

PIDs and repositories: experiences & practical issues of implementing DOIs for content

Perspectives from Strathclyde (Strathprints)

George Macgregor
University of Strathclyde
<https://purl.org/g3om4c>



The 'technical' bit: what are PIDs?

PID = *persistent identifier* 😊

- PIDs come in many flavours, e.g. DOI, ORCID, RAiD, ROR, etc. -- *and beyond!*
- Enables long-term identification but also 'actionability' because they are (*generally*) formed as URIs
- PIDs usually underpinned by some form of registry which registers new PIDs and resolves them (and contains metadata!) (e.g. DataCite, CrossRef)

Why have PIDs?

PID = *persistent identification* 😊

- Provides unique and persistent reference to an *entity* normally accessible over the Internet
- Maintenance of the scholarly record - citability, verification, reproducibility, replicability
- 'Reference rot' / 'link rot' - e.g [Klein, Van de Sompel et al.](#)
- Uniquely identify entities on the web; enables discovery these entities -- happy times!
- *Encode PIDs within metadata to link and grow the utility of data*

The 'PID graph'...



TIB – Leibniz Information Centre for Science and Technology - PID Service (CC-BY)



The screenshot displays the Neo4j Browser interface. On the left is a sidebar with navigation icons and information about Neo4j, including version 4.4.6 and copyright details. The main area shows a Cypher query: `neo4j$ match (n:researcher:orcid)-[r]-(m) return * limit 100`. Below the query, a graph visualization shows a network of nodes and relationships. The nodes are represented by circles, with a central green node labeled 'David Moore' and another central green node labeled 'Charles Verel'. Numerous other nodes are connected to these central nodes, representing various research topics and publications. The graph is rendered in a dark theme with light blue and green nodes and edges.

Research Graph Foundation (Australia), National (PID) Graph (CC-BY)

The 'soft' bit: things to remember about PIDs

- Technical issues exist *but* persistent identifiers really a 'soft' issue because...

PIDs only persistent:

- while the PID registration service *commits* to resolving them, and;
- while the organization / publisher *commits* to updating the registration service (registry)

Commitments, commitments, commitments ...

Strathprints DOI journey...

Using Strathprints as the case study (based on EPrints)

Strathprints connected to Current Research Information System (CRIS) - Pure

- [Repository and CRIS interoperability issues...](#) to be overcome
But much of what will be stated applies beyond EPrints and Pure....

What do you need to begin minting?

- [DataCite](#) membership (subscription)
- Single subscription but can be used for different DOI prefixes and/or different services
- e.g. [Strathprints](#) (research publications & other animals - grey lit, overlay journals), [STAX](#) (theses), [KnowledgeBase](#) (research data), etc.
- Piggyback off research data account...?

Strathprints DOI journey...

DataCite member since circa 2015

Membership used for:

- 2015: Research data (KnowledgeBase)
- **2017: Open grey scholarly literature (Strathprints) - manual DOI minting**
- **2018: Research publications (Strathprints) - semi-automated DOI minting**
- **2021: Research publications (Strathprints) - fully automated DOI minting**
- 2022: Research theses (STAX) - fully automated DOI minting

Example

'Eigenvalue decomposition of a parahermitian matrix : extraction of analytic Eigenvectors'

- Example: <https://doi.org/10.17868/strath.00085235>
- And to the [big registry](#) in the sky....

DataCiteDoi integration with EPrints

Using 3.4+: <https://github.com/eprintsug/DataCiteDoiUsing>

Using 3.3+: <https://bazaar.eprints.org/1100/>

DataCiteDoi Configuration

Relatively simple configuration, with basic functioning out-of-the-box

- Define which content types to receive DOIs
- Define which database field to use for local DOIs
- Define the appearance of your DOIs (looks are everything! 😊)
- Define some essential metadata elements, inc. how content types from EPrints schema map to DataCite schema

From configuration to *customization*...

- *Customization* required if using a CRIS with EPrints (our experience)

The screenshot shows a web interface with a navigation bar at the top containing tabs for 'Details', 'Actions', 'History', and 'Issues'. The 'Actions' tab is active. Below the navigation bar, there are several blue buttons with corresponding descriptions:

- New version**: Use this to submit a new version of this item. It will create an exact copy, make changes to. This item and the new version will be linked.
- Use as template**: Create a new item using this item as a template. There will be no connections to other items.
- Edit item**: Edit item
- Destroy item**: Remove this item from the system forever.

