The PUNCH4NFDI Consortium

Particles, Universe, NuClei and Hadrons for the NFDI

Olaf Kaczmarek (Bielefeld University) for the PUNCH4NFDI Consortium

KHuK Annual Meeting 2022, 9. December 2022

Bad Honnef

NFDI - National Research Data Infrastructure

Gemeinsame Wissenschaftskonferenz GWK

DFG Deutsche Forschungsgemeinschaft

27 Consortia selected in three rounds (2020, 2021, 2023) with aims including

- Establishment of **data handling standards, procedures and guidelines** in close collaboration with the community of interest

- Development of **cross-disciplinary metadata standards**

- Development of **reliable and interoperable data management measures and services** tailored to the needs of the community of interest

- Increased **reusability of existing data**, also beyond subject boundaries

- Improved networking and collaboration with partners outside the German academic research system with expertise in research data management

- Involvement in developing and establishing generic, cross-consortia services and standards in research data management together with other consortia

PUNCH4NFDI - Who We Are

Universities, Helmholtz, Max Planck, Leibniz (20 co-applicant institutions + 22 partner institutions)
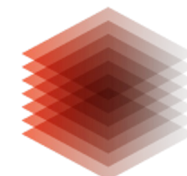
PUNCH4NFDI - Who We Are

Universities, Helmholtz, Max Planck, Leibniz (DESY + 19 co-applicant institutions + 22 partner institutions)
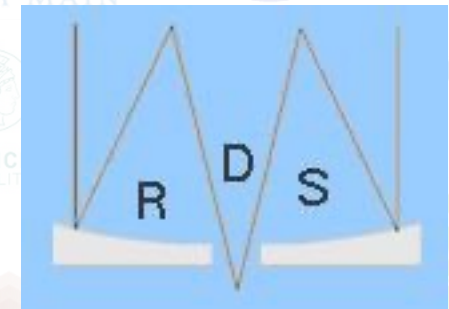
Task areas

TA 2: Data
management

TA 3: Data
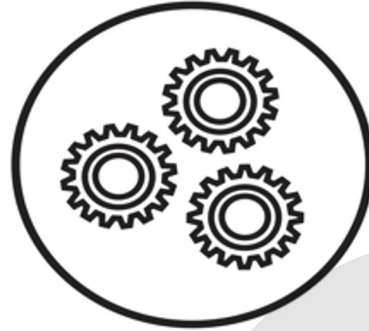transformations

TA 1: Management
and governance

TA 4: Data
portal

TA 5: Data
Irreversibility

TA 7: Education,
training, outreach,
citizen science

TA 6: Synergies
& services

FAIR data management

Findability, Accessibility, Interoperability, and Reusability of digital datasets are cornerstones of PUNCH4NFDI
The FAIR guiding principles (https://www.nature.com/articles/sdata201618.pdf):

