Synchronisation Workshop 2024 Persistent Identifiers

October 1st, 2024

Chair: Josefine Nordling, CSC
Rapporteur: Liisa Marjamaa-Mankinen, CSC
Co-rapporteur: Elizabeth Newbold, UKRI STFC

Funded by the European Union

Synchronisation Force

ONLINE WORKSHOP

eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

2024 Edition

3 September - 7 November 2024

Save the Date!

Welcome to Session Three: Persistent Identifiers

Josefine Nordling (CSC) - chair
Liisa Marjamaa-Mankinen (CSC) - rapporteur
Elizabeth Newbold (UKRI STFC) - co-rapporteur

- The goal of this workshop is to discuss **collaboration opportunities on Persistent Identifiers**, both within EOSC as well as beyond
- To understand what is needed to realise an inter-connected coordination among PID actors, we will also discuss **sustainability of PIDs**
- One of the building blocks of sustainable PIDs stems from having comprehensive and well-planned PID policies in place for guiding PID implementation. So in this session, we will also look into **PID policies** and pay some special attention to exit strategies
- This is an **interactive workshop**; discussion and adding content is warmly welcomed.

Goal of the workshop and intended output

2022 Workshop

8 November - 12 December

2023 Workshop

2 November - 8 February 2024

2024 Workshop

3 September - 7 November 2024

- Goal: chart developments in six areas from the represented projects and initiatives.
- Output of this workshop:
  - Spreadsheet with your input
  - **Concise workshop report**
    (to be **finalised by December 2024**)
- **Output of the *three* workshops**: White Paper with recommendations for how to encourage impact, alignment and synchronisation around FAIR and EOSC
  (to be **finalised by March 2025**)
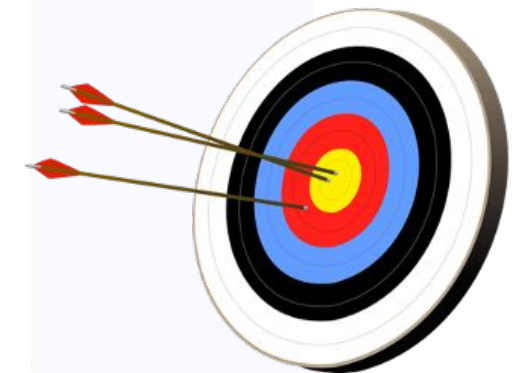
What we expect from you

Provide information about topic-related developments and plans in your project, initiative or domain:

1. During the topic sessions, use a shared **Google spreadsheet** to complete the information provided;
2. Help in reporting on workshop findings and recommendations in the concluding plenary session.

Synchronisation Force workshop series - 2024 edition - preliminary survey

Welcome on board!

Between 3rd September and 7th November 2024 you'll join us for four online collaborative sessions to assess the implementation of the FAIR principles across research projects and initiatives in the EOSC framework. **Six online collaborative sessions between 26 September and 8 October 2024** will focus around strategic FAIR-related recommendations, how they are contributing to the EOSC and FAIR vision in Europe as well as the current state of implementation. The novelty of this edition is a **session dedicated to sustainability of the project outputs**, with the aim to ensure their long-term usability.

In order to ensure an interactive meeting, we kindly ask you to complete the light survey below by Thursday 12th of September. Select the sessions that you will join and complete the questions that will appear: your inputs will constitute the starting point for our discussions!

The survey template will also be available as an offline document upon request for you to prepare your submission.

| 1	2	3	4	5	6	7	8
Project or initiative	Session 1 September 26: Metrics and assessing FAIRness	Session 2 September 30: Sustainabili of project outputs	Session 3 October 1: CET Persistent identifiers	Session 4 October 3: Legal & organisatior interoperab	Session 5 October 3: Trustworthy and FAIR-enabling repositories	Session 6 October 8: Metadata, semantics and interoperab	Complete

Project or initiative *

Please indicate which project or initiative you are providing information about. Add only one, if you represent more projects / initiatives please submit another survey

Program for today

Presenter	Topic
Josefine Nordling - CSC	Introduction to the workshop
Tibor Kalman - GWDG	PID collaboration opportunities and sustainability of PIDs
Wim Hugo - DANS	Creating EOSC compliant PID Policies
Josefine Nordling - CSC	The future of the EOSC PID Policy
Gabriela Mejias - DataCite	Coordination Mechanism for EOSC PID Service Providers
Everyone	Additional reflections on the discussion starters
Everyone	Discussion about the four pre-asked questions
Josefine Nordling - CSC	Concluding words and end of workshop
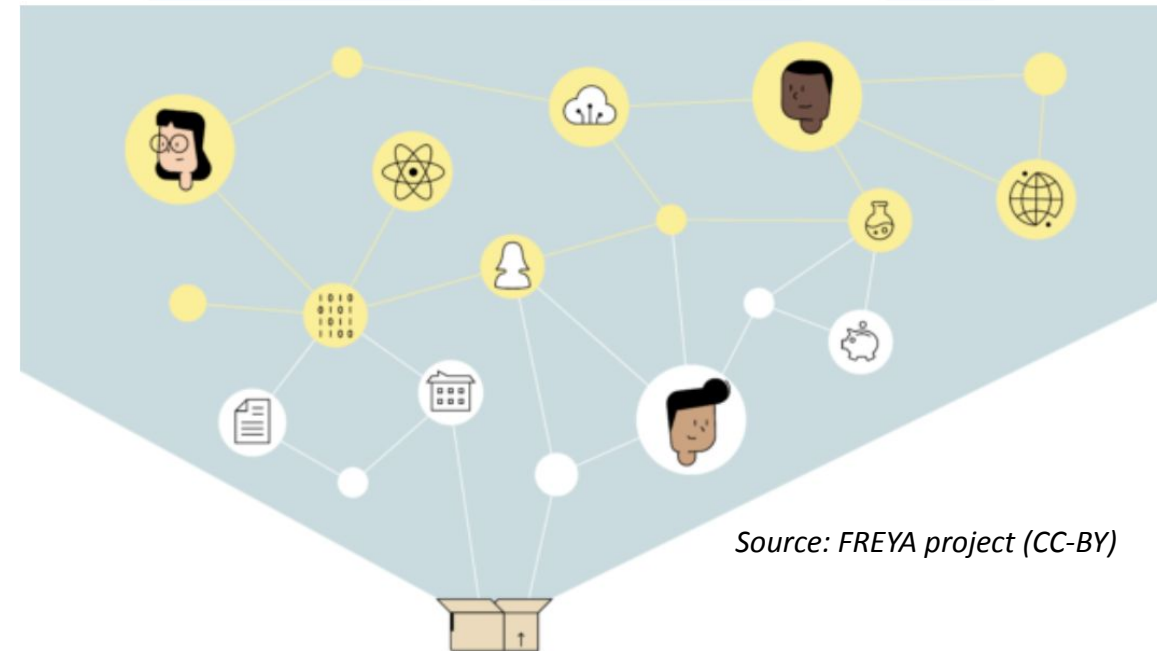
Source: FREYA project (CC-BY)

Some details for this session.

- The session will be recorded but only for internal use for the rapporteur and the report writing.
- Shared **spreadsheet** and **note taking document** for use in the session (and afterwards).
  - In the spreadsheet, please keep information factual, short and include links wherever possible.
  - In the note taking document you may add more detail and background.
  - Please be careful when editing spreadsheet cells.
  - Please refrain from editing other people's information - even typos.

Some more details for this session.

- You may wish to add more information later, the spreadsheet & note taking document will be open for further contributions until EOB **Thursday November 7th**

- Survey responses have been added to the spreadsheet

- Four questions but you may not have answers for them all

The main questions

1. What does your project or initiative do to implement PIDs? Please provide any relevant links.
2. Which Data/PID policies and recommendations do you follow, and why?
3. Do you know if your PID Service Providers have an exit strategy in place, or are developing one, for ensuring sustainability?
4. How do you think the lack of PID coordination mechanism would impact the EOSC community?

The journey…

EOSC-A "OA1 Expert Group"

- Oct 2024 — EOSC Symposium
- 2024 — Operation
- Jan/Feb 2024 — Winter School
- During 2023 — Mature idea
- Start…

Image credit: Image by starline on Freepik

EOSC-A TF "PID"

2024

- End of 2023 — End of mandate
- End of 2023 — Task force deliverables
- End of 2023 — Charter Update
- End of 2021 — Started
- 2022 — Re-organisation of the TF // Focus Groups running

