The PUNCH4NFDI Consortium

Particles, Universe, NuClei and Hadrons for the NFDI

Andreas Haungs (KIT), Thomas Schörner (DESY)

DPG Spring Meeting, 16 March 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

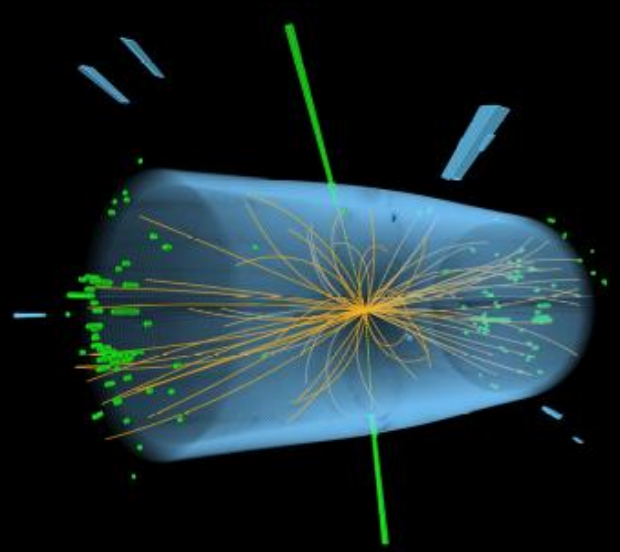
See also DFG.de/nfdi and nfdi.de

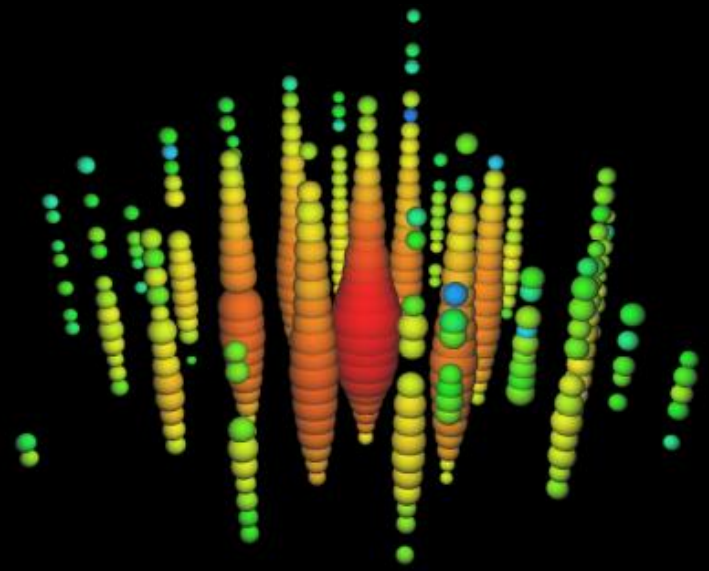
Gemeinsame Wissenschaftskonferenz GWK

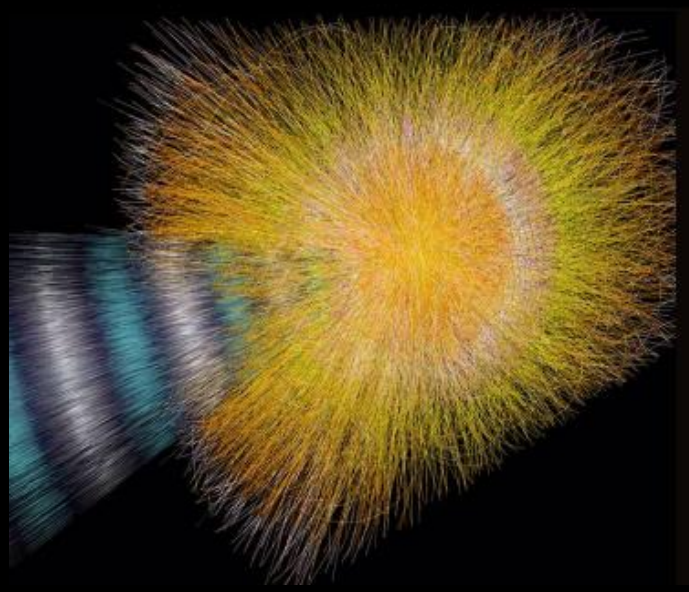
DFG Deutsche Forschungsgemeinschaft

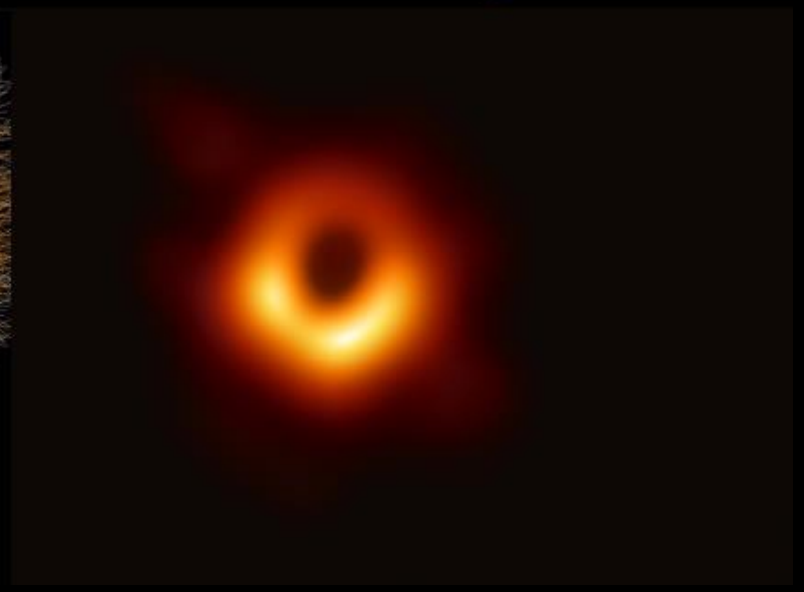
Particles, Universe, NuClei and Hadrons for the NFDI

See also DFG.de/nfdi and nfdi.de



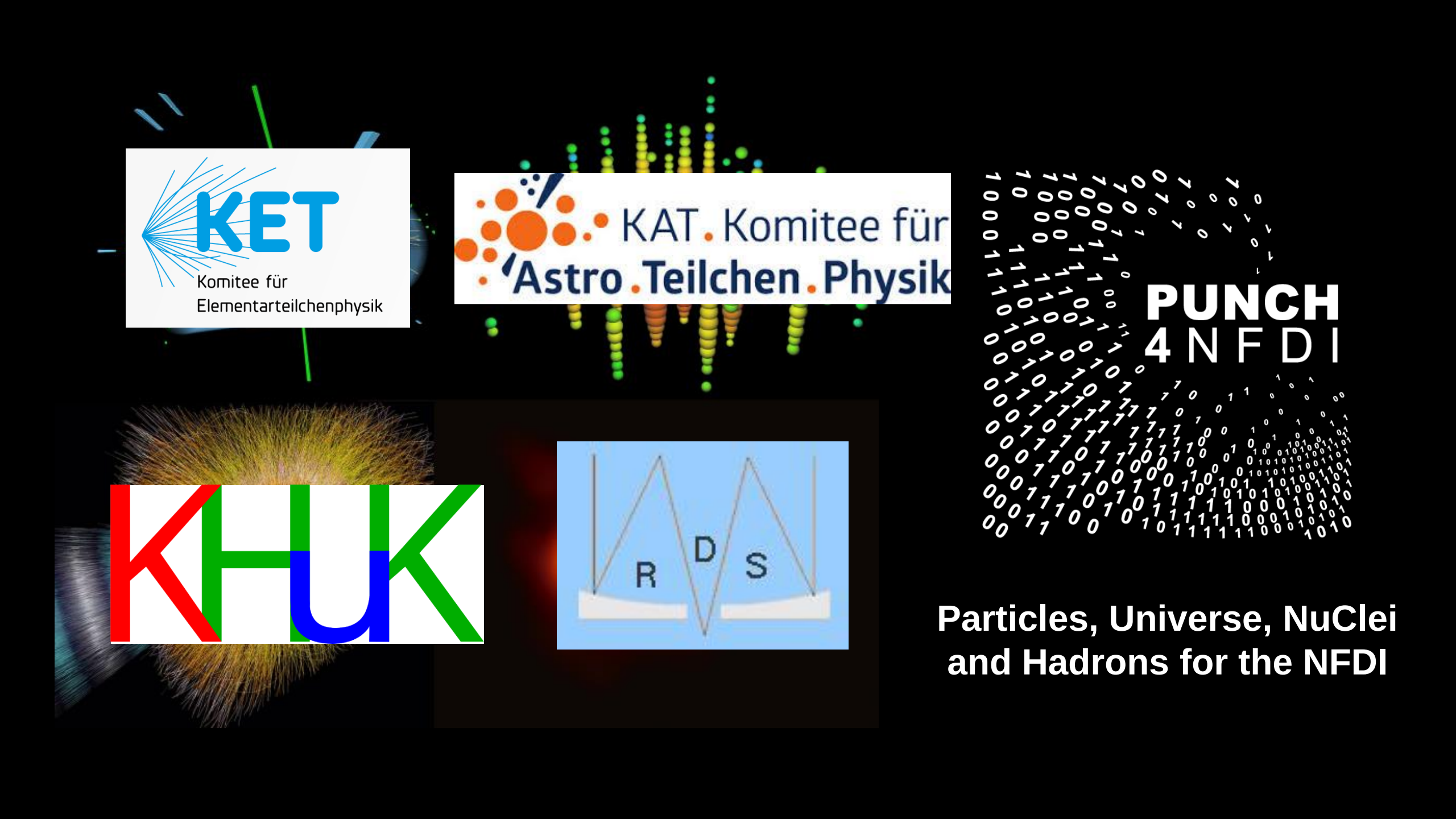






PUNCH
4NFDI

Particles, Universe, NuClei and Hadrons for the NFDI



Particles, Universe, NuClei and Hadrons for the NFDI

PUNCH4NFDI

Particles, Universe, NuClei and Hadrons for the NFDI

How FAIR is PUNCH?

