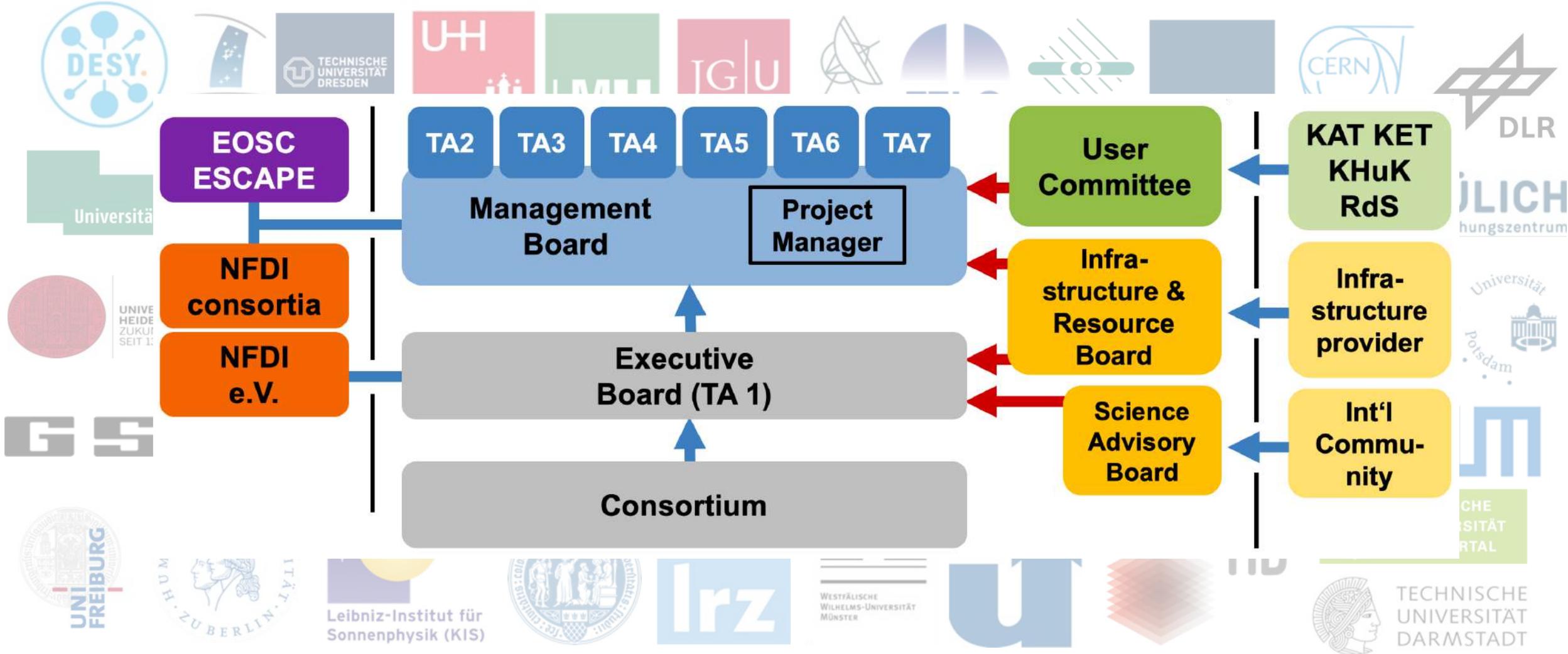
**Particles, Universe, NuClei  
and Hadrons for the NFDI**

