Project Title	Expanding FAIR solutions across EOSC
Project Acronym	FAIR-IMPACT
Grant Agreement No.	101057344
Start Date of Project	2022-06-01
Duration of Project	36 months
Project Website	https://fair-impact.eu/

1 MS3.3 - Aligning requirements for onboarding PID providers into EOSC, including emerging PIDs

Work Package	WP 3, PIDs
Lead Author (Org)	Gabriela Mejias (DataCite)
Contributing Author(s) (Org)	Helena Cousijn (DataCite), Liisa Marjamaa-Mankinen, Josefine Nordling Lassi Lager (CSC), Natascha van Lieshout (SURF), Elizabeth Newbold (UKRI-STFC)
Due Date	2024-05-31
Date	2024-05-21
Version	V1.0
DOI	https://doi.org/10.5281/zenodo.11232175

Dissemination Level	
X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)

2   Versioning and contribution history

Version	Date	Author	Notes
0.1	15.04.2024	Gabriela Mejias (DataCite)	V0.1
0.2	30.04.2024	Natascha van Lieshout (SURF), Liisa Marjamaa-Mankinen, Josefine Nordling, Lassi Lager (CSC), Elizabeth Newbold (STFC-UKRI), Gabriela Mejias (DataCite)	V.02
0.3	07.05.2024	Natascha van Lieshout (SURF), Liisa Marjamaa-Mankinen, Josefine Nordling, Lassi Lager (CSC), Elizabeth Newbold (STFC-UKRI), Gabriela Mejias (DataCite), Rene van Horik, Wim Hugo (DANS), Sabringa Granger (INRIA)	
0.4	08.05.2024	Helena Cousijn (DataCite)	
0.5	15.05.2024	Gabriela Mejias, Helena Cousijn (DataCite), Natascha van Lieshout (SURF), Liisa Marjamaa-Mankinen, Josefine Nordling, Lassi Lager (CSC), Elizabeth Newbold (STFC-UKRI), Rene van Horik, Wim Hugo (DANS), Sabrina Granger (INRIA)	
1.0	21.05.2024	Vasso Kalaitzi (DANS) Renato C. Juacaba Neto (EMBL-EBI), Gabriela Mejias (DataCite)	

Funded by
the European Union

3   Table of Contents

TERMINOLOGY

Terminology/Acronym	Description
EC	European Commission
EOSC	European Open Science Cloud
EU	European Union
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, and Reusable
PID	Persistent identifier
SP	Service Provider
SRIA	EOSC Strategic Research and Innovation Agenda
WP	Work Package
WG	Working Group

4 Introduction

Persistent identifiers (PIDs) are crucial to enabling and achieving the goals of the FAIR principles, which describe how research data and other entities within the research lifecycle should be made Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). Some of the goals of the FAIR-IMPACT project relate to working with PID service providers and infrastructures to better understand and meet user needs, aligning with the European Open Science Cloud (EOSC) PID policy, and maximising PID adoption. Within the context of the project, these goals are addressed under a specific Work Package (WP) dedicated to PIDs, which is also contributing to the relevant activity area of the EOSC SRIA.[1] To achieve these goals, more coordination between PID service providers and EOSC is needed, resulting in PIDs for a diverse range of research outputs including publications, data, software, instruments, services, people, and organisations. After establishing a joint value proposition by PID providers,[2] and proposing a coordination mechanism for PID Providers across EOSC[3] the next step is to streamline the requirements to onboard PID Service Providers into EOSC. This report outlines the proposed requirements and provides an overview of the current PID landscape, including emerging PIDs. The current report is intended to inform the FAIR-IMPACT project and generate discussion within the broader EOSC community on the requirements to onboard PID providers into EOSC.

