Welcome to Session Three: Persistent Identifiers

Josefine Nordling (CSC) - chair
Liisa Marjamaa-Mankinen (CSC) - rapporteur

• The goal of this workshop is to discuss trust in and sustainability of PID systems, when defining PID Policy assessment criteria
• This is an interactive workshop; discussion and adding content is warmly welcomed.
Program for today

- Compliance Assessment Toolkit (CAT) - Wim Hugo (DANS)
- CAT in research workflows - workshop outcomes - Natascha van Lieshout (SURF)
- Presentation on PID use cases in FAIR-IMPACT and trust in PID infrastructures (KE report) - Josefine Nordling (CSC)
- Reflections from PID use case partners on trust and sustainability in PID systems - Everyone - feel very free to join in on the discussion!
  - Hervé l’Hours (CESSDA ERIC - PIDs in sensitive data)
  - Renato Juacaba Neto (EMBL-EBI - PIDs in data production workflows & sensitive data)
  - Parham Ramezani (LifeWatch ERIC - PIDs in complex data citation)
  - Nick Juty (UNIMAN - PIDs in data production workflows)
- Discussion about the three pre-asked questions
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, spreadsheet & note taking document will be available until December 10th
• Survey responses have been added to the spreadsheet
• 3 questions but you may not have answers for them all
The main questions

1. What does your project or initiative do to implement PID(s)? Please provide any relevant links.
2. What are the guiding principles for you when choosing a PID system or service?
3. What does in your opinion constitute trust in a PID service?
Compliance Assessment Toolkit

Wim Hugo (DANS)
Context and Process


Work in Progress: FAIR-IMPACT WP3
Community Expectations
Use Cases
Workflows
PID Policies
Best Practices

PID Knowledge Base integrated with CAT
## C1: Minimum Operations Are Available

<table>
<thead>
<tr>
<th>#</th>
<th>Principle or Objective</th>
<th>Suggested Criterion</th>
<th>Description</th>
<th>Metric</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Preferred Unambiguous Interoperability</td>
<td>Minimum Operations Are Available</td>
<td>Service providers SHOULD provide a common Application Programming Interface to interact with PIDs, supporting a minimum set of operations (create, resolve and modify PID and PID Kernel Information)</td>
<td>Σ T1, n</td>
<td>=0 → 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;1 → 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Test</th>
<th>Description</th>
<th>Type</th>
<th>Method</th>
<th>Guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1,1</td>
<td>CREATE</td>
<td>Create a PID and provide kernel information: API exists and evidence (URL) is available</td>
<td>Binary</td>
<td>Yes = 1</td>
<td>G1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No = 0</td>
<td></td>
</tr>
<tr>
<td>T1,2</td>
<td>UPDATE</td>
<td>Update kernel information for existing PID: API exists and evidence (URL) is available</td>
<td>Binary</td>
<td>Yes = 1</td>
<td>G1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No = 0</td>
<td></td>
</tr>
<tr>
<td>T1,3</td>
<td>Resolution Service</td>
<td>Resolution API (URL) or URI Pattern exists, evidence is provided</td>
<td>Binary</td>
<td>Yes = 1</td>
<td>G1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No = 0</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>One may extend the tests to recognise typical and popular standards for API implementation, such as REST, SmartAPI, and the like.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Resolvable | Mutable

Suggestion: Provider SHOULD provide a common Application Programming Interface to interact with PIDs, supporting a minimum set of operations (create, resolve and modify PID and PID Kernel Information).
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, …)
PID Stacks

PID Stacks are best described by looking at some examples. The table on the right shows a number of PID Stacks built from the Handle System. It includes popular stacks such as the DataCite and CrossRef DOIs, as well as ePIC.

There are many other stacks in the ecosystem.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Authority</th>
<th>MPA</th>
<th>Provider (Registration Agencies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>Corporation for National Research Initiatives (CNRI)</td>
<td>Not investigated</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>Coalition for Handle Services – China</td>
<td>Not investigated</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>GDWG/ ePIC</td>
<td>See A.4.6</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>CTIC</td>
<td>Not investigated</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>MISADI</td>
<td>Not investigated</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>Smart Africa Alliance</td>
<td>Not investigated</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>Tunisian Internet Alliance</td>
<td>Not investigated</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>RosTelecom</td>
<td>Not investigated</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>Airiti</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>BSI Identify</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>Chinese DOI</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>CNKI</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>CrossRef</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>EIDR</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>HAND</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>JaLC</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>KISTI</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>mEDRA</td>
</tr>
<tr>
<td>Handle System</td>
<td>DONA Foundation</td>
<td>International DOI Foundation</td>
<td>EU-OP</td>
</tr>
</tbody>
</table>
Compliance Assessment Toolkit

EOSC PID Policy compliance for all actors in the ecosystem.

Workflows, Use Cases, Best Practices, Entities, ...

(René van Horik, Wim Hugo, DANS)
The Value of EOSC PID Policy Compliance

**For Authorities and Providers:** It is important to back up public claims of EOSC Policy Compliance and of performance/features with publicly available information.

