FAIR-IMPACT
Expanding FAIR Solutions across EOSC

T3.2 Integration of PID practices into FAIR data management

T3.2.1 PIDs in workflows

UNIMAN: Carole Goble,
Stian Soiland-Reyes, Nick Juty

2022-10-06

Cite as: https://doi.org/10.5281/zenodo.7157647

Funded by
the European Union

FAIR Workflow registry

WorkflowHub

Workflow-system **agnostic**

Search for and **discover** workflows

Metadata standardization
(CWL, schema.org, custom tags, RO-Crate)

DOI publication, citation & credit

Collections

Teams, Organizations and **Communities**

Programmatic access: GA4GH TRS API, RO-Crate

Registry, not repository

Workflows can live elsewhere, e.g. GitHub

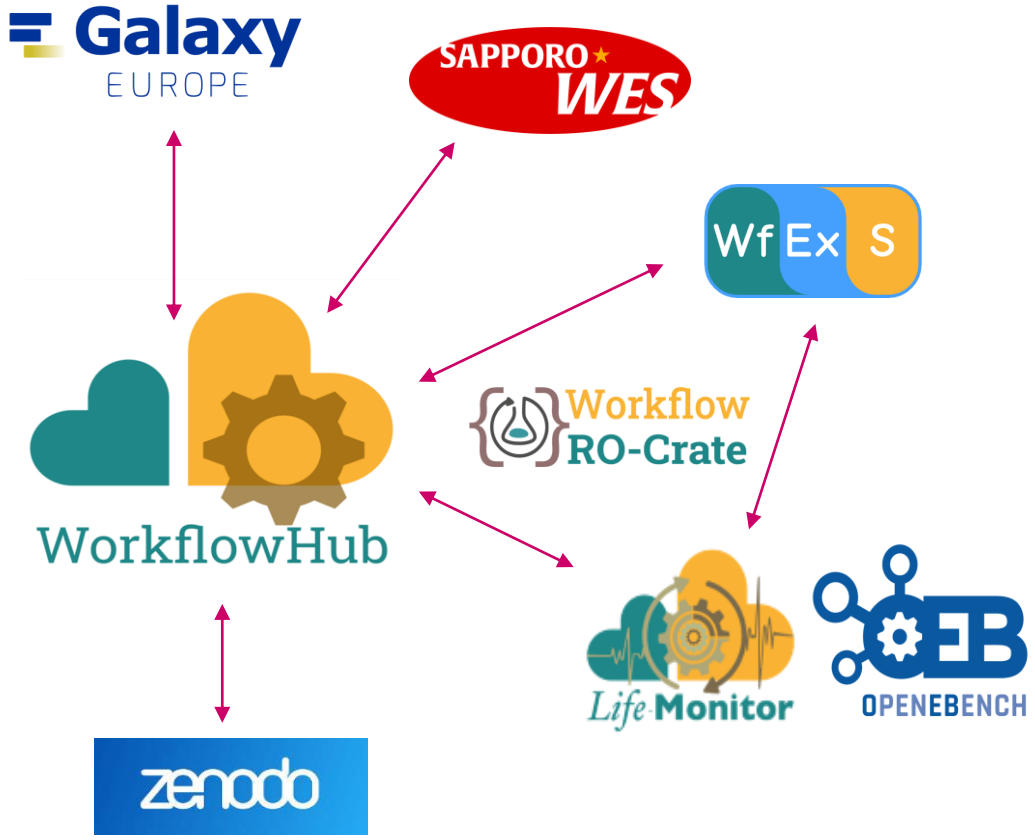
Integration with **execution** platforms (incl. usegalaxy.eu)

Challenges: PID before workflow registration is ready to be promoted to DataCite?

Encourage workflow citations.

https://workflowhub.eu/

EOSC-Life ecosystem



The services in the *Workflow Collaboratory* exchange digital objects as Workflow RO-Crates

