PUNCH4NFDI Consortium

Science Data Portal and Digital Research Product

Particles, Universe, NuClei and Hadrons for the NFDI
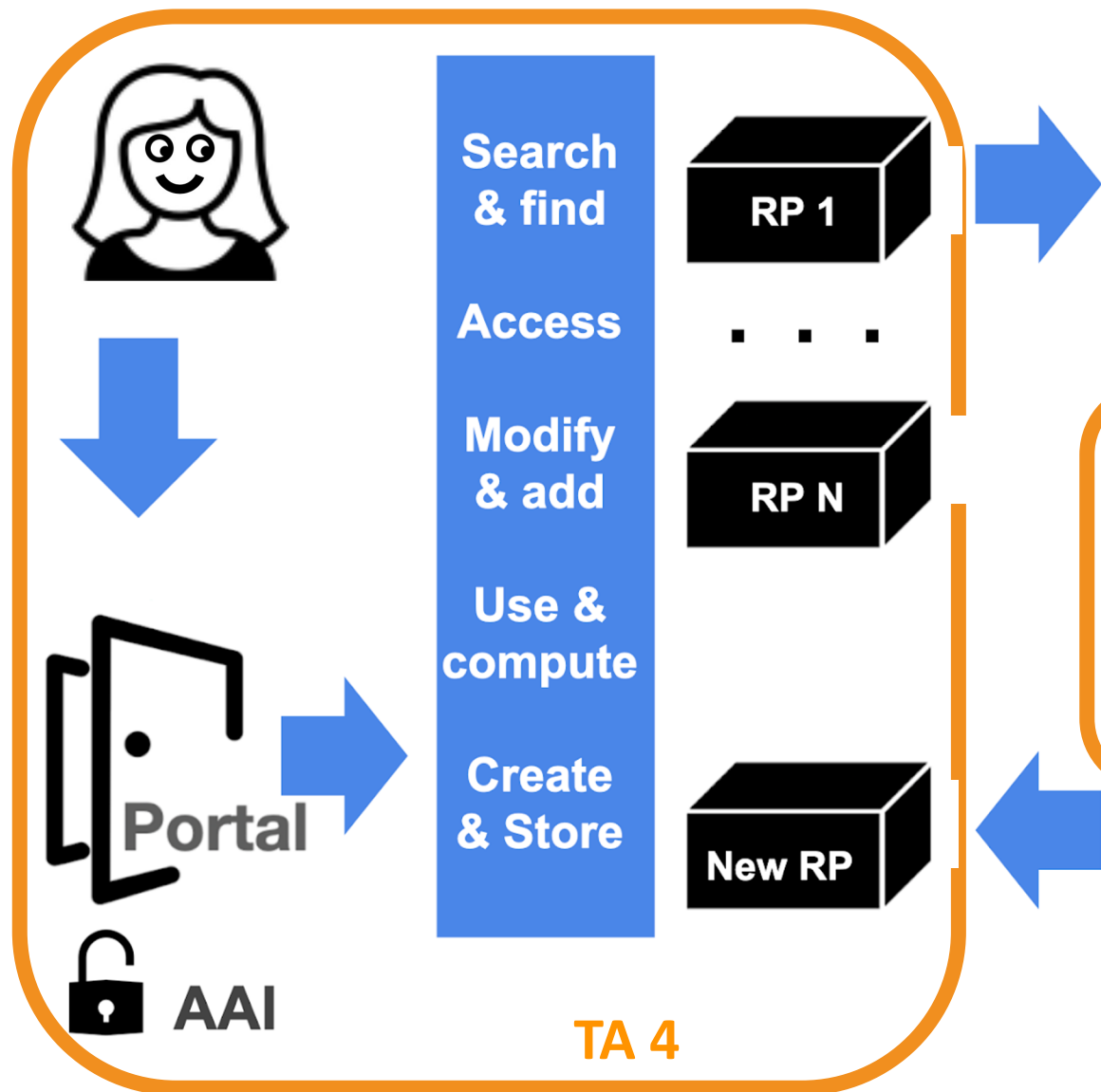
Harry Enke (AIP) TA4

15.09.2022

Science Data Platform and Digital Research Product

- the PUNCH4NFDI science data portal is a central element (hub) and main external access point for the PUNCH and broader communities to the PUNCH4NFDI science data platform services
- the creation of the Digital (dynamic) Research Product combines the available advanced developments in our communities

