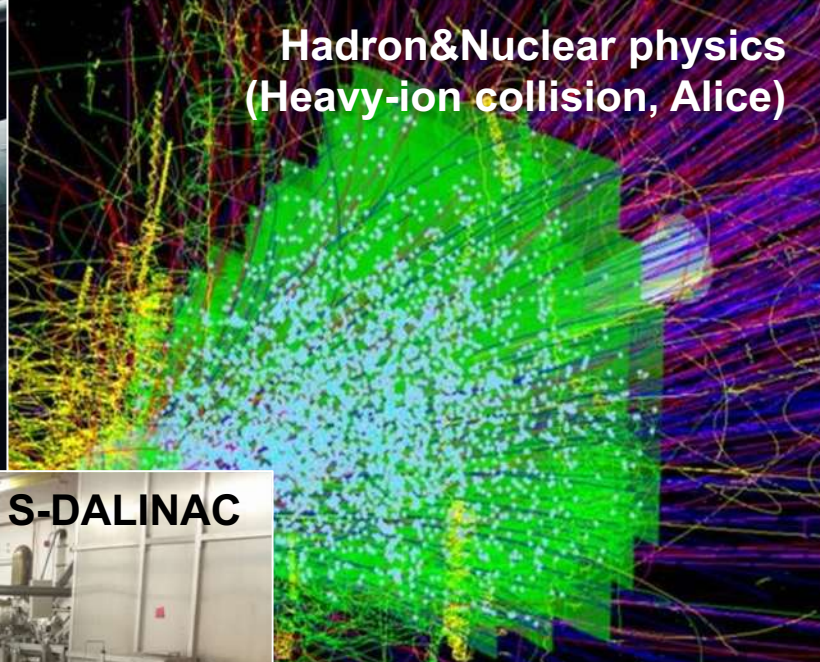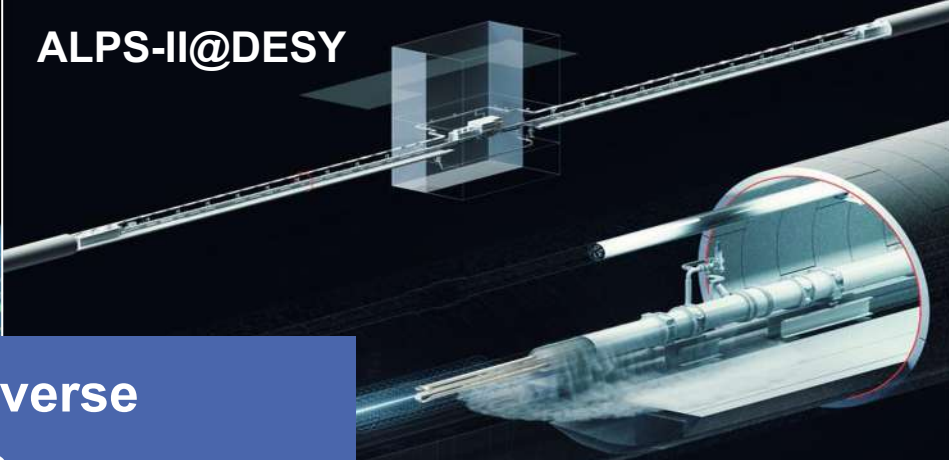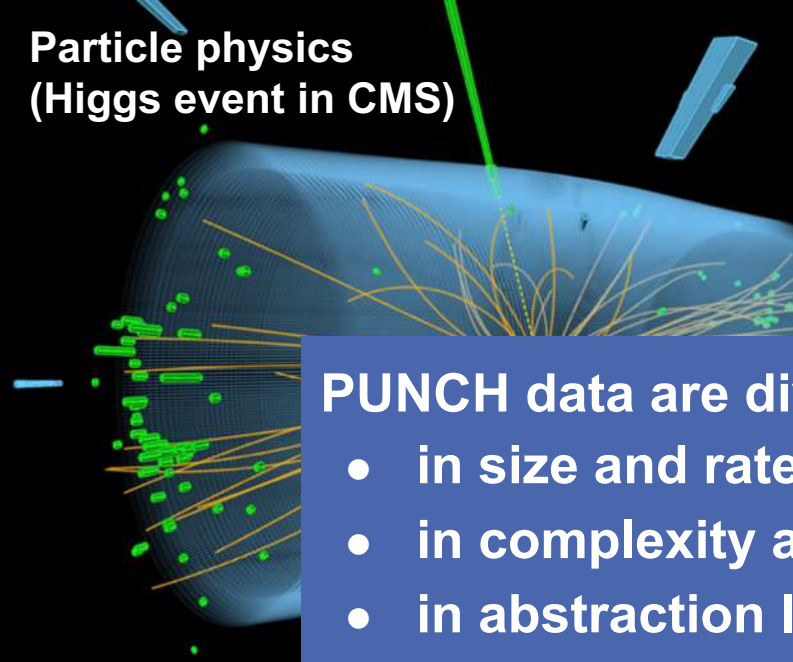# Particles, Universe, NuClei and Hadrons for the NFDI