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

TA 1

Management and governance
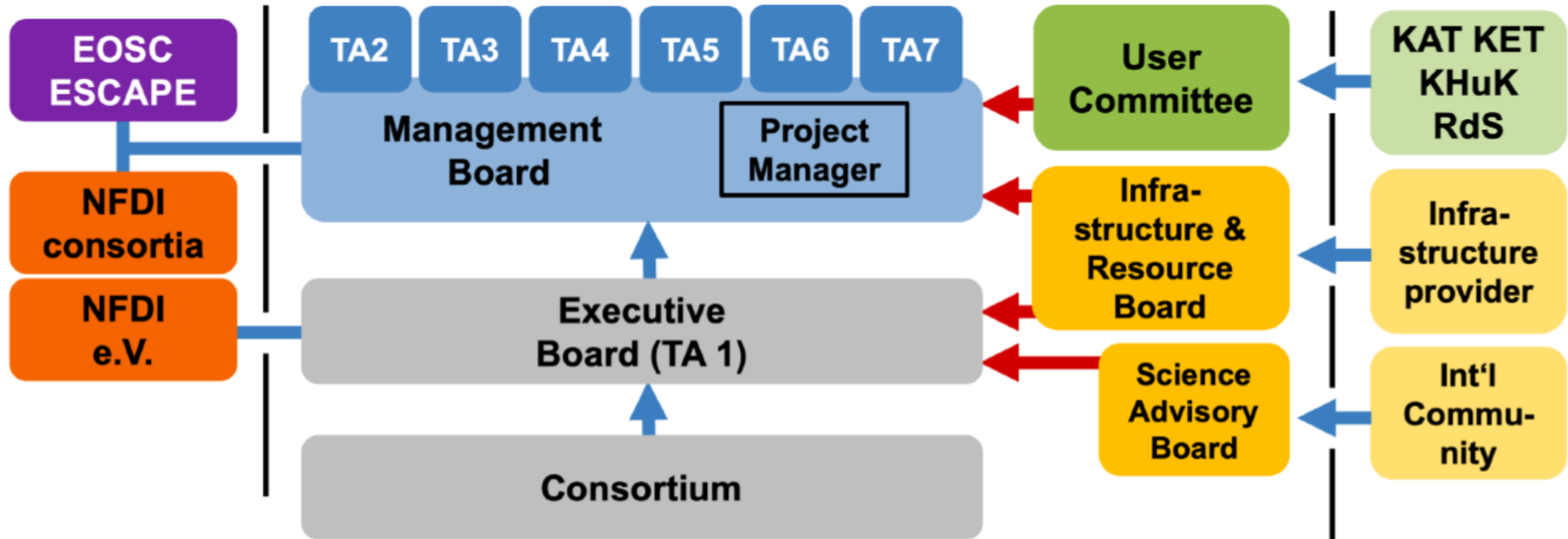
TA 2: Data management

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

Common PUNCH-AAI (based on Helmhotz-AAI) to access the services

Storage4PUNCH

- dCache based system at DESY

- XRootD based system at U Bonn

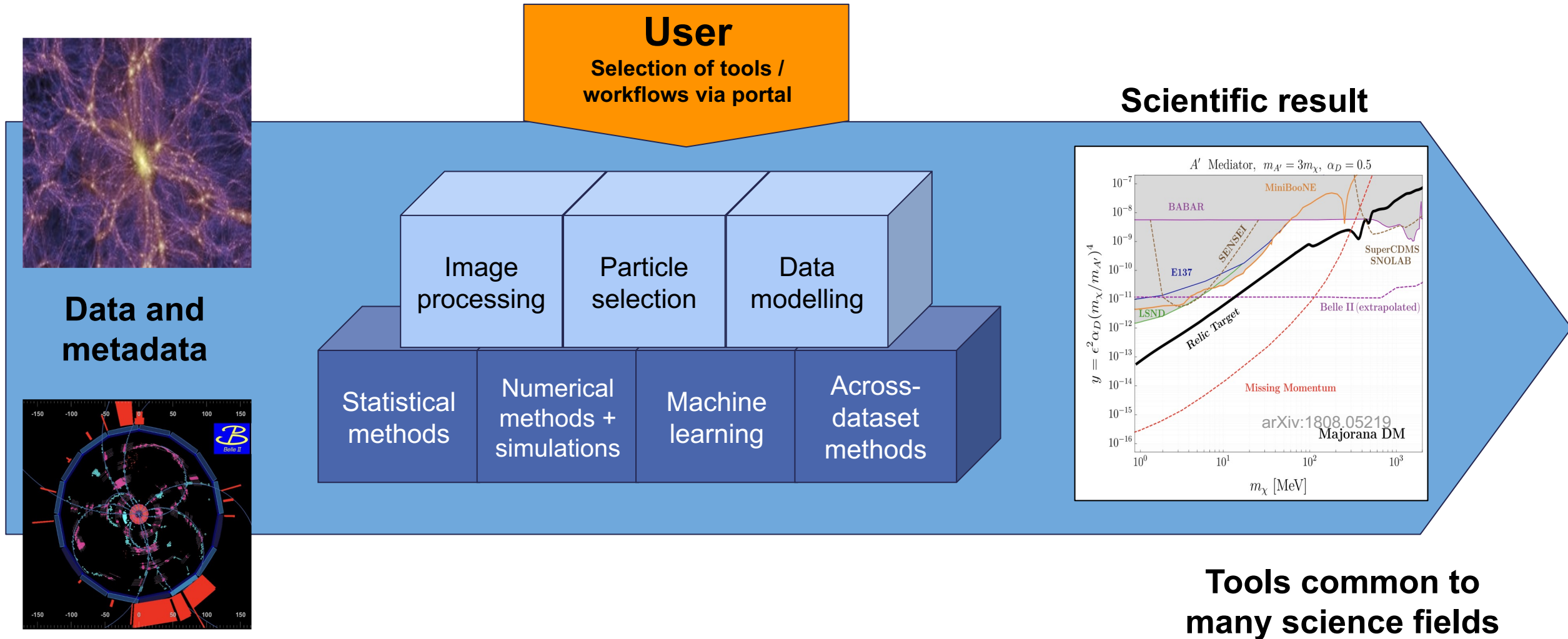- Another system at GSI is being prepared

