

PIDs

Preliminary information gathered by workshop participants

Tuesday 22nd November, 11.00-12.30 CET

Theme/Workshop Chair: Jessica Parland-von Essen, CSC

Rapporteur: Liisa Marjamaa-Mankinen, CSC

Event link: <https://fair-impact.eu/events/synchronisation-force-events/synchronisation-force-1st-workshop-november-2022>

Context: The EOSC Persistent Identifier (PID) policy has been written for senior decision makers within potential EOSC service and infrastructure providers and is of interest to all EOSC stakeholders. It defines a set of expectations about what persistent identifiers will be used to support a functioning environment of FAIR research. In this session we will discuss the key concepts of the EOSC PID policy and how PID policies and implementations currently look in different contexts. The session will be an interactive workshop to facilitate discussion and collect data by using concept board and mentimeter.

Expected outcomes. The cross-workshop expected outcome is to assess the status of adoption of FAIR practices with respect to the theme. Based on the data we will present an evaluation of the maturity of PID implementations. This will not be done in a comparative way, but to find the areas that need most focus in developing PID policy work and implementation.

Recommendations assessed during the session:

- **Recommendation 1:** Implement the EOSC PID policy and architecture, including the development of a global PID resolver

Relevant EOSC-A Task Forces

- [EOSC PID Policy and Implementation.](#)

Useful references

- [EOSC Multi-Annual Roadmap \(2023-2024\).](#)
- [Turning FAIR into Reality report Rec. n 2 and 3](#)

Preliminary questions

1. What does your project or initiative do to implement PIDs? Please provide any relevant links. 2
2. Do you use persistent identifiers in your project / community? 4
3. Which are the roles and responsibilities in your organisation regarding PID management? 6
4. Do you have a PID policy in your organisation? 7
5. People who contributed to the preliminary information gathered 8

1. What does your project or initiative do to implement PIDs? Please provide any relevant links.

project or initiative	Input
Life in Kyrgyzstan Study	We post the panel study data for public use and we provide the DOI for the data source. We also provide IDs for households and individuals within the data which are anonymised.
CESSDA	CESSDA has a PID policy. My organisation is a CESSDA Service Provideres and implements PIDs.
BlueCloud	https://marketplace.eosc-portal.eu/services/seadatanet-doi-minting-service?q=SeaDataNet+DOI+minting+service
Open Biological and Biomedical Ontologies Foundry	Require GUPRIs for datasets and all entities used in datasets (ontology terms) PIDs are always URIs, but we work a lot with compact URIs as well
EOSC-A Long Term Data Preservation task force	We will liaise with the TF PID Policy and Implementation to include preservation specific requirements concerning PIDs. Important here are mature PID infrastructures, services supporting versioning and tracking throughout the data life-cycle, machine actionable PIDs.
EuroScienceGateway	We are planning to submit data from our analysis workbench to invenio and Zenodo. For workflows we submit them to workflowhub.eu - both of them provide DOIs.
TRIPLE	As an aggregator (of publications and datasets) we mostly reuse PIDs, not create new ones. We support our providers so that they use PIDs.
The project of French national catalogue of individual health data collections (FReSH, for France Recherche en Santé Humaine) is currently in preparation phase	<p>The metadata provided describe several entities that can be identified by PIDs and thus enable to connect records across different identifier systems. Some identifiers will be submitted by dataset producers upon the direct registration, others will be harvested from external repositories. Whenever possible, we will implement the usage of authority files and PIDs, using one or more identifier systems.</p> <ul style="list-style-type: none"> - DOI for datasets: Since the catalogue will not host the data, the DOI will be filled in only for the data retrieved from other repositories. - DOI for associated papers will be retrieved from available sources. - ORCID will be used if available. If there is no ORCID record, IdRef (ID developed and maintained by the French Bibliographic Agency of the Higher Education) can be considered as an alternative identifier: https://www.idref.fr/