---

[1] https://eosc.eu/sria-mar/

[2] Mejias, G., Cousijn, H., Marjamaa-Mankinen, L., van Lieshout, N., Tatum, C., & Lambert, S. (2023). M3.1 - Joint value proposition by relevant PID providers. Zenodo. https://doi.org/10.5281/zenodo.7798215

[3] Mejias, G., Cousijn, H., Marjamaa-Mankinen, L., Nordling, J., van Lieshout, N., & Gonzalez-Beltran, A. (2023). M3.2 - Proposal for an EOSC PID Service providers coordination mechanism. Zenodo. https://doi.org/10.5281/zenodo.8405818

5   Description of the Milestone

This report describes the achievement of Milestone *3.3 Aligning requirements for onboarding PID providers into EOSC*, including emerging PIDs, in the context of the FAIR-IMPACT project. It provides an overview of the PID landscape, including emerging PIDs, and proposes the EOSC related onboarding requirements for PID providers and services as well as associated verification measures.

5.1   Role of the milestone

The alignment and proposal of requirements for PID providers to be onboarded into EOSC is a critical milestone to increasing the effectiveness of the project. The current report is intended to inform the FAIR-IMPACT project and generate discussion within the broader EOSC community on the requirements to onboard PID providers into EOSC. The proposal was drafted in collaboration with the FAIRCORE4EOSC project, and more specifically the Compliance Assessment Toolkit WP. This work was also guided by the Survey of PID Services Landscape[4] from the FREYA project[5]. This milestone represents the third step in achieving a shared-long term vision for PID service providers in EOSC (Deliverable 3.1).

Shared long-term vision for PID service providers in EOSC (M34)

2) EOSC PID providers coordination mechanism proposed (M16)

3) Align requirements for onboarding PID providers into EOSC, including emerging PIDs (M24)

1) Joint value proposition by relevant PID providers (M10)

5.1.1   Means of verification

The means of verification for achieving this Milestone are:

● Discussions and work with task members during monthly task meetings.

---

[4] Christine Ferguson, Jo McEntrye, Vasily Bunakov, Simon Lambert, Stephanie van der Sandt, Rachael Kotarski, Sarah Stewart, Andrew MacEwan, Martin Fenner, Patricia Cruse, René van Horik, Tina Dohna, Ketil Koop-Jacobsen, Uwe Schindler, & Siobhan McCafferty. (2019). D3.1 Survey of Current PID Services Landscape - Revised (Version 2). Zenodo. https://doi.org/10.5281/zenodo.3554255
[5] https://www.project-freya.eu/en

- In person discussion at the [FAIR-IMPACT 2023 All Hands Meeting](#) (including members of the FAIRCORE4EOSC project)

- Discussions during the [EOSC 2024 Winter School](#) as part of the *Opportunity Area 1: PIDs*

- Discussions at the [IDCC Workshop: Defining Criteria for Assessing PID Policies and Services](#)

5.1.2 The PID Landscape

There is currently a well-developed PID landscape across the EOSC (and the broader research community) that includes PIDs that are widely used to identify individuals, organisations, and a range of research outputs including research data. To date, there are more than 107 thousand ROR IDs[6], 157 million Crossref DOIs[7], 55 million DataCite DOIs[8], 8 million yearly active ORCID iDs[9], over 345 million ePIC Handles[10].

Based on previous work[11] produced by the FREYA project and the EOSC PID Policy Compliance Assessment Toolkit work package in FAIRCORE4EOSC[12], we have compiled the following list of PIDs and their current operational status. The status is defined following the FREYA project maturity levels:

- Established: infrastructure in common research community use; regular use within a research discipline.

- Implementation: some demos, infrastructure not in common research community use.

- Emerging: PID in discussions or specific working groups convened, infrastructure/services being actively planned. Alternatively, infrastructure is established (perhaps because it uses an existing infrastructure such as DOI) but the PID has not gained much traction yet, or there are a diversity of approaches with little clarity (or consensus) on which to use in any given situation.

---

[6] Source: ROR at https://ror.org/search accessed 3 April 2024.

[7] Source Crossref at https://www.crossref.org/06members/53status.html accessed 3 April 2024.

[8] Source: DataCite statistics https://stats.datacite.org/ accessed 3 April 2024.

[9] Source: ORCID at https://info.orcid.org/resources/orcid-statistics/ accessed 3 April 2024.

