Particles, Universe, NuClei and Hadrons for the NFDI

Open data in the PUNCH landscape and the PUNCH4NFDI proposal

Thomas Schörner for the consortium

Organisational Details

- Please mute your microphone
- Use the “raise hand” tool in the discussion parts
- Speakers please stick to your times (we have 15-minute slots INCLUDING discussion)
  - After 8 and 10 minutes we will fire a warning shot and cut you out after 12 minutes!
- We are NOT recording the event …
- … but we are taking notes – especially of the discussions – and intend to make them available after one round of editing with the speakers.
- Finally: We can’t do full justice to all fields and topics today – please consider this workshop as a starting point for deeper discussions and for the work of PUNCH4NFDI in the field of open data.

*Particles, Universe, NuClei & Hadrons for the NFDI

PUNCH4NFDI* in one Slide

A consortium for the NFDI

PUNCH4NFDI

Represents **(astro)particle, astro, hadron & nuclear physics** in the NFDI.

Specific strengths: **big data and open data**; ready to take leading role in NFDI.

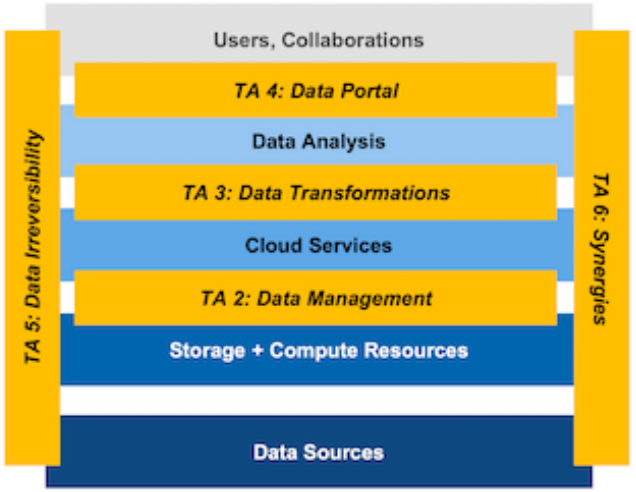
Broad community representation: > 40 partner institutes

Our offer:

A **layered model of data management** with scalability that allows for easy FAIRification

Numerous services to develop community-specific approaches in this direction

The **PUNCH science data platform** evolving around **advanced research products**



Task areas:

Data management

Data transformations

Data portal

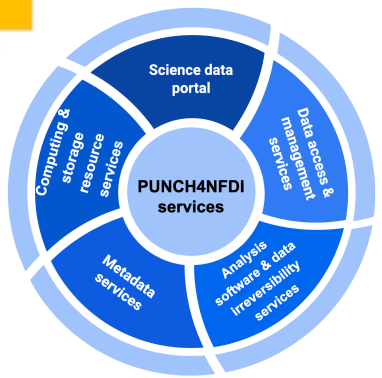
Data irreversiblity

Synergies&services, Teaching&outreach

Timeline and general situation:

9 consortia (out of 30 max in NFDI) funded in first NFDI round – none from physics. Now competing e.g. with FAIRMat, DAPHNE4NFDI

Submission of proposal 30 Sep 2020; evaluation by review panel 10 Dec 2020; grants by July 2021; funding start 1 Oct 2021



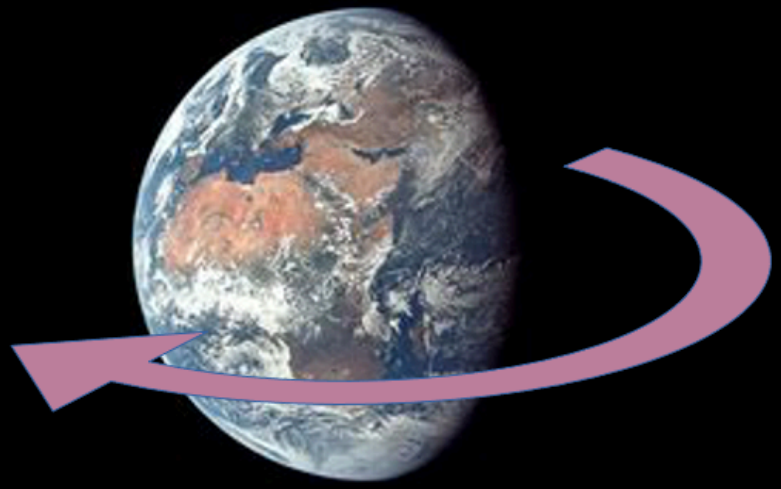
Services:

Evolving around research products and their dynamic life cycle

Connecting to entire NFDI – a cornerstone of research data management in D.