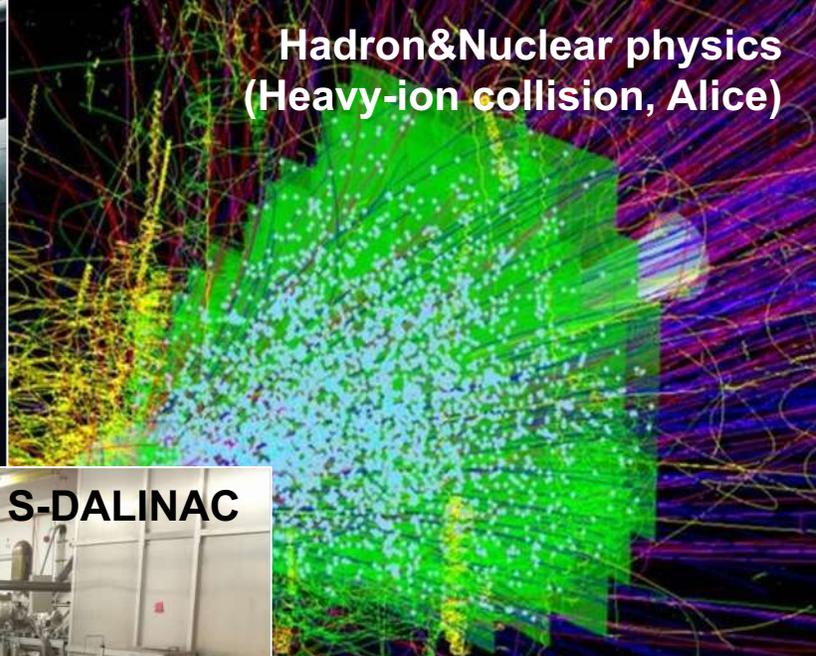
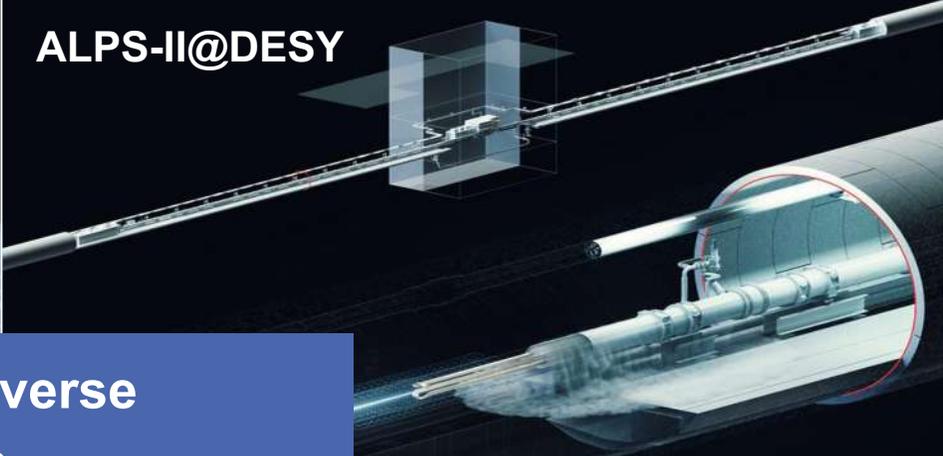
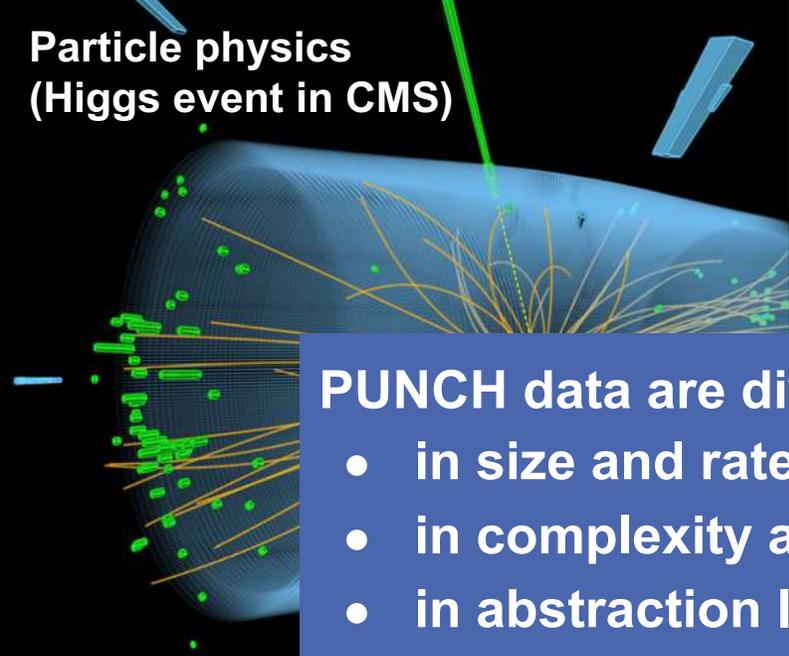
# PUNCH4NFDI Genesis



# Our Partners

Universities, Helmholtz, Max Planck, Leibniz





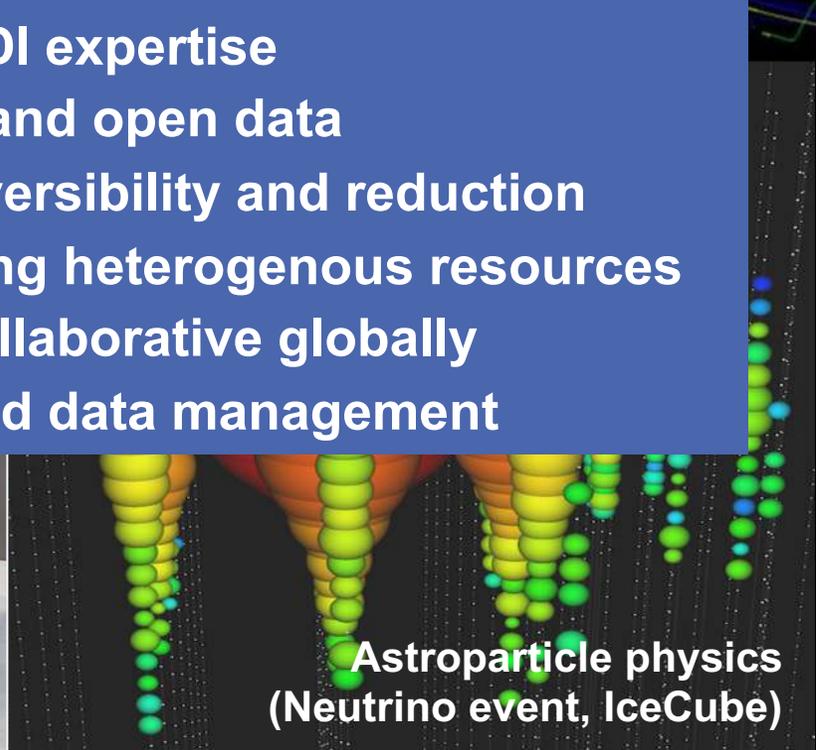
**PUNCH data are diverse**

- 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



data generated by SKA

Global Internet Traffic

FAIR@GSI  
Online data  
30 EB

FAIR@GSI  
Physics  
300 PB

searches on Google  
98PB

upload to facebook  
180PB

business emails sent worldwide



3,000PB  
(3EB)