Data Management in the PUNCH Landscape

Harry Enke (AIP), Stefan Typel (TUDa), Thomas Schörner (DESY)
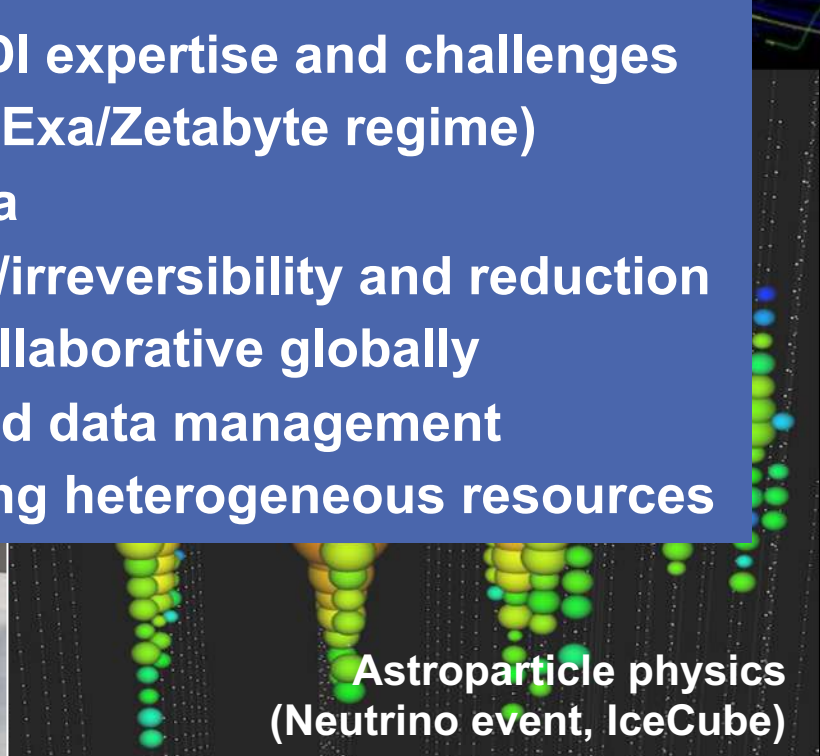**Data Management Plan Workshop, 2 March 2021**

PUNCH
4 N F D I

Particle physics
(Higgs event in CMS)

ALPS-II@DESY

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

S-DALINAC

PUNCH data are diverse
- **in size and rate**
- **in complexity and purpose**
- **in abstraction level**

PUNCH4NFDI expertise and challenges
- **Big data (Exa/Zetabyte regime)**
- **Open data**
- **Data loss/irreversibility and reduction**
- **Highly collaborative globally distributed data management**
- **Harnessing heterogeneous resources**

Astronomy
(Black hole, Event Horizon Telescope)

JUWELS@FZJ, GridKa ...