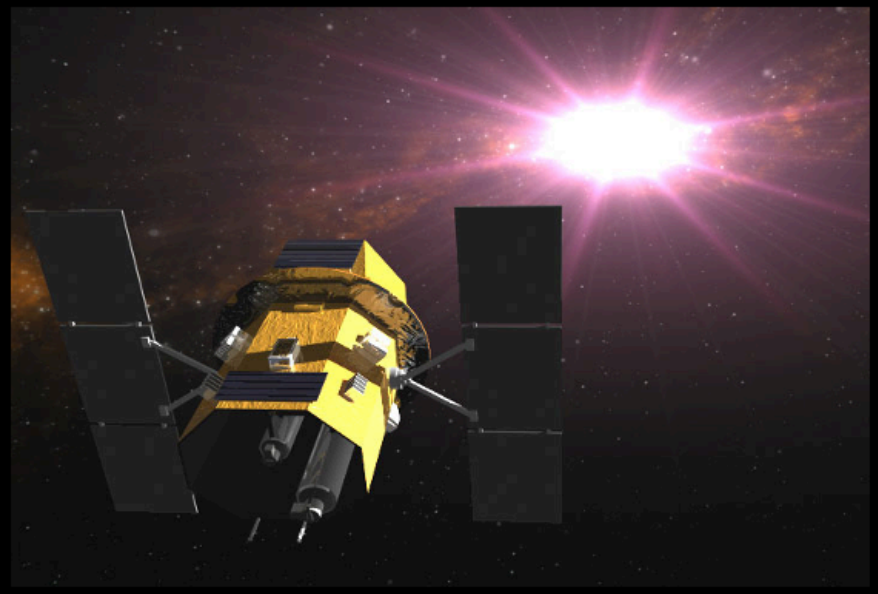
Open Data

Yesterday morning, a cosmic explosion was discovered and announced.



This morning everyone can download the data and combine them with own/other observations.

Follow-up satellite observations were proposed yesterday evening.



The measurements were made last night.