420EB



4EB  
SKA1 mid  
archive

urvey

Low

Mid



15.6ZB

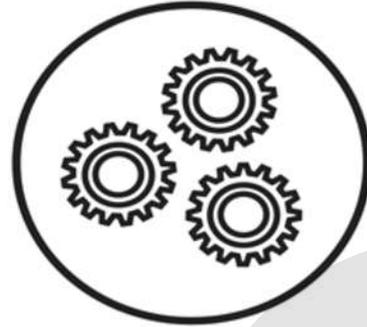
HL-LHC

### Challenges

- FAIRification: data & workflows
- "big data" and "open data"
- Irreversibility challenge and data loss
- heterogeneous data & infrastructures
- transfer of knowledge

# Task Areas

TA 2: Data management



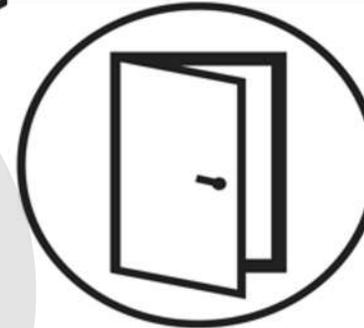
TA 3: Data transformations



TA 1: Management and governance



TA 4: Data portal



TA 7: Education, training, outreach, citizen science



TA 5: Data Irreversibility



TA 6: Synergies & services



...