… and how to make it even FAIRer

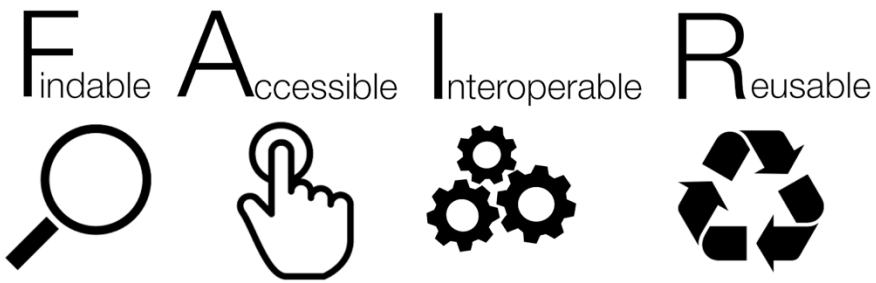
PUNCH data very heterogeneous – and so is the degree of FAIRness. Examples:

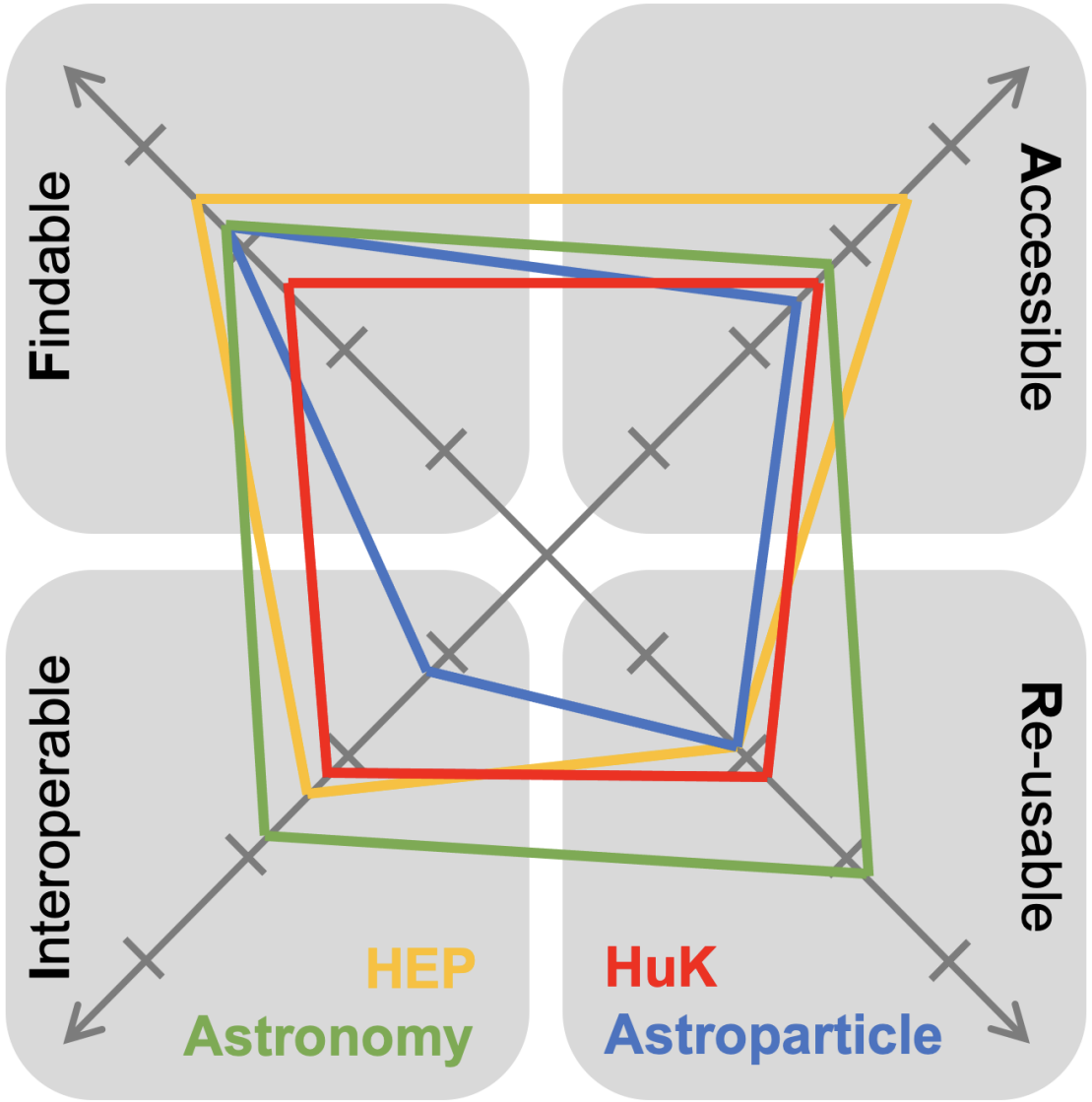
- **Astronomy**: tradition of data sharing and re-using (2018: 50% of publications on re-used data!)
- **Particle physics**: data findable & accessible in large collaborations; older data & software via archives

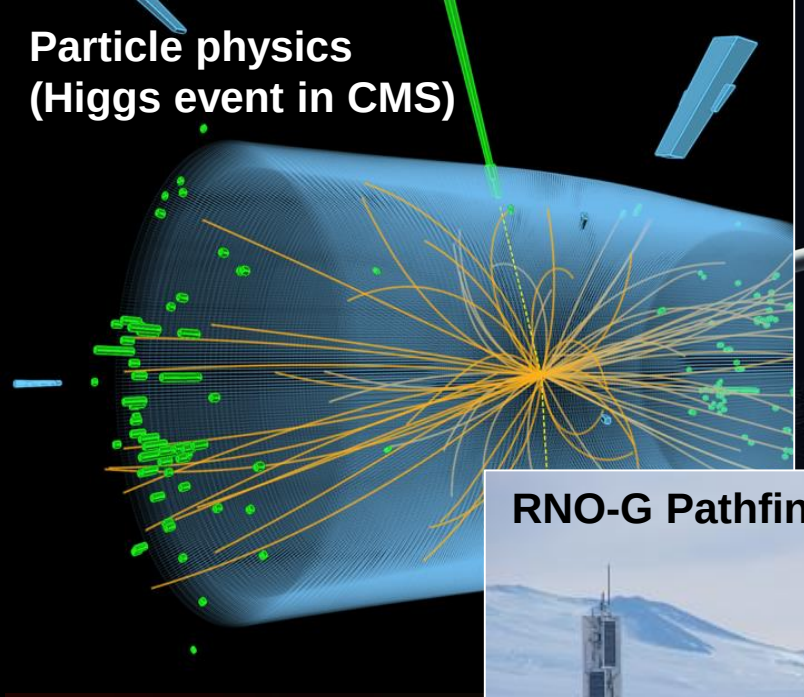
Further FAIRification - some examples:

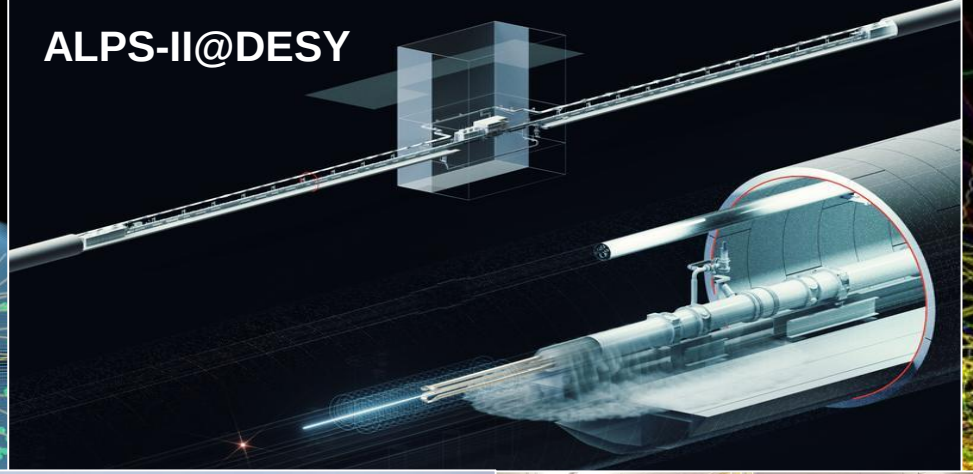
- pragmatic transformation routines between existing metadata systems
- digital research products in science data platform (later)
- education & training

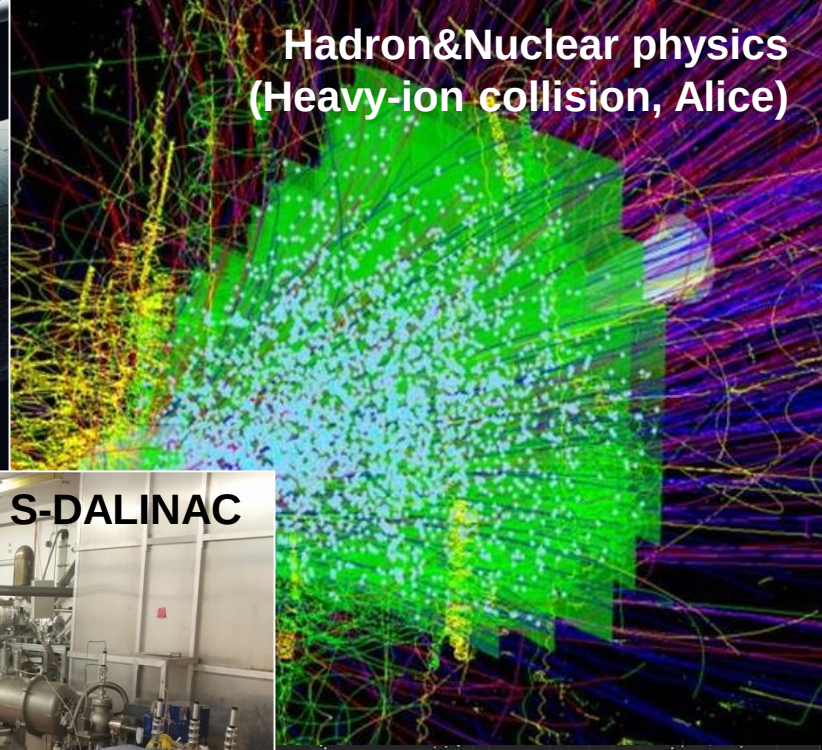
Detailed work programme for each of the four letters:






Particle physics
(Higgs event in CMS)


ALPS-II@DESY

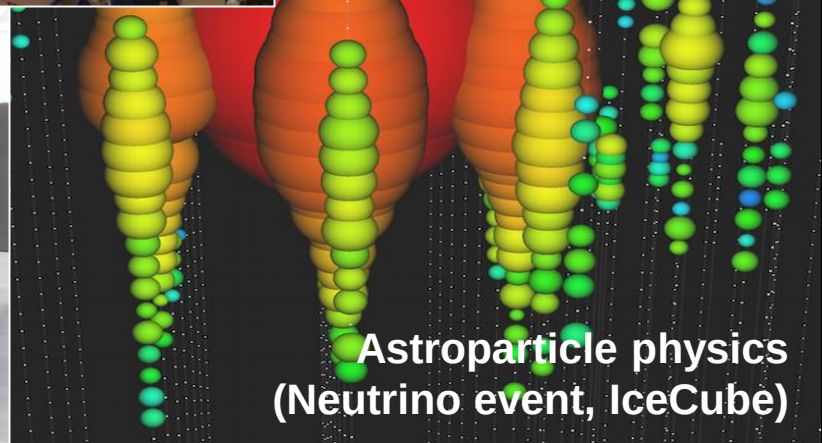

Hadron&Nuclear physics
(Heavy-ion collision, Alice)