Image credit: Image by starline on Freepik

Synchronisation Force
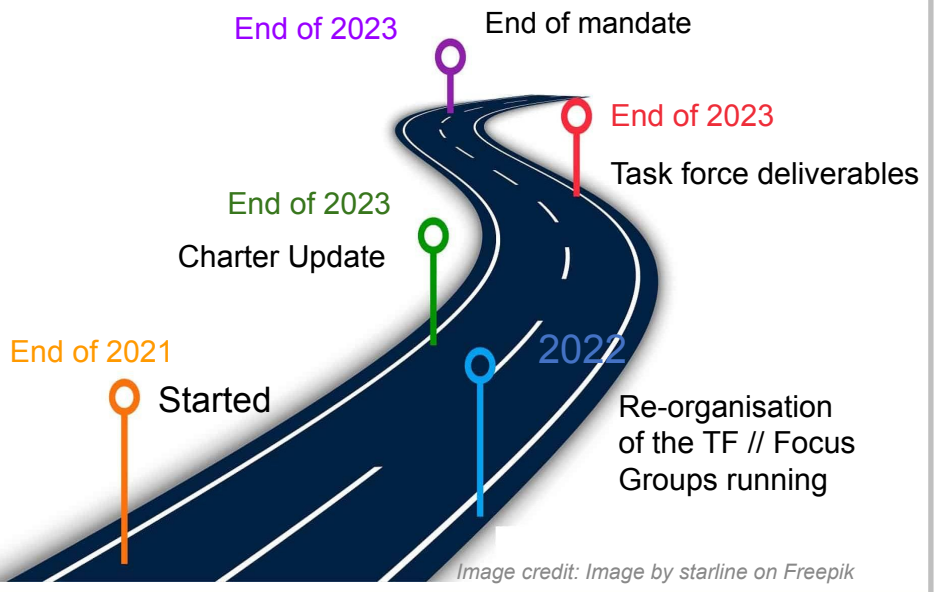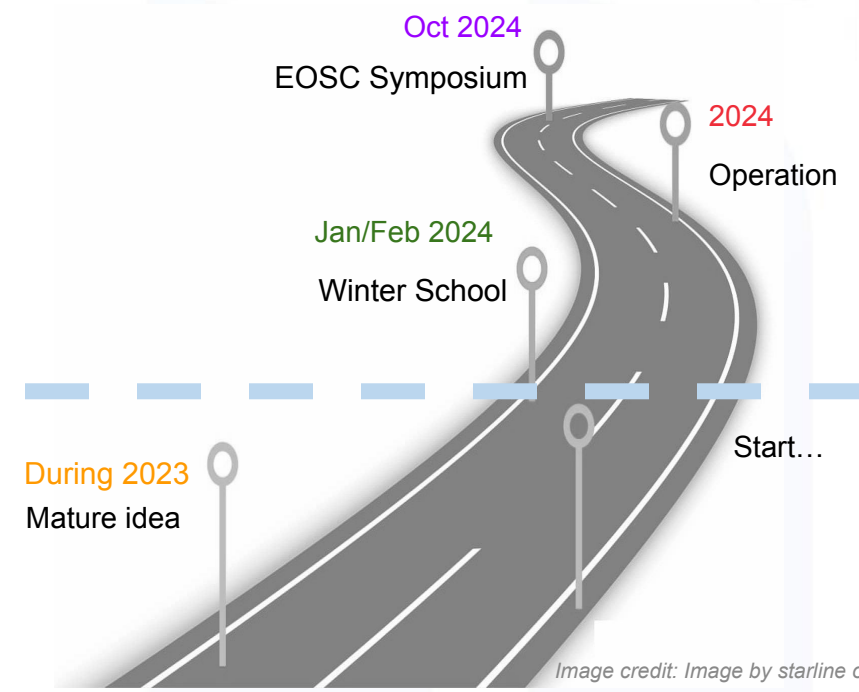
Collaboration opportunities around PIDs, PID policies, and sustainability of PIDs

Organisational point of view:

- "*within*" EOSC:
  - ~~TF "PID";~~ → TF X; TF Y, TF Z, etc
  - OA1
    - Focus?, governance?, "power"?, etc? Many things are less clear *(than TF)*
  - "nodes"
    - "Federation Handbook" might be an instrument
    - National-, E-INFRAS-, Thematic-, etc → *s. below*
- National Policies
- E-INFRAS, PID Providers
- Global Data Initiatives
- *(projects)*

⇒ for all involved parties: this might end up as a multi-layer issue
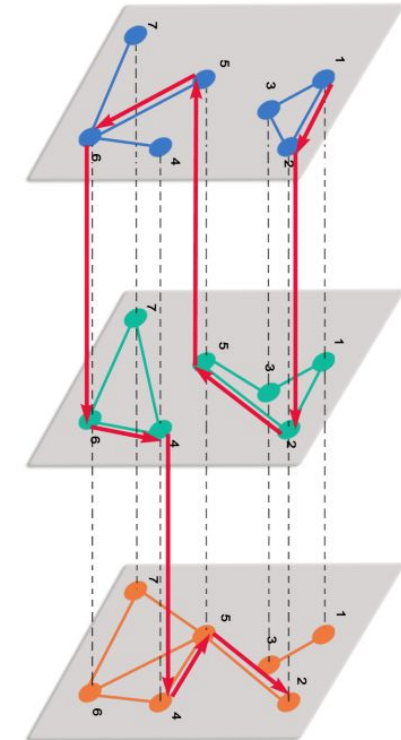
Image credit: Wikimedia:Multilayer Networks

Collaboration opportunities around PIDs, PID policies, and sustainability of PIDs

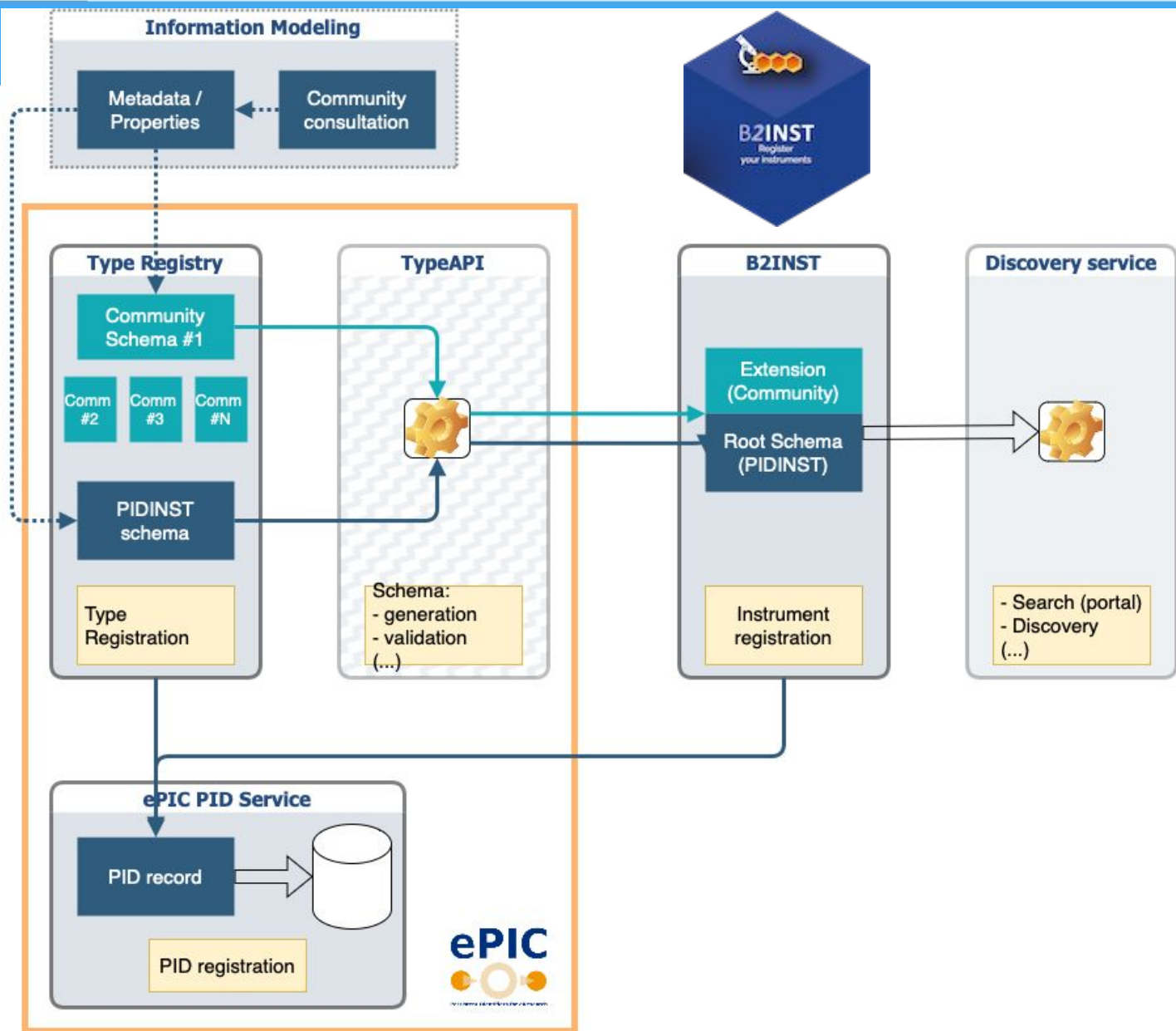**Technological** point of view:

- Example **#1: Handle System & certificates**
  - Trust is nowadays handled differently (web browsers & CA certificates)
  - New features required (PKI). Aligning those separate worlds?
  - Interoperability != Interoperation

- Example **#2: FAIRifying instruments. The policy landscape.** *(next slide)*
  - TypeRegistry & B2INST
  - prototype —> uptake/fine tuning —> new productional service

- Example **#3:** *(next slide)*
  - MetaResolver
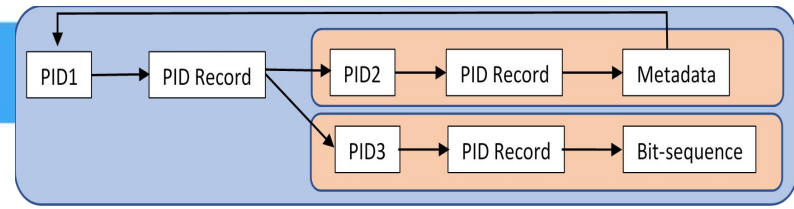 - Meta*Maintainer?*

Example#2: FAIRify Instruments. The policy landscape.

- FAIRify information on instrument by
  - registering its metadata and
  - assigning a PID to the instrument

- Instruments are non-digital objects.

- The challenge:
  - how to FAIRify non-digital objects
  - and make these discoverable

- B2INST service:
  - s. Figure on the right

Challenge:
- Who defines what?
- Who is responsible for which policy?
- RDA, FDO-F, ePIC. But EOSC?

