# Research Analysis Identifier - RAId **EOSC-RAISE Project**

### Gorka Epelde Unanue

### Madrid, Spain

### #EOSCSymposium2023

တeosc







Funded by the European Union



#### **Technical Representatives**

Evdokimos Konstantinidis & Despoina Petsani Aristotle University of Thessaloniki

Gorka Epelde

Vicomtech



Project name: Research Analysis Identifier System

Start date: 1 October 2022

End date: 31 January 2026

meos

**Coordinator:** Aristotelio Panepistimio Thessalonikis

RAISE project aims to provide the mechanisms for a distributed crowdsourced data processing system, moving from open data to data open for processing. To do so, RAISE will attempt to adapt open data to the culture of the research community, ensuring FAIR principles. The vision of the project is the EOSC Web of FAIR Data and Services for Science is an open, fair and reliable Research Community where every researcher will be accredited for their work and all research data will be equally accessible for processing without violating data protection regulations.



Source: European Commission, Directorate-General for Research and Innovation, European Research Data Landscape: final report, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/3648

0% 20% 30% 40% 10%50% 60% 70% 80%

Data protection 41% 39% 9%

33%

36%

Obstacles to the management and sharing of research data

37%

35%

37%

44%

39%

31%

27%

15%

16%

16%

13%

10%

#### **RAISE** works on moving from open data to data open for processing







Research data depositing



Lack of data repositories

Misinterpretation / falsification risk

meos

Commercial use

Lack of recognition

Time and effort

Skills needed

**Financial costs** 

Legal restrictions

3

50%

70%

60%





### EOSC SRIA

meos

### • [GAPS]

- PIDs and Sensitive data
- New PID technologies
- [Priority] Develop standardised identifiers for resource types that have not as yet become standard practice.
  - Identified a gap in reproducible SW experiments persistent identifies supporting protected / controlled datasets





## Proposed PID concept



- <u>Research Analysis</u> Identifier (RAI) Id, to guarantee the reproducibility of a <u>software analysis</u>, reflecting authorship, source dataset (open dataset or protected dataset), algorithm and results.
- Foreseen benefits:
  - Allow a researcher, journal or patent reviewer, or company that wants to purchase an IPR, to reproduce the experiments (including protected datasets), and have the certainty/guarantee of the experiment (reflecting authorship, timestamp, source dataset, algorithm) that has given rise to the results that are intended to be judged.
  - Track shared (protected) dataset usage and acknowledge providers
  - Track shared (protected) script usage and acknowledge providers





### Implementation

- (Protected) datasets are uploaded, given an internal PID and hosted in trusted lab / RAI nodes
- (Protected) scripts are uploaded, given an internal PID
- Upon a successful experiment run (script/code against a given dataset), the registration request is composed
- A defined hashing strategy in steps is applied (first different assets and metadata later)
- Final hash is used as a PID and is written to Blockchain (Implemented BigChainDB & Hyperledger Fabric)
- Services to resolve RAI in the RAISE portal, identify composing assets internal PID, reproduce experiment
  being developed

meoso



## RAISE service ecosystem

meosc





++++++

+ -

## Potential collaboration at PID level

•

- EOSC HE technical
  - FAIRCORE4EOSC
    - Research Activity Identifier (RAiD) PID for research projects and related activities
      - To enable the integration performed research
    - PID Meta Resolver
      - To make it easier and more accessible to users resolving RAI Ids
  - RDA TIGER
    - Engagement for RAI Id, within <u>RDA PID Interest Group</u>
      - <u>To contrast RAI Id novelty and disseminate</u>
- EOSC Association Task Force PID Policy and Implementation
  - For the alignment of RAI Id developments with Task Force activities
  - Contrast EOSC PID policy guidelines implementation within RAI PID
- Analyse established PIDs reuse within RAI Id
  - Orcid for researcher PID, Datacite for dataset PID, DOI for object (object = RAI Id)
    - <u>To make it as compatible interoperable as possible</u>



## RAISE expected outcome of PID session

•

- Feedback on how to improve RAI Id concept
- Feedback from the community on the relevance of the proposed PID concept
- Feedback from the community on the novelty of the proposed PID concept
- Start a discussion on
  - How to connect with established PIDs
  - Possible collaboration with other PIDs



meos









# Some more discussions about PIDs?

• • • •

+ + + +





### Stay tuned symposium23.eoscfuture.eu

**Session Chair contacts:** 

- Themis Zamani GRNET (themis@grnet.gr)
- Tibor Kalman GWDG (tibor.kalman@gwdg.de)

. . . . .

+ + + + + +