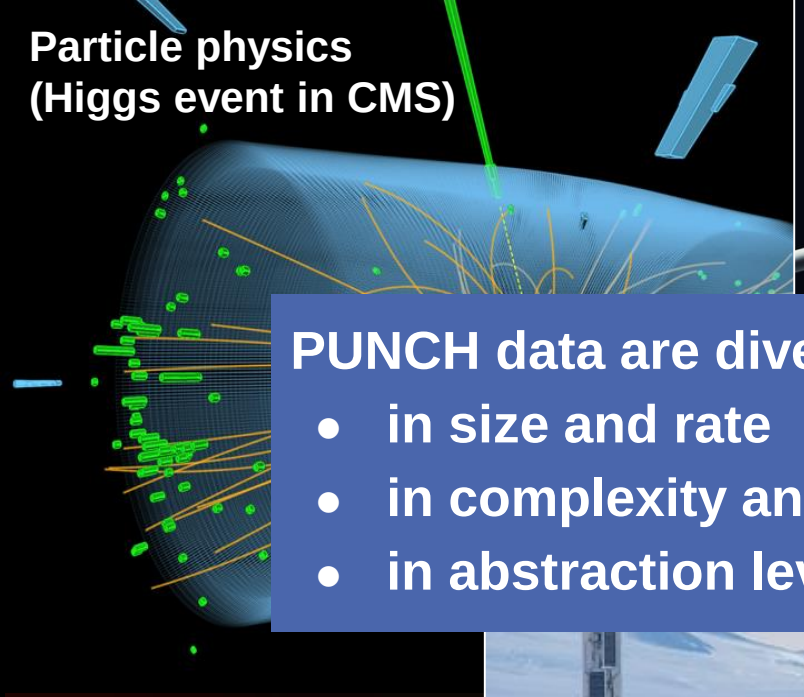
RNO-G Pathfinder

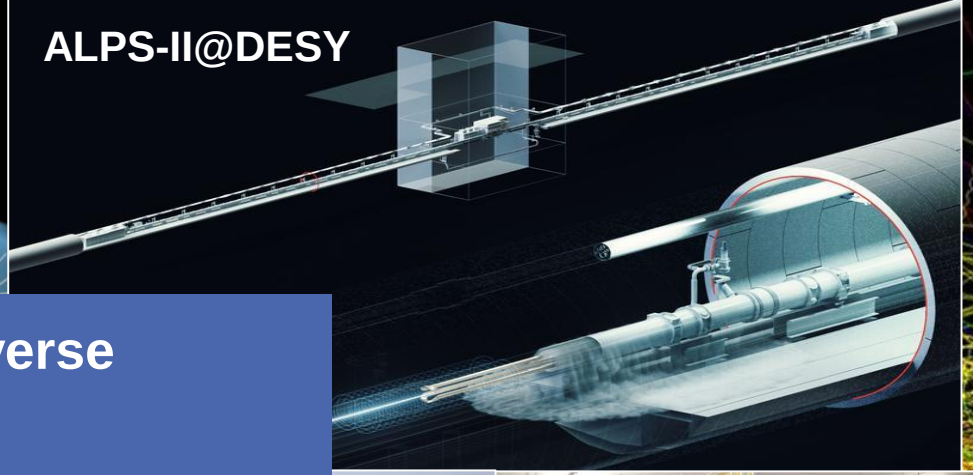
S-DALINAC

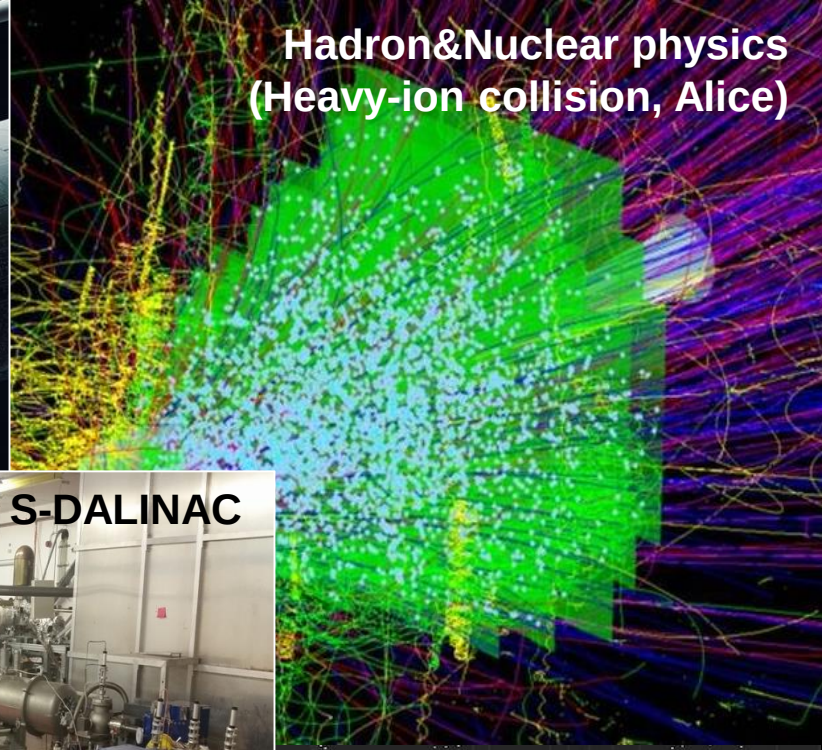
Astronomy
(Black hole, Event Horizon Telescope)

JUWELS@FZJ, GridKa ...


Astroparticle physics
(Neutrino event, IceCube)