PID Meta Resolver

PIDMR
EOSC PID
Meta Resolver

Issue: multitude of systems are used to create and maintain PIDs.

Challenge :

- to know which system is responsible for the resolution process
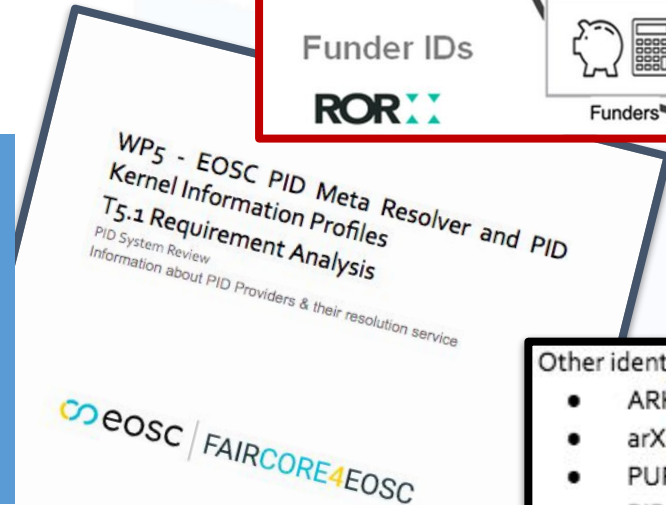- the process that provides the referenced metadata for a PID.

Need:

a uniform interface that allows PIDs from different systems to be resolved ("one place to resolve PIDs")

PID Meta Resolver is such an interface and integrates different systems.

- knows where to route different types of identifiers (eg DOI, URN)
- improves machine based data processing
- allows getting digital object information without in-depth knowledge of the resolution mechanism of different PID systems.

New Challenge + Need:

- very similar, but for *maintaining* PIDs!

WP5 - EOSC PID Meta Resolver and PID Kernel Information Profiles
T5.1 Requirement Analysis
PID System Review
Information about PID Providers & their resolution service

Supported providers:
- Handle [Handle.net]
- DOI [DOI]
- ORCID [ORCID]
- SWHID [SoftWare Heritage]
- URN:NBN [URN:NBN]
- RoR [ROR]
- ZbMatch
- RAID [RAID]

Other identifiers considered for later on-boarding:
- ARK https://arks.org/
- arXiv https://arxiv.org/
- PURL https://purl.archive.org/
- BIBCODE https://adsabs.harvard.edu/abs_
- EAN-13
- eISSN
- IGSN
- ISSN https://portal.issn.org/ and URN:ISSN
- ISBN https://www.iso.org/standard/65483.h
- ISTC https://www.iso.org/standard/41603.h
- IISSN
- LSID
- PMID
- UPC
- RRID https://www.rrids.org/

What else?

- Great, if we manage to upgrade the EOSC PID Policy document.
  - This includes not just the content, but also clarifying ownership, etc

- We have some more…
  - EOSC PID Architecture document
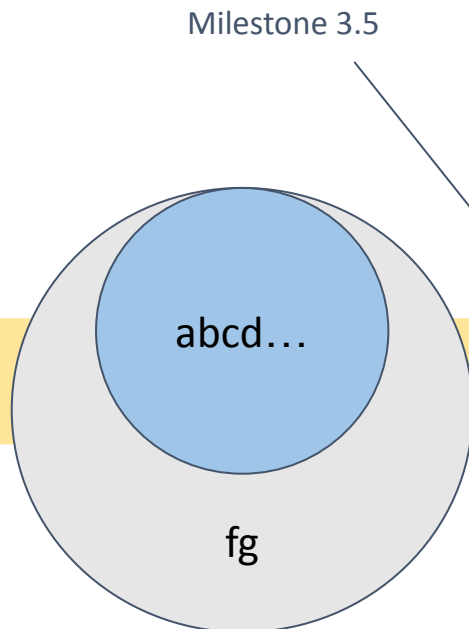 - EOSC Federation Handbook
  - ???

Creating EOSC compliant PID Policies

Wim Hugo (DANS)

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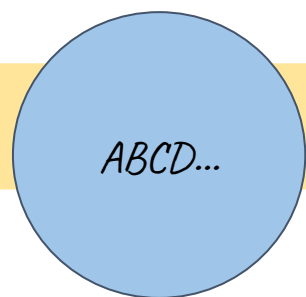
European Commission, Directorate-General for Research and Innovation, Hellström, M., Heughebaert, A., Kotarski, R., et al., **A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC),** Publications Office, 2020, https://data.europa.eu/doi/10.2777/926037

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

Hugo, W., Steinhoff, W., Lieshout, N., Buys, M., Zamani, T., van Rijsselberg, F., & Märkälä, A. (2024). D2.2 – Compliance Assessment Toolkit. Zenodo. https://doi.org/10.5281/zenodo.12683218
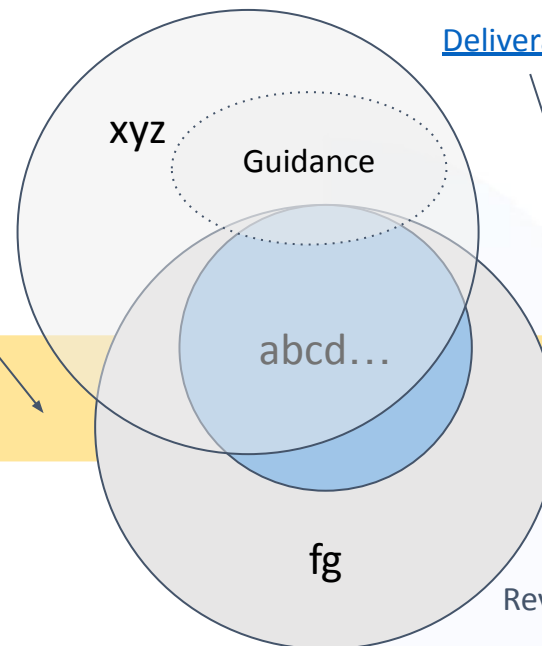
Work in Progress: FAIR-IMPACT WP3

**Community Expectations
Use Cases
Workflows
PID Policies
Best Practices**

van Horik, R., & Hugo, W. (2024). D3.3 - Guidelines for creating a user tailored EOSC Compliant PID Policy (V1.0 DRAFT NOT YET APPROVED BY EC). Zenodo. https://doi.org/10.5281/zenodo.11354246
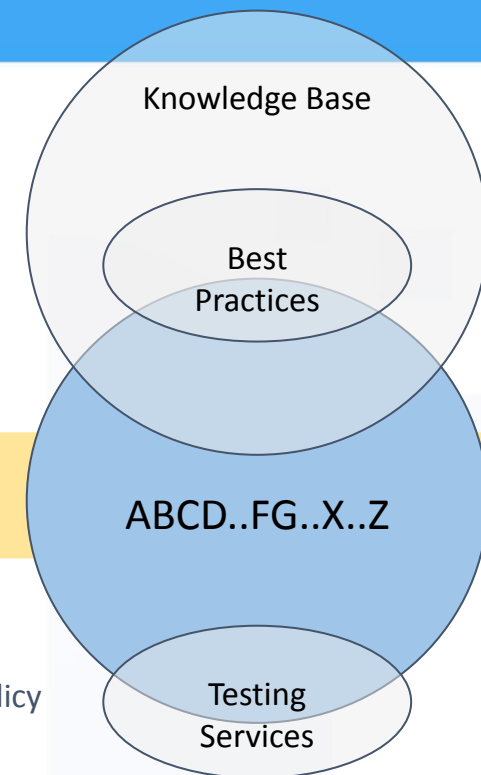
Work in Progress: FAIR-IMPACT WP3, FC4E WP2