Compute4PUNCH

- Login node at KIT using tokens (no local account required)

- CPU resources included located at KIT, Bonn, Münster – other locations being prepared

Metadata Catalogue

- Catalogue with flexible schema

- Initial focus on LQCD metadata and related applications
    - Development system now setup
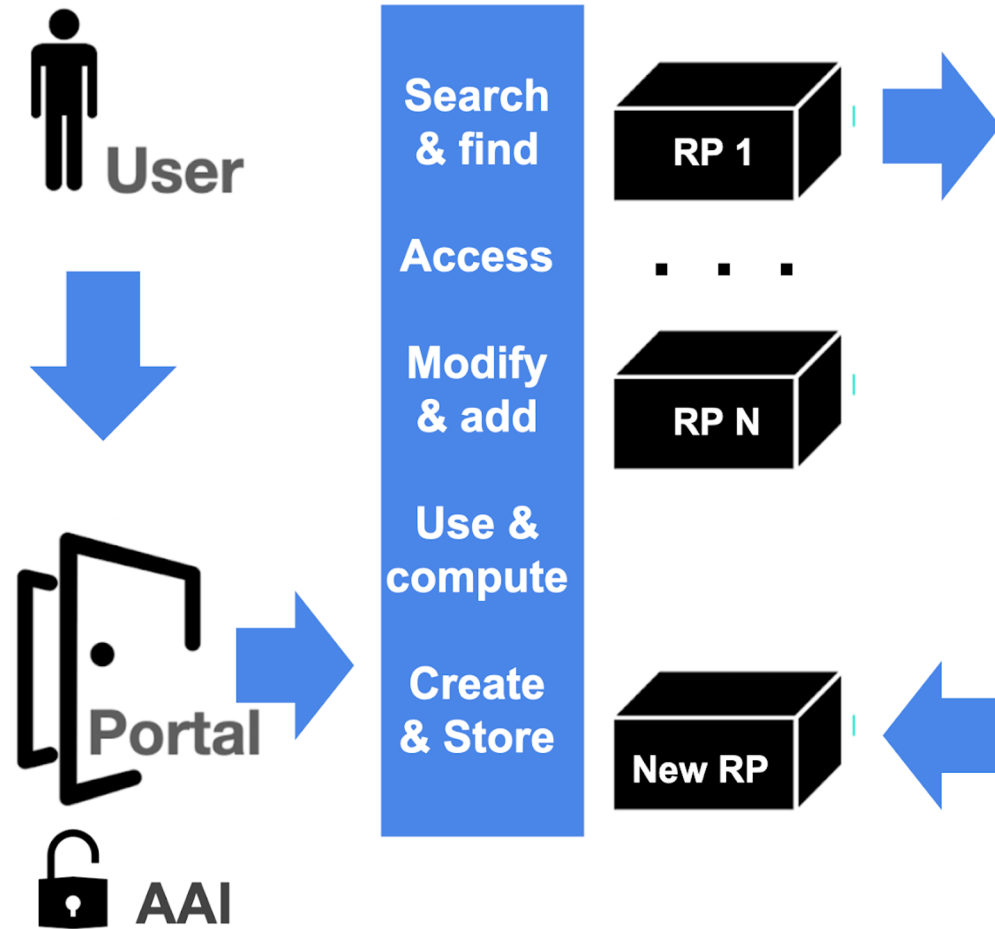   - Planning for other applications beyond LQCD