[10] Source: ePIC Handle consortium

[11] Christine Ferguson, Jo McEntrye, Vasily Bunakov, Simon Lambert, Stephanie van der Sandt, Rachael Kotarski, Sarah Stewart, Andrew MacEwan, Martin Fenner, Patricia Cruse, René van Horik, Tina Dohna, Ketil Koop-Jacobsen, Uwe Schindler, & Siobhan McCafferty. (2019). D3.1 Survey of Current PID Services Landscape - Revised (Version 2). Zenodo. https://doi.org/10.5281/zenodo.3554255

[12] https://faircore4eosc.eu/

Table 1 - List of PID Schemes and current operational status

PID Scheme	Long Name	Entity identified	Status
Accession number		Publication, Conference, Data, Project, Sample, Cultural Artefacts	Established
ADSbibcode	Astrophysics Data System	Publications	Established
ARK	Archival Resource Keys (ARKs)	Data, Information Objects	Established
arXiv	arXiv identifier scheme (en)	Preprints	Established
DiD	W3C Decentralized Identifiers (DIDs):	Any entity (e.g., a person, organization, thing, data model, abstract entity, etc.)	Emerging
Crossref DOI	Digital Object Identifier	Publications and other outputs	Established
DataCite DOI	Digital Object Identifier	Data, Preprints, Samples, and other outputs	Established
ePIC Handle	European PID Consortium (ePIC) Handle	Data objects and collections	Established
Handle	Handle System	Publications, data, and more	Established
ISBN	International Standard Book Number	Books	Established
ISLRN	International Standard Language Reference Number	Language resources	Emerging
ISNI	International Standard Name Identifier	Persons and organizations	Established
ISSN	International Standard Serial Number	Newspapers, Magazines, Journals,	Established

		Monographic Series	
IGSN	IGSN (International Geo Sample Number)	Samples	Established
ORCID	Open Researcher and Contributor Identifier	Researchers and Contributors	Established
PMID	Medline PMID (PubMed Identifier)	Publications	Established
PURL	Persistent Uniform Resource Locators	Publications and data	Established
RAI ID	Research Activity Identifier	Any result along with the dataset information and the processing script, without disclosing any source code or raw data	Emerging
RAiD	Research Activity Identifier	Projects	Implementation
ROR	Research Organisation Registry	Organizations	Established
SWHID	SoftWare Hash Identifier	Software	Implementation
URI	Uniform Resource Identifier	Archival/Storage facility, Historical or mythical person, Workflow	Established
URN	Uniform Resource Names	Data, Cultural Artifact	Established
zbMATH ID	zbMATH Open (formerly known as Zentralblatt MATH)	Researchers and Publications	Established

5.1.3 Emerging PIDs

Based on the previous table and after discussions with WP members, we have identified the following as emerging PIDs or PIDs in implementation within the EOSC context[13]:

- B2INST[14], a service that enables researchers and institutes to register their instruments and to be able to persistently refer to these registrations in other services or publications. It is one of the FAIRCORE4EOSC demonstrators.

- identifiers.org[15], a resolution service that provides consistent access to life science data using compact identifiers. identifiers.org via EMBL-EBI is part of the FAIR-IMPACT project. It is important to note that this service is well used in life sciences, specially in the Elixir context.

- ISLRN[16], a new, unique and universal identification schema for Language Resources which provides Language Resources with unique names using a standardized nomenclature. It has been onboarded by FAIRCORE4EOSC as a WP5 provider.

- RAiD[17], a unique PID developed by the Australian Research Data Commons (ARDC) to identify research projects and activities for access by research communities worldwide. RAiD will be one of the EOSC Core Components, developed by the FAIRCORE4EOSC project. RAiDs are based on DataCite DOI infrastructure.

- RAI ID[18], a unique identifier of any result along with the corresponding dataset information and processing script. It provides this information without disclosing any source code or raw data. This identifier is currently being developed as part of the RAISE project.

- SWHID[19], a persistent and intrinsic identifier for software source code artefacts such as source code files, source trees, commits, and other objects typically found in version control systems. Software Heritage participates in FAIR-IMPACT and FAIRCORE4EOSC. In June 2023, the version 1.1 of the SWHID specification was released.