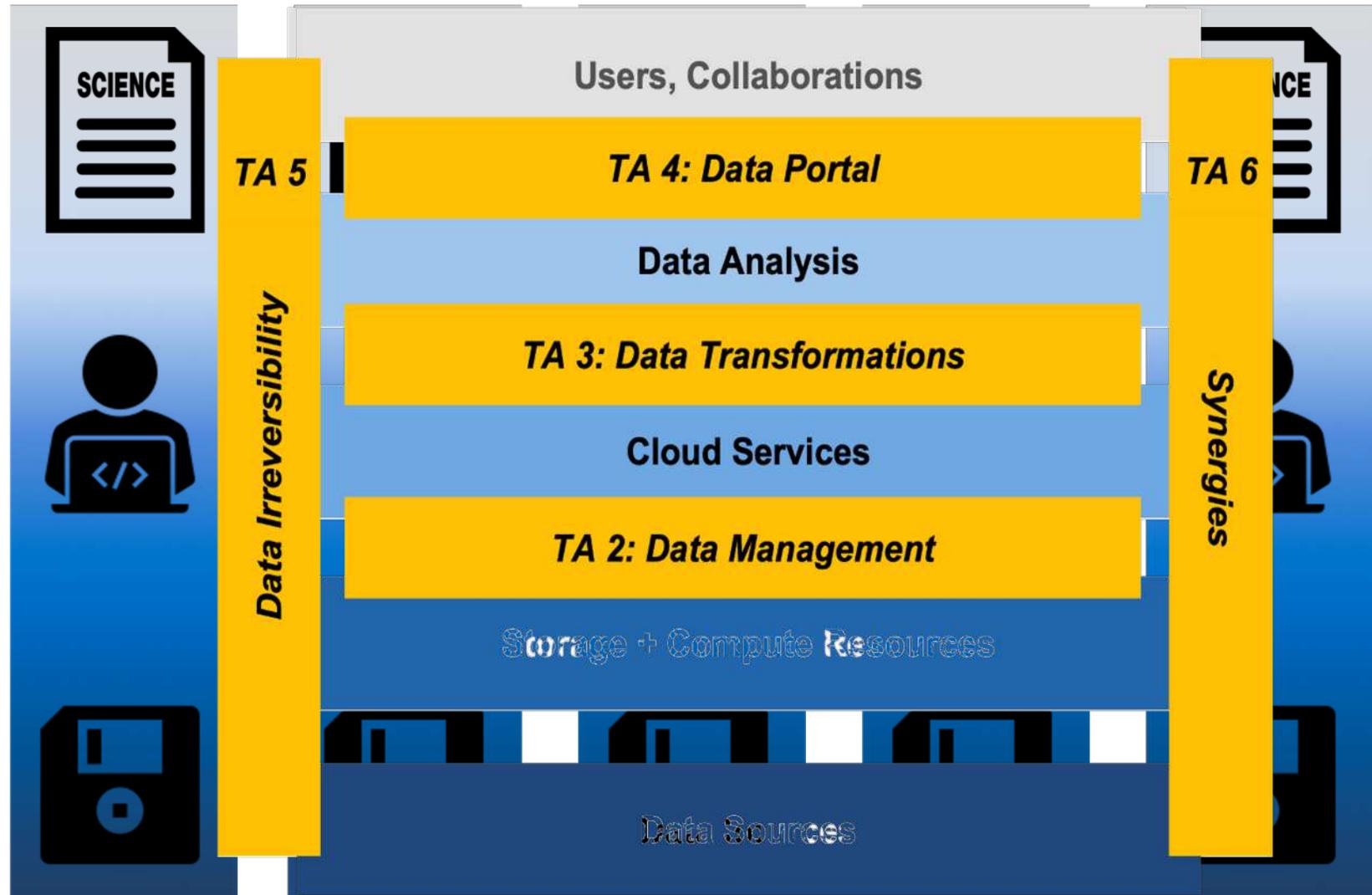
Analysis tools

Compute resources

Data access

Data storage

Data sources



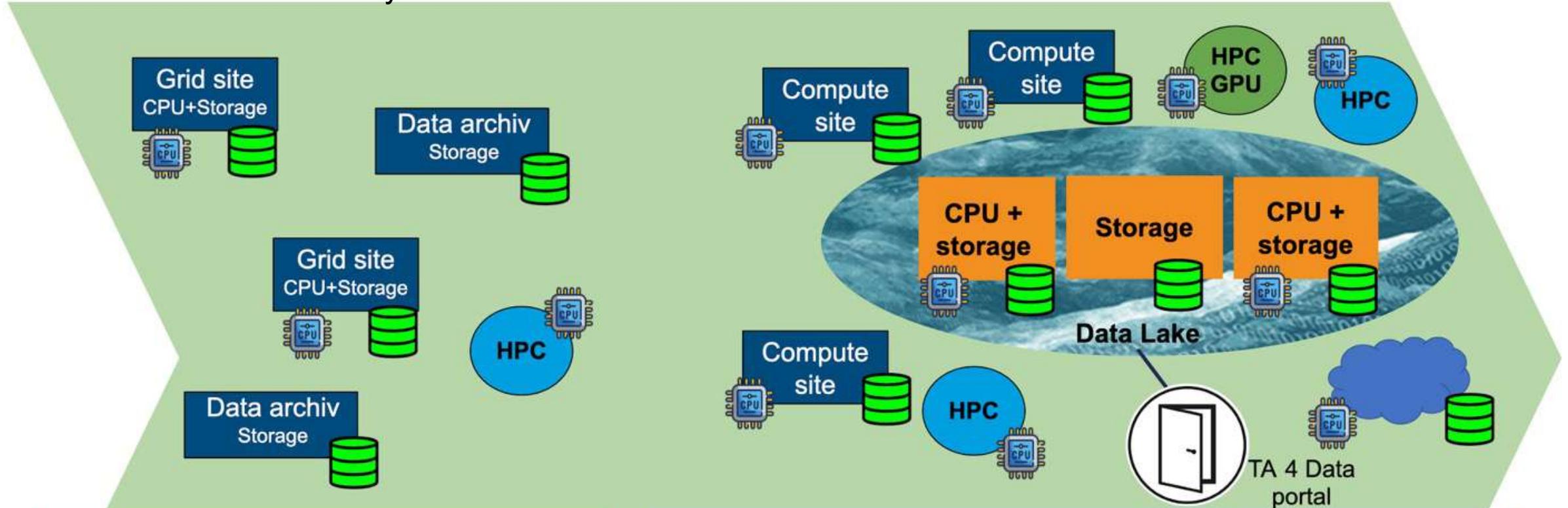
**Flexibility, efficiency, scalability:**  
**data volumes, number of users and analyses, heterogeneous resources; data combinations**