TA 3: Data transformations

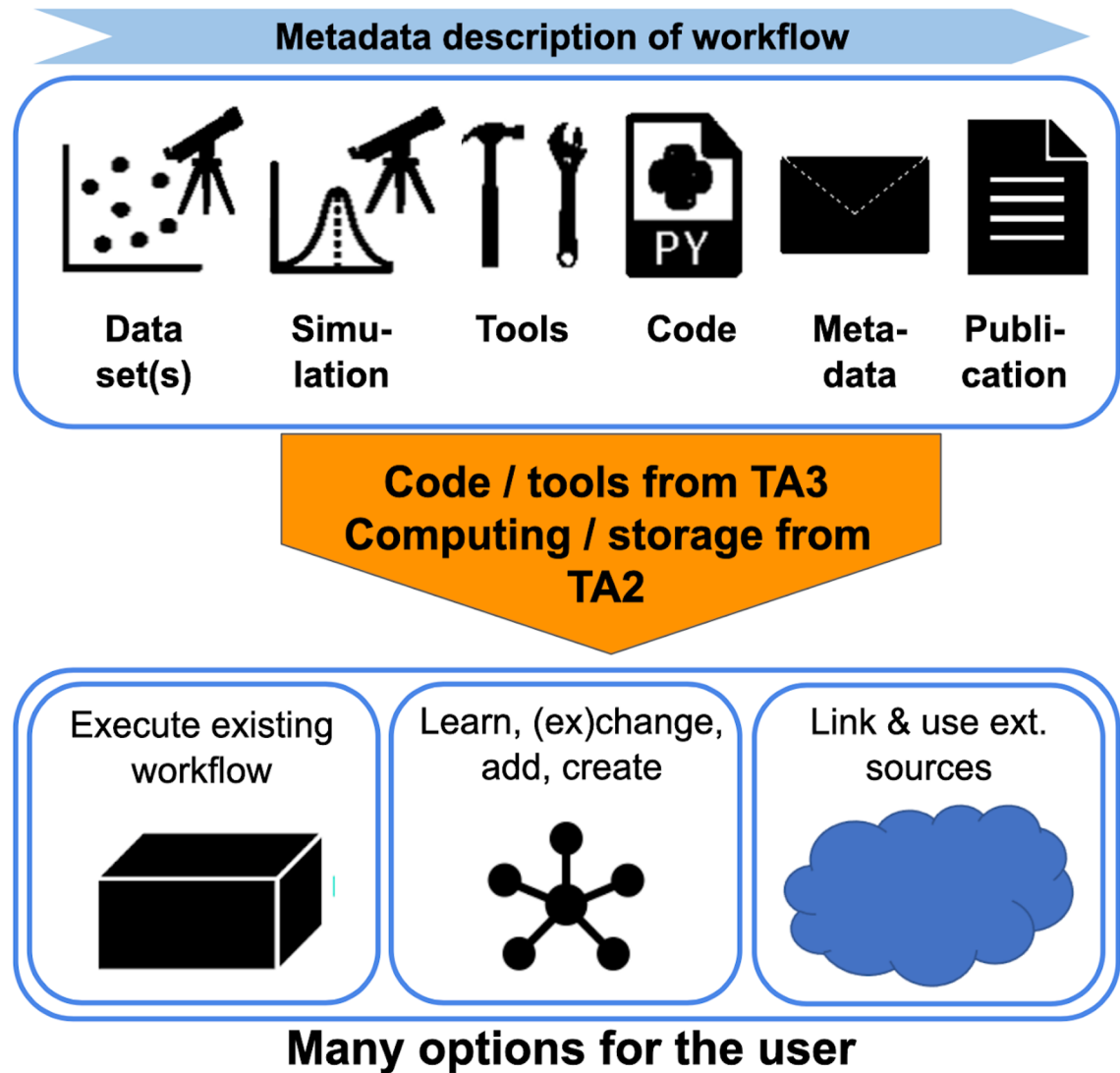**Tools for parallel processing of