5.1.4 The EOSC Landscape

Currently, the EOSC landscape consists of a dynamic ecosystem of stakeholders, infrastructures, and initiatives working towards a common vision. As EOSC continues to evolve, its different governance mechanisms, groups, and tools play a crucial role in driving its growth, sustainability, and impact on the European research and innovation landscape.

---

[13] Other emerging PIDs not within the EOSC PID infrastructure are: DIDs and DPIDs
[14] https://b2inst.gwdg.de/
[15] https://identifiers.org/
[16] https://www.islrn.org/
[17] https://raid.org/
[18] https://cordis.europa.eu/project/id/101058479
[19] https://docs.softwareheritage.org/devel/swh-model/persistent-identifiers.html

Tools like the EOSC Portal serve as an entry point for users to discover and access services and resources for their research workflows.

Within the EOSC Portal, EOSC Providers can be onboarded to provide resources such as services, data sources, and research products such as actual data, publications, software, etc[20]. The onboarding process requires a provider to supply information about itself and about the resources that would need to be included in the EOSC Catalogue and EOSC Marketplace.

To be onboarded into the EOSC Portal, service providers must ensure their services are accessible beyond their original community. They should describe their services using a standardised template which emphasises the service's value proposition and functional capabilities. At least one service instance must be operational and available to users while adhering to the FAIR principles for data. Providers must offer release notes and comprehensive documentation as well as an accessible helpdesk channel for user support and feedback[21].

While endeavouring to ascertain the PID providers currently onboarded into the EOSC Portal, we noticed that there is no standardised way to identify onboarded PID Providers, as defined by the EOSC PID Policy[22]. It is possible to find services that integrate PIDs but not PID providers specifically. Consequently, there is a pressing need for improved labelling mechanisms to facilitate the identification of PID providers within the portal infrastructure.

In 2024 it was announced that the EOSC Providers Portal, EOSC Catalogue and EOSCMarketplace will be phased out as of 15 April 2024. The European Commission will pivot the EOSC architecture to the EOSC EU Node, which will be managed and operated by the Commision[23]. The EOSC EU Node operators will ensure the discoverability and findability (search functions) of existing services from mid-April. Full features of the EOSC EU Node are expected to go live in the second half of 2024. According to the newest website, the EOSC EU Node will offer the following services (currently under development):

File Sync & Share, Virtual Machines, Interactive Notebooks, Cloud Container Platform, Large File Transfer, Bulk Data Transfer.

5.1.5 The EOSC PID Policy

The EOSC PID Policy sets expectations about the use of persistent identifiers to support FAIR research practices. The policy outlines requirements for PID providers and their services that

---

[20] https://eosc-portal.eu/eosc-providers-hub/how-become-eosc-provider/how-become-eosc-provider-a-general-overview#Who_can_Onboard_Providers_and_Resources

[21] https://providers.eosc-portal.eu/becomeAProvider

[22] A PID Service Provider is defined as an organisation which provides PID services in conformance to a PID Scheme, subject to its PID Authority. PID Service Providers have responsibility for the provision, integrity, reliability and scalability of PID Services, in particular the issuing and resolution of PIDs, but also lookup and search services, and interoperability with a generic resolution system.

[23] https://eosc-portal.eu/news-and-events/news/updates-transition-eosc-eu-node

have been approved by the EOSC Governance Board for implementation oversight. Recommendations for the PID Technical Architecture should guide implementation, as provided by the EOSC Architecture Working Group. Authored by representatives of the EOSC FAIR Working Group and EOSC Architecture Working Group, drafts of these documents were released in December 2019 and May 2020 for public consultation. The final policy[24] was delivered to the EOSC Executive and Governance Boards and published in October 2020.

In 2021 the EOSC Association launched the EOSC PID Policy and Implementation Task Force, among 12 other task forces meant to address key areas of EOSC implementation[25]. The goal of this group was, among others, to provide different kinds of recommendations on PID management and to set up criteria and certification of PIDs[26].