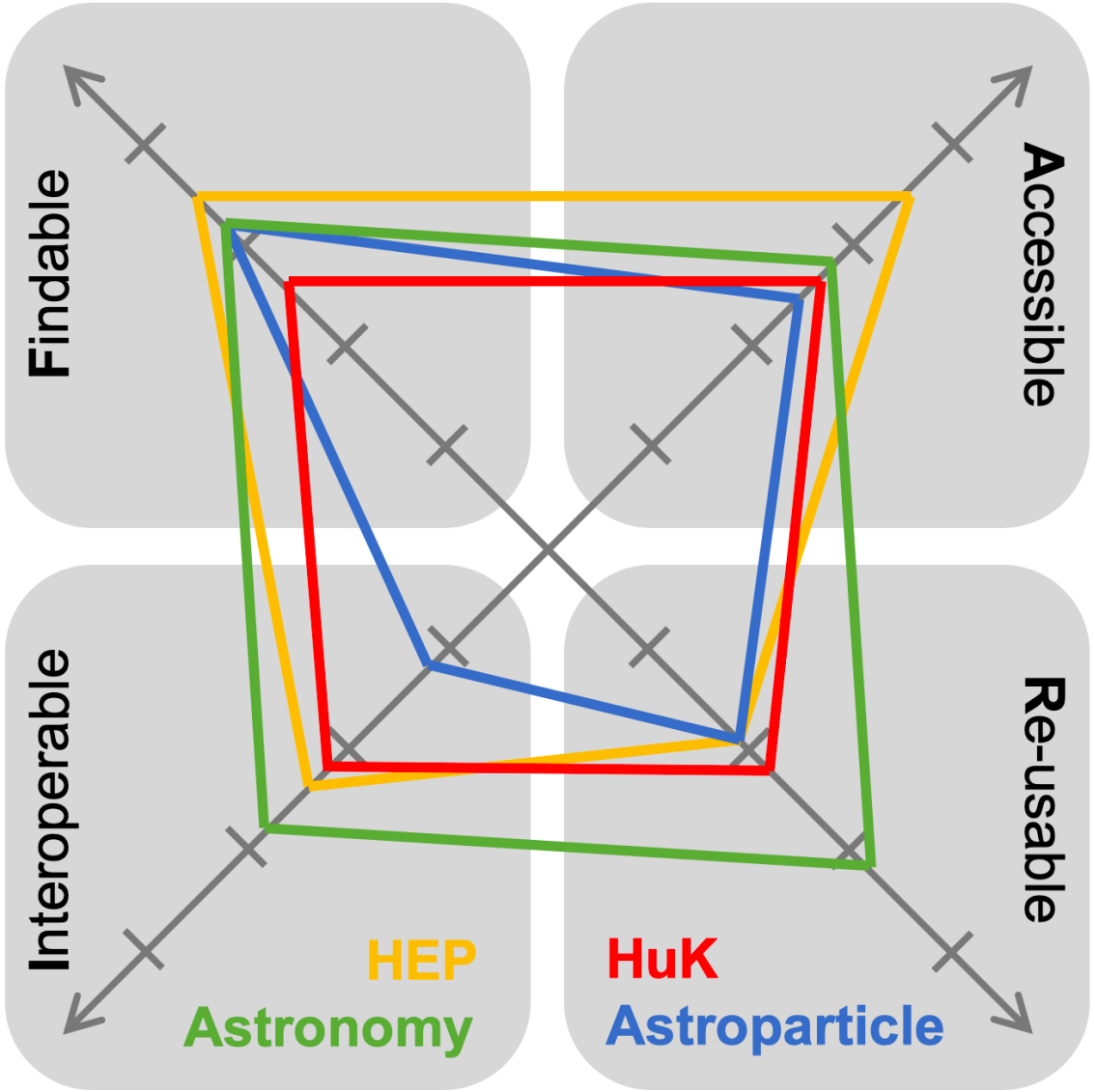
PUNCH4NFDI

Heterogeneous data landscape

"Open" = a_1×Findable + a_2×Accessible +
a_3×Interoperable + a_4×Re-usable

PUNCH: a long history of "open" aspects

- Sharing tools and data since long (esp. astronomy satellite missions, …)
- Particle physics and hadron&nuclear physics: data completely open with in collaborations; aiming for usability by non-collaborators; often release data at higher abstraction levels (histograms, likelihoods, …), role of archives (Inspire, arXiv, …)
- Astro community: sharing and using of data collections widely spread (FITS format, VO, short embargo periods, service perspective, …)
- Astroparticle: recent developments like KCDC ….



Today

... Starting a survey of PUNCH and other sciences

- PUNCH quite advanced in open data issues
  – openness as a central objective of our planning
    - Emphasis on digital research products and the PUNCH science data platform

- Openness as central requirements for
    - … new science
    - … data combination
    - … citizen science

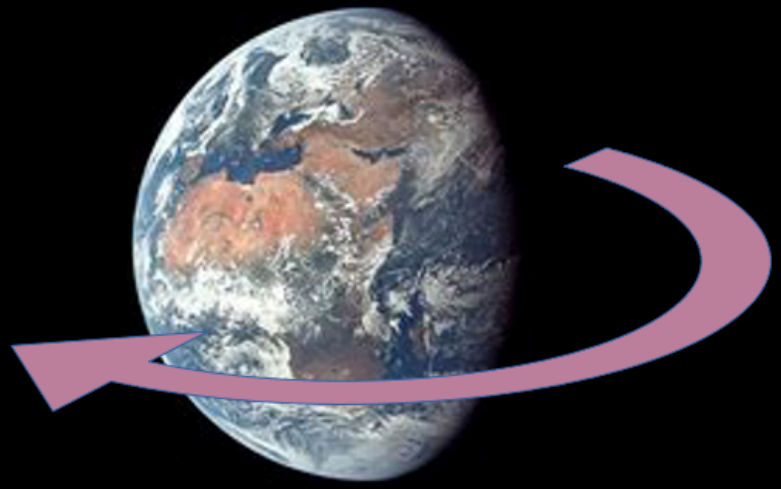
- Punch als vermittler / enabler fuer Kleiner communities

- Today: discussing experiences from and situation of communities
    - What is there? What is working? What are problems? Long-term visions?
    - How can PUNCH4NFI react to the different needs? Where and how to start?