huge data sets on heterogeneous resources**
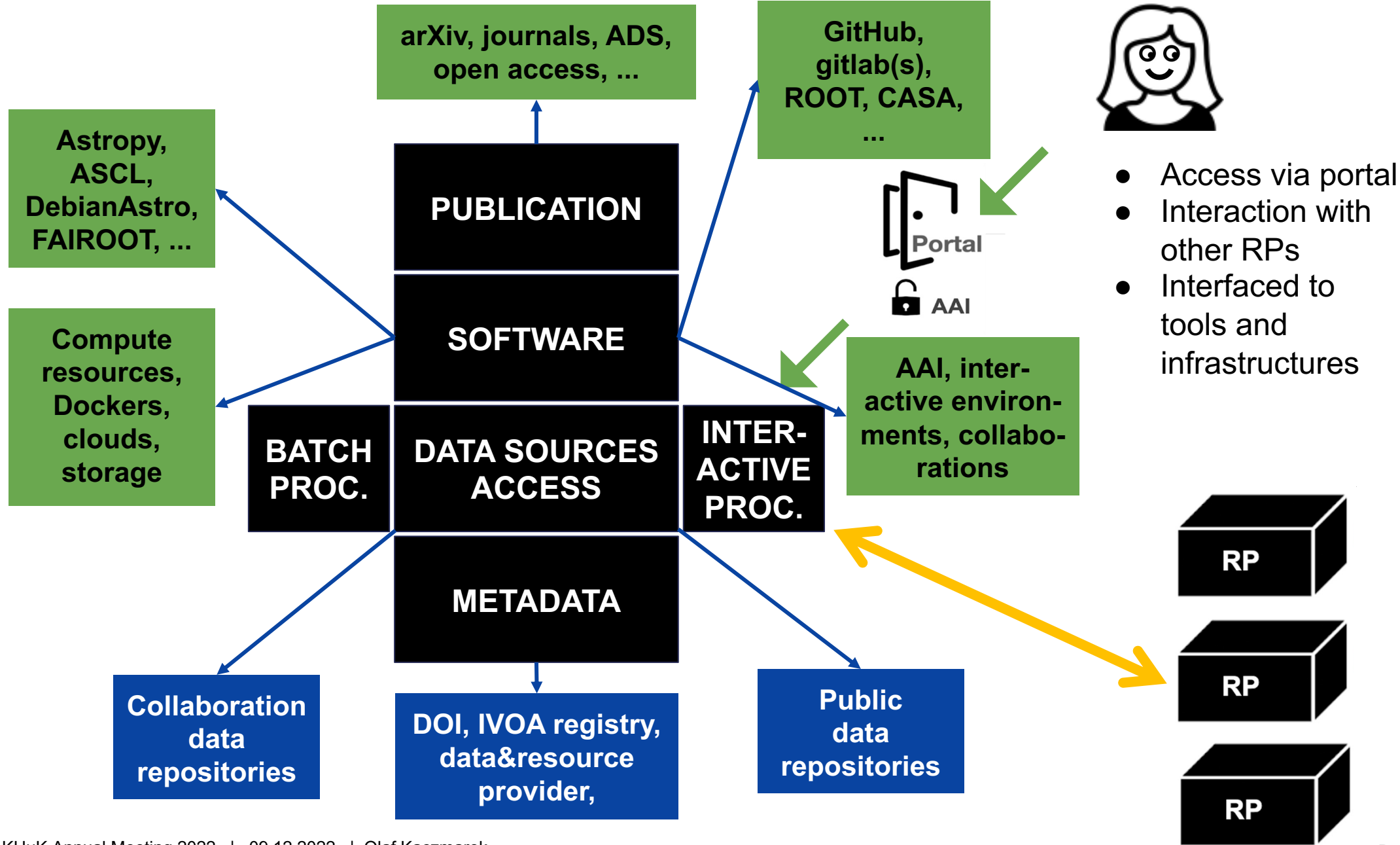
TA 4: Data portal