5.1.6 Requirements to be onboarded into EOSC

In collaboration with FAIRCORE4EOSC, the EOSC PID Policy criteria for Service Providers were analysed and operationalised into compliance assessment regimes. Following the Compliance Landscape and Use Cases - Vocabulary and EOSC PID Policy[27], the task members streamlined the EOSC PID Policy criteria to align requirements for PID Providers to be onboarded into EOSC.

Table 2 - EOSC PID Policy Requirements

Requirement	Original Policy Criteria	Comments
PID ownership MUST be visible to other actors in the ecosystem.	It must be clear at all times who is the PID Owner and thus who is allowed to make changes to the attributes.	Ownership is defined in the policy as "the authority to create a PID, assign a PID to an entity, and provide and maintain accurate Kernel Information for the PID" and this should be visible through the metadata. It should be required that PID Service Providers enable this. The PID Owner will need clear guidance on this as well. In some cases, RPOs

---

[24] European Commission, Directorate-General for Research and Innovation, Hellström, M., Heughebaert, A., Kotarski, R. et al., A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC), Publications Office, 2020, https://data.europa.eu/doi/10.2777/926037

[25] https://eosc.eu/eosc-task-forces/

[26] https://eosc.eu/advisory-groups/pid-policy-implementation/

[27] Hugo, W., Steinhoff, W., Turner, D., Buys, M., & Zamani, T. (2023). D2.1 Compliance Assessment Specification. Zenodo. https://doi.org/10.5281/zenodo.10067253

		are the owners of outputs created by individuals working for them but not always. Ownership is tied up with IPR and often causes confusion.
The PID Provider and PID Manager MUST provide the functionality required to maintain PID attributes.	It must be clear at all times who is the PID Owner and thus who is allowed to make changes to the attributes.	The PID Service Provider should enable the PID Manager to offer functionalities to update the attributes, e.g. through an API or user interface. The PID Owner will need clear guidance on this. In some cases, RPOs are the owners of outputs created by individuals working for them but not always. Ownership is tied up with IPR and often causes confusion.
The PID Provider SHOULD engage the end user community to determine changes in needs and practices and adjust their services and guidance accordingly.	(And so) levels of granularity appropriate to community best practice and use cases should be provided, while allowing for flexibility to respond to how those needs and practices will evolve.	PID providers engaging with their user community is key to developing responsive and community driven PID infrastructures and ecosystems.
PID Providers SHOULD support versioning.	PID services should support versioning. PID Service Providers and repositories must have clear policies and guidelines on how to manage versioning in case the FAIR Digital Object or entity changes.	PIDs, where applicable, should support versioning and/or provide when the PID was last updated. E.g. ORCID iDs do not have versions but publicly show when an ID was registered and last updated. DataCite DOIs support versioning and updates are visible on the metadata.
PID Providers and PID Managers SHOULD have clear versioning policies.	PID services should support versioning. PID Service Providers and repositories	PID Service Providers should provide guidance on how to manage versioning in case

	must have clear policies and guidelines on how to manage versioning in case the FAIR Digital Object or entity changes.	the underlying entity changes.
PID Providers MUST ensure that the PID issued by the PID Authority cannot be deleted in its records.	PIDs should not be re-assigned or deleted. In case that the entity being identified is deleted or ceases to exist, tombstone information needs to be in the PID attribute set.	PID Providers should ensure that PIDs cannot be deleted and/or re-assigned, and promote/implement "tombstone" best practices.
PID Providers MUST engage the community to develop one or more Kernel Information Profiles appropriate to the use cases addressed by their services.	Classes of digital objects may need different attribute sets a PID is resolved to. It is the responsibility of a community of practice to define and document these attribute sets (PID Kernel Information Profiles).	According to the RDA "PID Kernel Information is defined as the set of attributes stored within a PID record" which for many established PIDs could be equivalent to "open metadata records". E.g. DataCite metadata schema is a community governed schema by the DataCite Metadata WG.
PIDs MUST be available/resolvable to all researchers in the EU.	A PID Service Provider should offer services that integrate well with European Research Infrastructures, but not at the exclusion of the broader research community.	
PID Providers SHOULD allow integration with European Research Infrastructures.	A PID Service Provider should offer services that integrate well with European Research Infrastructures, but not at the exclusion of the broader research community.	
The basic services of PID registration and resolution	The basic services of PID registration and resolution	