Astroparticle physics
(Neutrino event, IceCube)

How FAIR is PUNCH?
… and how to make it even FAIRer

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

- **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

- „**data management plan**" for each experiment / activity (developed in parallel to design & construction)

Further FAIRification - some examples:

- pragmatic transformation routines between existing metadata systems
- digital „research products" on „science data platform"
- education & training

Detailed work programme for each FAIR letter

- Focus: FAIRification of workflows (data lifecycle), gain of science by data combination and interoperability
- Generic solutions for entire NFDI
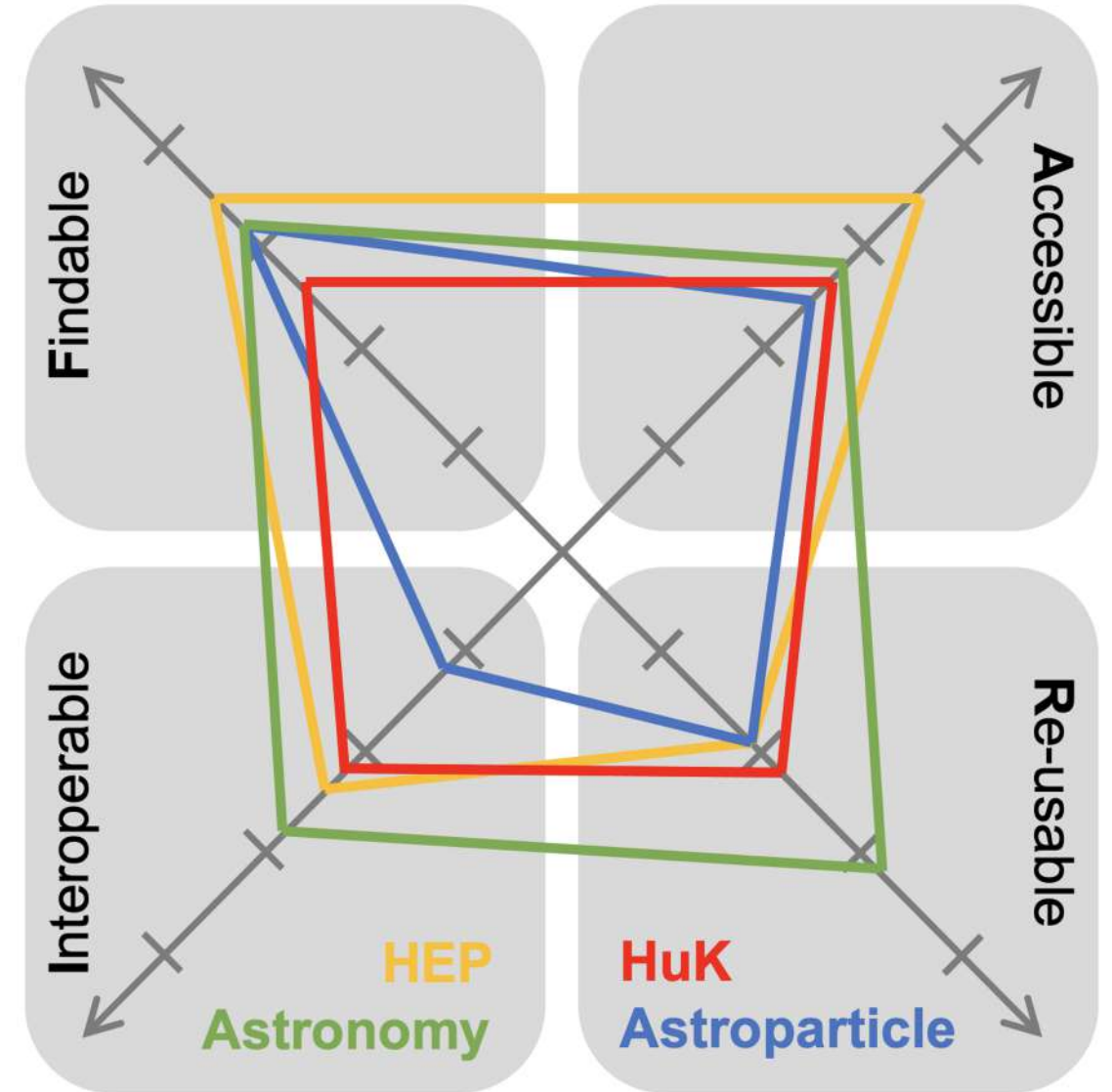
Particle Physics