➔ Devise actions to be taken

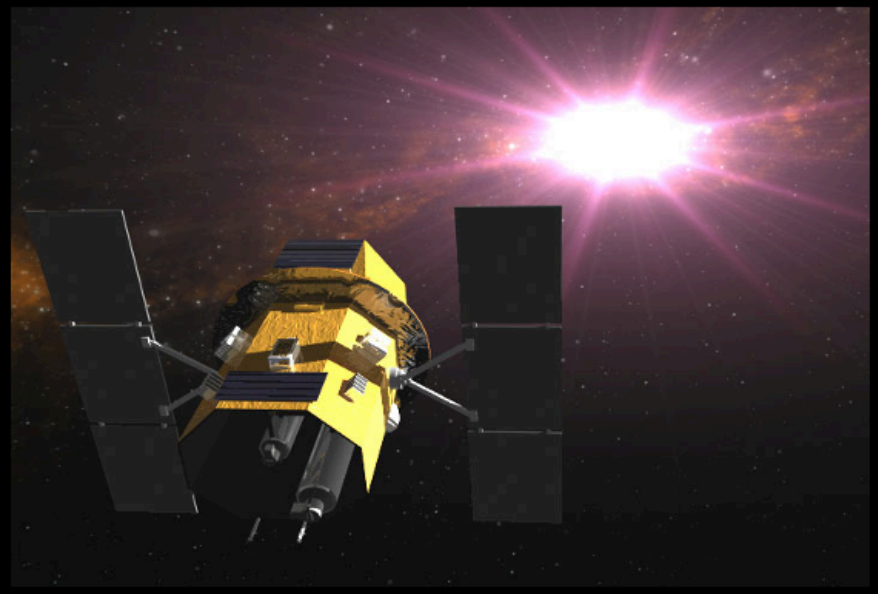
Open Data

Yesterday morning, a cosmic explosion was discovered and announced.



This morning everyone can download the data and combine them with own/other observations.

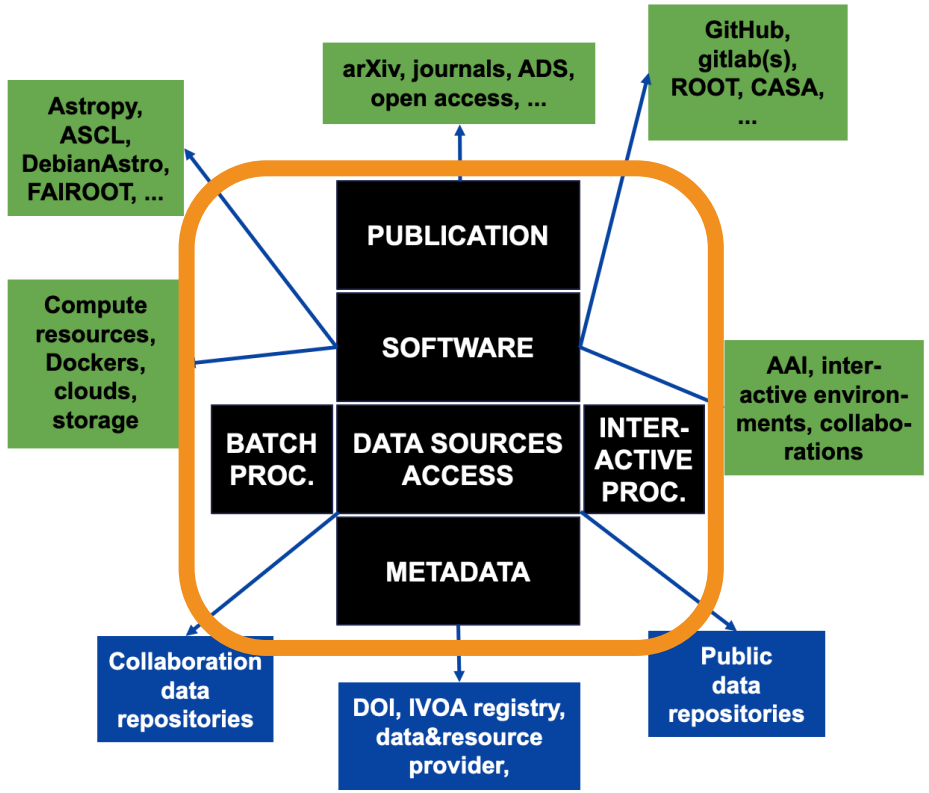
Follow-up satellite observations were proposed yesterday evening.