Packaging workflow files & companion objects

Submission / download

Exchange between services & systems

Reproducibility & Testing

Citation

Challenge: PID propagation and mutability

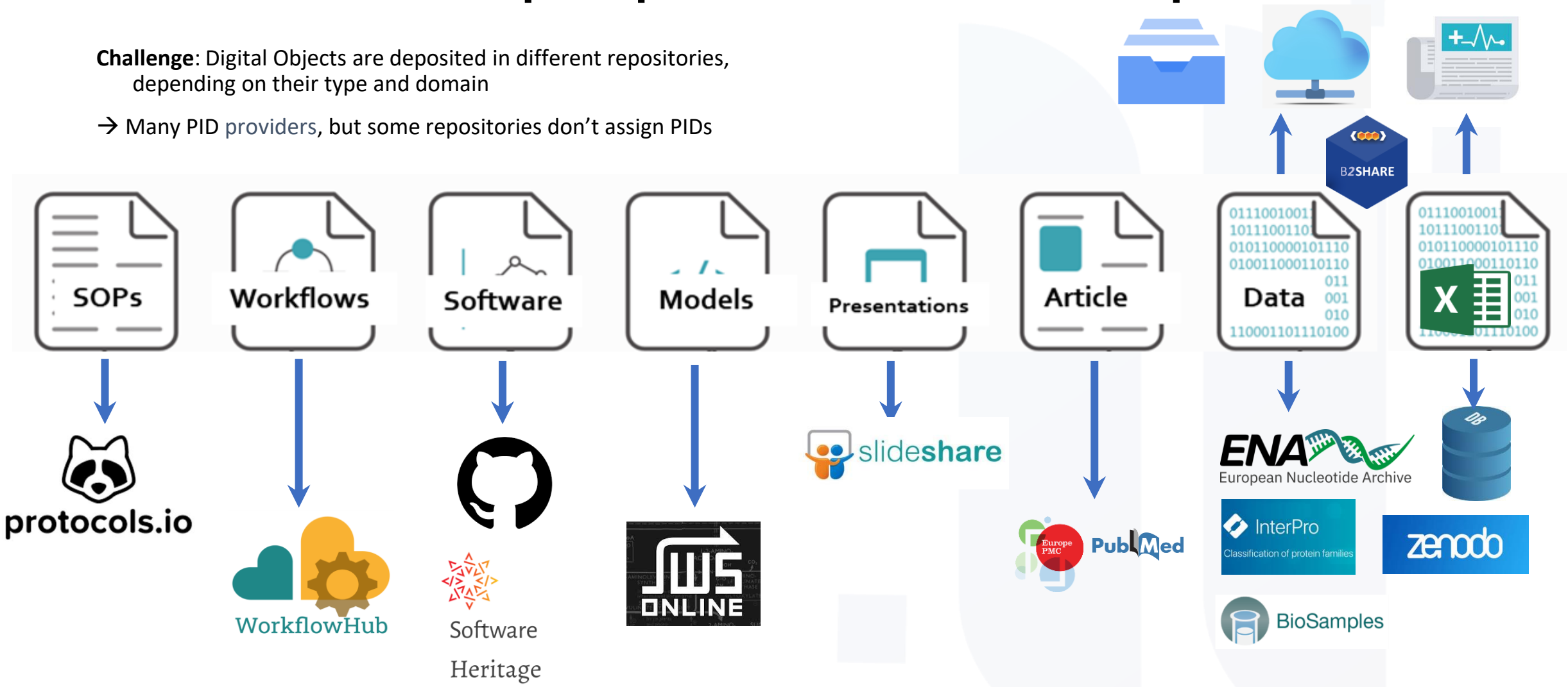
https://doi.org/10.5281/zenodo.4605654

FAIR-IMPACT

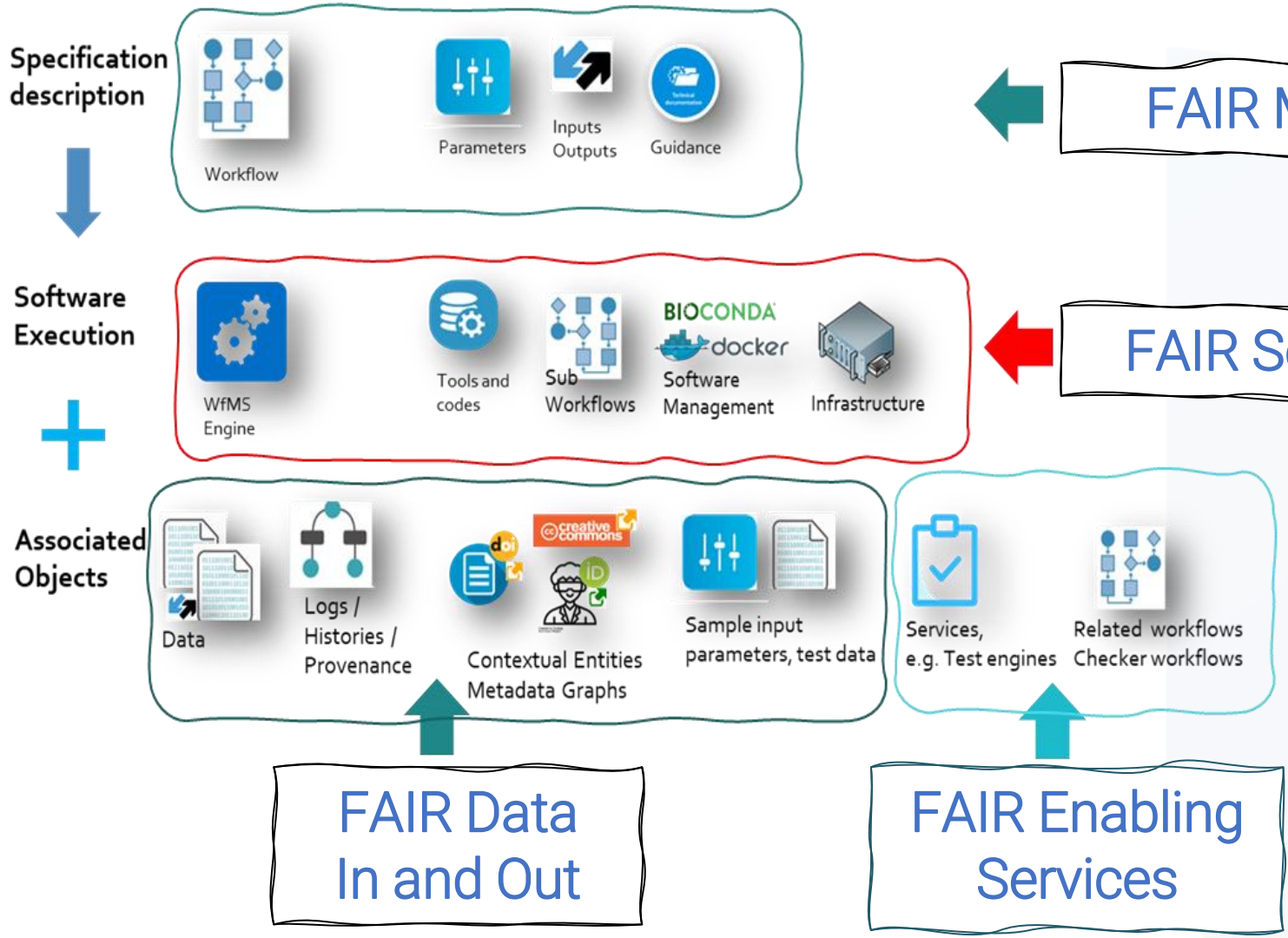
Multiple platforms and repositories

Challenge: Digital Objects are deposited in different repositories, depending on their type and domain