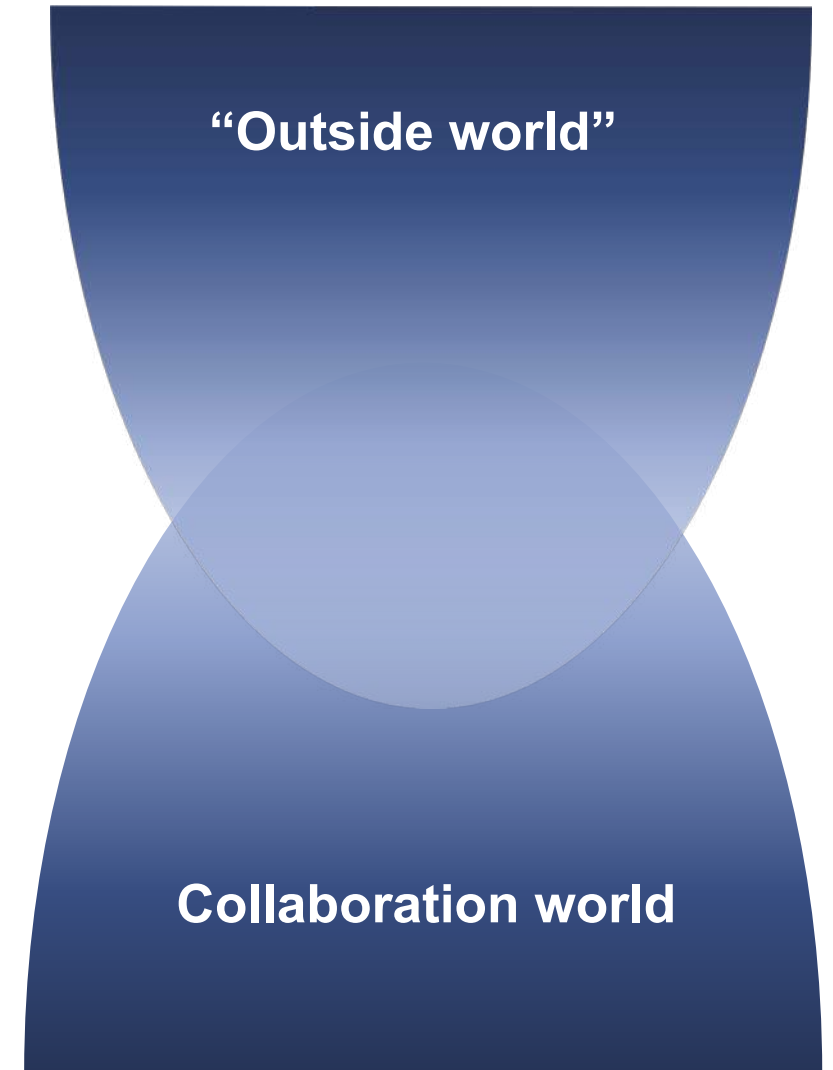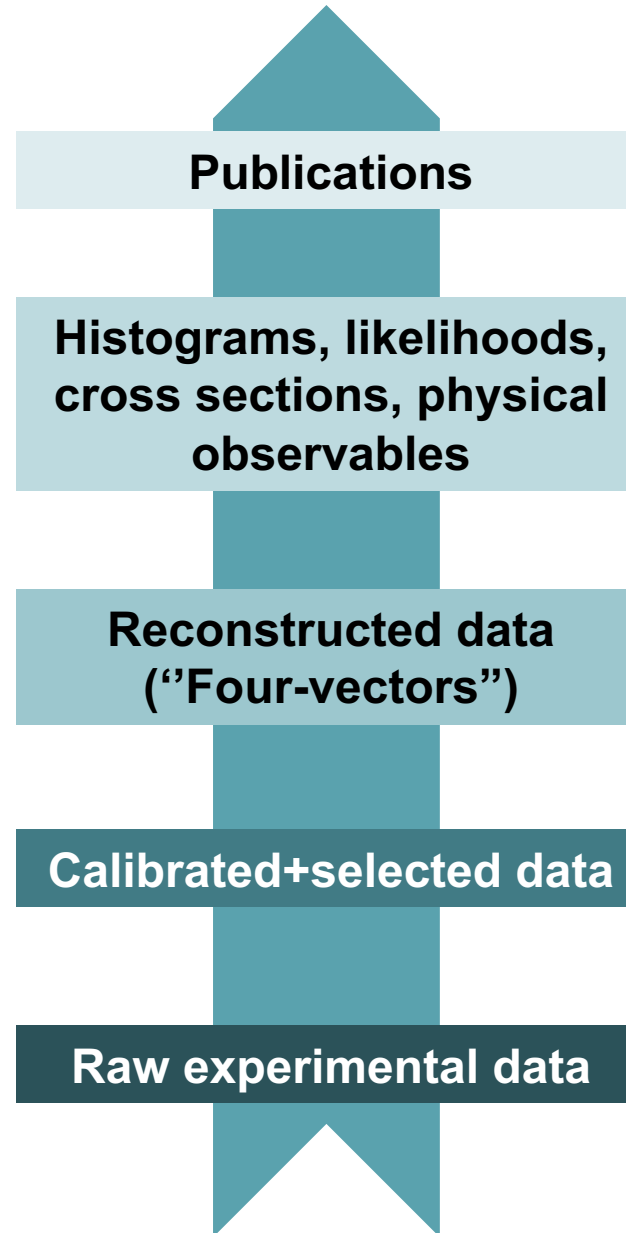## In large international collaborations

