IC_2 Provenance implementation

1. Background

1.1 Short description

In close cooperation with EUDAT, a cross RI provenance model shall be developed (B2PROF) and usage of this model by workflow engines (Taverna, dispel, GEF) and services like B2HANDLE (service for persistent identifier), B2NOTE (service for data and metadata annotation), B2SHARE (data repository service together with a service for persistent identifiers and a service for metadata) and others shall be enabled.

1.2 Contact

Background	Contact Person	Organization	Contact email	
_ICT	e-Infrastructure] (Use Case proposer, Agile Group Leader) _	*Barbara Magagna*	*Umweltbundesamt GmbH*	*Barbara.magagna @umweltbundesamt. at*
_ICT	e-Infrastructure]_	Marco Rorro	EUDAT	m.rorro@cineca.it
_ICT	e-Infrastructure]_	Giovanni Morelli	EUDAT	g.morelli@cineca.it

1.3 Use case type

Implementation case

1.4 Scientific domain and communities

Scientific domain

[Atmosphere | Biosphere | hydrosphere | geosphere]
Cross all domains.

Community

[Data Acquisition | Data Curation | Data Publication | Data Service Provision | Data Usage | Or others]
All of them

Behavior

Data product generation/ Data Replication/Data Publication/Semantic Harmonisation/Data Discovery and Access/Data Citation

Roles

Data curator/Data publication repository/Service Provider

2. Detailed description

Objective and Impact

Enables the tracking of the complete data life cycle and the access of the data user to this information
This helps to identify datasets which are derived from the same source but displayed or published or processed in different ways or on different data services.
This also helps to acknowledge any step in the life cycle and the corresponding actor in this step, it supports the citation of them und thus the willingness to publish the data.

Challenges

1. Find a common provenance schema (e.g. PROV)
2. Harmonize existing schemata
3. Map existing schemata to the common schema

Detailed scenarios

tbd

Technical status and requirements

In development based on existing provenance tracking in different projects (e.g. EUDAT, myExperiment, BioVel, …)

Implementation plan and timetable

1. Timeline: tbd

2. Milestones:

-
    - (1) Define minimum information that has to be tracked
    - (2) Find a conceptual model for provenance which conforms to the needed information
    - (3) Map existing models to the common model
    - (4) Find repository to store the provenance information
    - (5) Establish a definition and implementation plan for services which should provide provenance information
    - (6) Establish a definition and implementation plan for publication services in order that provenance information is delivered together with data and metadata

3. Involved RIs: LTER, ICOS, …
4. Links to ENVRI Plus workpackages/tasks: 5.3, 5.4, 6.1, 8.1, 8.2, 8.3
5. Allocation of resources: tbd

Expected output and evaluation of output
(1) Provenance Information delivered together with data and metadata.
(2) Showcase for selected datasets coming from selected RIs (e.g. LTER)

External Links

1. IC_2 notebook: {+}https://envriplus.manageprojects.com/projects/wp9-service-validation-and-deployment-1/notebooks/636/+