PID Knowledge Base integrated with CAT

Knowledge Base
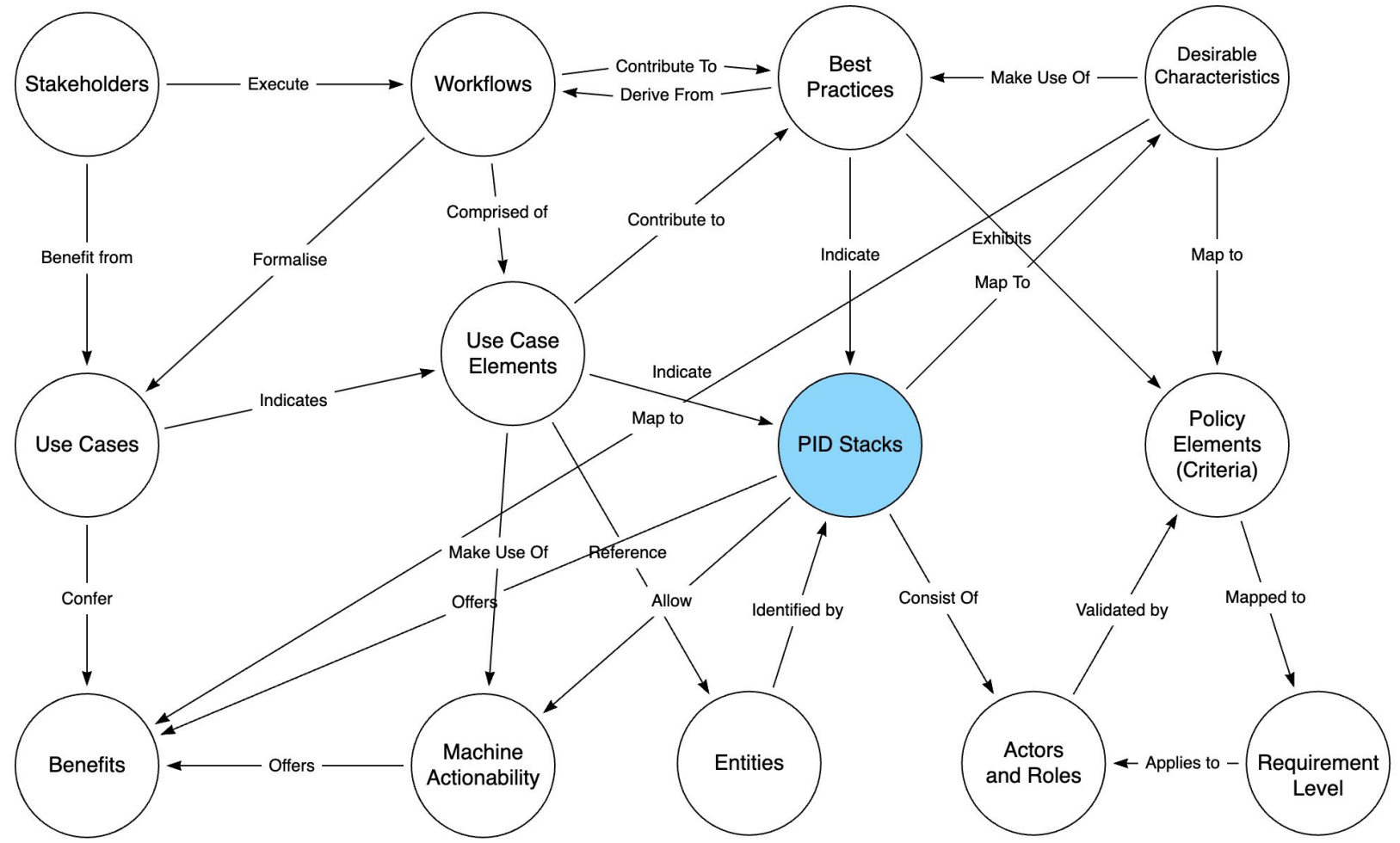
PID Stacks

We have generalised the ecosystem that provides PID Stacks by looking at a number of PID families (Handle/ DOI, ARK, URNs of various kinds, ORCID, IGSN, and SWHID - and many others).

Note that an *Owner*, in some cases, can obtain PIDs directly from an Authority (e.g. ARK, SWHID, PURL, …)

To develop sound PID Policies:

- Select the appropriate PID Stack (the supply side) for your use cases and the benefits desired. Not under your control, hence selection is key.
- Develop policy for the 'demand-side' - this is under your control.

Supply

PID Ecosystem
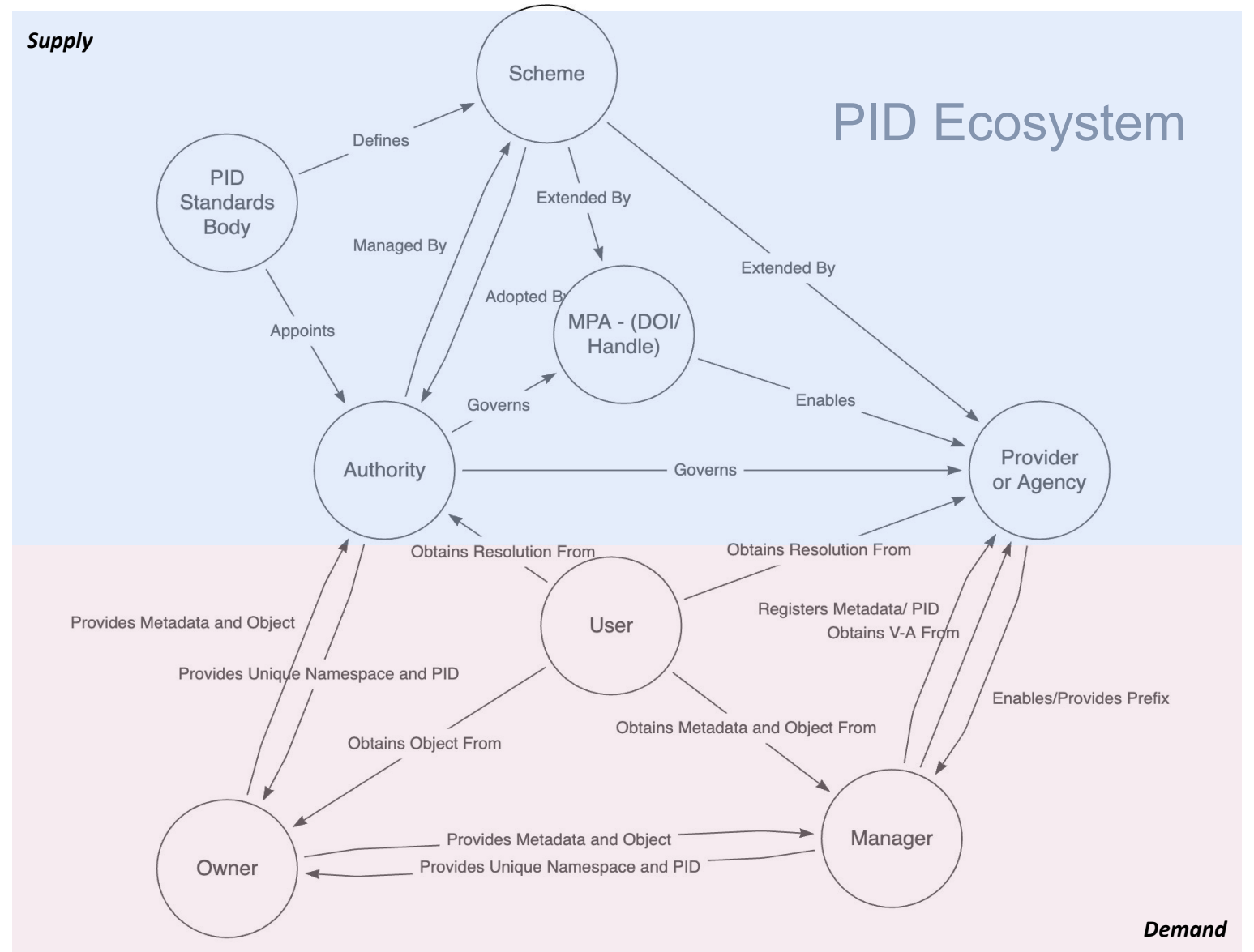
Demand

Best Practices for Managers (Demand for PID Services)

Work in FAIR-IMPACT, published in D3.3

16 Best Practices/ Guidelines identified to date, linked to Features, Characteristics and

Attributes of PID Stacks, and linked to EOSC PID Policy

Sources of Best Practice

- EOSC PID Policy

- FAIR-IMPACT

  - Review of Data and PID Policies (National, Institutional)

  - RDA Outputs and recommendations

  - Review of PID Stack documentation (recommended practices from providers, …)

  - Published use of PIDs in workflows and specific use cases

- FAIRCORE4EOSC

  - Support 'cloning' of EOSC PID Policy for custom policies
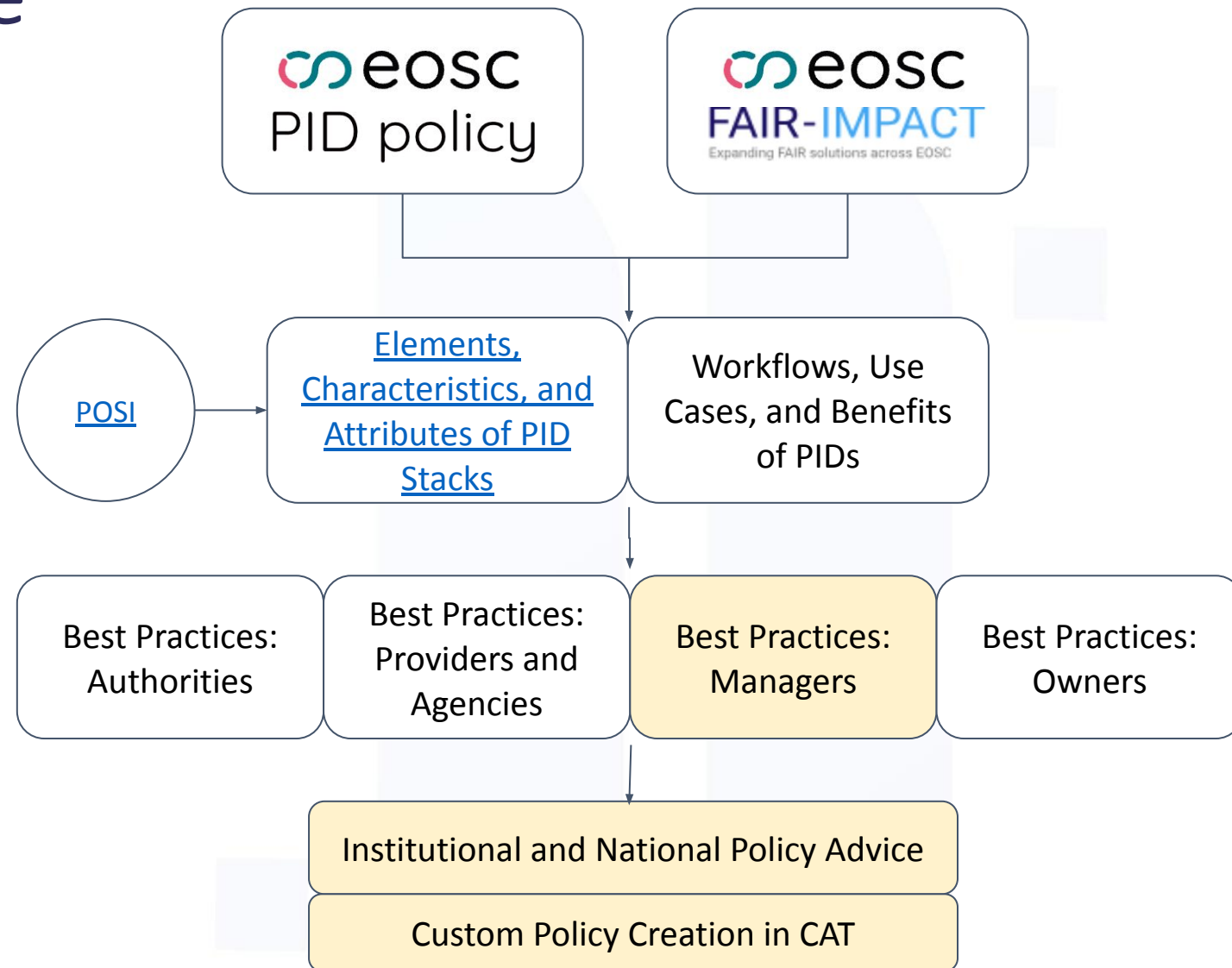
Sources of Best Practice

Largely applies to mainstream, object-referencing identifiers.