Of limited use without FAIR data?

More and more available in archives; in principle high degree of FAIRness, interoperability as critical issue (➔ Science Data Platform, Research Products)

Easily understandable data ("particles") in reduced data format ➔ "open data", find-/accessible, limited interoperability

Reduced (triggered&selected) data volumes, not suited for public, exp.-specific event data model (findable and accessible in collaborations)

Huge data volumes, uncalibrated data, not suited for science/public, exp.-specific event data model

Publications

Histograms, likelihoods, cross sections, physical observables

Reconstructed data ("Four-vectors")

Calibrated+selected data

Raw experimental data

"Outside world"

Collaboration world

Hadron and Nuclear Physics

In small collaborations (e.g. universities DA, F, K, M, MZ, … in PUNCH4NFDI)

Arbeitsgruppen nicht Teil 'großer' Kollaborationen/Forschungseinrichtungen
- sehr breites Spektrum an Arbeitsgebieten in Experiment und Theorie
- heterogene Art/Struktur der Daten, nicht unbedingt umfangreich

Probleme
- langfristige Aufbewahrung von Forschungsdaten
- Umsetzung von FAIR-Standards
- professionelle personelle Unterstützung

Datenmanagement
- oft individuelle, lokale Lösungen
- zum Teil Nutzung der Uni-Angebote (Rechenzentren/Bibliotheken) für Erstellung von Datenmanagementplänen (z.B.: TUdmo@TUDa, RDMO@GUF)
- zunehmend gefordert von Geldgebern, expliziter Teil von Projekten (SFB, . . .)