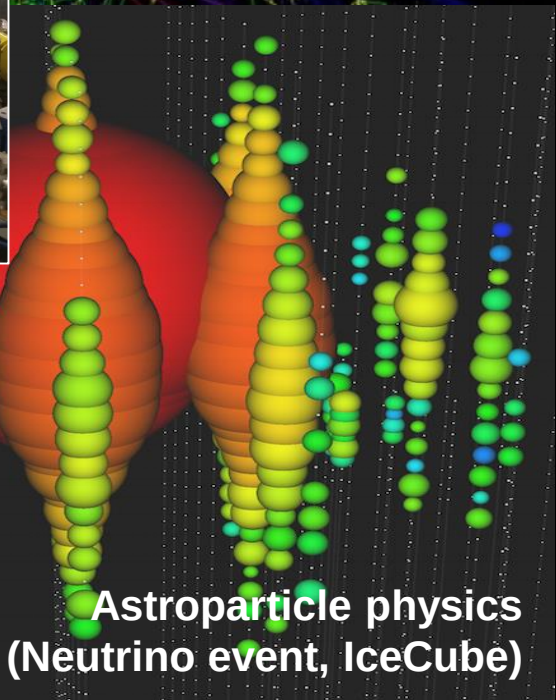


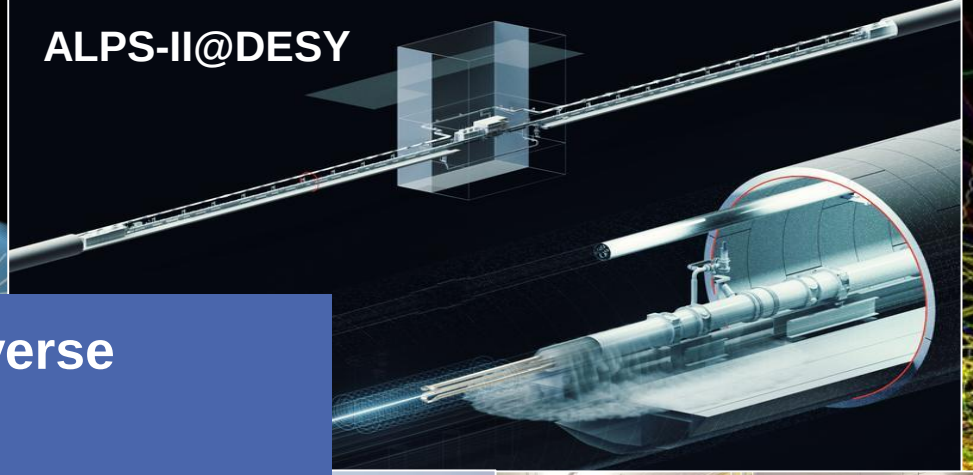


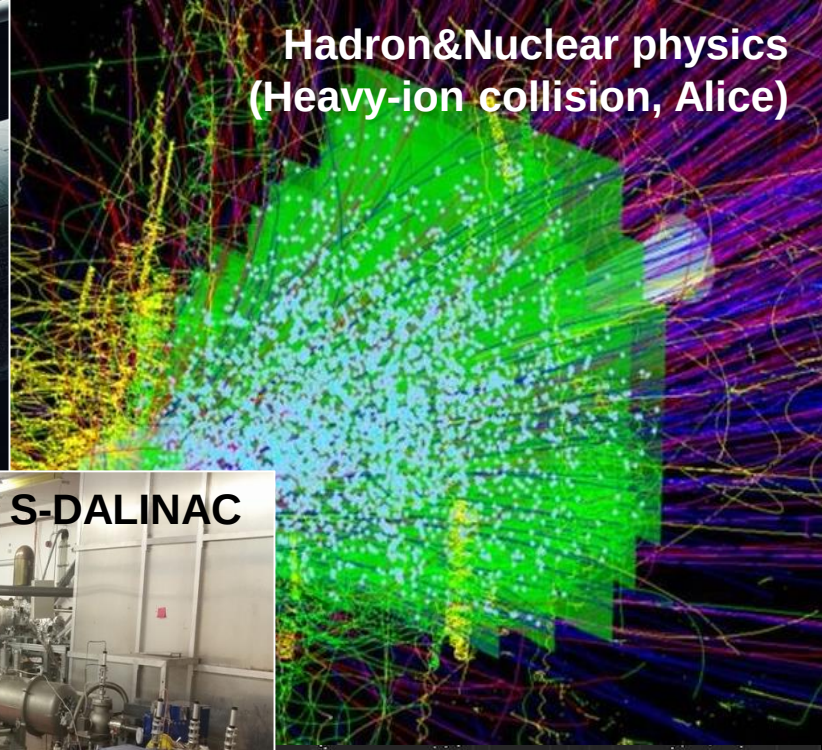


PUNCH data are diverse

- in size and rate
- in complexity and purpose
- in abstraction level





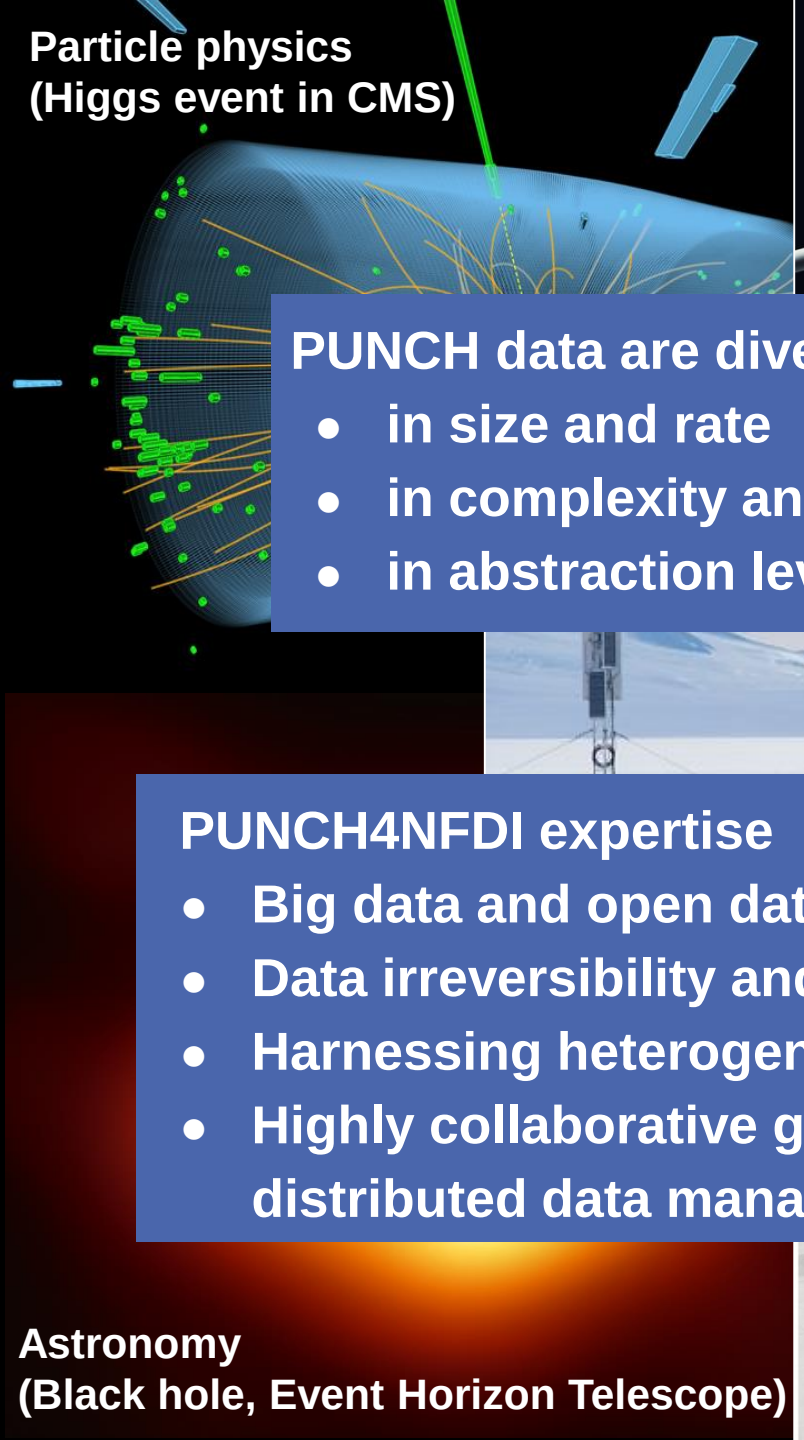


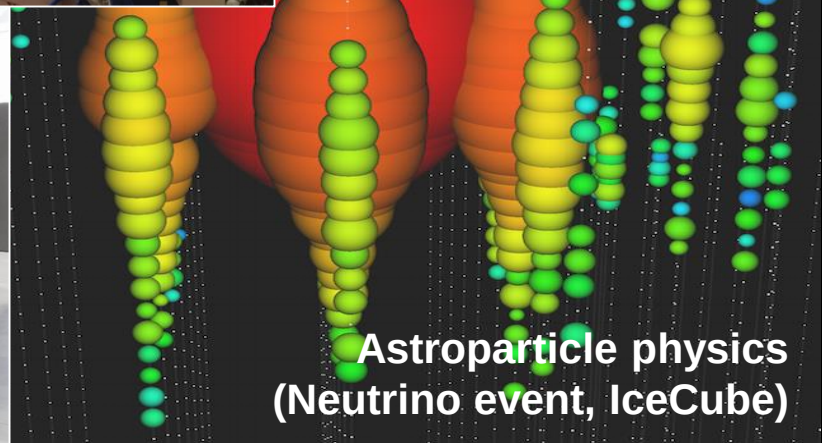
PUNCH data are diverse

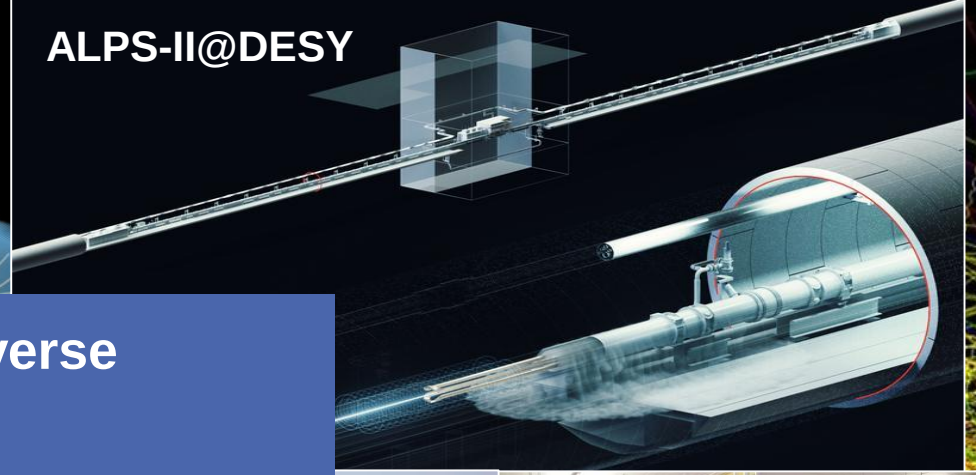
- in size and rate
- in complexity and purpose
- in abstraction level

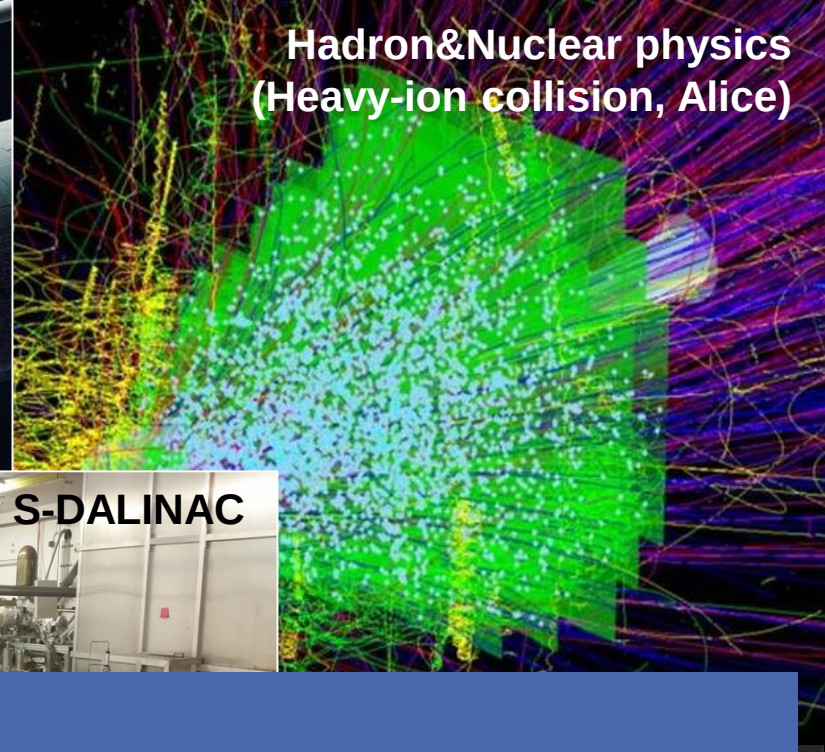
PUNCH4NFDI expertise

- Big data and open data
- Data irreversibility and reduction
- Harnessing heterogenous resources
- Highly collaborative globally distributed data management