Below these actions is a grey bar labeled 'Editorial Actions'. Underneath, there are more blue buttons:

- Move to Review**: Move item to review.
- Delete Item**: Removes the item from the archive.
- Reindex Item**: Queue this item for re-indexing.
- Change Owner**: Change the owner of this eprint
- Coin DOI**: This will add a coin-DOI event to the event queue (i.e. Will not happen immediately)
- Archivematica Transfer**: Archive this item via Archivematica

At the bottom of the interface is another grey bar labeled 'Export'. The 'Coin DOI' button is circled in green.

```
<rioxx2_project_input>  
<item>  
  <project>EP/S000631/1</project>  
  <funder_name>EPSRC (Engineering and Physical Sciences  
Research Council)</funder_name>  
</item>  
</rioxx2_project_input>
```

Eigenvalue decomposition of a parahermitian matrix : extraction of analytic Eigenvectors (eprint 85235 r8)

Modified by indexer at 21 April 2023 10:21:11 UTC

Before	After
	<deposit_pid>10.17868/strath.00085235</deposit_pid>
	<date>2023-04-18</date>
<rev_number>7</rev_number>	<rev_number>8</rev_number>
<creators> <item> <name> <family>Weiss</family> <given>Stephan</given> </name>	<creators> <item> <name> <family>Weiss</family> <given>Stephan</given> </name>

Example... walk the walk 😊

'PIDs and repositories: experiences & practical issues of implementing DOIs for content : perspectives from Strathclyde (Strathprints)'

Deposited here:

Let's mint now! 🍷

Things to ponder...when to mint?

Auto minting?

- [ZORA](#) - Zurich Open Repository Archive (University of Zurich)
- Example: <https://doi.org/10.5167/uzh-224344>

Scaling Native Language Identification with Transformer Adapters

Uluslu, Ahmet; Schneider, Gerold (2022). *Scaling Native Language Identification with Transformer Adapters*. In: 5th International Conference on Natural Language and Speech Processing (ICNLSP), Trento, 16 December 2022 - 17 December 2022, Cornell University. Copy

Abstract

Native language identification (NLI) is the task of automatically identifying the native language (L1) of an individual based on their language production in a learned language. It is useful for a variety of purposes including marketing, security and educational applications. NLI is usually framed as a multi-label classification task, where numerous designed features are combined to achieve state-of-the-art results. Recently deep generative approach based on transformer decoders (GPT-2) outperformed its counterparts and achieved the best results on the NLI benchmark datasets. We investigate this approach to determine the practical implications compared to traditional state-of-the-art NLI systems. We introduce transformer adapters to address memory limitations and improve training/inference speed to scale NLI applications for production.

[Find similar titles](#)

Citations Altmetrics

Google Scholar™



[View details on Altmetric's website](#)

Tweeted by 6

9 downloads since deposited on 04 Jan 2023

9 downloads since 12 months

[Detailed statistics](#)

Downloads

Additional indexing

Permanent URL: <https://doi.org/10.5167/uzh-224344> Copy

Green Open Access

Download



Content: Accepted Version
Language: English
Licence:

[Download PDF \(266kB\)](#)

Links

[View at publisher](#)

AKABER Export

Authors, Affiliations, Collaborations

Uluslu, Ahmet Schneider, Gerold

Source of ORCID ID: Author

...pondering continued

or rules-based minting?

- EPrints DataCite integration enables automatic but conditional DOI minting
- i.e. specific content types

...pondering extended

or discretionary minting?

- Manually initiated by repository team
- e.g. for RRS/UKRI compliance, overlay journal publication, grey literature, etc.

AAMs, versioning, & other animals

Versioning: Agustina and Alexia to say more 😊

- Support for DOI versioning in EPrints (?) - what's a minor or major change?
- Ensuring richest data sent to DataCite registry upon minting
- Preference for waiting for essential metadata to ensure relational linking

Ouch! Some pain points...

'Variable quality' (ahem!) project data causes DOI registration issues

- Local project data capture too removed from thinking globally
- Zero support for URIs in Pure (e.g. funders, grants)

Project data now suppressed in DataCite XML sent to DataCite by Strathprints

- Manual intervention needed -- either at Fabrica (DataCite) or directly on Strathprints prior to initiating minting
-issue obviated if we were using EPrints in isolation... 👍

Ouch! More suffering...

Integrations with CRIS software:

- Specific technical issues to be overcome in with EPrints/Pure ([learn more](#))
- Need to 'write protect' database fields in EPrints

Other PIDs and enriching the PID graph

- Imposing relational links challenging when Strathprints a slave system to Pure
- See above RE project data 😊
- Manual intervention at Strathprints side but need to avoid double-keying (and over writing)

Minting, but not as we know it, Jim...

Important to remember

- There are always options available to comply with UKRI / Plan S, even with DOIs.

DataCite, for example:

- Register and mint your DOIs manually, using DataCite Fabrica UI
- Export DataCite compliant XML from your repo, and upload to registry using DataCite Fabrica UI

Think out of the box! Use a parallel solution!

- Grab a PID from an open system that supports them, e.g. Zenodo
Or, consider [CORE OAI ID](#), *potential* distributed alternative

Useful links

Some useful links you may have missed in the presentation:

- [DataCite](#)
- [DataCite Fabrica](#)
- [Jisc Persistent Identifiers](#)
- [CORE OAI ID Resolver](#)

DataCiteDoi integration (EPrints)

- [For 3.4+](#)
- [For 3.3+](#)

Questions.....?!

If you think of questions later... contact me here: <https://purl.org/g3om4c>