project or initiative	Input
	<ul style="list-style-type: none"> - Health Studies (clinical or epidemiological): no single mature PID for the moment; the same study can be registered on different platforms, each issuing authority assigning its own identifier. - Organisations: since the catalogue will cover the national scope, the focus is the fine-grained identification of the national players. For this purpose, we consider using RNSR (Répertoire national des structures de recherché, https://appliweb.dgri.education.fr/rnsr/ChoixCriteres.jsp?PUBLIC=OK) which assigns the identifiers to the research labs. For non-academic players or international partners we consider using ROR.
OpenAIRE	OpenAIRE Guidelines are based on Dublin Core, Datacite and CERIF schemas. PIDs are an important and essential rule for the aggregated records on the compliance of the data sources with the OpenAIRE Guidelines.
LifeWatch ERIC	We are a member of DataCite and provide DOIs <ol style="list-style-type: none"> 1. for metadata records in our metadata catalogue (https://metadatalogue.lifewatch.eu) 2. for semantic artefacts published on our portal (EcoPortal - http://ecoportal.lifewatch.eu/).
ENVRI-FAIR	SeaDataNet uses the SEANOE (SEA scieNtific Open data Edition) service to facilitate scientists to publish their research data in the field of marine sciences as citable resources. https://www.seanoe.org/
FAIRtracks (also part of EuroScienceGateway)	We are interested in setting up a new PID generator for track data files, which are currently lacking
Bioregistry	The Bioregistry is a catalog of standard prefixes for biomedical, life sciences, and clinical vocabularies/ontologies. It is a meta-resolved for CURIEs like GO:0032571 for the Gene Ontology (GO) entry on "response to vitamin K" with https://bioregistry.io/go:0032571 . It is deployed at https://bioregistry.io/ and is fully open source / open data at https://github.com/biopragmatics/bioregistry . The underlying service can also be re-deployed with a registry of prefixes for other domains.
FAIRsharing	We use DataCite to mint DOIs for our resource records. We cross-reference against ORCIDs and ROR. We will use PIDs in our metadata where provided
	https://elixir-europe.org/services/interoperability/identifiers https://elixir-europe.org/events/elixir-webinar-referencing-data-using-identifiers Bio.tools: https://elixir-europe.org/internal-



project or initiative	Input
	<p>projects/commissioned-services/registry-tools identifiers.org / N2T.net: https://elixir-europe.org/news/identifiersorg-n2tnet Recommended interoperability resources: https://elixir-europe.org/platforms/interoperability/rirs FAIRsharing provides descriptions of 29 different types of identifier schemas, and relationships to those resources that utilise them. https://bioregistry.io (https://github.com/biopragmatics/bioregistry) - a fully open source, sustainable alternative to Identifiers.org / N2T that harmonizes with other resolvers/PID systems (e.g., FAIRsharing, BARTOC, OBO Foundry)</p>
	<p>CLARIN B Centres host data repositories and as such are required to properly manage the deposited data and metadata. This includes assigning and managing PIDs (Handles and DOIs are used) Centrally we offer services, such as the virtual collection registry, that offers datasets which are assigned PIDs as well Centrally we offer services that consume PIDs to retrieve metadata or bit streams to digital objects (Switchboard)</p>
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	<p>LifeWatch Italy provides access to different resources (e.g. dataset, services, VREs). All the dataset hosted on our national data portal are assigned a PID (handleID). In addition, we provide DOIs for certain dataset that are well described, with rich metadata, and are licensed with CC-BY 4.0 or more open licenses, by exploiting the agreement LifeWatch ERIC-DataCite The DOI/internal PID are used to link the metadata records present on the head quarter metadata catalogue (https://metadatalogue.lifewatch.eu). An example of a dataset that is provided with a DOI and an PID is here https://dataportal.lifewatchitaly.eu/view/urn%3Auuid%3A092761f7-1910-4d15-9ba7-db3b3e35c493</p>

2. Do you use persistent identifiers in your project / community?

Project or initiative	Input
Life in Kyrgyzstan Study	Yes. As the data we collect is of longitudinal nature, we provide persistent panel identifiers for the respondents.

CESSDA	Yes. CESSDA has a PID Policy: https://doi.org/10.5281/zenodo.6607000 and my organisation uses PIDs (urn and orcid).
BlueCloud	DOIs Handles for SeaDataCloud
Open Biological and Biomedical Ontologies Foundry	100%
EuroScienceGateway	yes
TRIPLE	In our project we reuse existing ones. In our community (social sciences and humanities), not 100%, but growing.
The project of French national catalogue of individual health data collections (FReSH, for France Recherche en Santé Humaine) is currently in preparation phase	DOI, ORCID, RNSR, ROR
OpenAIRE	We utilise the Persistent Identifiers on building the OpenAIRE Research Graph for discovery, enrichment, and linking of research results.
LifeWatch ERIC	We do, but we need to extend their use in the semantic artefacts we produce, particularly once these become populated with instances (Linked Data). A strategy for such a large amount of PIDs is of utmost and urgent importance.
ENVRI-FAIR	
FAIRtracks (also part of EuroScienceGateway)	In our metadata schema, we are allowing and to a certain extent require the use of globally unique PIDs (including DOI as well as bioinformatic-related ones). See this discussion on our website (same as previous).
Bioregistry	The biomedical community uses PIDs for referencing molecular entities (genes, proteins, small molecules, complexes), higher order processes (pathways, diseases, symptoms), taxa, experimets (clinical trials), bibliometric entities (patents, research articles, organizations, grants), and more. These are either written as URIs/IRIs (as is common in the ontology community) or as compact URIs (CURIEs; mediated by a registry like the Bioregistry)
	Yes, see the previous point as well: to identify metadata and digital objects and to resolve to digital objects in machine actionable workflows.

LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products

Yes

3. Which are the roles and responsibilities in your organisation regarding PID management?

Project or initiative	Input
Life in Kyrgyzstan Study	There was no systematic approach as the data collection was done based on different organisations and funders. We have done a basic PID management, but that is to be systematised if the project is to be continued.
CESSDA	We have a metadata team that is responsible for our PID policy, technical team is responsible for the implementation.
Open Biological and Biomedical Ontologies Foundry	We are running a PURL registry for OBO ontologies (ontology and term level PURLs).
EOSC-A Long Term Data Preservation task force	NA
TRIPLE	The tech team check if the aggregated resources have PIDs and they are well-formed.
The project of French national catalogue of individual health data collections (FReSH, for France Recherche en Santé Humaine) is currently in preparation phase	One person is the correspondent of the RNSR directory. She creates and updates the records corresponding to the institute's laboratories. No other activities regarding PID management, as far as I know
LifeWatch ERIC	For EcoPortal, the owner/creator of the semantic artefact may request a DOI upon publication of his resource on the portal. In the Metadata Catalogue, the creator of the metadata record can request a DOI.
<u>FAIRtracks (also part of EuroScienceGateway)</u>	Developer and management of the metadata schema, which restricts the PIDs that are allowed. Possibly developer of new PID generator in the future
Bioregistry	The Bioregistry does not take a primary role in managing PIDs themselves - it is not a nomenclature authority. Rather, it is a collection of metadata about other PID-issuing authorities (e.g., here's what it says about DOI https://bioregistry.io/doi). The

Project or initiative	Input
	metadata can be updated in a community-driven way as it becomes out of date, e.g., a resource switches its website address for resolving PIDs. It also acts as a meta-resolver (i.e., dispatches to other PID resolvers)
FAIRsharing	see previous answer
CLARIn-ERIC	Creating and managing PIDs in the repositories hosted by the CLARIN centres, so mainly PID ownership and PID management.
FAIR-IMPACT/FAIRCORE4EOSC	In the FAIRCORE4EOSC project, DANS with others (GDWG, SURF, GRNET, ...) is establishing services for PID policy compliance assessment. The current EOSC PID Policy implies the role of 'Compliance Assessor' or 'Authority'.
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	PIDs are maintained so that they remain stable and resolvable for the long term. Indeed they point to the new version of the record (if exists) or state that the record has been archived. The DOI is managed by LifeWatch ERIC and DataCite and follows the rules of LW ERIC

4. Do you have a PID policy in your organisation?

Project or initiative	Input
Life in Kyrgyzstan Study	Not sure.
CESSDA	My organisation is a CESSDA Service Provider and we follow the CESSDA PID Policy. We have our own (internal) PID Policy that is maintained by our metadata team.
Open Biological and Biomedical Ontologies Foundry	https://obofoundry.org/principles/fp-003-uris.html
EuroScienceGateway	Not at university level afaik.
TRIPLE	it's a requirement for indexing
The project of French national catalogue of individual health data collections (FReSH, for France Recherche en Santé Humaine) is currently in preparation phase	Soft incentives to implement DOI for data sets, namely through the implementation of the institutional model of DMP. ResearcherID (WoS, Clarivate Analytics) is required rather than ORCID ID by the evaluation department.
OpenAIRE	No, but OpenAIRE participates in projects, initiatives and working groups for PID policies.
LifeWatch ERIC	Not an explicit one.
<u>FAIRtracks (also part of EuroScienceGateway)</u>	Possibly in ELIXIR, but not fully aware of this

Project or initiative	Input
CLARIN-ERIC	Section 7 in https://office.clarin.eu/v/CE-2013-0095-B_checklist-v7_3_1.pdf
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	No

5. People who contributed to the preliminary information gathered

project or initiative	Name	Surname	Affiliation
Life in Kyrgyzstan Study	Damir	Esenaliev	Leibniz Institute of Vegetable and Ornamental Crops (IGZ)
CESSDA	Mari	Kleemola	Tampere University
BlueCloud	ALEXANDRA	KOKKINAKI	NOC-BODC, Blue Cloud
Open Biological and Biomedical Ontologies Foundry	Nicolas	Matentzoglou	Semanticly
EOSC-A Long Term Data Preservation task force	Roxanne	Wyns	KU Leuven
EuroScienceGateway	Bjoern	Gruening	Uni-Freiburg
TRIPLE	Arnaud	Gingold	OPERAS-Aix Marseille University
The project of French national catalogue of individual health data collections (FRéSH, for France Recherche en Santé Humaine) is currently in preparation phase	Baudoin	Lesya	Inserm
OpenAIRE	Leonidas	Pispiringas	OpenAIRE
LifeWatch ERIC	Xeni	Kechagioglou	LifeWatch ERIC
ENVRI-FAIR	Katrin	Seemeyer	Forschungszentrum Juelich
	Katharina	Lauer	ELIXIR
CLARIN-ERIC	Willem	Elbers	CLARIN-ERIC
FAIR-IMPACT/FAIRCORE4EOSC	Wim	Hugo	DANS



project or initiative	Name	Surname	Affiliation
LifeWatch Italy, the National Hub for biodiversity and ecosystem data and research products	Ilaria	Rosati	National research Council
EOSC4Cancer	Sergi	Aguiló	BSC