Several use cases require PIDs that have different characteristics and features

documentation (recommended practices from providers, …)
- Published use of PIDs in workflows and specific use cases
- FAIRCORE4EOSC
  - Support 'cloning' of EOSC PID Policy for custom policies

eosc PID policy

eosc FAIR-IMPACT
Expanding FAIR solutions across EOSC

POSI → Elements, Characteristics, and Attributes of PID Stacks

Workflows, Use Cases, and Benefits of PIDs

Best Practices: Authorities

Best Practices: Providers and Agencies

Best Practices: Managers

Best Practices: Owners

Institutional and National Policy Advice

Custom Policy Creation in CAT

Elements, Characteristics and Attributes of PID Stacks
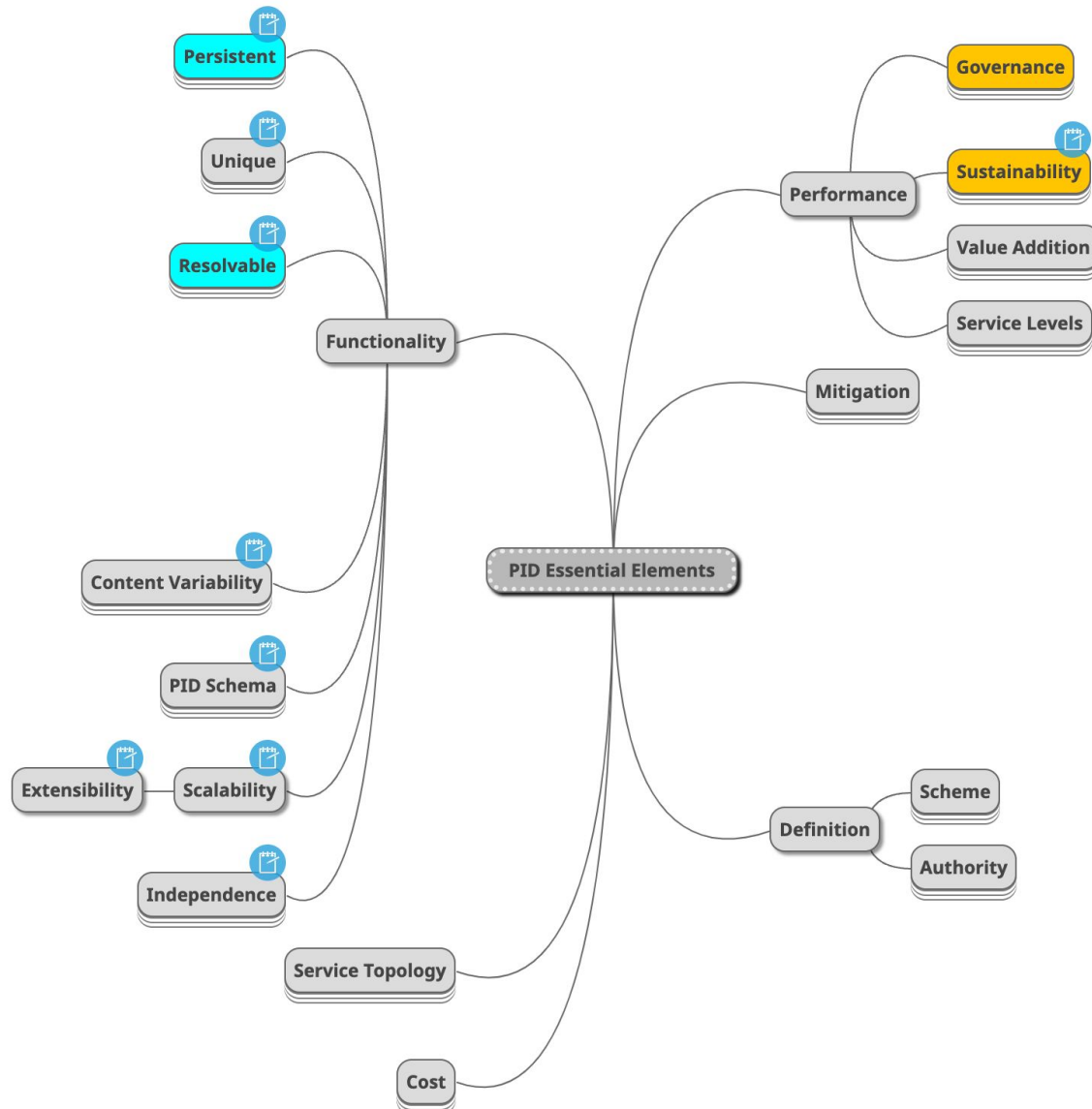
This inventory and classification is in development by the FAIR-IMPACT project. It consolidates the expectations of the community, assertions and features advertised by PID Stacks, and the content of the EOSC PID Policy into a multi-level hierarchy.

Live version

Note: work in progress!

Criteria Applicable to PID Managers

#	Principles	Criterion	Description
C4	Unambiguous FAIR	Attribute Maintenance	The PID owner SHOULD maintain PID attributes.
C5	Unambiguous FAIR	Update Functionality	The PID manager MUST provide the functionality required to maintain PID attributes.
C6	Unambiguous FAIR Independence Sustainable	Ownership Transfer	The PID manager SHOULD provide policies and contractual arrangements for transfer of ownership should the owner no longer be able to assume responsibilities in compliance with the policy.
C7	Unambiguous FAIR	Resolution Integrity	The PID Manager MUST maintain the integrity of the relationship between entities and their PIDs, in conformance to a PID Scheme defined by a PID Authority.
C11	Unambiguous FAIR	Versioning - Procedure	PID services and PID Managers SHOULD have clear versioning policies.
C14	Persistence FAIR Sustainable	Resolution Authenticity or Efficiency	PID Manager MUST ensure that the entity remains linked to the PID. In case that the entity being identified is deleted or ceases to exist, tombstone information needs to be included in the PID attribute set.
C16	Diversity FAIR	Digital Representation	Physical and conceptual entities MUST be represented via a digital representation (e.g. landing page, metadata, attribute set, database index) to have a presence in the digital landscape.
C19	Ecosystem FAIR	Accurate Entity Metadata	The PID Manager MUST maintain entity (custom) metadata as accurately as possible in collaboration with the PID Owner. This copy is the authoritative version.

Criteria Applicable to PID Managers

#	Principles	Criterion	Description
C20	Preferred Integrated Viable, Trusted	Openly Available Services	Services MUST be available to all researchers in the EU.
C22	Preferred Services	No End User Cost	The basic services of PID registration and resolution SHALL have no cost to end users.
C23	Services Maturity Sustainable	Basic Service Maturity	A PID Service infrastructure MUST be at a minimum technology readiness level of 8. This applies to basic services (registration, resolution).
C24	Services	Maturity - Value Added Services	Added value services MAY be offered at technology readiness levels lower than 8. OR Added value services SHOULD be offered at technology readiness level 8.
C28	Preferred Governed Viable, Trusted	Certification	PID Authorities and Services MUST agree to be certified with a mutually agreed frequency in respect of policy compliance. *Managers MAY be certified through self-assessment.*
C29	Services	Agreed Responsibilities	PID Services SHOULD agree with PID Managers the responsibilities for Kernel Information maintenance, preferably via contract. The same should be done between Authorities and Providers.
C34	Persistence	Persistence Median and Distribution	PID Services ecosystem SHOULD aim for a persistence median time that is acceptable to and aligns with community and dependency expectations.
C35	Resolution	Resolution Percentage	PID Service ecosystem SHOULD resolve at least p percent of PIDs in a randomised sample, where p is determined by community and dependency expectations.

Best Practices Applicable to PID Managers

Guideline 1: GUPRI

Select a PID Stack with persistence, uniqueness, and resolution characteristics appropriate to the use case. For this the acronym GUPRI can be used: the PID must be Globally Unique, Persistent with a Resolvable Identifier.