Einbettung in PUNCH4NFDI (Stichworte "*Science Data Platform*" und "*Research Products*")
- Standards für Metadaten
- Nutzung von entwickelten Werkzeugen
- Vereinheitlicher Zugang zu Daten

Astronomy
Data Management

Data Flow →

Observations with big facilities

Surveys:	Input Catalogs/Observation	Processing / Pipelinig	Catalog Production / Publication	Individual Exploitation
PI –Mode:	Plan / Observation	Processing	Individual Exploitation	partial Data Publication

Observations with small (institutional) facilities:

Surveys:	Input Catalogs	Processing	Individual Exploitation	Catalog Production / Publication
PI –Mode:	Observation Plan	Processing	Individual Exploitation	

Simulations

Big data (HPC):	Proposal	Simulation run	Analysis / Individual Exploitation	Data Publication
Small Facilities:	Idea	Simulation run	Analysis / Individual Explotation	

Software

Observation	Archival sources	using old code / complete rewrite	Analysis : mostly public av. code + commercial	IVOA, data center codes, Open. Source
Simulation	Models + Observational Data	using old code / complete rewrite	Analysis: mostly public av. codes + commercial	IVOA, data center codes, Open Source

Platform, Portal, Research Product

Complete data lifecycle

Portal: access to research products landscape and infrastructure for RP lifecycle ("platform").
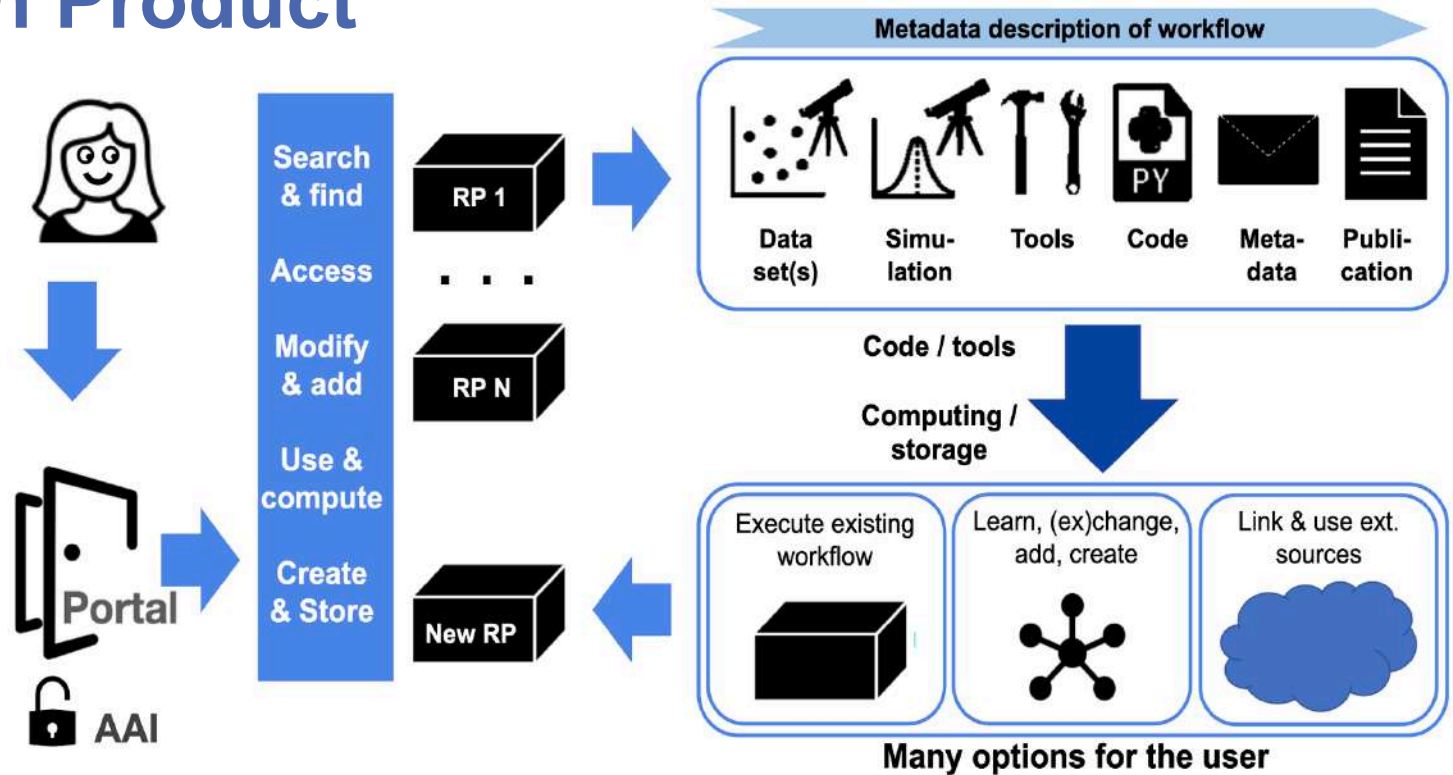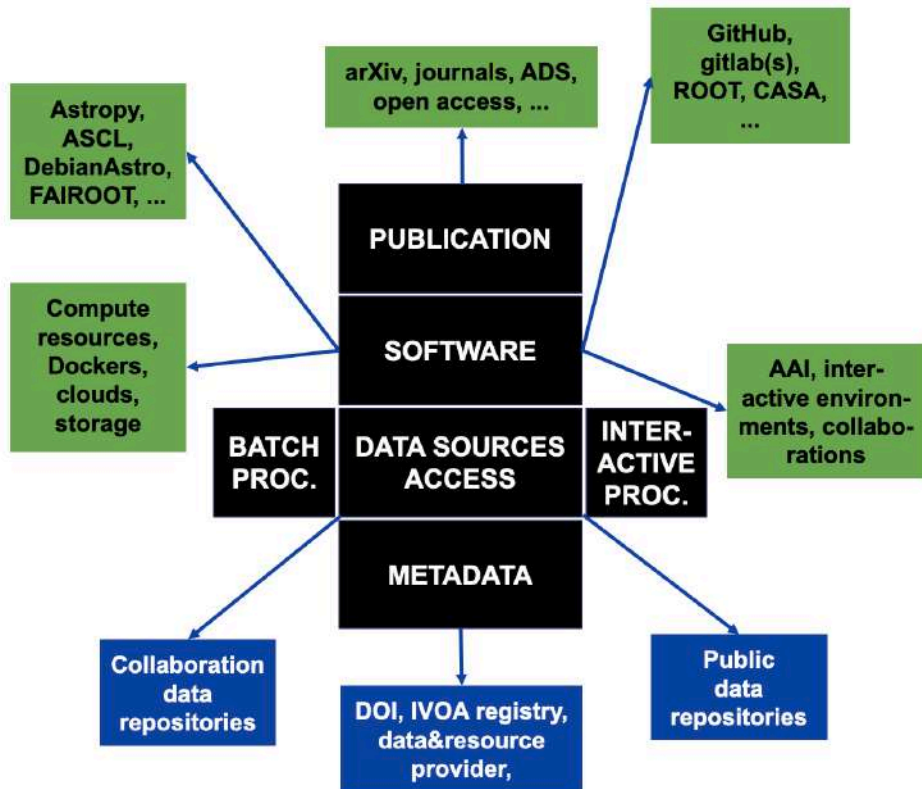
Findability: portal! Requires schemes for connecting data and publications (ADS, Inpire, ...), curation process, and enrichment of RPs with metadata.

Accessibility: requires AAI and open, free, easy-to-implement communication protocols.

Interoperability: prerequisite for PUNCH-SDP! Needs interfaces e.g. for combined analysis of data sets.

Re-usability: questions of open licenses, new concepts of metadata for reproducibility of results, provenance standards.

Our Partners

Thank you!

The PUNCH4NFDI Consortium

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

Contact:

Mail: punch4nfdi@desy.de

Web: www.punch4nfdi.de

Twitter: #punch4nfdi