# TA 2: Infrastructure Evolution

Access to data, federated computing, automation, data lake prototype

Today

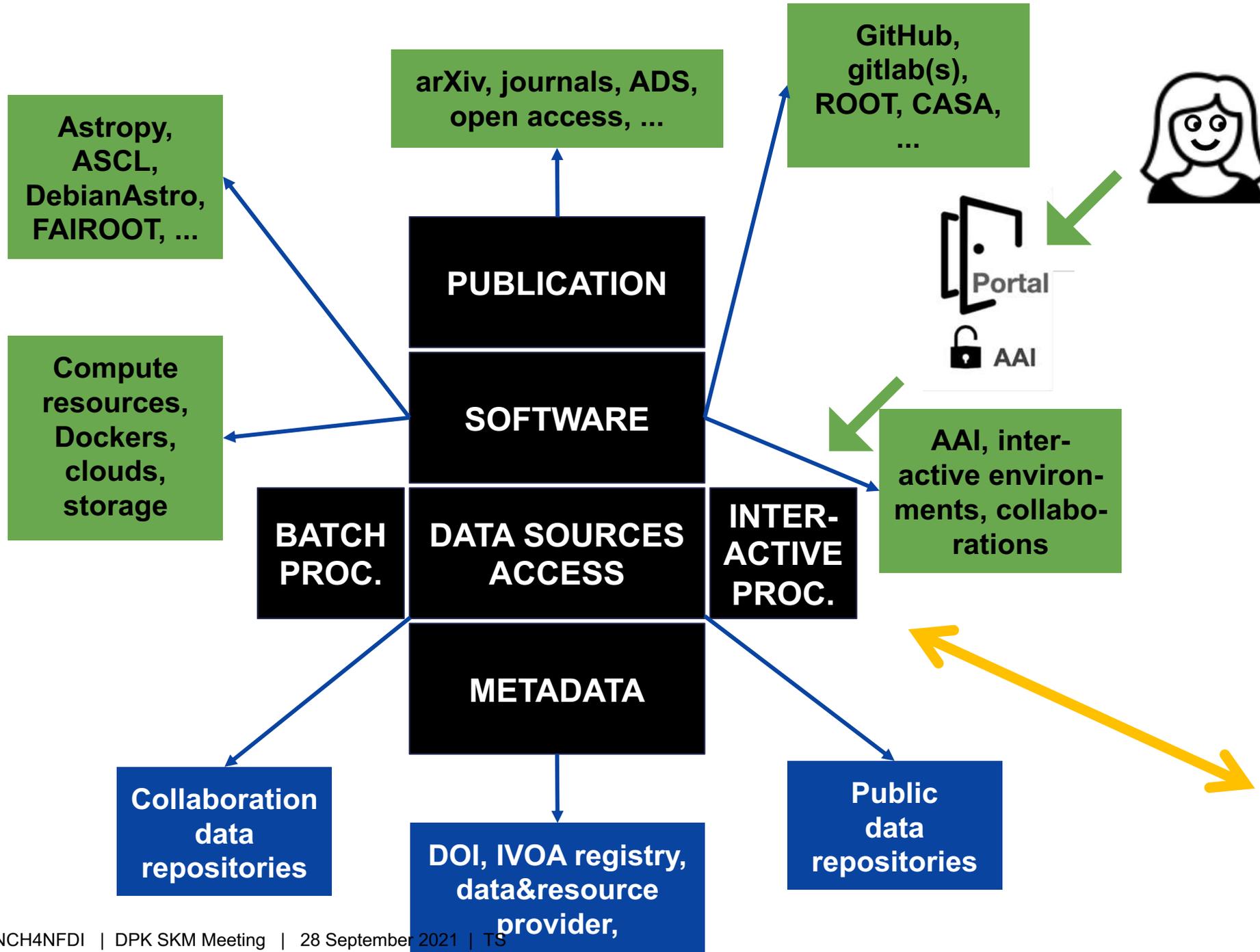
Future



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

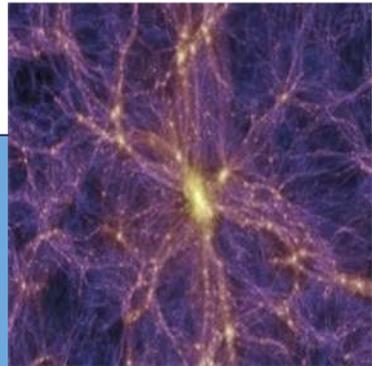
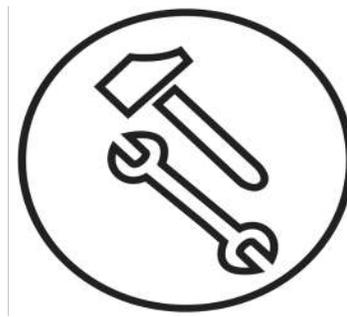
# Digital Research Product



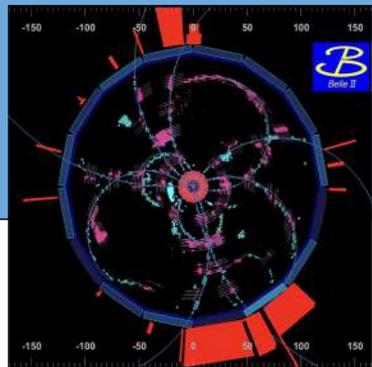
- Access via portal
- Interaction with other RPs
- Interfaced to tools and infrastructures

# Task Area 3: Data Transformations

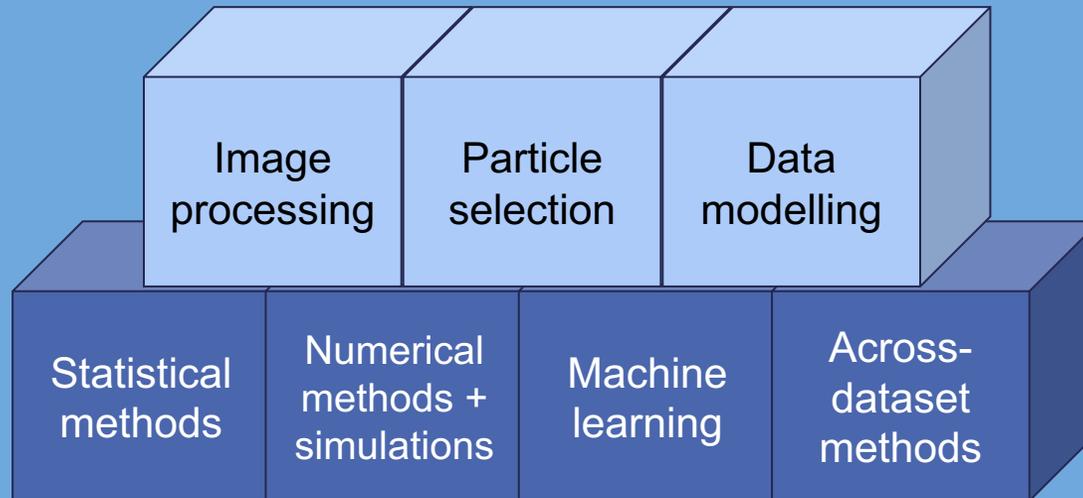
Integration of common tools into a data infrastructure based on code-to-data principle  
Provision of tools for parallel processing of huge data sets on heterogeneous resources