PUNCH Science Data Platform (SDP) for Research Products (RPs)

User

Search & find

Access

Modify & add

Use & compute

Create & Store

Portal

AAI

RP 1

RP N

New RP

Research product contains executable workflow

Metadata description of workflow

Data set(s) Simu-lation Tools Code Meta-data Publi-cation

Code / tools from TA3
Computing / storage from TA2

Execute existing workflow

Learn, (ex)change, add, create

Link & use ext. sources

Many options for the user

Digital Research Product

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

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

Astropy, ASCL, DebianAstro, FAIROOT, ...

PUBLICATION

Portal

AAI

SOFTWARE

Compute resources, Dockers, clouds, storage

BATCH PROC.

DATA SOURCES ACCESS

INTER-ACTIVE PROC.

AAI, inter-active environ-ments, collabo-rations

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

METADATA

Collaboration data repositories

DOI, IVOA registry, data&resource provider,

Public data repositories

RP

RP

RP

TA 5: Data irreversibility

only a small fraction of data can be stored long term
⇒ irreversible loss of information
⇒ Implications for discovery potential and reproducibility

- **dynamic filtering:**
  real time extraction of relevant information from huge data streams (e.g. with machine learning algorithms).

- **dynamic archiving:**
  feedback from offline analysis to sensor controls.

- **Scaling workflows:**
  analysis of huge online and offline datasets requires workflows for optimal use of resources

- **reproducability:**
  reconstruction of how and why specific decisions were made in real time. Simulations are critical for validation and understanding.

dynamic data life cycle
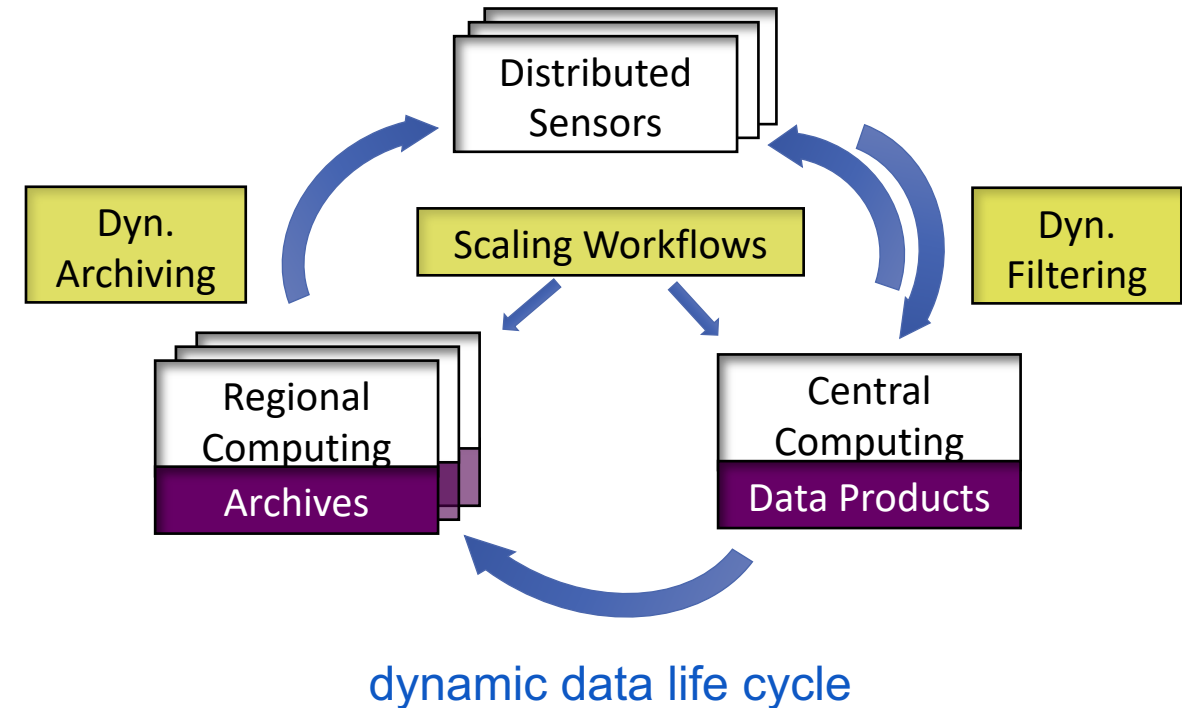
Synergies and services

we are not alone...