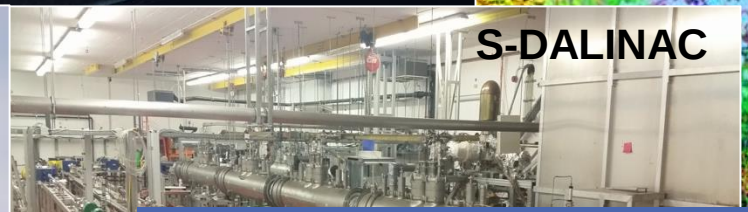


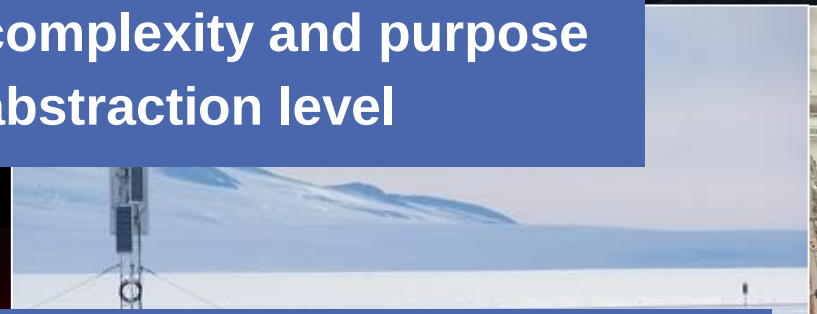




PUNCH data are diverse

- in size and rate
- in complexity and purpose
- in abstraction level



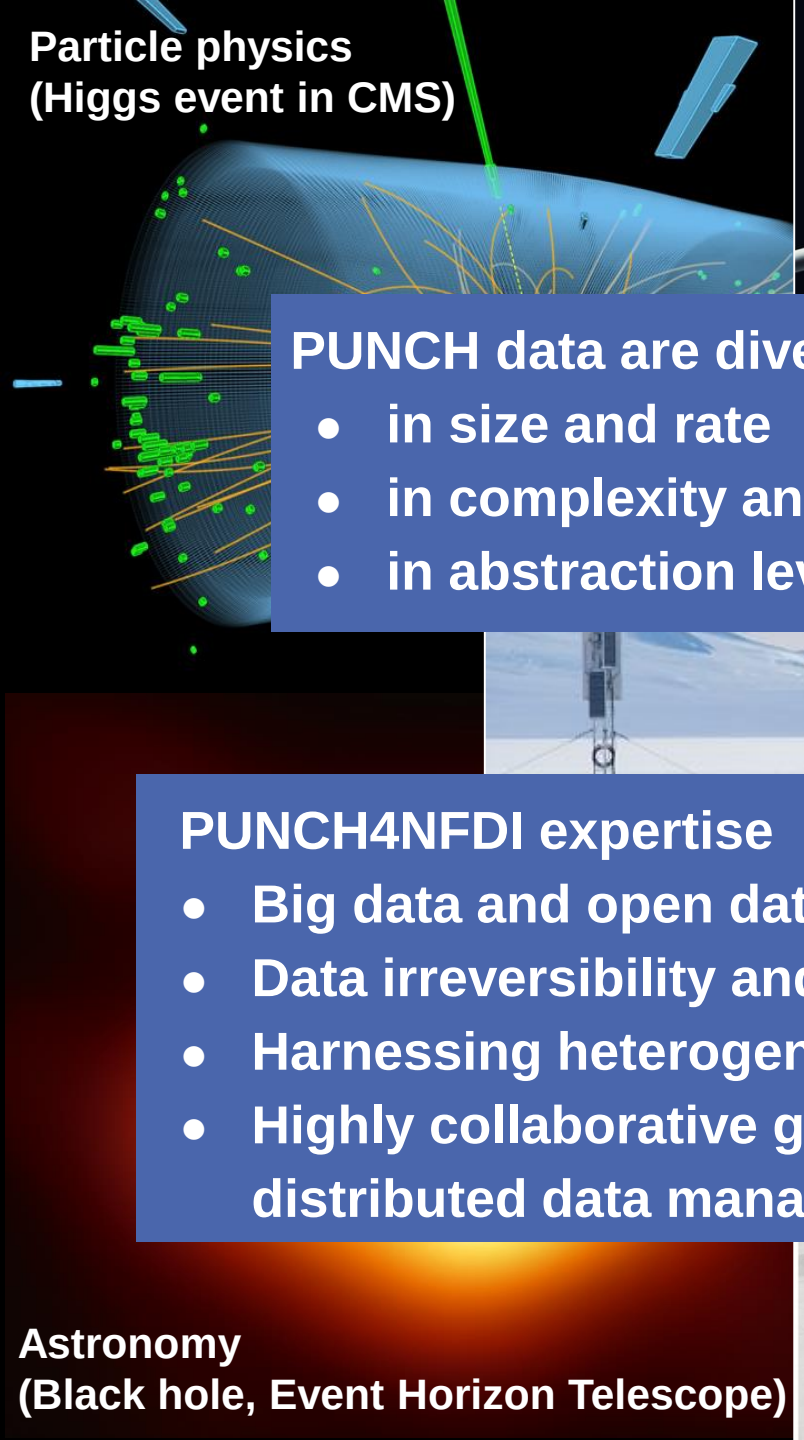


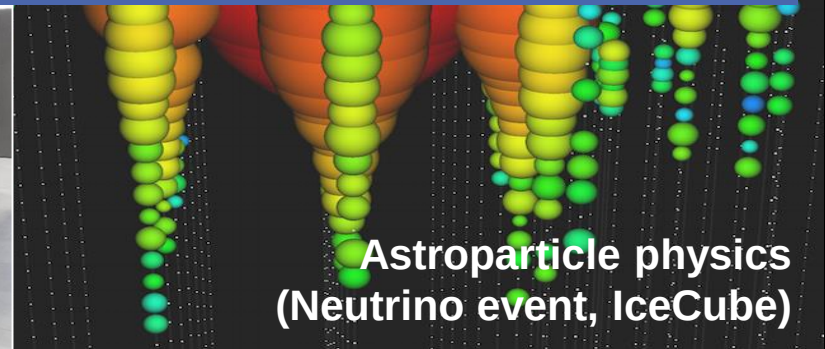
Challenges

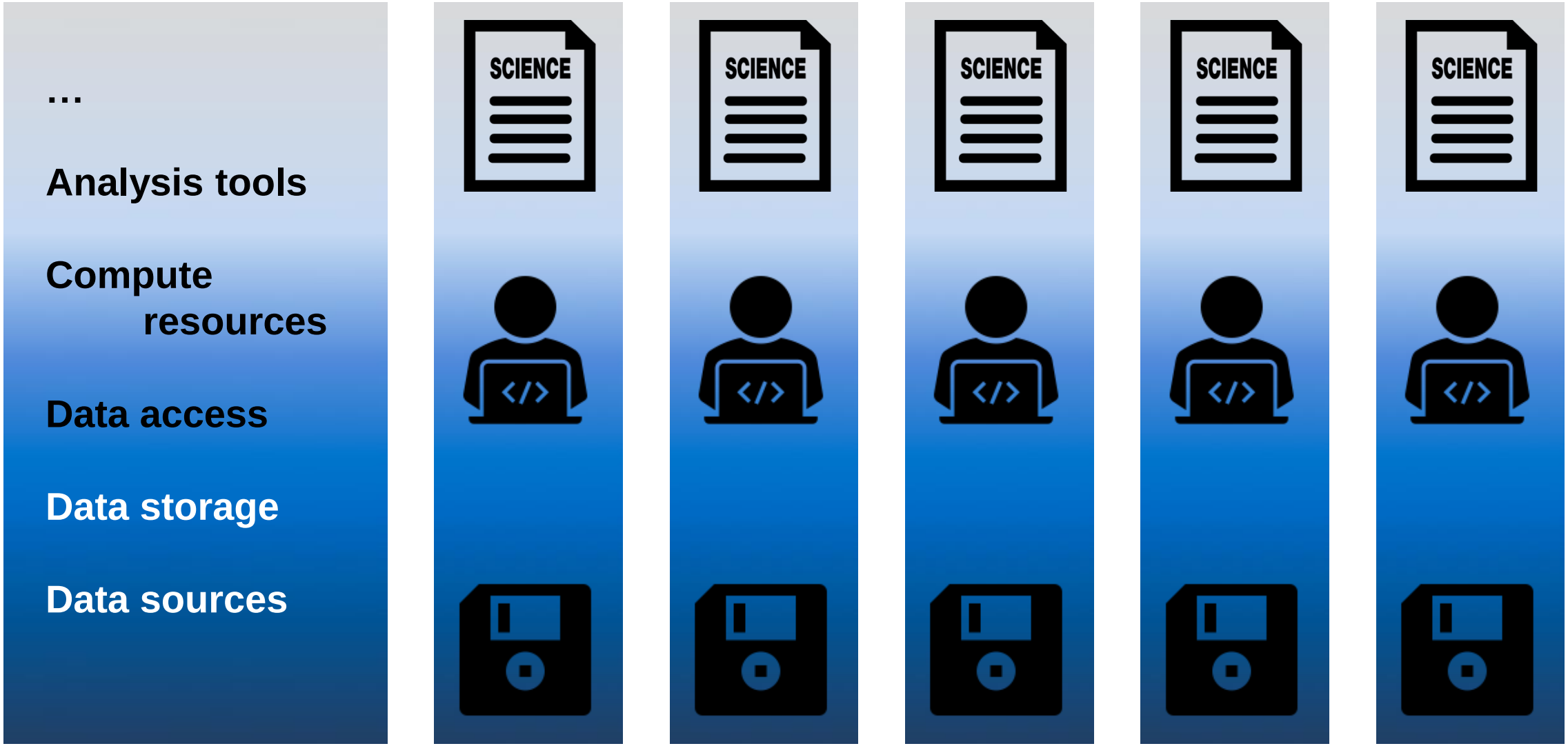
- FAIRification: data & workflows
- “big data” and “open data”
- Irreversibility challenge and data loss
- heterogeneous data & infrastructures
- transfer of knowledge

PUNCH4NFDI expertise

- Big data and open data
- Data irreversibility and reduction
- Harnessing heterogenous resources
- Highly collaborative globally distributed data management