*   Digital (dynamic) research products and their catalogues

*   Mapping and collating metadata

*   Implementation of interfaces (APIs)

*   Build and operate the science data portal
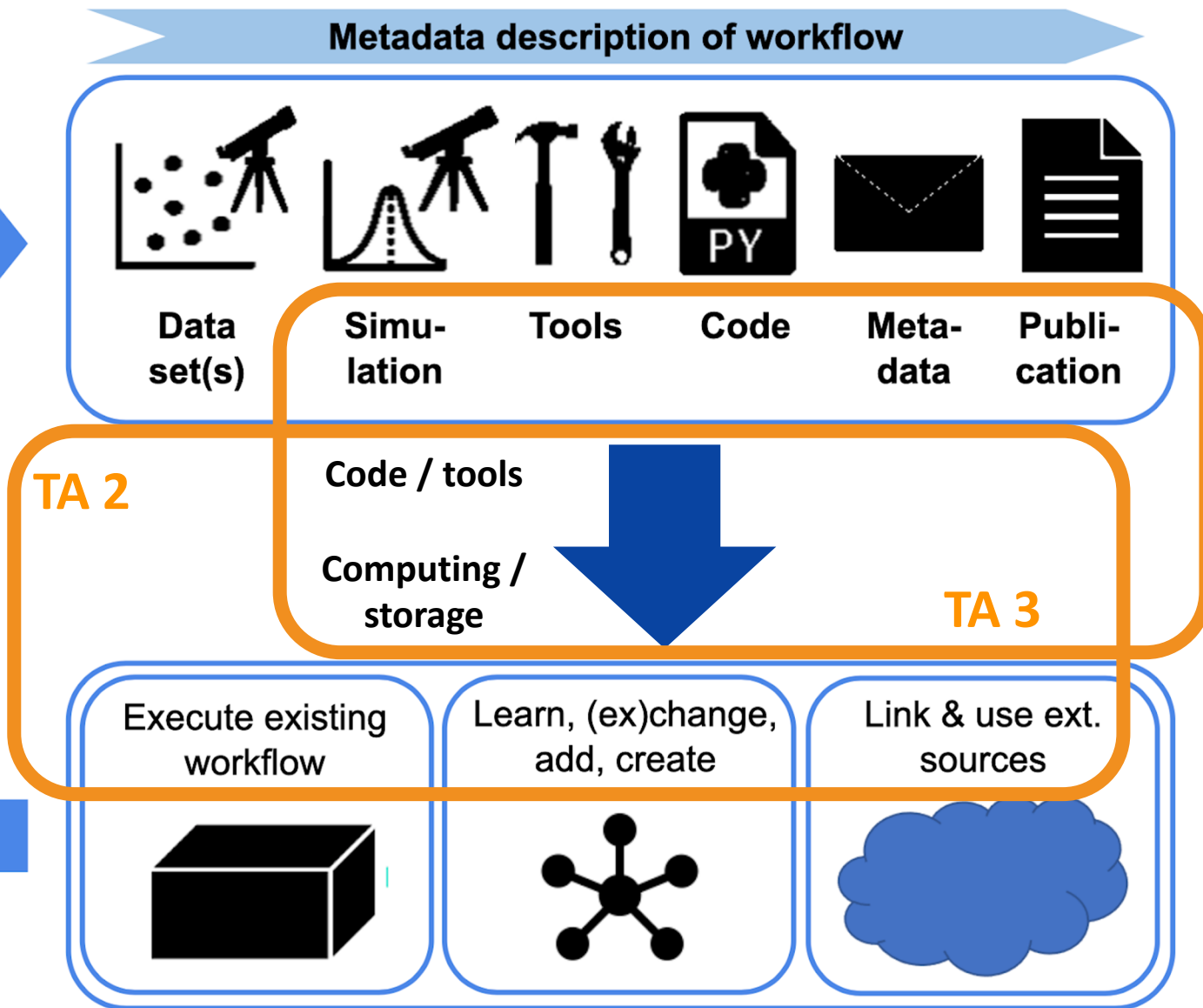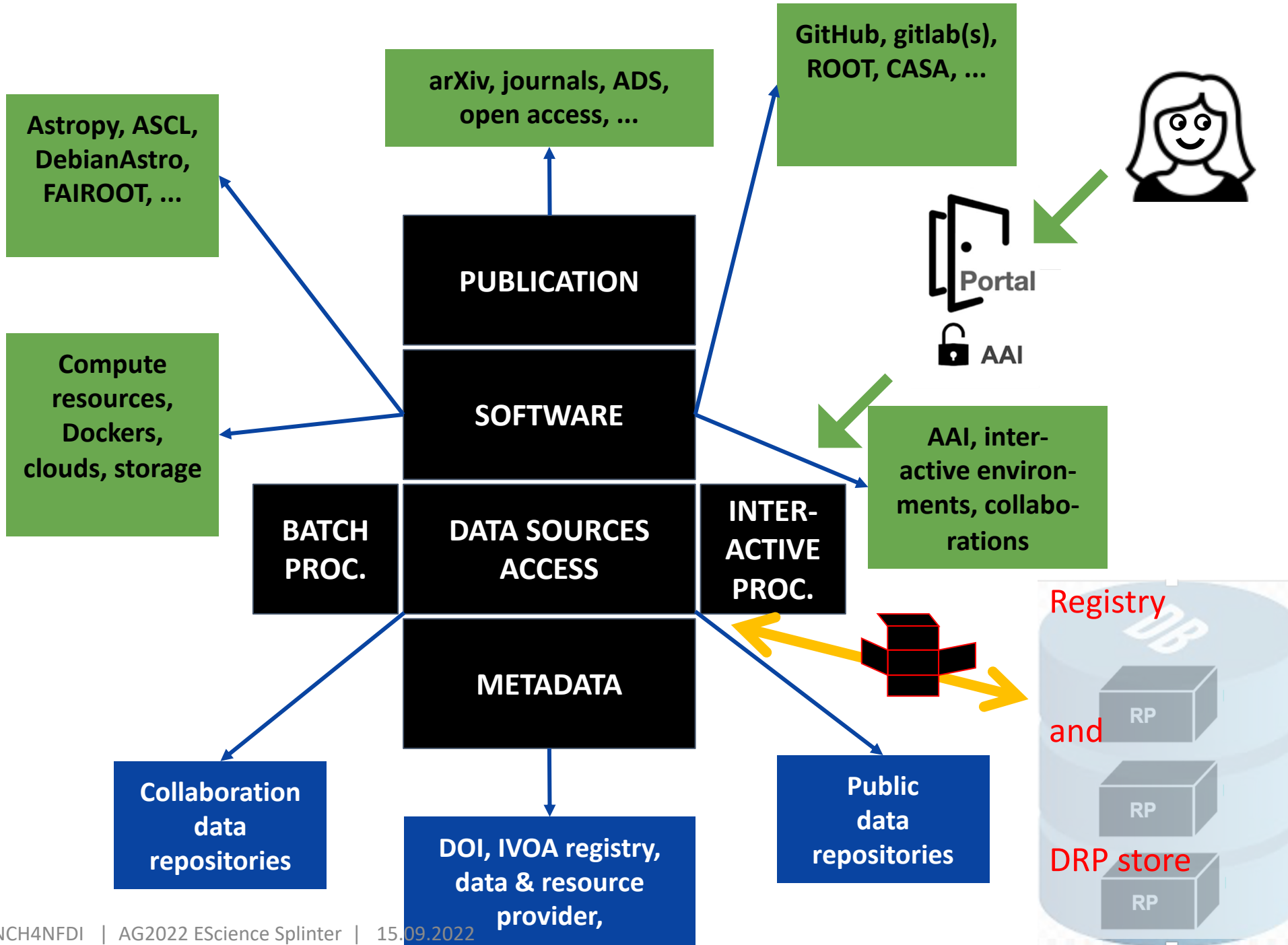
PUNCH-SDP