→ Many PID providers, but some repositories don't assign PIDs



Workflows are composite objects



FAIR Method Object

FAIR Software Objects

FAIR Data In and Out

FAIR Enabling Services

Challenges: How to assign identifiers to parts of workflows. Do they need to be PIDs or can be relative?

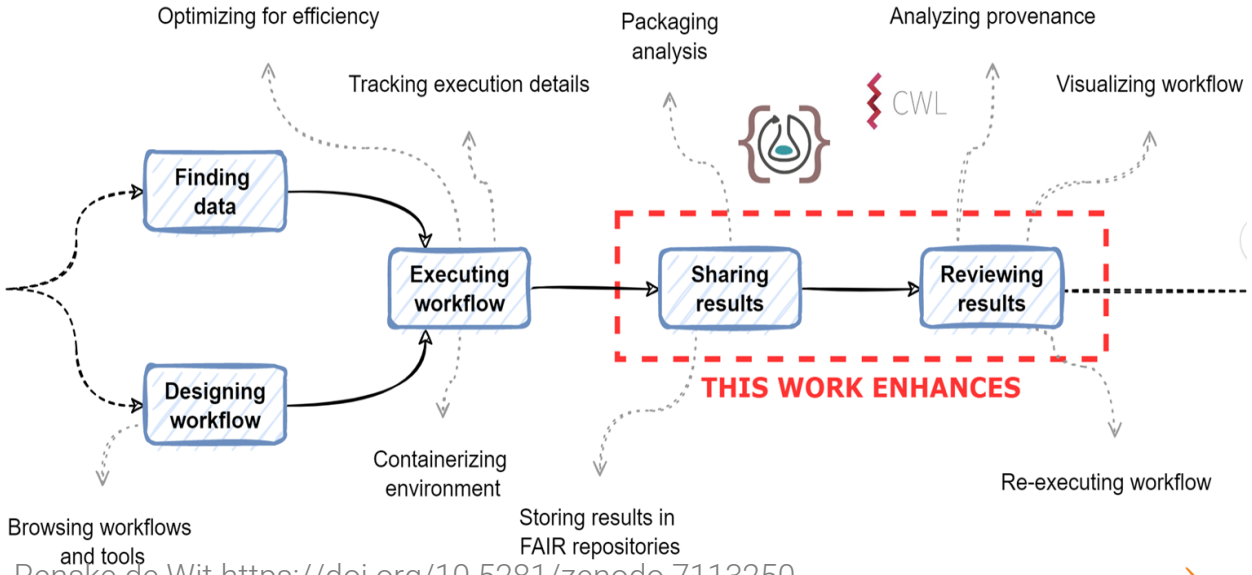
How to assign PIDs to existing objects (e.g. software used)