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

…

Analysis tools

Compute resources

Data access

Data storage

Data sources

Users, Collaborations

Data Analysis

Cloud Services

Storage + Compute Resources

Data Sources

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

…

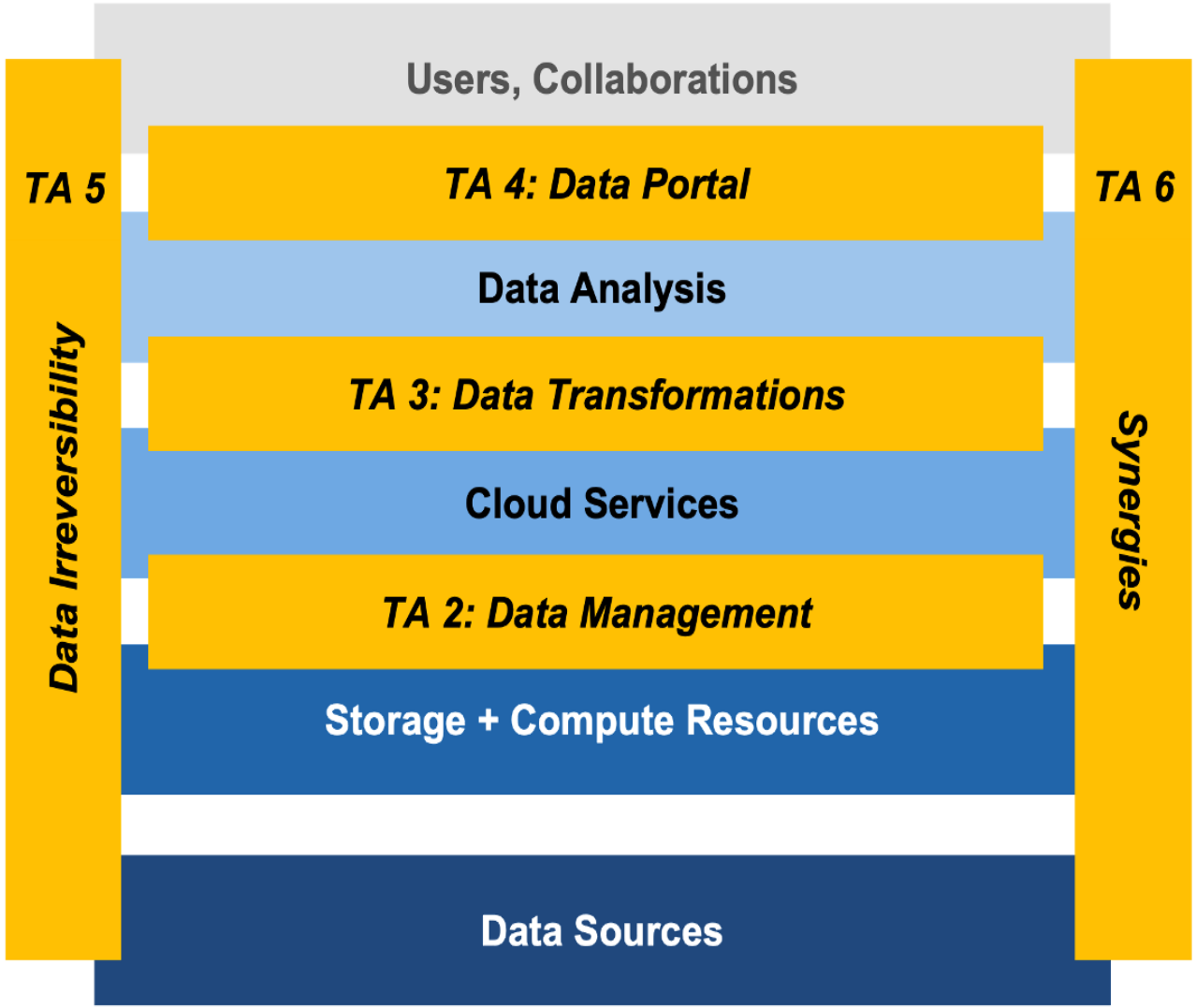
Analysis tools

Compute resources

Data access

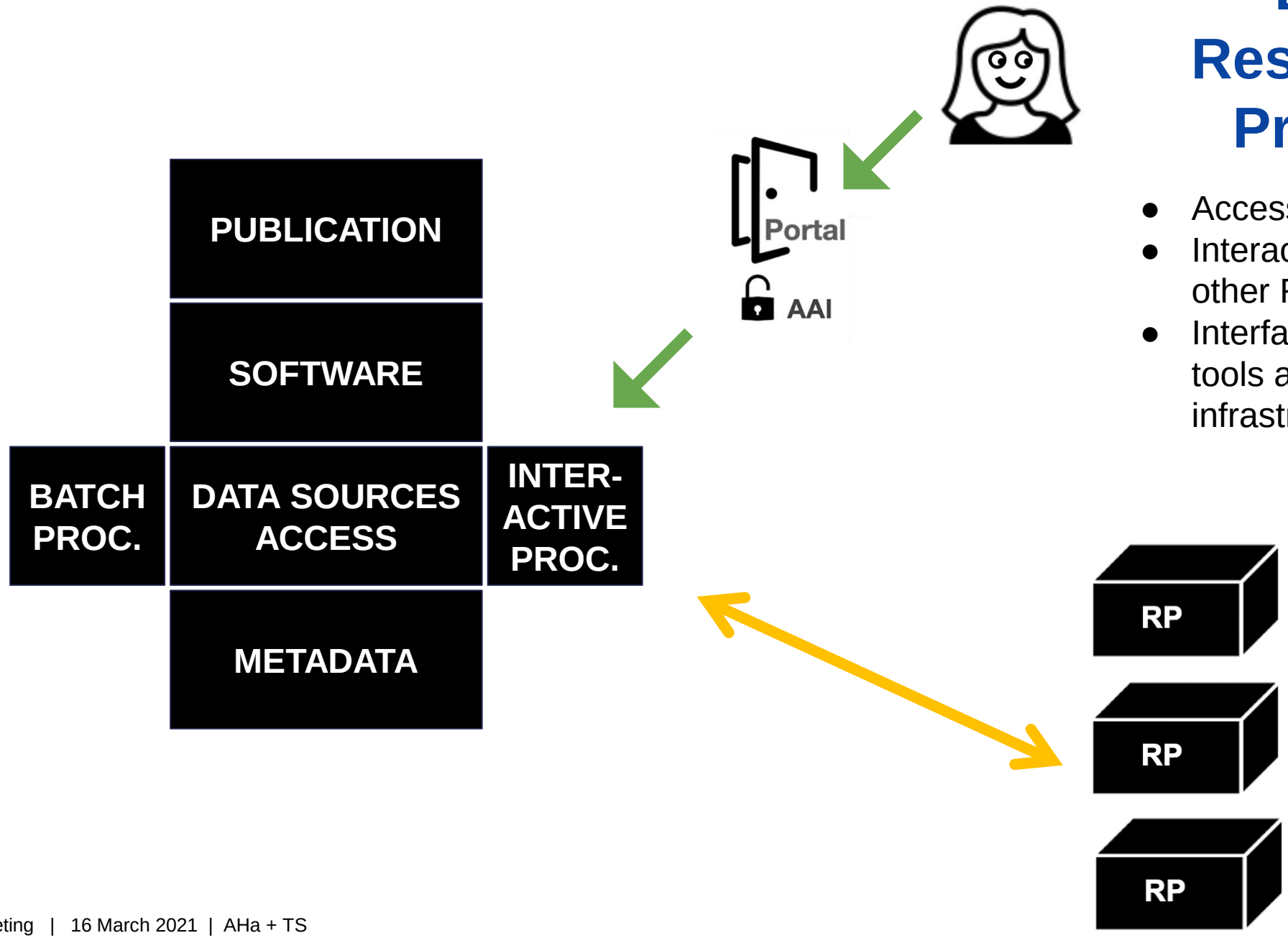
Data storage

Data sources



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

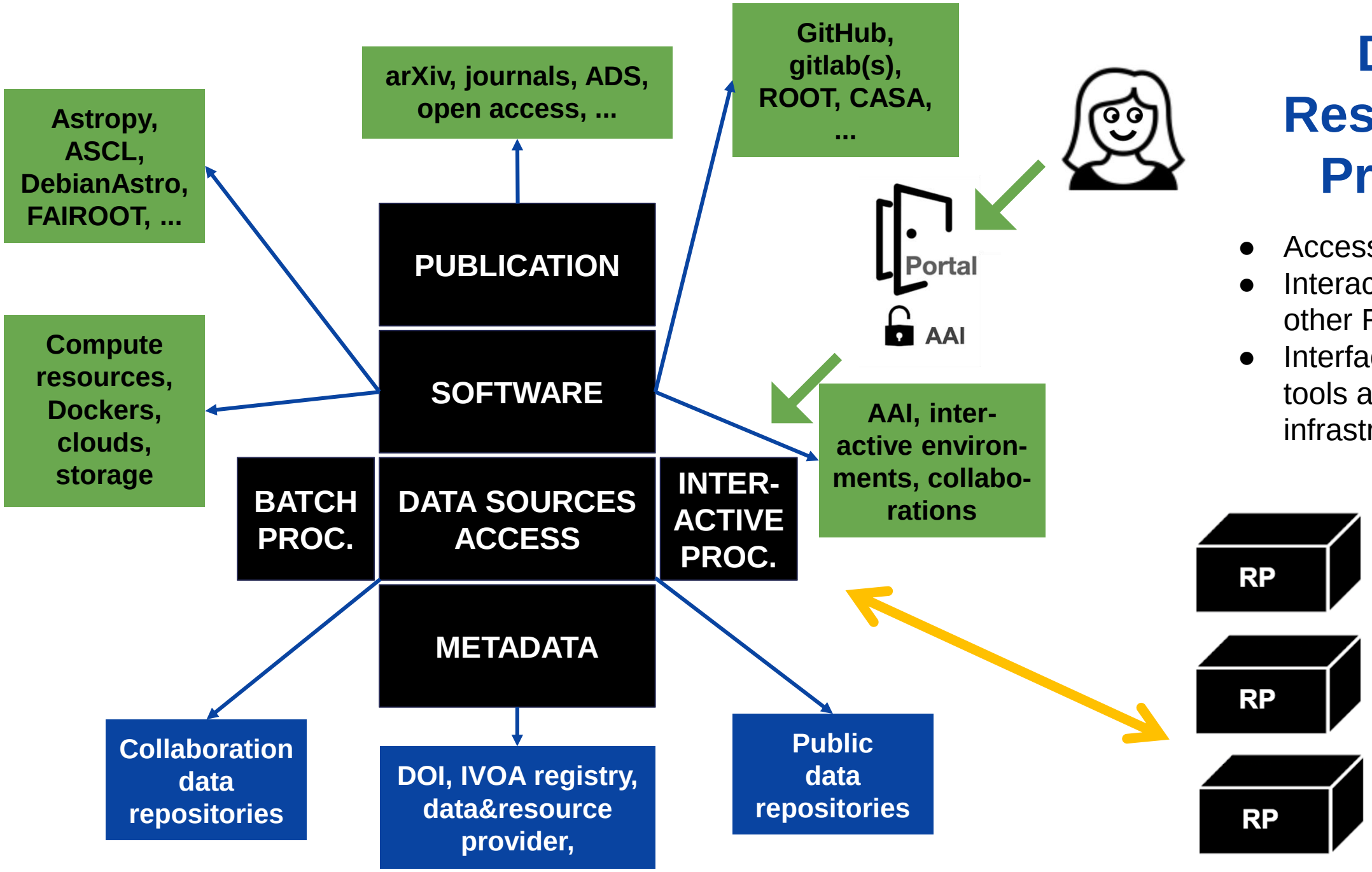
Digital Research Product



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

Digital Research Product

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



Task Areas

TA 1: Management and governance

TA 2: Data management

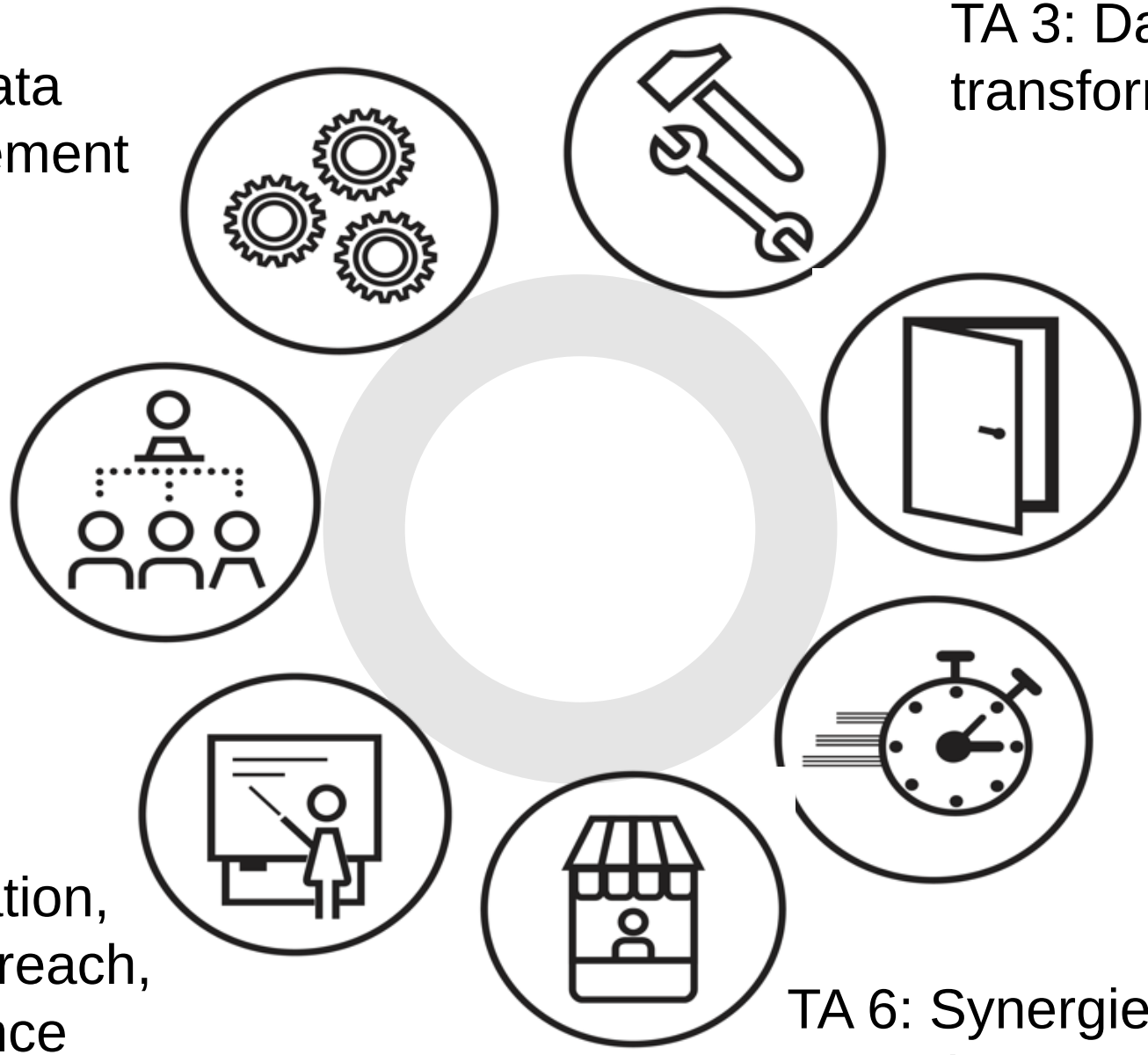
TA 3: Data transformations

TA 4: Data portal

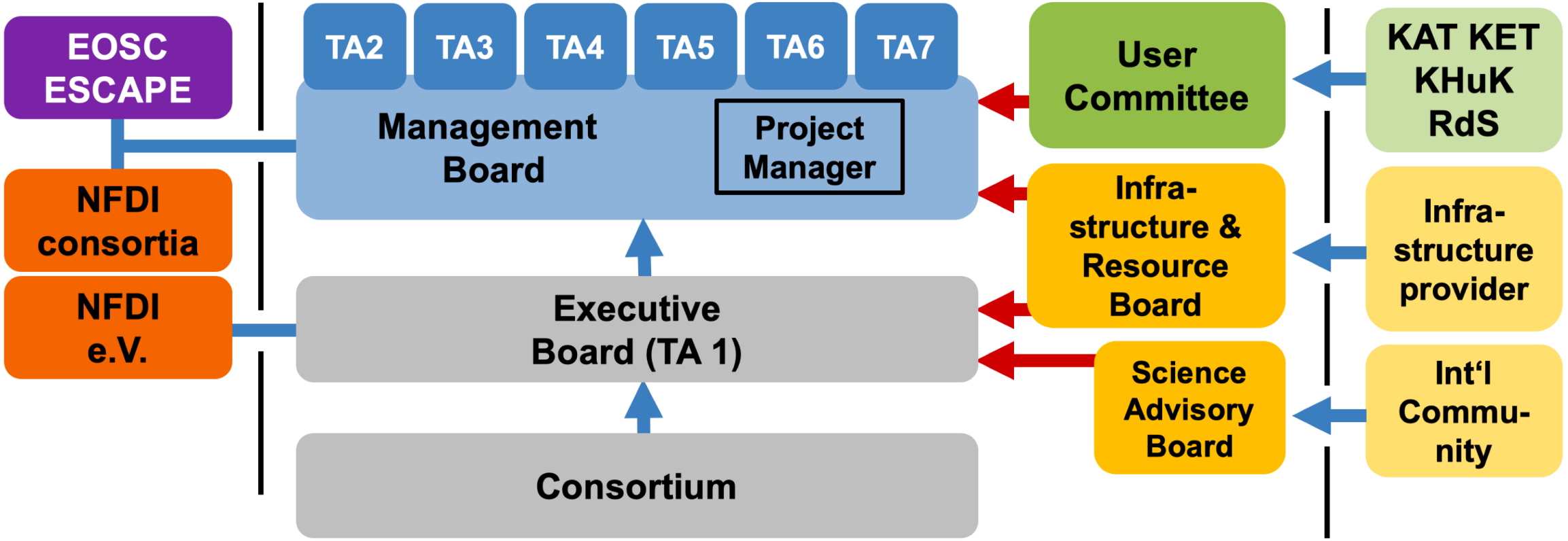