**For Managers:** Optional disclosure of policy compliance for purposes of trust and community acceptance.

**For Owners and Users:** Matching PID services with use cases and expected benefits.

All assessments are initially self-assessments - envisaged that we assist all interested PID Stacks with this. *The EOSC PID Policy envisages a compliance authority.*

All assessments are initially self-assessments and **there is no obligation to publish** these assessments for the foreseeable future - publish if it is useful.

Find PID services and assess their characteristics, features, and how it matches use cases - whether they are policy compliant or not.
Assessments

Read about different actors in the ecosystem before starting.

- PID Authority
  - View public assessments

- PID Service Provider
  - View public assessments

- PID Manager
  - View public assessments

- PID Owner
  - View public assessments

- PID Scheme
  - View public assessments

Contact and Service
- Contact
- Terms and Conditions
- Data Licenses
- Privacy

About CAT
- Our Team
- GitHub
- Disclaimer

Follow Us
- Newsletter
- Youtube
- Twitter

Version: 0.2.0  Commit: 6f9f6a23594e492804f3ca70ca74483a4580ea2b  2023-11-16T12:16:15Z
Create a New Assessment
Object or Service Being Assessed
Assessment Criteria

**P1 - Application:**
- **C5 - Update Functionality**
  - Principle P1: Application
  - PID application depends on unambiguous ownership, proper maintenance, and unambiguous identification of the entity being referenced.

**C6 - Ownership Transfer**
- **C7 - Resolution Integrity**
- **C11 - Versioning**

**P13 - Persistence:**
- **C14 - Resolution Authenticity or Efficiency**
- **C34 - Persistence Mean**

**P6 - Diversity:**
- **C16 - Digital Representation**

**P3 - Ecosystem:**
- **C19 - Accurate Entity Metadata**

**P7 - Services:**
- **C22 - No End User Cost**

---

**Test T2.1: Secure - Encrypted**

**Question:** Are the API services offered, encrypted using HTTPS?

**Answer:**
- Yes
- No

**Test Result:** 1
Mandatory and Optional Criteria

Principle P9: Resolution
PID Service SHOULD resolve at least p percent of PIDs in a randomised sample, where p is determined by community and dependency expectations.
Outcomes from looking at PIDs in research workflows workshop

By: Natascha van Lieshout (SURF)
Goal:
To demonstrate concrete tools, policies and developments within the realm of PID projects from the EOSC and use that to fuel further innovation through the FAIR IMPACT Open Support Call in January 2024.

EOSC Compliant PID Implementations - Practical Guidelines for Implementing Best Practices

09.00 - 13.00 CET
21 November 2023

Online
Motivation

Familiar and put them to practice: 18%
Familiar but do not currently put them to practice: 13%
Heard of them but not very familiar of what they mean: 31%
Never heard of the FAIR principles before: 38%

European Research Data Landscape Study
- Ran from 2021-2022
- 15,066 respondents

https://op.europa.eu/en/publication-detail/-/publication/03b5562d-6a35-11ed-b14f-01aa75ed71a1
https://zenodo.org/communities/erdl21
Familiarity with FAIR principles:

- Familiar and have put them into practice: 10
- Familiar but have yet to put them into practice: 6
- Aware of their existence but don't know what they mean: 2
- Completely unaware of that they are: 0

Familiarity with PIDs:

- An expert in: 3
- An owner or manager of: 4
- A user of: 6
- Aware but not a user of: 4
- Unfamiliar with: 0

Familiarity with EOSC PID Policy:

- Read and implemented: 1
- Read but not implemented: 2
- Aware of but not read: 9
- Never heard of it: 5
I came to the EOSC Compliant PID Implementations workshop to learn about…
Open questions about PIDs:

Where to obtain or create them

What the open questions are in EOSC

why everyone thinks DOI equals PID...

Relation between all PIDs

not confused, but metadata (minimal) requirements are often not obvious

long term persistence guarantees

Uniqueness

How to use the same PID across very different domains

Who is the ‘source’ for governance

what the EOSC PID policy means practically, especially if there is also a national PID policy in place.

Interoperability

Do we have a agreement about which PID to use for what (ORCID for authors, ROR for organisations, DOIs for paper, data and software, .).

why DOI is considered the only viable PID sometimes

Adressed

Developing
Breakout: Publications

- The perspective of owners and selection of appropriate PID services is also quite important but not really covered by EOSC PID Policy
- Reviewed considerations for owners to select an appropriate service
- PID Policy is ‘binary’ - there are some refinements necessary
Breakout: Software

- Intrinsic vs extrinsic PIDs
- A PID is an additional commitment and software licenses could potentially help define this commitment
- Human curation is necessary but we must standardize and track it through metadata
- Some things need to be left to specialists
Breakout: Datasets

- EOSC can define PID policies, best practices and guidelines and PID providers can enable compliance with these policies but this does not guarantee that end users will do so
- Interoperability between PIDs and other tools is crucial
- Data workflows and PID maintenance within those workflows remains complex
Thank you!
Use cases working on documenting best practices for PIDs