The FAIR-IMPACT Knowledge Base (under construction) will provide guidance in this respect and link to the Compliance Assessment Toolkit.
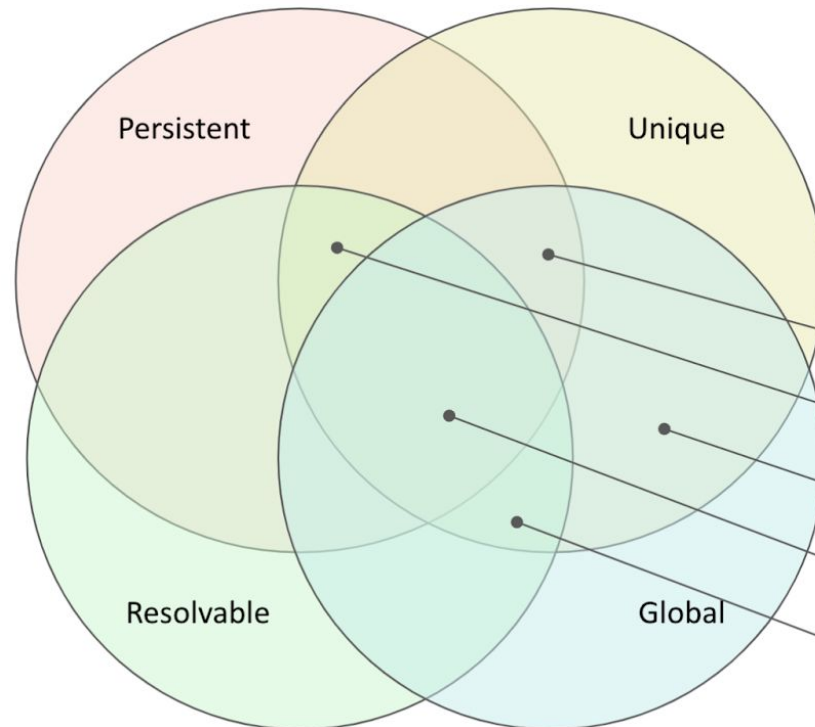
GUPRI: Globally unique, persistent, resolvable identifier
In practice:

- Globally unique: some namespace within DNS
- Persistent: requires effort and management
- Resolvable: HTTP
- Identifier: unambiguous within a namespace - implies an authority of some kind

URN:DEV, …

URN:NBN, …

LSID, …

ORCID, ARK, DOI, IGSN, …

Web URLs: DBPedia, GitHub, PURL, …

Persistent   Unique   Resolvable   Coverage   Standardised

Best Practices Applicable to PID Managers

BP2: Managing Persistence

C5   C34

Guaranteeing persistence requires effort - usually from the registry (**Authority**) or **Provider**, and from the **Manager**. Managers **MUST** develop policies and procedures to guarantee maintenance of the correct link between the identifier and the resolution target, and make sure the responsibilities are well defined.
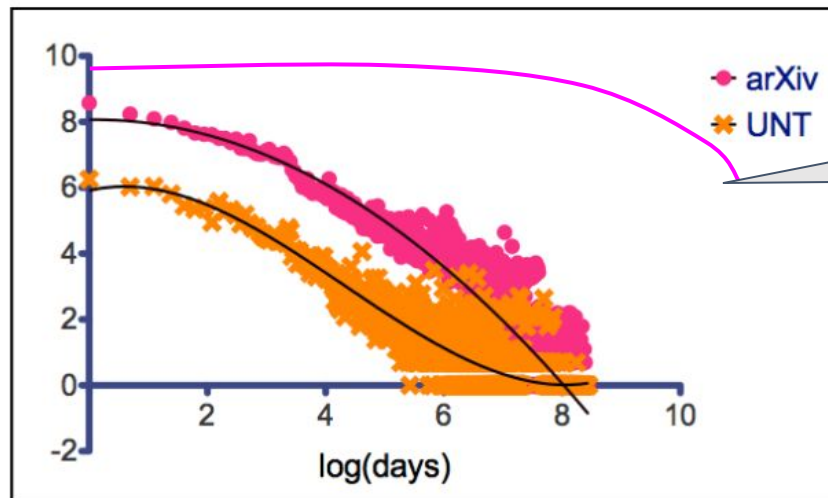
Figure 5. log() of publication/archive date difference vs log() of count of such URLs

One would expect to improve on this using well-managed persistent identifiers. Lack of persistence has two main components: *failure to resolve* and *content drift.*

[6]   Sanderson, R., Phillips, M., van de Sompel, H. Analyzing the Persistence of Referenced Web Resources with Memento, Open Repositories 2011 Conference, 2011
https://arxiv.org/abs/1105.3459v1
https://doi.org/10.48550/arXiv.1105.3459

Persistent   Living Will   Continuity Plan   Sustainability   Service Levels   Administrative Capacity

Best Practices Applicable to PID Managers

Guideline 3: Managing Versions

C5 C11

Managers MUST have a clear policy on version management, and the provisions of the policy depends on the purpose of referencing resources by the persistent identifier.

User expectations in respect of content variance is not a single concept, although there is often a perception that the resource referenced by a PID should 'remain unchanged forever'. There are several generic scenarios.

The table summarises the typical best practice appropriate for a number of scenarios.

Main Objective	Recommended practice
Stable Citation/ Reference	Metadata and non-critical data enhancements lead to minor versions with the same PID and provenance
Reproducibility and Authenticity	Data amendments that change the checksum of the referenced object leads to a new PID with provenance links to the previous version
Content Evolution and Manifestations	All previous versions must be available, and there is a choice 1. Same PID, resolving to the latest version but with previous versions easily available (e.g. Zenodo) 2. Each 'manifestation' has a unique identifier with version links to other manifestations.
Dynamic Content Growth	Community recommendations from RDA, and published formally, represents good practice

Versioning Policy

Best Practices Applicable to PID Managers

Guideline 3a: Naming Versions

C5   C11

Versioning semantics MAY be aligned with good practice in respect of software versioning [171], adapted as follows: Given a version number MAJOR.MINOR.PATCH, increment the:

- MAJOR version when you make changes that do not support reproducibility;

- MINOR version when you add content in a reproducibility-compatible manner

- PATCH version when you make backward compatible improvements

Versioning   Policy

Best Practices Applicable to PID Managers

Guideline 4: Stakeholder Involvement

Managers SHOULD make time and resources available to participate in governance structures of PID Stacks that they use. Typical activities in this respect are participation in governance through a board or oversight committee and the development of kernel, domain and custom metadata schema.

This is especially applicable in the following contexts:

- Managers that are also European Research Infrastructures and/ or e-Infrastructures
- National or multinational domain repositories

Typical activities

- Participation in governance through a board or oversight committee
- Development of kernel, domain and custom metadata schema

Stakeholder Involvement

Best Practices Applicable to PID Managers

Guideline 5: PID Stack Checklist

C21 C31

Managers SHOULD confirm the degree to which PID Stacks (providers/ agencies, authorities) support or conform to a number of important considerations. Some of these will be guaranteed by EOSC PID Policy Compliance and/ or alignment with POSI.

Aspect	Recommended practice
Certification and Compliance	EOSC PID Policy, POSI, ISO 27001
Continuity	Publication of a continuity plan, and/ or a 'living will'.
Sustainability	Financial, technical and social sustainability aspects to be taken into account.
Responsibilities	Responsibilities of actors in the ecosystem are well defined and preferably captured in formal agreements
Value-added services	CItation metrics, guidance and best practices, APIs, …

Certification   Living Will   Continuity Plan   Sustainability   Defined Responsibilities   Value Addition

Best Practices Applicable to PID Managers

Guideline 6: Select Appropriate Scale

C22

Managers MUST consider the scale at which PIDs will be used - this can range from 100s (for research outputs) to hundreds of millions (for graph-like nodes and relations with versioning and authenticity). Two interrelated considerations: scalability of the service, and the cost. Also consider future migration and annual growth.

Scale	As a Manager	As an Owner
Less than 1,000	Almost any infrastructure or service will do, provided other criteria are met. Migration can be manual if required.	
1,000-10,000	Consider registration with a provider, which may involve fixed and variable costs. Cost is not likely to be a decider.	Consider Managers that are registered with a stable provider and offers a migration possibility.
10,000-100,000	As above, but costs may start being significant and determine choice of provider. Migration readiness becomes a major consideration.	
100,000-1,000,000		Larger collections may result in a cost from Managers, depending on service. If so, consider becoming a Manager.
1,000,000-10,000,000	Consider becoming a Provider in an established stack.	Consider becoming a Manager.
10,000,000 and more	Consider own infrastructure and mirroring/ federation for performance and availability.	

Scalability   Sustainability   Business Model

Best Practices Applicable to PID Managers

Guideline 7: Select Appropriate Identifier Schema and Structure C6

Managers SHOULD consider the type of identifier and determine its stability (preferably a published and managed standard), as well as its implications for migration and its scope. Also consider *human readability* - important in some use cases!

Approach	Uniqueness	Usage and Resolution	Migration
A single namespace for all identifiers (directly globally unique)	Globally unique	Simple to use and resolve	Difficult to migrate to a new manager and/ or owner
A root namespace (prefix) with sub-namespaces for Providers		Simple to use, resolution requires additional registry information	Difficult to migrate to a new manager and/ or owner but provider might assist
A root namespace (prefix) with sub-namespaces for Managers (suffixes)		More complex resolution infrastructure	Simple to migrate managers, but not owners
A root namespace (prefix) with sub-namespaces for Owners (suffixes)		Complex to manage	Simple to migrate managers and owners
Multiple unique namespaces without a specific structure	Not guaranteed to be globally unique	Very complex to manage	Migration is simple.
No namespace in use (usually implied in internal systems)		Not suitable as external PIDs	Migration is controlled locally.

(Identifier) (Standardised) (Scope) (Definition) (Human Readable) (Transferable)

Best Practices Applicable to PID Managers

Guideline 8: Consider Resolution Options

C7   C14

Managers SHOULD consider the type of resolution mechanism offered to users and owners when selecting a service and creating their own infrastructure. Impacts on usability for humans and machines, and on interoperability.