TA 5: Data Irreversibility

TA 6: Synergies & services

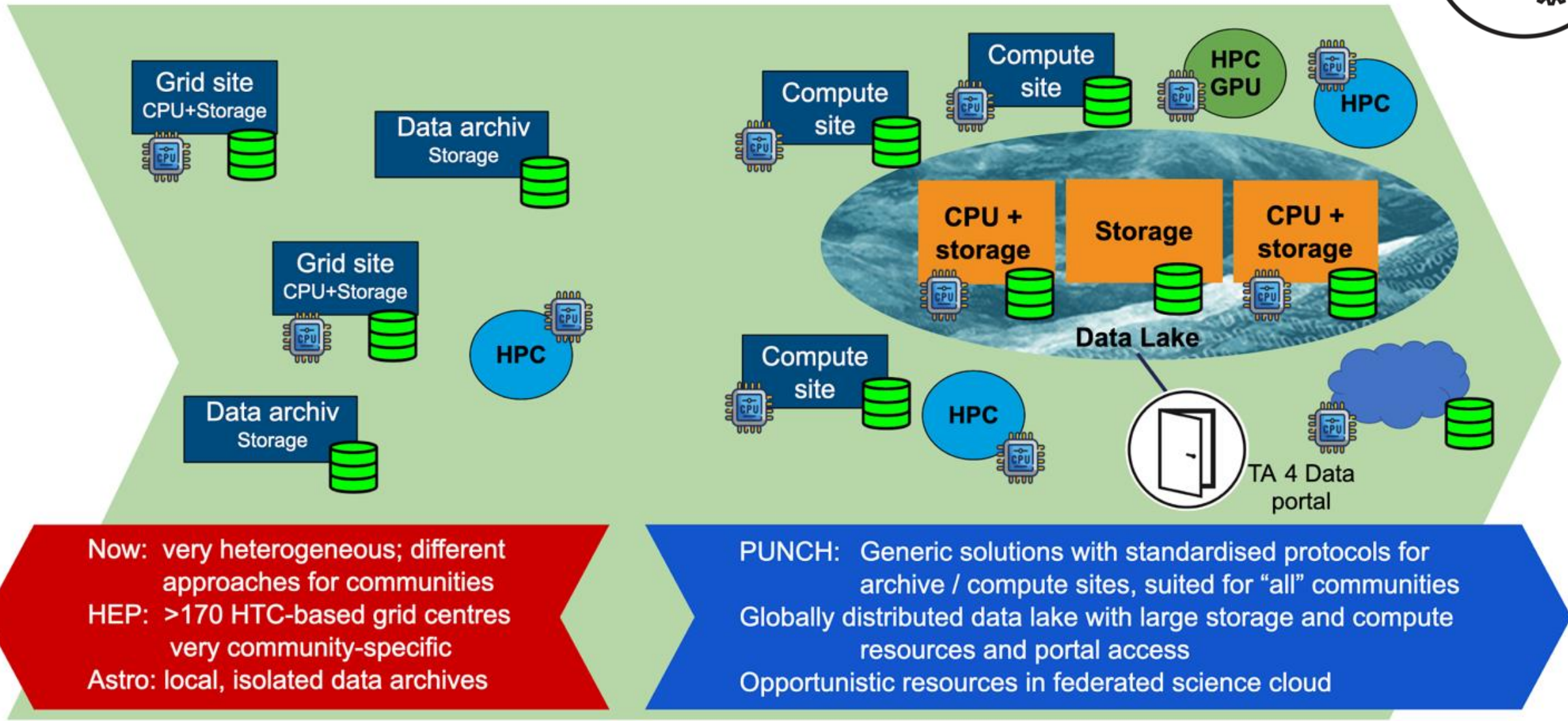
TA 7: Education, training, outreach, citizen science



TA1: Governance and Management

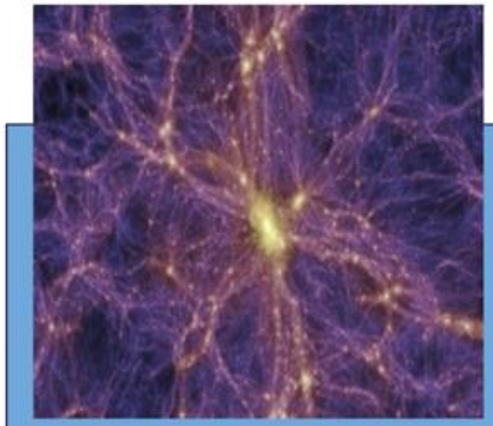


TA 2: Data Management

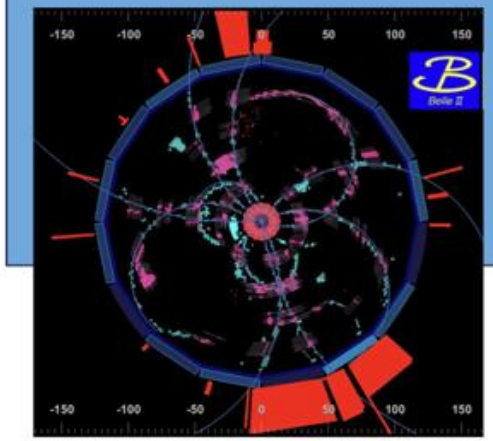


TA 3: Data Transformations

Integration of common tools into a data infrastructure based on code-to-data principle



Data and metadata

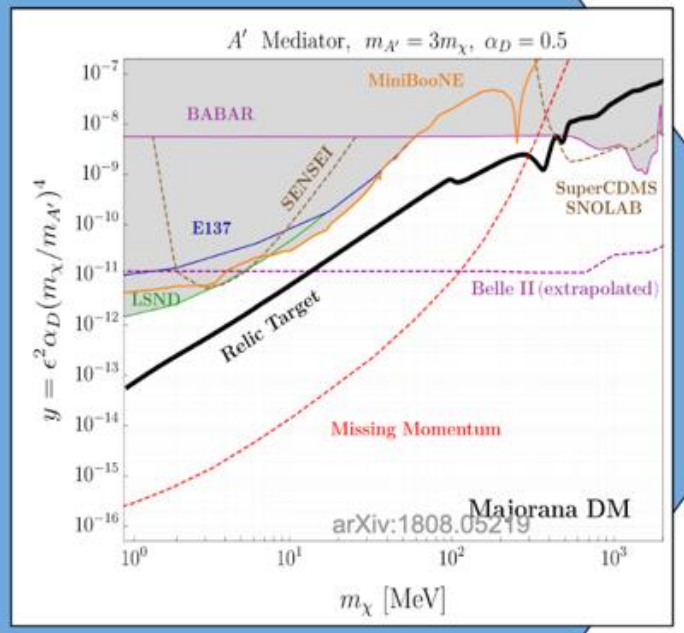


User

Image processing | Particle selection | Data modelling

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

Scientific result



Tools common to many science fields

TA4: Data Portal

Search & find

Access

Modify & add

Use & compute

Create & Store

RP 1

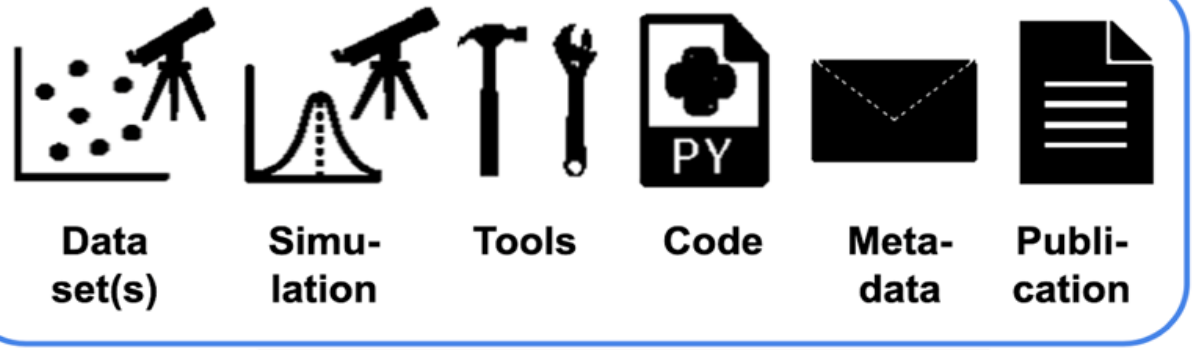
. . .

RP N

New RP

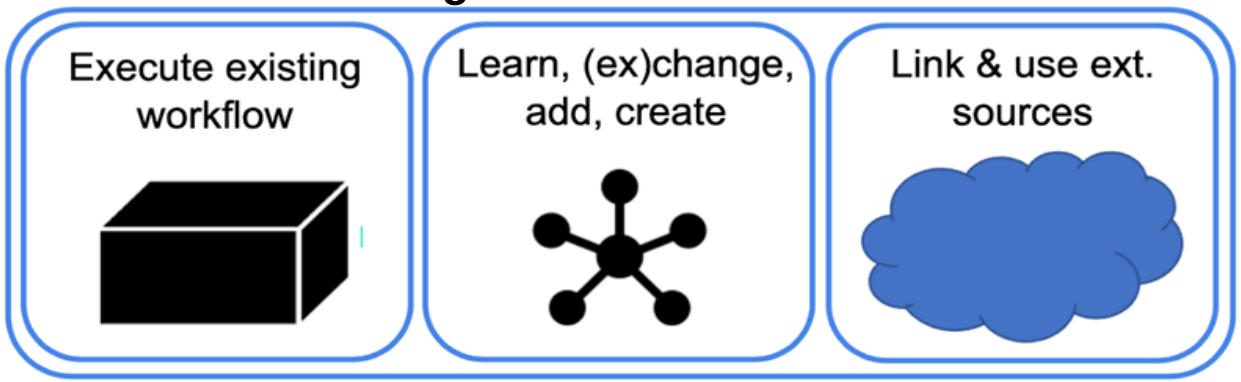
Research product contains executable workflow