The science data platform for RPs

Search & find

Access

Modify & add

Use & compute

Create & Store

RP 1

RP N

New RP

Portal

AAI

TA 4

Research product contains executable workflow

Metadata description of workflow

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

TA 2

Code / tools

Computing / storage

TA 3

Execute existing workflow

Learn, (ex)change, add, create

Link & use ext. sources

Many options for the user

Digital Research Product

Astropy, ASCL, DebianAstro, FAIROOT, ...

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

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

Portal

AAI

PUBLICATION

SOFTWARE

Compute resources, Dockers, clouds, storage

BATCH PROC.

DATA SOURCES ACCESS

INTER-ACTIVE PROC.

METADATA

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

Registry and DRP store

RP
RP
RP

Collaboration data repositories

DOI, IVOA registry, data & resource provider,

Public data repositories

- Improves reproducibility and re-usability
- guided by FAIR
- Access via portal
- Interaction with other RPs
- Interfaces to tools and infrastructures
- Built on available developments of our communities
- Use standards already implemented
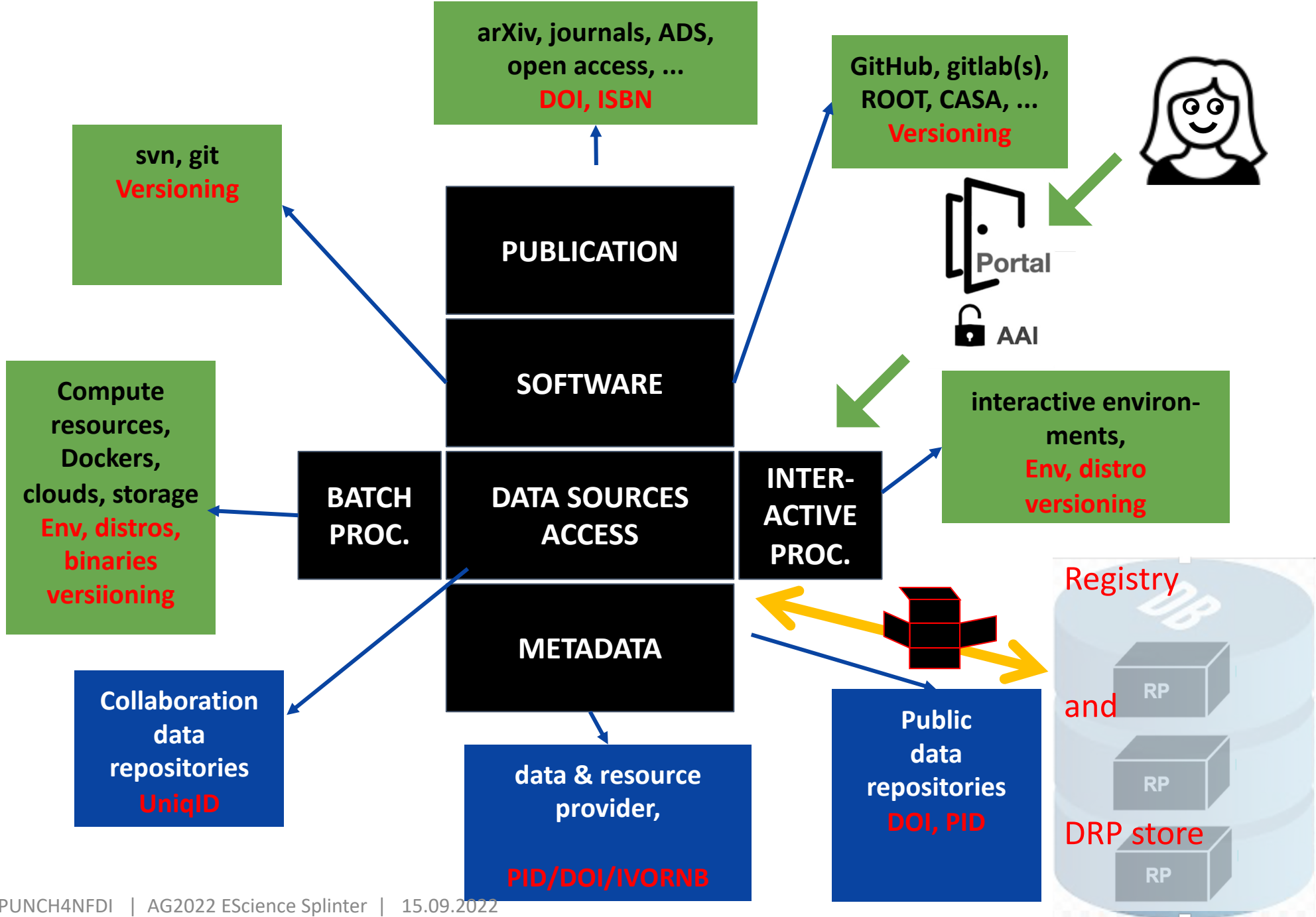- avoid extensive mappings, do the necessary