Multiple PIDs for same object

Using standards for workflow execution provenance

Challenges: Assigning PIDs (or not) to data in workflow runs.

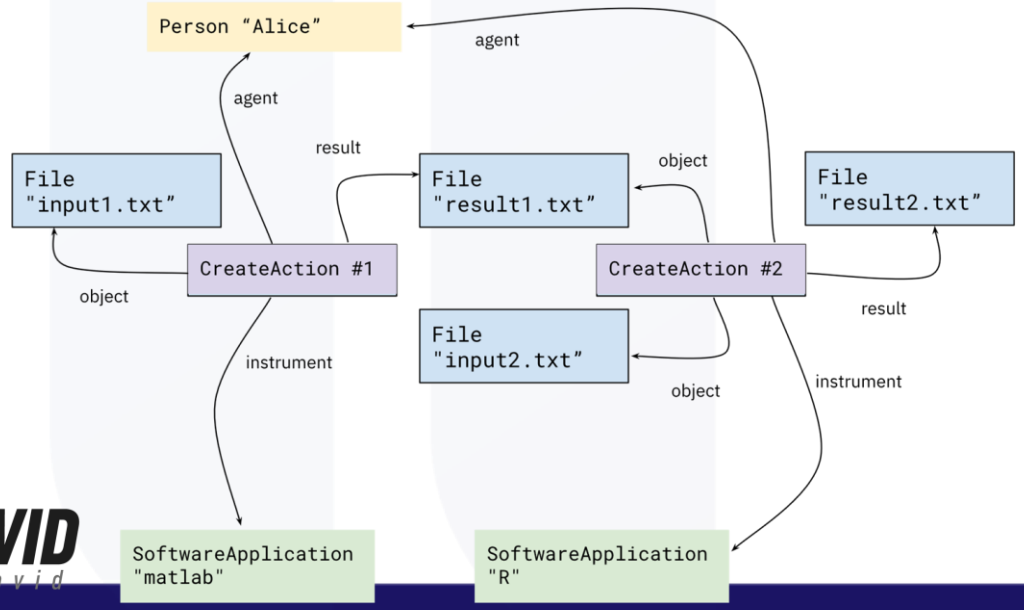
Workflow Engine that writes provenance does not know where data will go – can it mint PIDs?



Renske de Wit https://doi.org/10.5281/zenodo.7113250

Workflow Run RO-Crate

Simone Leo, Laura Rodríguez-Navas, et al.
https://www.researchobject.org/workflow-run-crate/