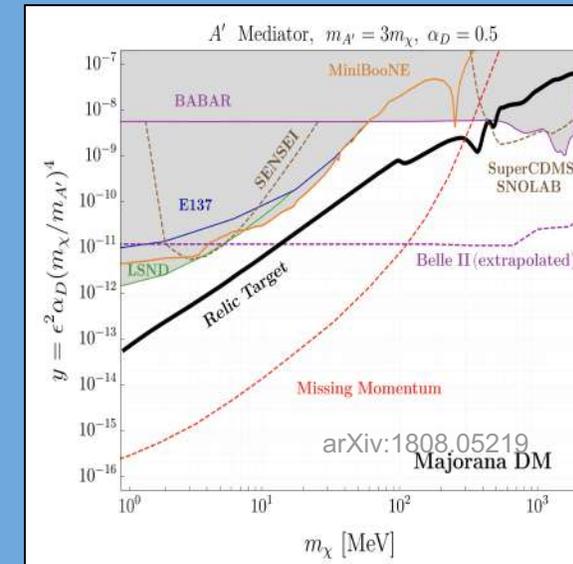
Data and metadata



**User**  
Selection of tools / workflows via portal



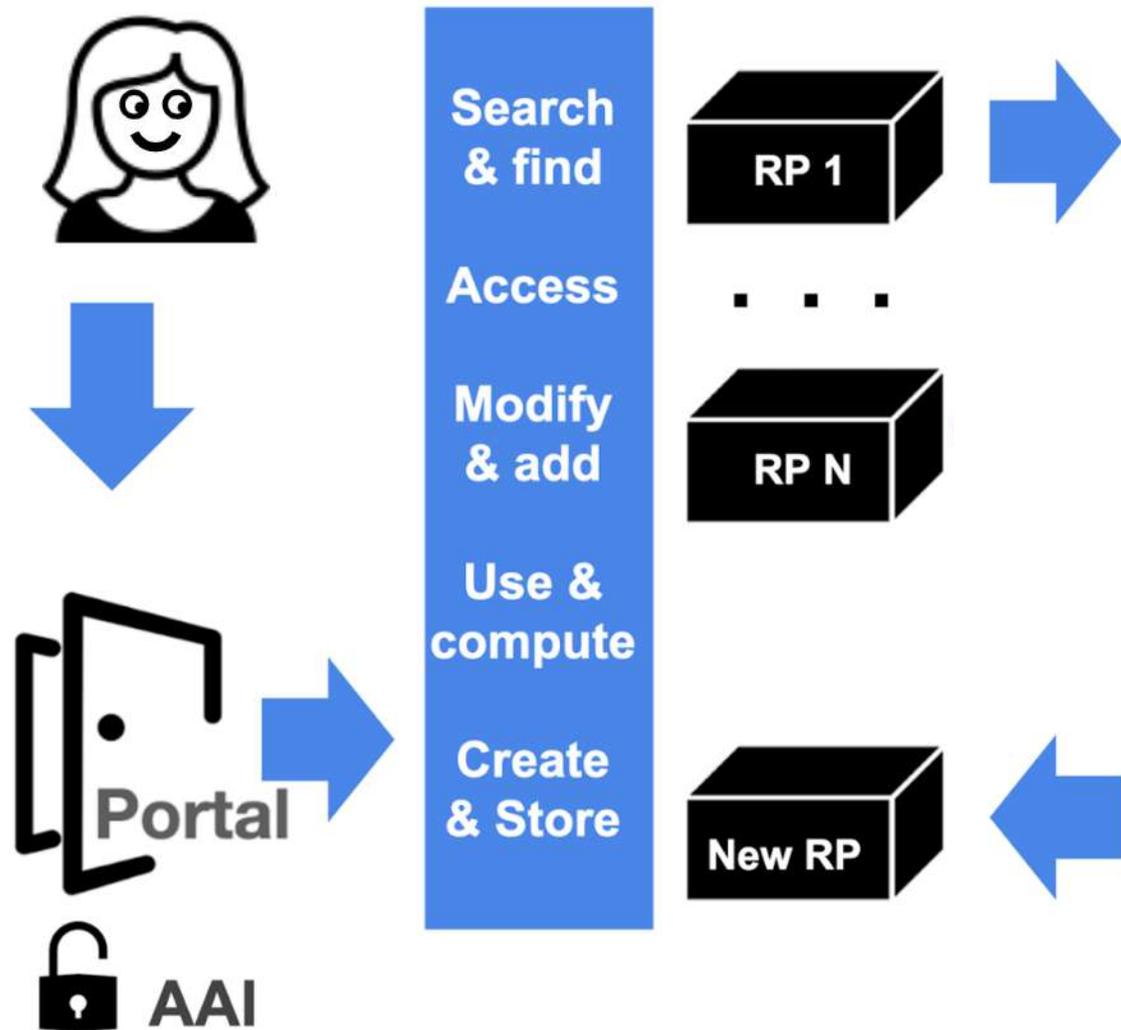
## Scientific result



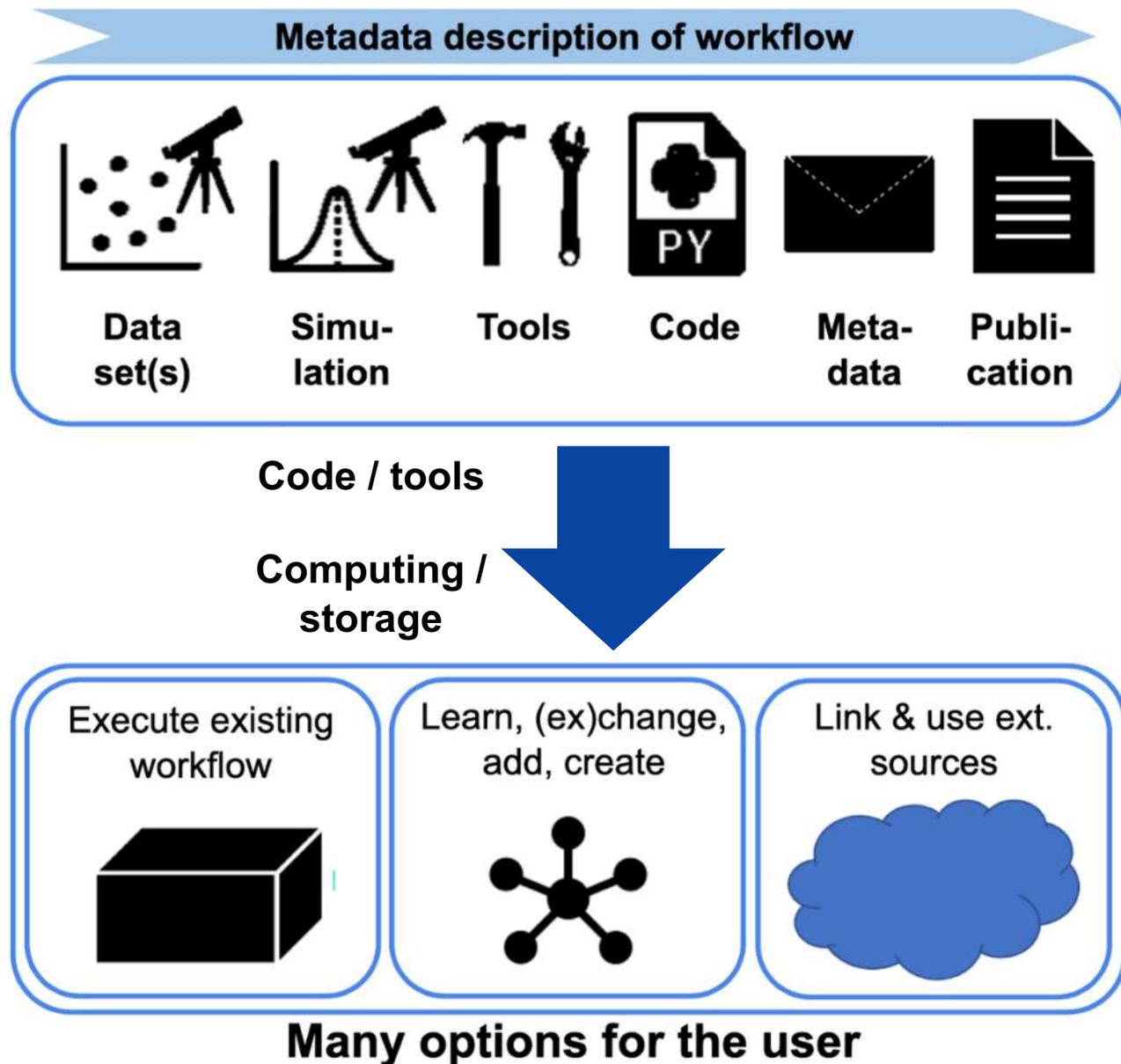
Tools common to many science fields

# PUNCH-SDP

The science data platform for RPs



## Research product contains executable workflow



**MaRDI: micro-metadata schemes, statistical procedures, use cases**

**NFDI4CS: 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 resusabiliy**

**DAPHNE4NFDI & FAIRMat & DPG: address physics-specific aspects**



## TA 6 Marketplace



**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**

# 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 Contreras<sup>1, 2, X</sup>, Jonas Dehning<sup>1, X</sup>, Matthias Loidolt<sup>1, X</sup>, Johannes Zierenberg<sup>1</sup>,  
F. Paul Spitzner<sup>1</sup>, Jorge H. Urrea-Quintero<sup>1</sup>, Sebastian B. Mohr<sup>1</sup>, Michael Wilczek<sup>1,3</sup>,  
Michael Wibrals<sup>4</sup>, and Viola Priesemann<sup>1,3\*</sup>

$$H_{\infty} = H_{\infty}^s \frac{\lambda_s}{\Gamma} \left( \frac{\Gamma}{\lambda_s} + \xi^{\text{ap}} \right) \quad (4)$$