NFDI sections on

- Common infrastructures (section-infra)
- Ethical, legal and social aspects (section-elsa)
- Metadata and provenience (section-metadata)
- Training & Education (section-edutrain)

Base4NFDI

- Basic services for NFDI

Exchange with

- International data management projects
- Data providers and digital libraries
- Researcher in the PUNCH communities
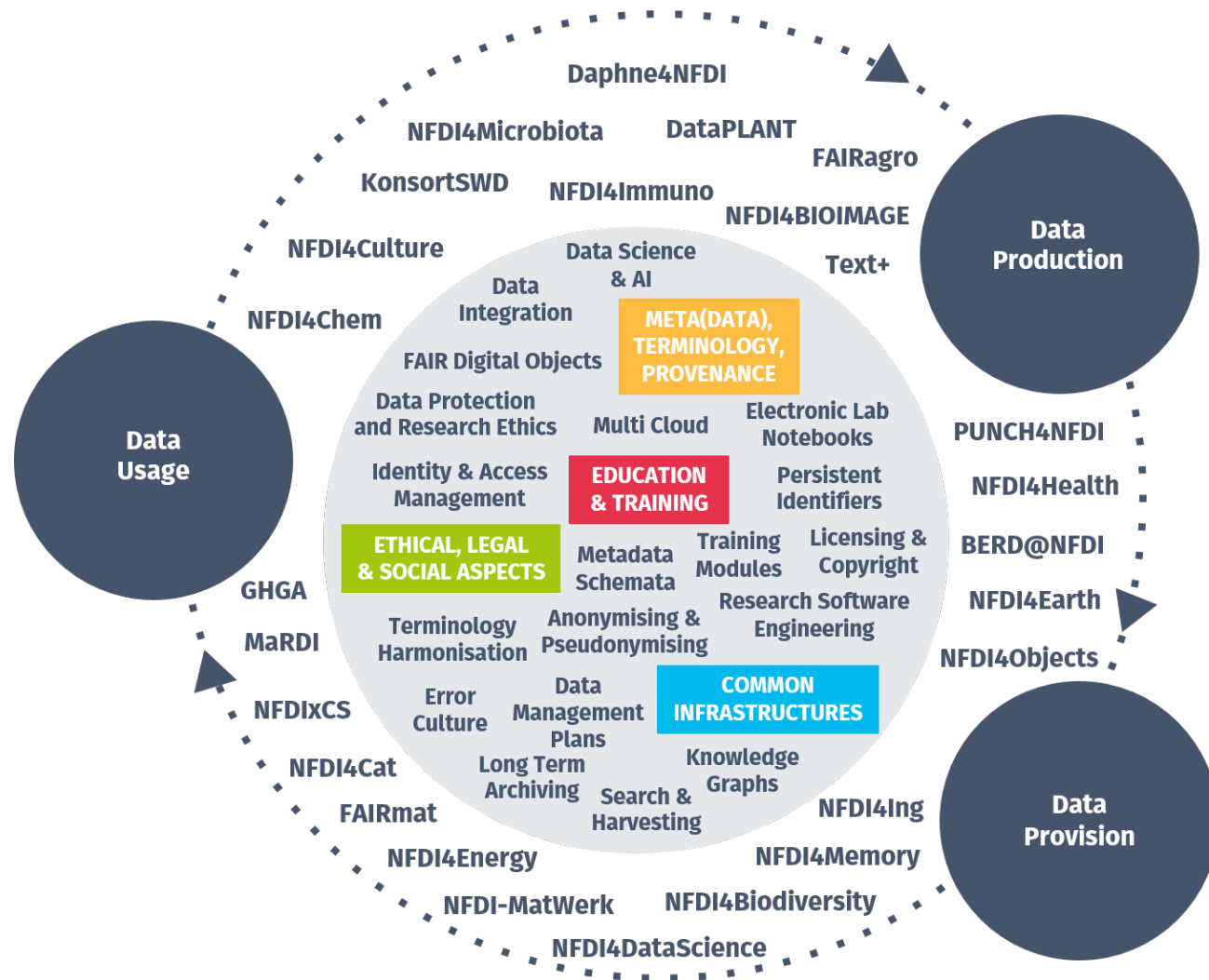
Education, training, outreach, citizen science

Training	Education	Outreach	Citizen Science
Experts	Universities: lecturers and students	Scientists, media, schools, public	Amateur, public
foster expertise and career prospects	focused education and career promotion	communicate, foster young talents, strengthen schools	foster commitment and deeper understanding, democratise science
provisioning of data and educational ressources	educational and data ressources, on-site and online seminars	training of communication, school-academy-network, events, ressources	online projects and campaigns