Approach	Machine Usability	Human Usability
Via HTTP Wrappers or Prefixes for the Identifier	Yes, but patterns must be machine discoverable from a registry.	Yes, but additional information is required with some technical capability to create URLs.
Via a Web Page	Not usable.	Yes, human-directed infrastructure.
Via API	Yes, but a registry of APIs and patterns is likely required for multiple PID stacks.	Not easy to use without some technical knowledge.
Supports Compact Identifiers	Some implementations require compact identifiers to be handled by the resolution mechanism. If not, it requires an implementation layer locally to resolve compact identifiers, or third-party services (such as e.g. a metaresolver).	
Via a MetaResolver Service	Yes, if an API is offered.	Yes, if a UI is offered.

Machine-Readable   Human Readable   Resolvable   Interoperable

Best Practices Applicable to PID Managers

Guideline 9: Resolution Integrity
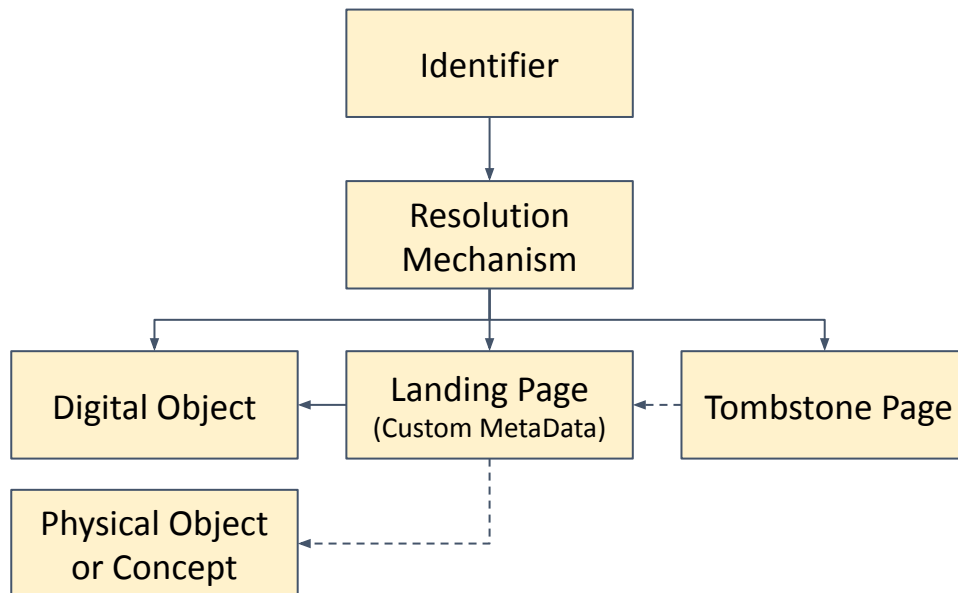
C5 C7 C14 C16

Managers MUST maintain the link between the identifier and its resolution mechanism, and the object or concept being referenced. In most cases, Managers offer custom metadata for the object or concept that is *authoritative*, and this MUST be maintained. Tombstones MUST be offered in cases where objects or concepts are no longer available, based on rational cases - see below.

Identifier

Resolution Mechanism

Digital Object

Landing Page (Custom MetaData)

Tombstone Page

Physical Object or Concept

The link between the identifier and its resolution mechanism, and the object or concept being referenced is usually maintained as a direct web reference to the *Digital Object,* a *Landing Page* that usually provides human- and machine readable custom metadata, and in cases where an object or concept is no longer available, a *Tombstone Page* ('Targets'). Physical objects and concepts must also have landing pages.

A landing page can (optionally) point to the object for access, and a tombstone page can optionally point to or be similar to the metadata landing page.

There are only a few reasons why a PID target is deleted - such as fraud or withdrawal or expiry of publication permissions for legal reasons - and hence Managers must make a strong commitment towards maintenance of the link.

Efficient   Interoperable   Actionable

Once registered with a Provider, PIDs cannot be deleted. For development and test purposes, provisional PIDs can be offered.

Best Practices Applicable to PID Managers

Guideline 10: Metadata Management

C4 **C5** **C11** **C19**

Managers SHOULD manage metadata in alignment with community and disciplinary standards, and MUST maintain an authoritative version of the metadata - either as kernel or custom metadata - in collaboration with the owner (depositor).

PID Stack	Authority	MPA	Provider	Manager	Owner
Metadata Scope	*Identifier and Kernel Metadata*			*Custom (Resource) Metadata*	
Handle System	Handle Metadata	N/A	N/A	N/A	Optional
DataCite DOI	Handle Metadata	IDF Metadata	DataCite	Manager	Optional
IGSN DOI	Handle Metadata	IDF Metadata	IGSN	Manager	Optional
ARK	ARK Metadata	N/A	Optional		Optional
URN:NBN	N/A	N/A	Optional	Optional	Optional
ORCID	ORCID Metadata	N/A	N/A	Optional	N/A
ePIC	Handle Metadata	ePIC Metadata	Optional	N/A	Optional

There must always be at least one authoritative version of metadata, and there are different candidates for this depending on the PID Stack in use. See examples.

Advice

Kernel metadata: used for citations and inventories of collections.

Custom metadata: good for findability, interoperability, and re-use.

Maintained

Shaded: grey: resource metadata authoritative copy, blue: identifier and kernel metadata authoritative copy

Guideline 11: Machine-Actionable Extensions

C20 C22 C24

Managers SHOULD consider implementation of content negotiation and machine-actionable links to improve the usability of the resource across the research enterprise ('mediations').

Approach	Description	Application
Content Negotiation	This is typically used to modify the response format of the web-based resource, and can be part of a header request or sometimes as a parameter or child node of the URI.	Example: instead of a human-readable metadata page (HTML) one could request a machine-readable one (e.g. XML or JSON).
Inflection and Multiple Resolution	Adding standard processors to a URI to request a different type of resource or format	Inflection implemented by ARK. Provides access to detailed metadata and to policy/ commitments. MR by Handle System
Signposting	A mechanism for redirecting machines to other resources in a named relation with the target, using the header. Can be summarised in a single linkset relation.	For example redirecting to author pages, project pages, query APIs, supplementary materials, linked publications, etc.
RO-Crate	Provides a mechanism for describing the research context of the object for reproducibility improvement.	Can be included in the Signposting linkset for convenience.
Affordances	Allows repository-level added services to be described and defined.	For example directling to query APIs or harvesting endpoints that apply to all resources.

Some recently developed approaches are potentially useful for this.

Content negotiation and inflection represent one approach class, where the URI used to resolve a PID is modified to add user-determined refinements and options to the resolution result. (more human-friendly)

The other approaches are aimed at provision of a standardised linkset whereby machines can discover more options.

Resolvable   Actionable   Mediated

Best Practices Applicable to PID Managers

Guideline 12: (Monitoring) Resolution Integrity

C20   C22   C24

Managers SHOULD consider implementation of mechanisms (procedures) to verify the integrity of resolution - for large collections this could be based on sampling. Integrity verification includes two elements: **link rot** and **content drift**, and the latter is partly dependent on versioning strategy.

Versioning Approach	Link Rot	Content Drift Considerations
Stable Citation	A sample of PIDs must resolve to the expected digital object or landing page. This can be achieved by comparing objects that do not resolve as expected to - for example - Memento snapshots of the object going back in time.	Any changes to a digital object that invalidates a citation needs to be versioned - e.g. corrections to a text or dataset. Improvements to metadata or supplementary materials need not be versioned or can be a minor version of the same PID.
Reproducibility		An entire digital object has to remain essentially unchanged: checksum-level similarity is required.
Content Evolution (Minor versions)		Ensure that a PID landing page provides links to previous and newer versions.
Authenticity (Major versions)		Ensure that each version has a unique PID, and that the landing page for each is linked to the next/ previous ones in the series.
Dynamic Content Growth		Several strategies can be followed, with a PID for the dynamic dataset, and linking PIDs for citable sub-sets of data used by researchers. Costly and difficult to maintain.

Resolvable   Actionable   Mediated

Best Practices Applicable to PID Managers

C2

Guideline 13 Sensitive Metadata

Managers SHOULD consider implementation of practices to deal with sensitive metadata in cases where it is required.

Approach	Kernel Metadata	Custom Metadata
Avoid Sensitive Metadata	Not Applicable	Owners are asked not to include sensitive metadata when describing an object or concept.
Compartmentalise Sensitive Metadata	Not Applicable	Sensitive metadata is submitted as an encrypted or protected file, and is not indexed. Access is granted on request by the Owner or Curator.
Explicit Support	Metadata can be marked as sensitive and encrypted, and access is granted on request.	Sensitive metadata can be accommodated as a separate metadata category with limited access, and possibly encrypted.
		Metadata can be marked as sensitive and encrypted, and access is granted on request.

Sensitive metadata cannot be indexed.

As such, it is not useful for discovery, but only for reuse.

The exception may occur when it is possible to grant access to an entire catalogue or collection.

For typical long-tail repositories, this means obtaining permission from thousands of owners (or even more) - not feasible.

FAIR-IMPACT is evaluating case studies involving sensitive metadata - guidance to be added as applicable.

Sensitive Metadata   Encryption

Best Practices Applicable to PID Managers

Guideline 14: (Monitoring) Resolution Integrity - Sampling Size

C20   C22   C24

Managers that curate a large number of PID-referenced resources MAY consider random sampling to verify resolvability. The tables below indicate sampling sizes required for specific error margins and certainty.

Statistically significant sample sizes (2% error margin)

Population (number of PIDs)	Certainty (Confidence)		
	90%	95%	99%
100,000	1663	2345	3980
1,000,000	1689	2396	4128
10,000,000	1691	2401	4143
100,000,000	1691	2401	4144
1,000,000,000	1691	2401	4145
Unlimited/ Unknown	1691	2401	4145

Statistically significant sample sizes (1% error margin)

Population	Certainty (Confidence)		
	90%	95%	99%
1,000,000,000	6764	9603	16577
10,000,000,000	6764	9604	16578
Unlimited/ Unknown	6764	9604	16578

Resolvable   Actionable

Best Practices Applicable to PID Managers

Guideline 15: Sustainability and Continuity

C2   C28

Managers MUST develop and implement mechanisms to ensure continued access should their services need to wind down or change, and SHOULD preferably have access to sustainable funding. If applicable, certification as a trustworthy repository ensures that adequate measures are in place.