SHALL have no associated cost to end users.	will have no cost to end users.	
A PID Service infrastructure MUST be at a minimum technology readiness level of 8. This applies to basic services (registration, resolution, etc).	Basic PID Services should be at Technology Readiness Level (TRL) 8 (system complete and qualified) or 9 (actual system proven in operational environment)	According to EOSC PID Policy[28]: basic PID Services should be at Technology Readiness Level (TRL) 8 (system complete and qualified). According to HORIZON 2020 WORK PROGRAMME 2014 – 2015 19. General Annexes TRL 8 – system complete and qualified[29]
PID Services MUST meet 999[30] availability and uptime	PID Service Providers need to ensure close to "24/7" availability of basic PID Services.	
PID Providers SHOULD document a summary of their maintenance and availability provisions publicly.	Responsibilities for service maintenance and availability need to be documented clearly.	Many PID providers already have this in place. E.g. https://support.datacite.org/docs/datacite-service-status
PID Providers MUST have a clear sustainability and succession plan with an exit strategy that guarantees the continuity of the resolution of the PIDs registered with the service.	PID Service Providers need to have a clear sustainability and succession plan with an exit strategy that guarantees the continuity of the resolution (as per 3.4.1) of its PIDs.	
PID Providers MUST agree to be certified at a mutually agreed frequency with respect to policy compliance.	A set of trusted registration PID Authorities and PID Service Providers is needed that are regularly certified based on agreed rule sets.	

---

[28] https://data.europa.eu/doi/10.2777/926037
[29]https://ec.europa.eu/research/participants/portal4/doc/call/h2020/common/1617621-part_19_general_annexes_v.2.0_en.pdf
[30] Three-nines availability -- 99.9% -- allows close to 9 hours of downtime per year.

PID Providers SHOULD agree with PID Managers on the responsibilities for Kernel Information maintenance, preferably via contract.	(Services) should clarify who is responsible for keeping the Kernel Information up-to-date, if necessary, by enabling third parties to modify it.	
PID Providers MUST include representatives of the EU research community within their organization.	PID Service Providers should apply appropriate community governance to ensure that their PID Services and Systems adhere to these policies, and are agile and responsive to the needs of research, Open Science and EOSC. As such, the active EU research community needs to be represented in the governance structure to be able to influence the activities and business models.	Some PID Providers already have community governance models in place, including ones with representatives from the EU research community.

The EOSC onboarding process has been performed by the EOSC Portal Onboarding Team (EPOT) to ensure that resources provided through the EOSC Catalogue and Marketplace meet the set requirements and offer the level of quality and interoperability that make them valuable to researchers.

Regarding the ongoing transition from the EOSC Portal to the EU Node, "Practical next steps on how to engage with the EOSC EU Node, for example concerning onboarding of resources, will be communicated through the EOSC EU Node front-office directly."[31]

---

[31] https://eoscfuture.eu/newsfuture/updates-on-the-transition-to-the-eosc-eu-node/

6   Conclusions and next steps

This milestone represents the last report preceding the delivery of the comprehensive report (Deliverable 3.1). The deliverable will provide the PID service providers with a vision on collaboration efforts needed to achieve a more coherent and technically aligned EOSC ecosystem. In parallel to this deliverable, user guidelines on PID implementation will be published, which will further provide the means of facilitated PID implementations within EOSC.

Despite the successful completion of preceding steps, some uncertainties have emerged due to the discontinuity of the EOSC Task Force PID Policy & Implementation, which might have effects on the recommendations proposed in the previous reports. As this Task Force no longer exists, it is unclear how the work will be continued in the future. The current line of thinking is that the Opportunity Area (OA) will function as a working group consisting of members of relevant EOSC projects, which will work together to further joint interests related to PIDs.There is also a transition phase going on between EOSC portal and EU Node which can have effects on the onboarding process. This will remain to be seen during the remaining months of the FAIR-IMPACT project duration.