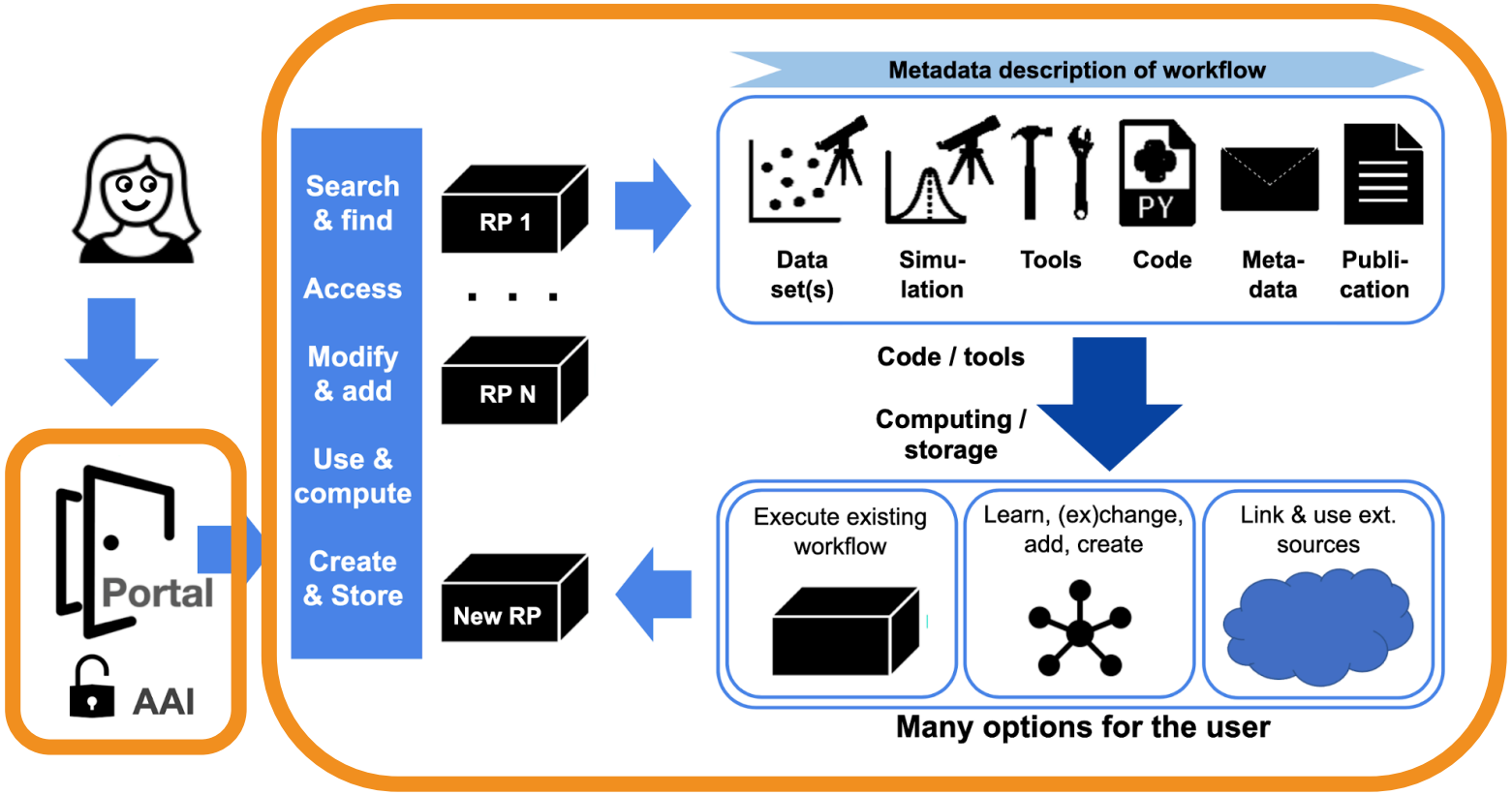


The measurements were made last night.

Platform, Portal, Research Product

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.

PUNCH4NFDI Work Programme and Open Data

Task areas, services, ...