Metadata description of workflow



Code / tools

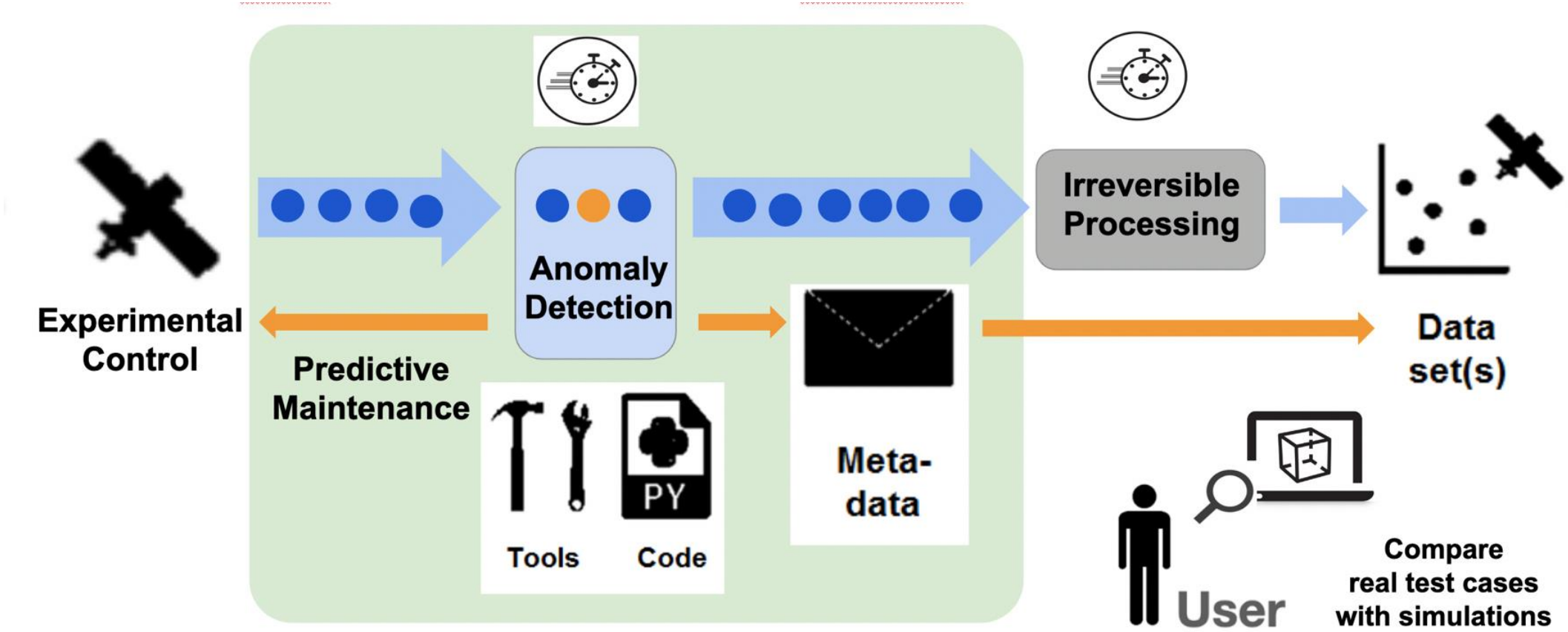
Computing / storage



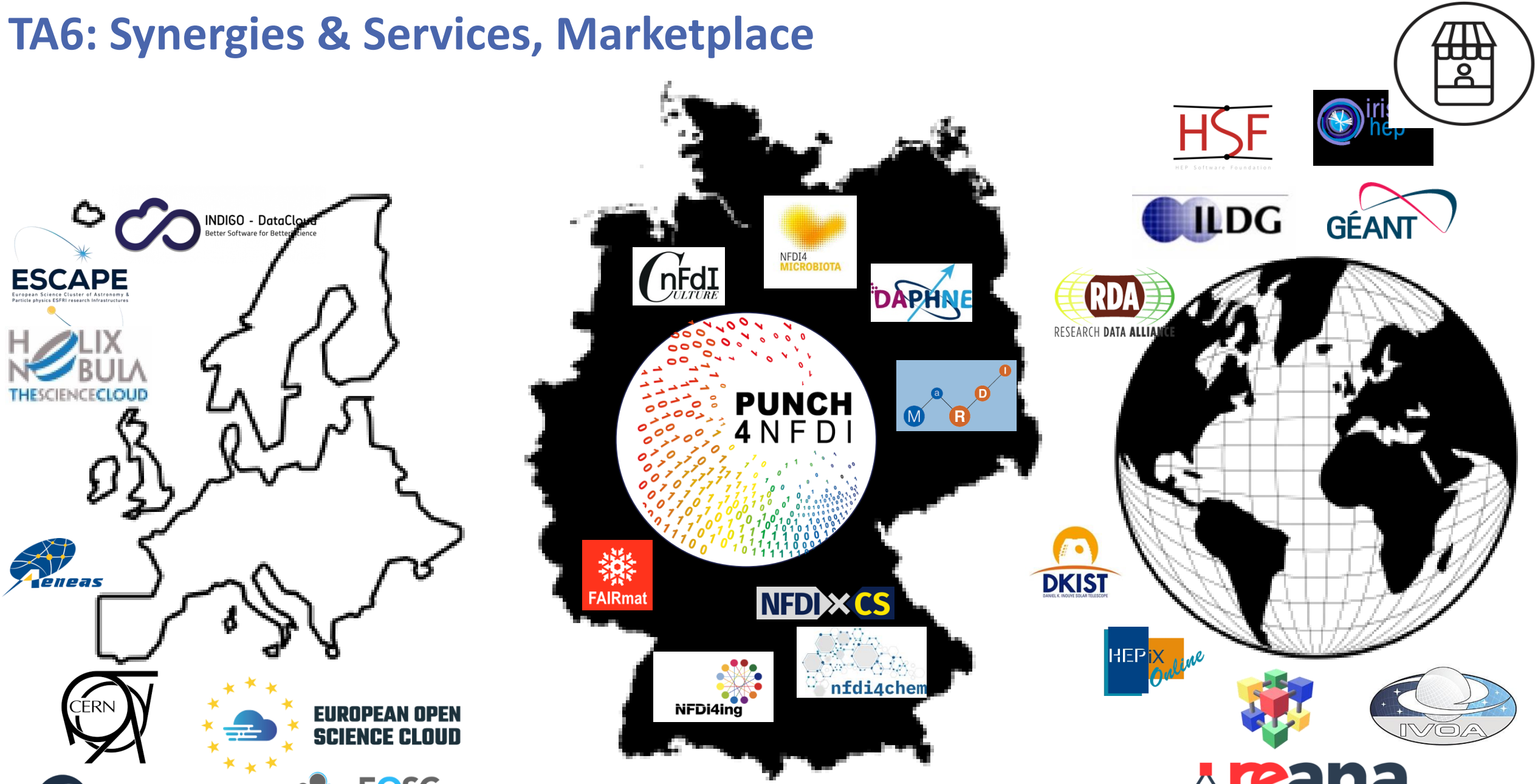
Many options for the user

TA5: Data irreversibility and real-time challenges

Impossible to store all data ⇒ irreversible loss of data - reality in e.g. HEP!
Challenges: implications for reproducibility, discovery potential & interpretation
Strategy: real-time dynamical filtering, dynamical archives & scalability



TA6: Synergies & Services, Marketplace



NFDI, ErUM-Data, HMC, KAT, KET, KHuK, RdS, ...

TA 7: Education, training, outreach, citizen science

Researchers

Promote expertise and career prospects

Training manager

communication training, school-academia network, events, resources

Schools

Communicate, support young talent, promote physics + informatics

PUNCH4NFDI Young Academy

Training resources
Courses, workshops
Standardized curriculum
Compute / data resources

Universities

Basic training and career advancement

Education manager

Public

Engage, promote deeper understanding, democratize science

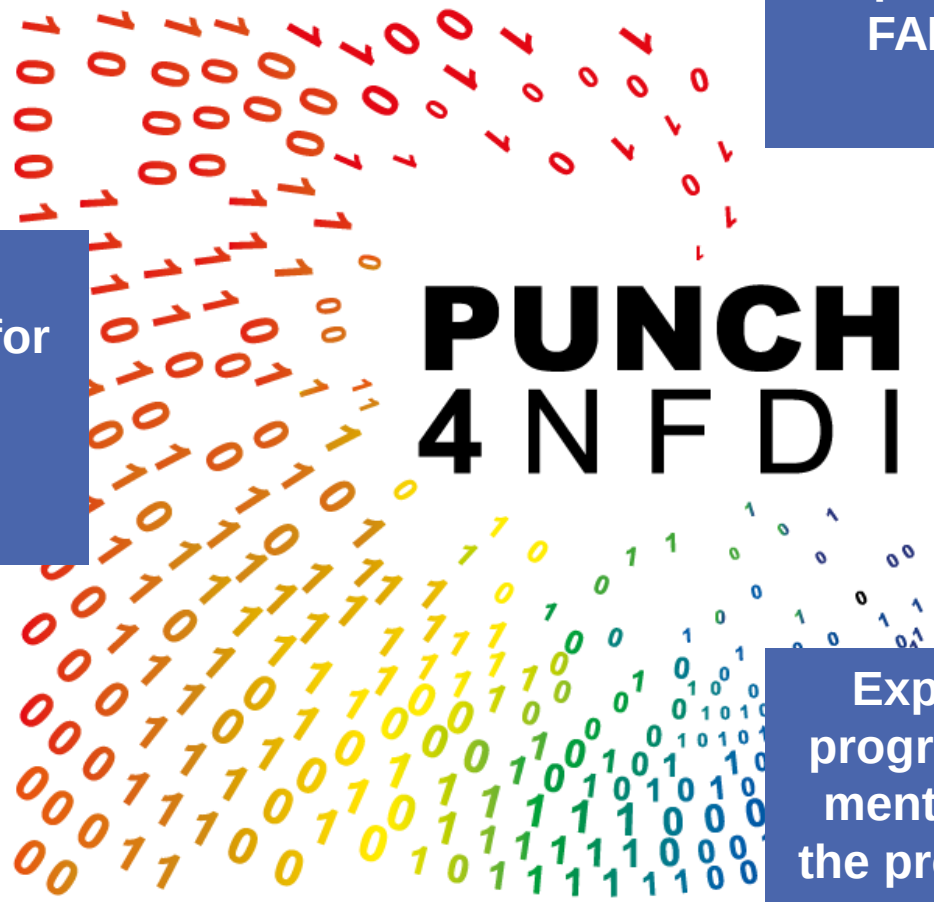
online projects & campaigns

“PUNCH”Lines

A leader in data management with expertise in big data, open data, FAIRification, use of hetero-geneous resources

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

Ready to go now!

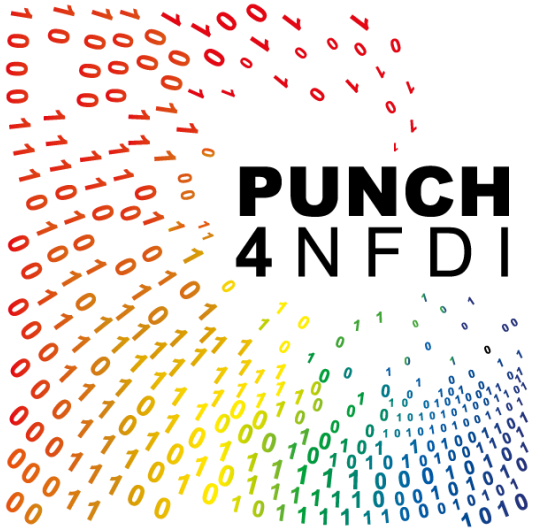


Experience, expertise & work programme to take data manage-ment to the next level and solve the problems that others will have tomorrow

An asset for the entire NFDI

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