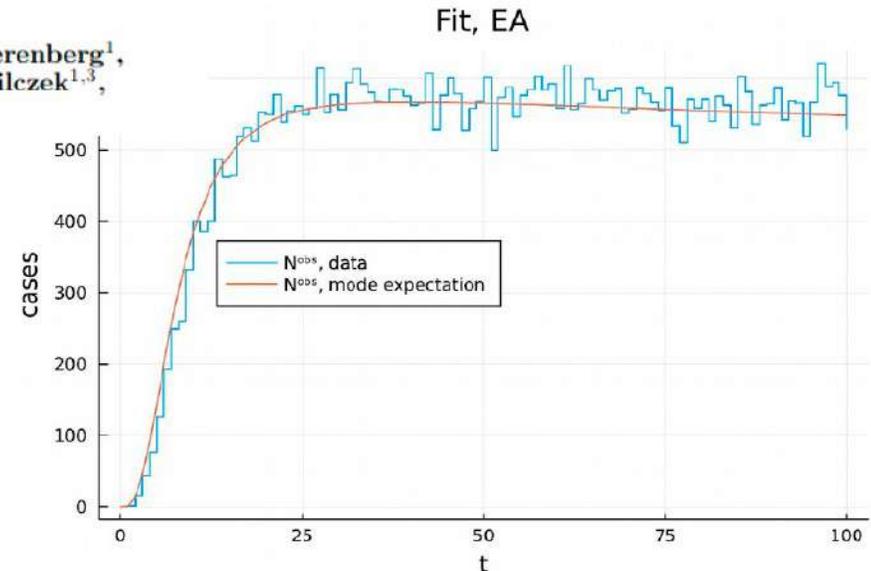
$$H_{\infty}^s = \frac{\Phi}{\lambda_s (1 + \eta R_t^H)} \left[ \left( \frac{\epsilon R_t^H}{\nu R_t^H - 1} + R_t^H \right) - (R_t^H - 1) \frac{\eta R_t^H + \frac{1 + \Gamma/\lambda_s}{1 - \xi^{\text{ap}}}}{\eta R_t^H + 1} \right]^{-1} \quad (5)$$

$$N_{\infty}^{\text{obs}} = \Gamma \nu R_t^H T_{\infty} + \lambda_s H_{\infty}^s (1 + \eta R_t^H). \quad (6)$$

To calculate  $N_{\infty}^{\text{obs}}$  in terms of the model's parameters, we insert equations (4)–(5) into equation (6):

$$N_{\infty}^{\text{obs}} = \lambda_s H_{\infty}^s (1 + \eta R_t^H) \left( \frac{1}{1 - \nu R_t^H} \mathcal{I} + \mathcal{I} \right), \quad (7)$$

$$= \lambda_s H_{\infty}^s \frac{1 + \eta R_t^H}{1 - \nu R_t^H}. \quad (8)$$



## 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

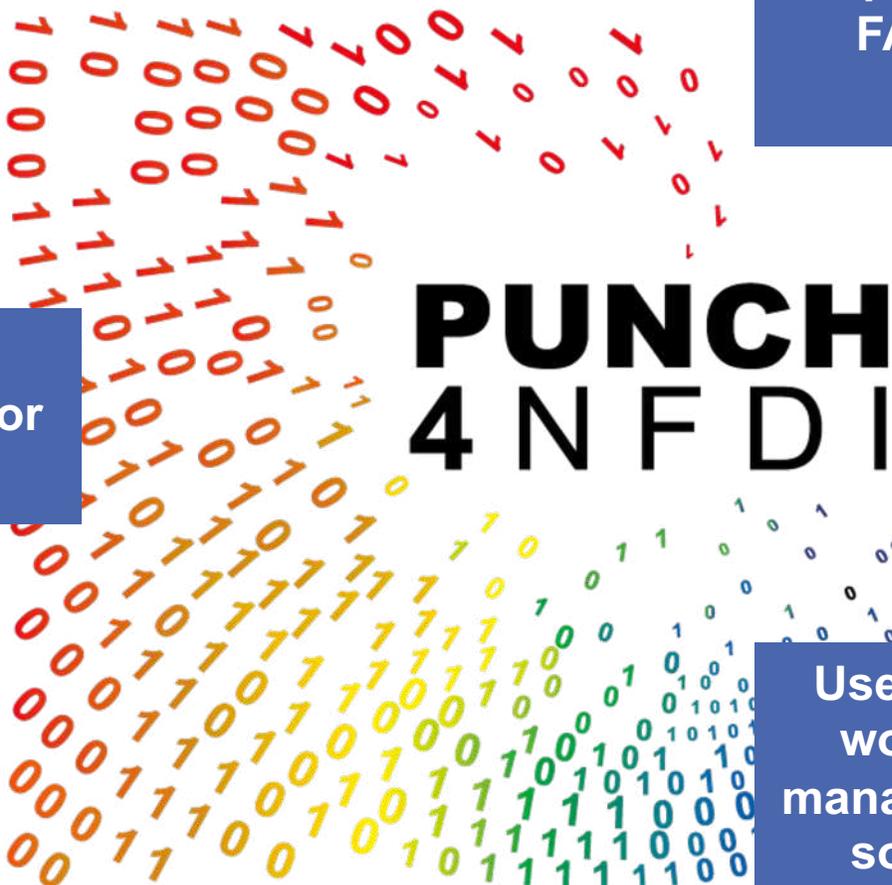
1. Map out potential research applications in *Citizen Science*
2. Prepare data sets and providing soft- and hardware infrastructure
3. Run pilot projects for 6-12 months by engaging schools and universities and evaluate results
4. Prepare and launch of further projects in collaboration with the physics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



# “PUNCH”Lines

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

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



**PUNCH  
4 N F D I**

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](mailto:thomas.schoerner@desy.de))  
DESY, Notkestr. 85, D-22607 Hamburg

### Contact:

Mail: [punch4nfdi@desy.de](mailto:punch4nfdi@desy.de)

Web: [www.punch4nfdi.de](http://www.punch4nfdi.de)

Twitter: [@punch4nfdi](https://twitter.com/punch4nfdi)