Digital Research Product

- main considerations for the DRP and SDP
  - all PUNCH communities have spent extensive work on
    - data storage
    - metadata systems
    - analysis software / workflows
    - data publication policies
  - the SDP needs to be inclusive and to offer added value
    - needs seamless integration of very heterogeneous
      - data collections
      - software stacks
      - resource providers
  - communities are international, NFDI Consortia are representing only a small fraction
  - heterogeneous resource providers
  - leading paradigm for our approach: building on microservices
  - wrap elements into a container which captures a working environment
    - include e.g. provenance information (limited to the processes captured)
  - provide suitable descriptive store and unpack procedures and environment
  - mint DOI or other PID to assign for such container
  - manage these in a (distributed) registry

DRP provide (pointers to) data , code and environment specs, publications

SDP provides the execution means
(AAI + required resources) , connects to
    community resources
    NFDI / EOSC etc.

Digital Research Product

svn, git
Versioning

arXiv, journals, ADS, open access, ...
DOI, ISBN

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

Portal

AAI

Compute resources, Dockers, clouds, storage
Env, distros, binaries versiioning

PUBLICATION

SOFTWARE

interactive environ-ments,
Env, distro versioning

BATCH PROC.

DATA SOURCES ACCESS

INTER-ACTIVE PROC.

METADATA

Registry and DRP store

Collaboration data repositories
UniqID

data & resource provider,
PID/DOI/IVORNB

Public data repositories
DOI, PID

RP
RP
RP

- Improves reproducibility and re-usability
- guided by FAIR
- Access via portal
- Interaction with other RPs
- Interfaces to tools and infrastructures
- Built on available developments of our communities
- Use standards already implemented
- avoid extensive mappings, do the necessary

Digital Research Product
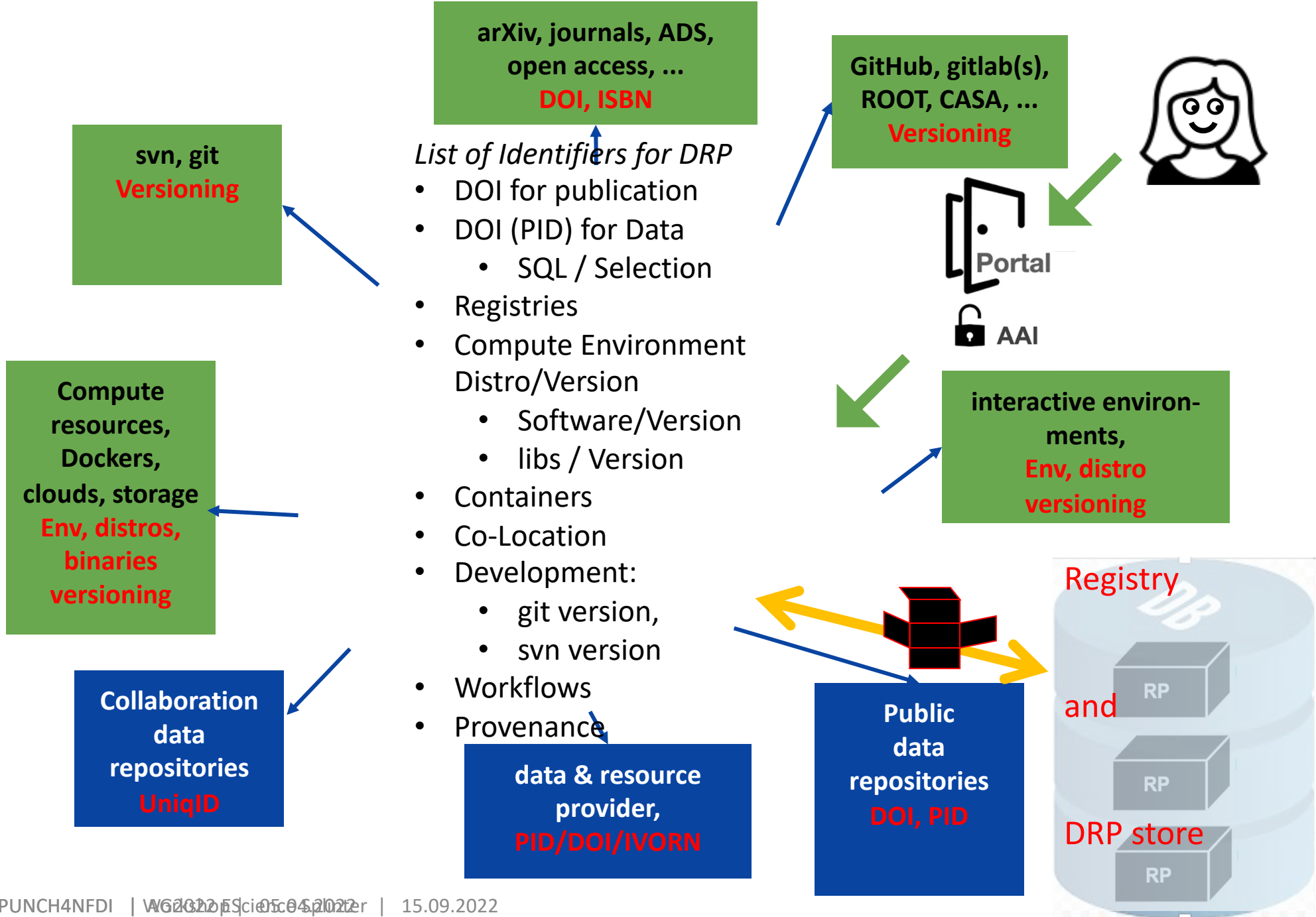
arXiv, journals, ADS, open access, ...
DOI, ISBN