Sustainability Aspect	Description
Technical	Data must be open, accessible, and adequately mirrored and backed up. Software used for metadata and PID management should preferably be open source.
Financial	Managers should have a sustainable business model.
Social and Governance	A continuity plan that makes provision for transfer of custom metadata, digital objects, and associated supplementary materials to a suitable custodian environment should be a strong consideration.

Options for certification are limited to dataset managers (Repositories) via CoreTrustSeal, nestor seal, or ISO 16363.

CoreTrustSeal guidance provides good additional best practice.

Continuity options vary, and depend on the nature of the digital objects.

1. Open digital content, with simple content types, can typically just be exported as static web resources that require little further curation.
2. If the digital objects are large or complex and need specialised technology to be maintained, the continuity options also become more complex.
3. If some of the digital objects and metadata are sensitive, active management of access requests will be required.

Sustainability   Governance   Certification

Best Practices Applicable to PID Managers

Guideline 16: Maturity and Availability of Services

C1 C23 C24 C25

Managers SHOULD adopt the EU Technology Readiness Level classification for services and web resources. Infrastructure and maintenance SHOULD aim for a level of availability that is acceptable to end users.

Readiness Level	Description and Applicability
TRL9	*System proven in operational environment .* All main services (creating and updating PID metadata, resolution targets) must be at this level.
TRL8	*System complete and qualified.* Demonstration systems and full beta releases. Applicable for releases of value-added, non-critical services.
TRL7	*Prototype demonstration in operational environment.* Alpha releases. Releases of value-added, non-critical services, but not advised.
TRL6	*Demonstrated in relevant environment.* Experiments and labs. Applicable to all services with appropriate disclaimers.
Lower Levels	Not applicable.

Availability expected of Managers depend on their context, and can vary from commercial-level expectations and agreements to the typical research consensus (no guarantees of service after hours, and the services are free).

Irrespective of the context, service levels should be published explicitly, even for free services.

Availability is expressed in a standard notation that can be considered.

Availability   Maturity

Consider automated monitoring of service availability and uptime.

Deliverable D3.3

zenodo

Search records...

Communities   My dashboard

FAIR-IMPACT   FAIR-IMPACT
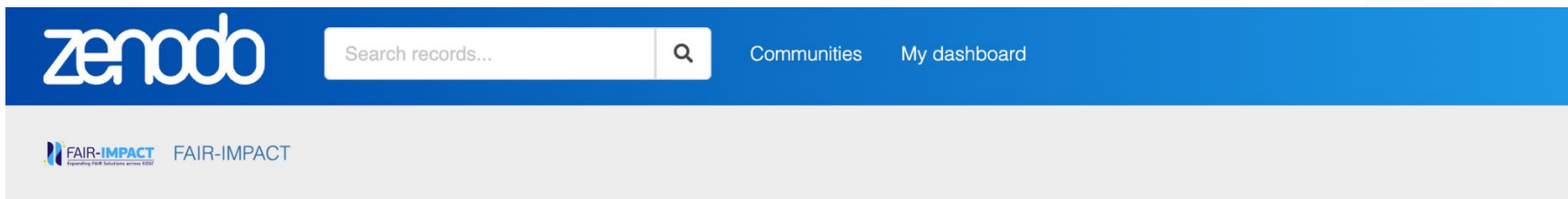
Published May 30, 2024 | Version V1.0 DRAFT NOT YET APPROVED BY THE
EUROPEAN COMMISSION

Project deliverable   🔓 Open

D3.3 - Guidelines for creating a user tailored EOSC Compliant PID Policy

van Horik, Rene (Project member)[1] iD;   Hugo, Wim (Project member)[1] iD

Show affiliations

Contributors

Project members:

Joy, Davidson[1]; Nordling, Josefine[2] iD; Marjamaa-Mankinen, Liisa[2] iD; Lager, Lassi[2] iD; Mejias, Gabriela[3] iD; Newbold, Elizabeth[4]

Show affiliations

The main topic of this report concerns the formulation of guidelines for PID Managers regarding the creation of a "user tailored EOSC compliant PID policy." Central to the work are the following questions. How can we determine the quality of the EOSC PID policy and - in collaboration with all stakeholders - how can we implement a user tailored PID policy? In other words, what principles, roles and criteria can we distinguish when it comes to the use of PIDs in the EOSC. To answer this question, we need to thoroughly and systematically analyze the PID landscape (of which PID Managers are a part). For this, we are working closely with the FAIRCORE4EOSC project in which a Compliance Assessment Toolkit (CAT) is being developed. The CAT will support PID Managers and others with services to encode, record, and query compliance with the policy.

In total 16 guidelines for formulating PID policies are compiled and presented in the report. On this basis, targeted support can be provided for the formulation of EOSC compliant PID policies as provided by the FAIR-IMPACT project.

Files

CAT Support for Tailor-Made Policies

CAT will soon support the following use case:

- Clone the EOSC PID Policy as a basis for an institutional, funder, project, infrastructure, or initiative-related PID policy
- Add specifics as criteria or tests: e.g. add a requirement that certain PIDs must be used for specific entities
- Refine or tighten EOSC PID Policy benchmarks

Complying with the derived policy will also result in EOSC PID Policy compliance if criteria and tests are not removed, or benchmarks are not relaxed.

Feedback on the guiding principles

No suggested changes to the content of the guiding principles.

However, multiple principles of the EOSC PID Policy are not elaborated upon in the policy criteria definitions, e.g. :

- **PID Services should be interoperable**
- **Technology independence of PIDs**
- Encouragement of **new and innovative services and tools**

Source: ntaskmanager.com

PID application

- The **PID Service Provider** SHOULD provide a **feedback mechanism** for users of the system
- **Providers MUST support versioning and have clear versioning policies in place**, and by implication that is for both Kernel Metadata and for the object or concept being referenced by the PID

PID service and PID service providers

- "PID Authorities and PID Service Providers MUST agree to be certified" —> better term for certified is to be **subject for review**, or to be **assessed or appraised**
- **Frequency of assessment** should be defined, compared to e.g. the CTS, completed every three years
- **Dedicated consultation for PID Managers** would be highly beneficial

Persistence is a promise

Governance and sustainability

- Clear definition of the meaning of **global governance**
- Decision to be made on **whom constitutes the EOSC Governance**
- Clearer conditions for **compliance assessment and continuity**
- The **ownership of this EOSC PID Policy** should be made clear, as well as the actor for conducting the frequent certifications / reviews.

Why PIDs?

- PIDs and metadata are key to the FAIR principles
- The goal is to enable and support a sustainable implementation of PIDs across EOSC by working together with PID service providers and infrastructures to meet user needs, align with EOSC PID Policy and to promote best practices.

WP 3 tasks

Task 3.1

Setting up a coordination mechanism for EOSC PID service providers

CSC
SURF
UKRI-STFC

Task 3.2

Integration of PID practices into FAIR data management

KNAW-DANS
INRAE
INRIA
SURF
DataCite, UESSEX-UKDS
LifeWatch
CNR EMBL-EBI
UKRI-STFC
UNIMAN

Task 3.3

EOSC PID Policy alignment & support

UEDIN
CSC
DataCite
UKRI-STFC

Task 3.4

PID implementation programme

KNAW-DANS
UEDIN
CSC
DTU-DeiC
DataCite

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)

Joint value proposition of PIDs

1. PIDs and metadata are needed to enable FAIR research;

2. PIDs enable global scaling of research through unique and standardized identification of scholarly entities;

3. PIDs improve understanding of research impact through interoperability and connectedness;

4. PIDs help stakeholders save money and time through automation ;

5. PIDs improve trust in research by facilitating recognition and preservation of diverse range of outputs;

6. PIDs improve equity across disciplines and countries by increase recognition of research contributors;

7. PIDs support long-term preservation and sustainability of research outputs through community governance

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 M3.1 - Joint value proposition by relevant PID providers

Work Package	WP3, Persistent identifiers
Lead Author (Org)	Gabriela Mejias (DataCite)
Contributing Author(s) (Org)	Helena Cousijn (DataCite), Liisa Marjamaa-Mankinen (CSC), Natascha van Lieshout (SURF), Clifford Tatum (SURF), Simon Lambert (STFC-UKRI).
Due Date	2023-03-31
Date	2023-03-30
DOI	https://doi.org/10.5281/zenodo.7798215
Version	V1.0

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)
	CO: Confidential, only for members of the consortium (including the Commission)

https://doi.org/10.5281/zenodo.7798215

Proposal for an EOSC PID Service providers coordination mechanism

1. Co-operation between PID Service Providers and EOSC

2. Emerging PIDs

3. Ensuring EOSC Compliance by PID Service Providers

4. Channeling needs from the EOSC community

5. Sharing outputs from relevant projects

https://doi.org/10.5281/zenodo.8405818

Aligned requirements for onboarding PID providers into EOSC, including emerging PIDs

- Requirements as per the EOSC PID Policy (in coord with FAIRCORE3EOSC)
- Emerging PIDs identified in the context of EOSC

https://doi.org/10.5281/zenodo.11232175

Open questions & next steps

- PIDs in EOSC
- EOSC Nodes
- Gather feedback from PID Providers

Thank you!

eosc | FAIR-IMPACT

@fairimpact_eu   /company/fair-impact-eu-project

Synchronisation Force