PUNCH4NFDI for you
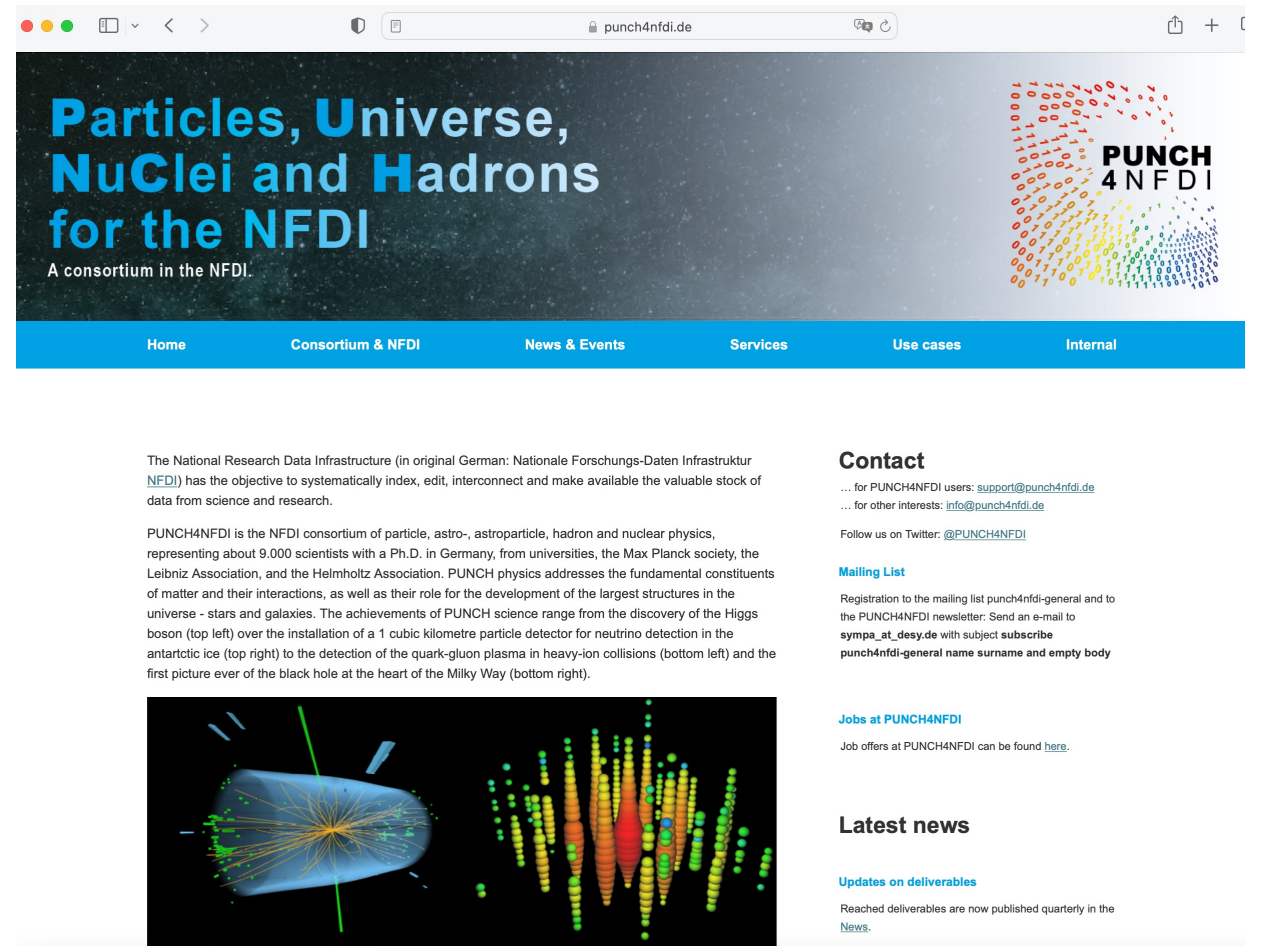
getting informed and involved

- Web page: www.punch4nfdi.de

- Mail: info@punch4nfdi.de

- Twitter: #punch4nfdi

- mailing list: general@punch4nfdi.de

  newsletter, PUNCHLunch, general meetings, Women4PUNCH, etc.

  (send mail with "subscribe punch4nfdi-general" to sympa@desy.de)

- Contact your local spokes-person