Synchronisation Force PID workshop
30 November, 2023

Josefine Nordling, CSC - IT Center for Science
WP lead on Persistent Identifiers
Tasks on PIDs

Task 3.1

Setting up a coordination mechanism for EOSC PID service providers

Task 3.2

Integration of PID practices into FAIR data management

Task 3.3

EOSC PID Policy alignment & support

Task 3.4

PID implementation programme
PID practices in FAIR data management

- Identifying and documenting best PID practices for; managing workflows, datasets and research objects
- Meeting user needs and a more coherent implementation of PIDs, leading to more exact data citation and a broader and more targeted use of PIDs
- Sharing the results across research communities to achieve alignment in PID practices
- Interacting with other communities through workshops

Source: Mindmeister.com
Focus on data products, automatic workflows, documenting data provenance in processes as well as PIDs for instruments, software etc. (UKRI-STFC, CNR, INRIA, UNIMAN)

Cross-cutting themes

Subtasks

Persistent Identifiers

- PIDs in data production workflows
- PIDs in complex data citation
- PIDs and sensitive data

Scientific reproducibility & machine actionability

Research object type definitions for PIDs

Granularity, versioning, identifier syntax and relations

Focus on PIDs for data citation with different types of entities, including versioning, collections and hierarchies (LifeWatch, EMBL-EBI, INRAE)

Focus on specific needs regarding kernel metadata and related owner rights (UESSEX-UKDS, EMBL-EBI)

Topics to be addressed by all subtasks
Knowledge Exchange report: PIDS & related risks and trust elements

Josefine Nordling, Senior Open Science Specialist
CSC – IT Center for Science
30 November, 2023
Risks related to PIDs

- misunderstood value
- no community engagement
- lack of funding
- lack of commitment
- people dependencies
- centralised solutions
- western-orientation
- lack of strategic comms
- lack of support
- discontinued services
- commercial stakeholders
- contingency funds
- lack of human resources
- poor metadata
- no interoperability
- membership fees
- non-conformity with gdpr
- lack of uptake
- sustainability
- poor scalability
- poor management
Organisational change - Who is in control of the data? What about personal data transfer? - How to draw a consensus about the future of a given identifier? How to prevent increasing control from publishers? - If we want to build a reliable infrastructure, organisations need to figure out how to best involve all countries - Hard to argue for the importance of PIDs at a policy level.
Risks related to PIDs - Economic

- No stable funding to ensure sustainability – the biggest risk!
- Calculating the benefits for membership, difficult for smaller infrastructures and organisations to justify multiple memberships.
Risks related to PIDs - Social

- Social
  - Misunderstood value
  - No community engagement
  - Lack of funding
  - Lack of commitment
  - People dependencies
  - Lack of support
  - Lack of strategic comms

- Interoperability
  - Membership fees
  - Non-conformity with GDPR
  - Centralised solutions
  - Western-orientation

- Sustainability
  - Poor scalability
  - Poor management
  - Discontinued services

- Centralisation of PID infrastructure and people
  - Unrealiable data, lack of control and curation

Awareness-raising crucial for commitment and showcasing value for supporting PID usage and uptake – People infrastructure - community agreement – Centralisation of PID infrastructure and people – Unrealiable data, lack of control and curation
Risks related to PIDs - Technological

- misunderstood value
- no community engagement
- lack of funding
- lack of commitment
- people dependencies
- centralised solutions
- western-orientation
- lack of strategic comms
- lack of support
- discontinued services
- commercial stakeholders
- contingency funds
- lack of human resources

Risks concerning the metadata associated with a PID, such as quality, richness, completeness and risks around PID systems, especially their interoperability and scalability.
Figure 3. Framework of the analysed risk and trust variables

- Trusting in PID systems
  - Constant availability
  - Backup and fail-safe plans

- Services
  - Correctness and quality of metadata
  - Documentation of responsibilities
  - Certification
  - Security and authorization functions
  - Not for profit
  - Transient government

- Technology
  - Scalability
  - Open API

- Organizations
  - Transparency of scheme
  - Exit and succession strategies/contracts
  - Inclusion of community

- Validation of interlinking and metadata
- Sustainable business model
State-funded, more traditional information management organisations, such as libraries, most sustainably reliable - Long existence (relationship with and knowledge of their users) - Agility and marketing experience important - Independent organisations (user-oriented, collaborative and democratic approach) – Transparency (open documentation, close communication, democratic decision-making) – Inclusion - Integrity (transition plans and protections against commercial takeovers in place)

The technology part of PID infrastructure is seen as mostly reliable and trustworthy - PIDs have to be technically usable, usability has to be demonstrated through thorough use cases to communicate value and benefits – Metadata quality and standards – Robustness of PID technology, referred to in connection with long-term preservation of digital content

Trusting relationships between PID users and PID providers - General trust into PID providers, uncommon with fallback plans among PID users - Contracts, policies, risk management workflows are important - Contingency plans needed - Risk management plans increase reliability
Overall, there were very few doubts about the competence and ability of PID providers. Early and ideally favourable experiences with the PID are very important to establish trust!
Thank you!

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