Galaxy

EOSC-Life

elixir

BY-COVID beyond covid

Resolving WorkflowHub PIDs to RO-Crate FDOs

FAIR Signposting – HTTP headers for PID, metadata, ++

Common **types** from schema.org

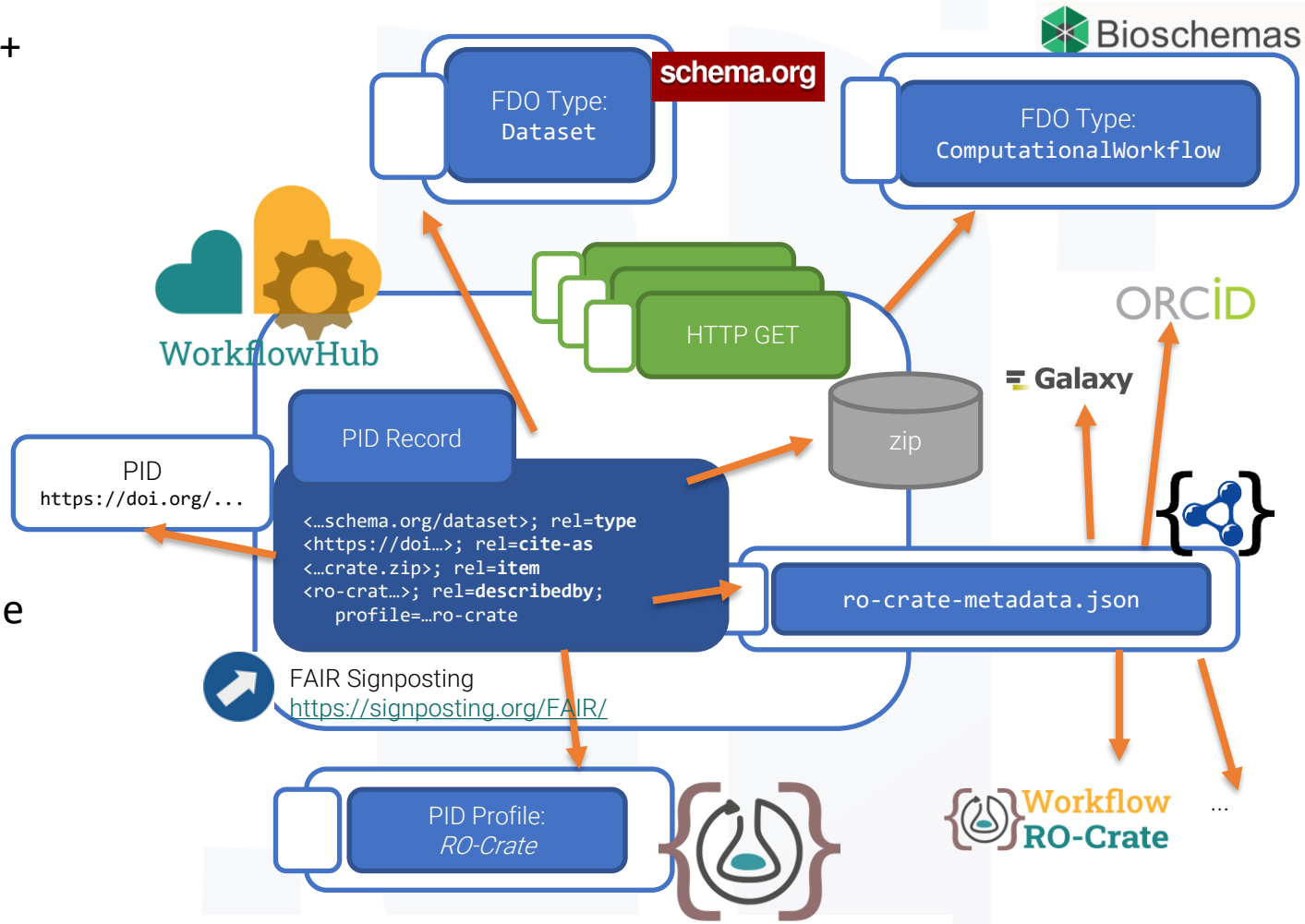
Multiple **metadata files** incl. JSON-LD, DataCite XML

Domain-specific **profiles** (e.g. Workflow Crate)

Conforming to existing Community **APIs** (e.g. GA4GH)

Challenges: Exposing individual parts of RO-Crate while also in ZIP file, git repo etc? Multiple PIDs.

Formalizing this convention, more implementations



Updating Linked Data practices for FAIR Digital Object principles
& Creating lightweight FAIR Digital Objects with RO-Crate.
Research Ideas and Outcomes 8:e93937 , 1st Intl Conf on FAIR Digital Objects

T3.2.1 contributions

- Formalize/document FAIR RO-Crate PID conventions for workflows
- Capture experiences in BY-COVID on workflow provenance PIDs
- Augment RO-Crate: Retrospectively assign PIDs
- Explore decentralised PIDs:
  - arcp generate global identifiers for parts of a package
  - Content-based addressing including Naming Things With Hashes (RFC6920)
  - RO-Crate as glue between local and global identifiers

FAIR-IMPACT integrations

- WP3
  - Contribute practical PID experiences from FAIR workflows (T3.3, T3.4)
  - Lobby for more workflow details into DataCite graph (e.g. workflow language, software dependencies)
- WP4
  - “Stress-test” FAIR Signposting convention/tooling (T4.3, T4.4)
  - Ensure PID practices are part of RO-Crate tooling (T4.2)
- WP5
  - Contribute FAIR Signposting benchmarks (T5.3)
  - WorkflowHub as exemplar for metrics testing (T5.1, T5.3)
- WP6
  - PID + FAIR Digital Object + RO-Crate (T6.1)
  - EOSC integration of FDOs (w/ EuroScienceGateway) (T6.3)

FAIR-IMPACT

Expanding FAIR Solutions across EOSC

@twitter

Linkedin

Funded by
the European Union