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

svn, git
Versioning

List of Identifiers for DRP
- DOI for publication
- DOI (PID) for Data
  - SQL / Selection
- Registries
- Compute Environment Distro/Version
  - Software/Version
  - libs / Version
- Containers
- Co-Location
- Development:
  - git version,
  - svn version
- Workflows
- Provenance

Portal

AAI

interactive environ-ments,
Env, distro versioning

Compute resources, Dockers, clouds, storage
Env, distros, binaries versioning

Collaboration data repositories
UniqID

data & resource provider,
PID/DOI/IVORN

Public data repositories
DOI, PID

Registry and DRP store

- Improves reproducibility and re-usability
- guided by FAIR
- Access via portal
- Interaction with other RPs
- Interfaces to tools and infrastructures
- Built on available developments of our communities
- Use standards already implemented
- avoid extensive mappings, do the necessary

Starting to collect elements of SDP

- PUNCH AAI
- Compute4Punch (C4P)
- Storage4Punch (S4P)
- S3 storage
- Docker and Kubernetes infrastructures
- Gitlab + Continuous Integration
  - Code Repository
  - Container Registry (in operation)
  - Package Registry (tbc)
  - integration with
- REANA (with support of Jupyter notebooks)
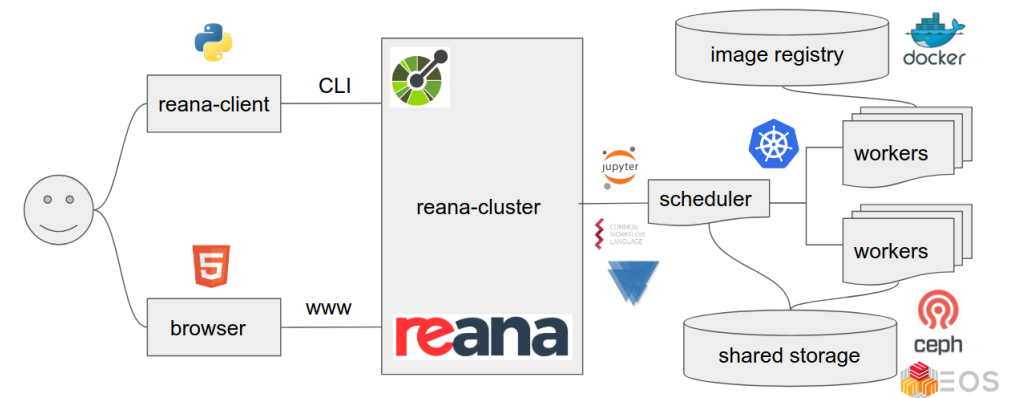- Dashboard + Intranet (based on gitlab)

Reproducible research data analysis platform

Starting to collect elements of the DRP

- Collecting Metadata schemas

  IVOA (Astronomy)

  ILDG (Lattice community)

  CERN Open Data

  Astro Particles

- Currently: 'Findability' aspect of FAIR
  - evaluating usage of DOI and the Metadata Kernel of DataCite
  - looking at registry concepts and implementations