TA 2 "Data management": M. Hoeft, *C. Wissing*
WP 2.1: Standardised access to data and metadata
WP 2.2: Compute4PUNCH
WP 2.3: Automatisation and optimisation of big data management workflows
TA 3 "Data transformations": *M. Brueggen*, *T. Kuhr*
WP 3.1: Statistical methods
WP 3.2: Numerical methods and simulations
WP 3.3: Machine learning methods
WP 3.4: Methods for analyses across datasets
TA 4 "Data portal": P. Bechtle, *H. Enke*
WP 4.1: Digital (dynamic) research products and their catalogue
WP 4.2: Mapping and collating metadata
WP 4.3: Implementation of interfaces
WP 4.4: Build and operate the science data portal

TA 5 "Data irreversibility": M. Kramer, A. Redelbach
WP 5.1: Implications for discovery potential and reproducibility
WP 5.2: Dynamic filtering
WP 5.3: Dynamic archiving
WP 5.4: Scaling workflows
WP 5.5: Evaluation and validation of instrument response ...
TA 6 "Synergies & services": K. Schwarz, S. Wagner
WP 6.1: Marketplace
WP 6.2: Authorisation and authentication infrastructure
WP 6.3: FAIRness
WP 6.4: Open-source data analysis tools
WP 6.5: Services in big data management
TA 7 "Training, education, outreach, citizen science": *F. Bertoldi*,
WP 7.1: Training of scientists - Young Academy
WP 7.2: Education of students
WP 7.3: Public outreach
WP 7.4: Support for citizen science

Broad spectrum

- of efforts towards individual FAIR / open data aspects (single deliverables and services) → many on short-term timescale
- of efforts towards realisation of overarching platform and research product infrastructure vision → longer term

Thank you!

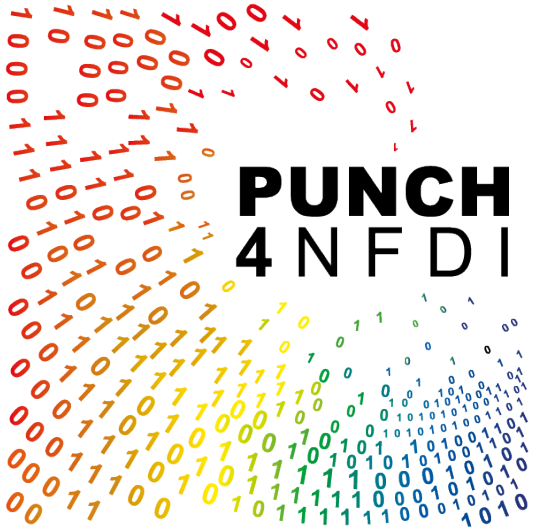
The PUNCH4NFDI Consortium

Spokesperson:

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

Contact:

Mail: punch4nfdi@desy.de
Web: www.punch4nfdi.de
Twitter: #punch4nfdi



Backup

Individual Efforts

Task areas, services, ... (examples)

TA 2, WP 1 "Standardised access to data and metadata":
e.g. infrastructure with standardised access to enable analysis of distributed datasets

- Prototype data lake setup (D-TA2-WP1-1)
- Inclusion of data (D-TA2-WP1-2,4)
- Metadata catalogue reference implementation for LQCD (D-TA2-WP1-3)

TA 3, WP 4 "Methods for analysis across datasets":

- Framework for conversion / reading of data for combined analyses … (D-TA3-WP4-1)
- Standardised interface for the publication of likelihoods (D-TA3-WP4-3)

TA 4, all WP ("Data portal"):

- WP 1: dynamic digital RPs and their catalogue: implementation of PUNCH DRP database (D-TA4-WP1-1); ingestion of selected RPs (D-TA4-WP1-2); enrichment of DRPs (D-TA4-WP1-3), prototype demonstrators for the SDP (D-TA4-WP1-4)
- WP 2: mapping and collating metadata: technical interfaces to external resources (D-TA4-WP2-1); integration of platform services and interfaces (D-TA4-WP2-2); interfaces allowing combined analysis of data from different sources (D-TA4-WP2-3)
- WP 3+4: implementation of interfaces, and build and operate the data portal

TA 6 +7, all WPs, in particular:

- TA 6, WP1: marketplace: information broker
- TA 6, WP 2: AAI
- TA 6, WP 3: reference guides, customising metadata frameworks for integration into the SDP, making Effelsberg data openly available, … demonstration of formats for metadata extensions
- TA 6, WP 5: services in big data management (e.g. access to open data archives D-TA6-WP5-1).