EUNIS VIGO 2023

The Role of Universities in the Implementation of PIDs

University of Strathclyde Glasgow

Pablo de Castro
euroCRIS Secretary

Open Access Advocacy Librarian
University of Strathclyde
pablo.de-castro@strath.ac.uk
http://orcid.org/0000-0001-6300-1033
@pcastromartin

The PID Graph

- Most recent (Aug 2020) detailed PID Graph snapshot at hand...
- ... but things have evolved quite a bit ever since: RORs/Ringgold IDs, DOI-based grant IDs, PIDINSTs, IGSNs, RAiDs...
- Some feeling of “building the plane as we fly it” – is it possible to keep track of this evolving snapshot?

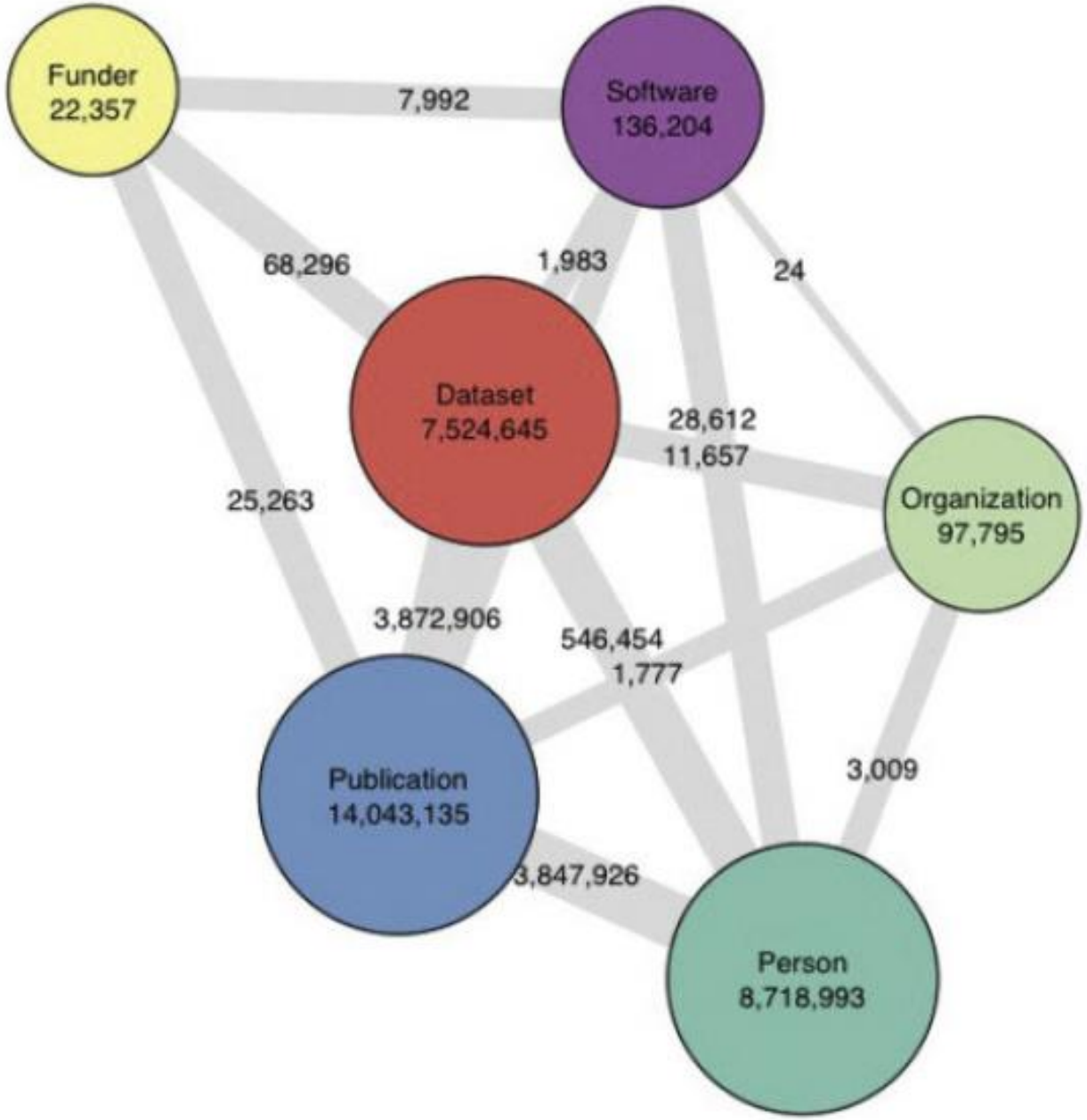


Figure: PID graph as described by FREYA in Aug 2020 – with a strong focus on research outputs

https://doi.org/10.1016/j.patter.2020.100180

The “Building the Plane” report – and associated case studies

- “... to identify, through investigation, analysis and recommendations, the best possible **strategic and operational paths** to achieve a well-functioning PID infra for KE member states and beyond”

- “... to identify the **main risks** when pursuing a well-functioning PID infrastructure for research, and to better understand the most important elements of **trust** in creating said infrastructure”

https://zenodo.org/record/7258286

The “Building the Plane” report – and associated case studies

Report:

Building the plane as we fly it: the promise of Persistent Identifiers, https://zenodo.org/record/7258286

Case studies:

- Adoption of the DAI in the Netherlands and subsequent superseding by ORCID/ISNI, https://zenodo.org/record/7327505
- The gradual implementation of organisational identifiers (OrgIDs), https://zenodo.org/record/7327535
- PIDs for research instruments and facilities: an emerging PID domain in need of coordination, https://zenodo.org/record/7330372
- IGSN – building and expanding a community-driven PID system, https://zenodo.org/record/7330498
- RePEc Author Service: An established community-driven PID, https://zenodo.org/record/7330516
- Failed PIDs and unreliable PID implementations, https://zenodo.org/record/7330527
- The role of research funders in the consolidation of the PID landscape, https://zenodo.org/record/7258210

The team behind the study

- Pablo de Castro

  Physicist. Open Access Advocacy Librarian at the University of Strathclyde in Glasgow since Jan 2017. Technical Secretary of the Dutch non-profit association euroCRIS since Jan 2018. Former OpenAIRE project officer. Member of the EOSC Association Task Force for PID Policy and Implementation. Associate of scidecode science consulting.

- Ulrich Herb

  Sociologist & Information Scientist, since 2001 Open Access expert/ project manager/ head of the Publication and Research Support Department at Saarland University, board member of the learned society for Information Science in the German-Speaking countries. Associate of scidecode science consulting.

- Laura Rothfritz

  Research assistant and PhD candidate at the Berlin School of Library and Information Science at Humboldt University Berlin. Associate of scidecode science consulting.

- Joachim Schöpfel

  Professor for Information Science at the University of Lille and independent consultant.

Plus the KE PID T&F Group: https://www.knowledge-exchange.info/event/pids-risk-and-trust

The study

- **Analysis** of the current state of the PID landscape in the six Knowledge Exchange partner countries with a focus on the e-infrastructure for *the currently available PID entities* (*eg* researchers, institutions, outputs, etc) and *new PIDs* (*eg* conferences, research equipment, facilities)

- Data collection via **literature study** & **expert interviews**

- These fed into:

    - the production of the seven **case studies** highlighting issues of risk and trust in the PID infrastructure, and

    - the formulation of **recommendations** for good practice and on the best possible strategic and operational paths to achieve a well-functioning PID infrastructure

The interviewees

Mathias	Astell	Hindawi	GBR	PID Manager
David	Aymonin	ABES	FRA	PID Authority
Geoffrey	Bilder	CrossRef	GBR	PID Service Provider
Matt	Buys	DataCite	GBR	PID Service Provider
Maria	Cruz	NWO	NL	PID Manager
John	Doove	SURF	NL	PID User
Nathalie	Fargier	CNRS	FRA	PID Owner
Martin	Fenner	formerly Technical Director at DataCite, involved in the FREYA project	GER	PID Manager
Stephanie	Hageman-Wilholt	TIB Hannover/ConfIDent	GER	PID Authority
Juha	Hakala	URN representative, National Library of Finland	FIN	PID Service Provider
Lars	Holm Nielsen	Zenodo	CHE	PID Owner
Karen	Hytteballe Ibanez	DTU - Technical University of Denmark	DNK	PID User
Jens	Klump	IGSN	GER	PID Service Provider
Rachael	Lammey	CrossRef	GBR	PID Service Provider
Dan	Smith	Wellcome Trust	GBR	PID Owner
Mark	van de Sanden	SURF, systems architect	NL	PID Authority
Herbert	Van de Sompel	DANS	NL	PID User
Peter	Verhaar	Leiden University	NL	PID Owner

Some (selective) findings

- PIDs are considered **socio-technical** infrastructures. Trust in organisations or individuals seems to be more important for the acceptance of PIDs than the technology used, as the risks associated with the technology are considered amorphous
- Predominantly mentioned: **well-established PIDs** such as DOI, ORCID and ROR, to a lesser extent emerging PIDs (funder and grant IDs, RAiDs, ConfIDs), standards like URN and schemes like ARK
- Main benefits: **Interoperability, value-added services, availability/interconnectivity** of rich metadata
- Dichotomy of **'technical'** (bottom-up, researcher driven) and **'admin-oriented' PIDs** (top-down, uptake driven by institutions, publishers and research funders)
- **Open source** and **open data** are a key feature for trust and reliability
- Establishing a **community of PID** users is a key factor for success and trustworthiness
- The **implementation of PIDs** requires a strategic analysis
- Significant **PID landscape fragmentation** and **competing initiatives**, though this is not necessarily seen as a major issue

Rationale for Grant ID adoption by funders

Why should a research funder like the Wellcome Trust, the NWO or the DFG with its own internal and functional ID adopt a new, global ID?

Even if a specific funder has its own well-maintained identification system for its grants, there are three main factors that justify adopting DOI-based, Crossref-issued grant IDs:

I. With so many funders out there there is always a risk of duplicate grant numbers across funders. Even if the combination of funder name + full grant number will typically be unique, this is potentially problematic and easy to fix via persistent grant IDs
II. Authors often make mistakes when including the full grant numbers for their projects in their manuscript acknowledgements. The more complex the grant number structure, the more frequent such typos are
III. Self-managed grant numbers issued by funders do not allow the full power of the PID Graph to be exploited. Once DOI-based grant IDs start regularly featuring in published papers that can be shared in the article metadata by publishers, it will be simple to link publications and datasets to a funded project in an automated, machine-readable way. This is already theoretically possible on the basis of self-managed project grant numbers issued by funders, but it's much more difficult because there are hardly any links between internal grant numbers and the landing pages containing the project information.

- Risk of duplicate full grant numbers across funders

Grant finder

Europe PMC funder grants | COVID-19 grants

Europe PMC funder grants search results

Export

Grant ID: 207467

Search again

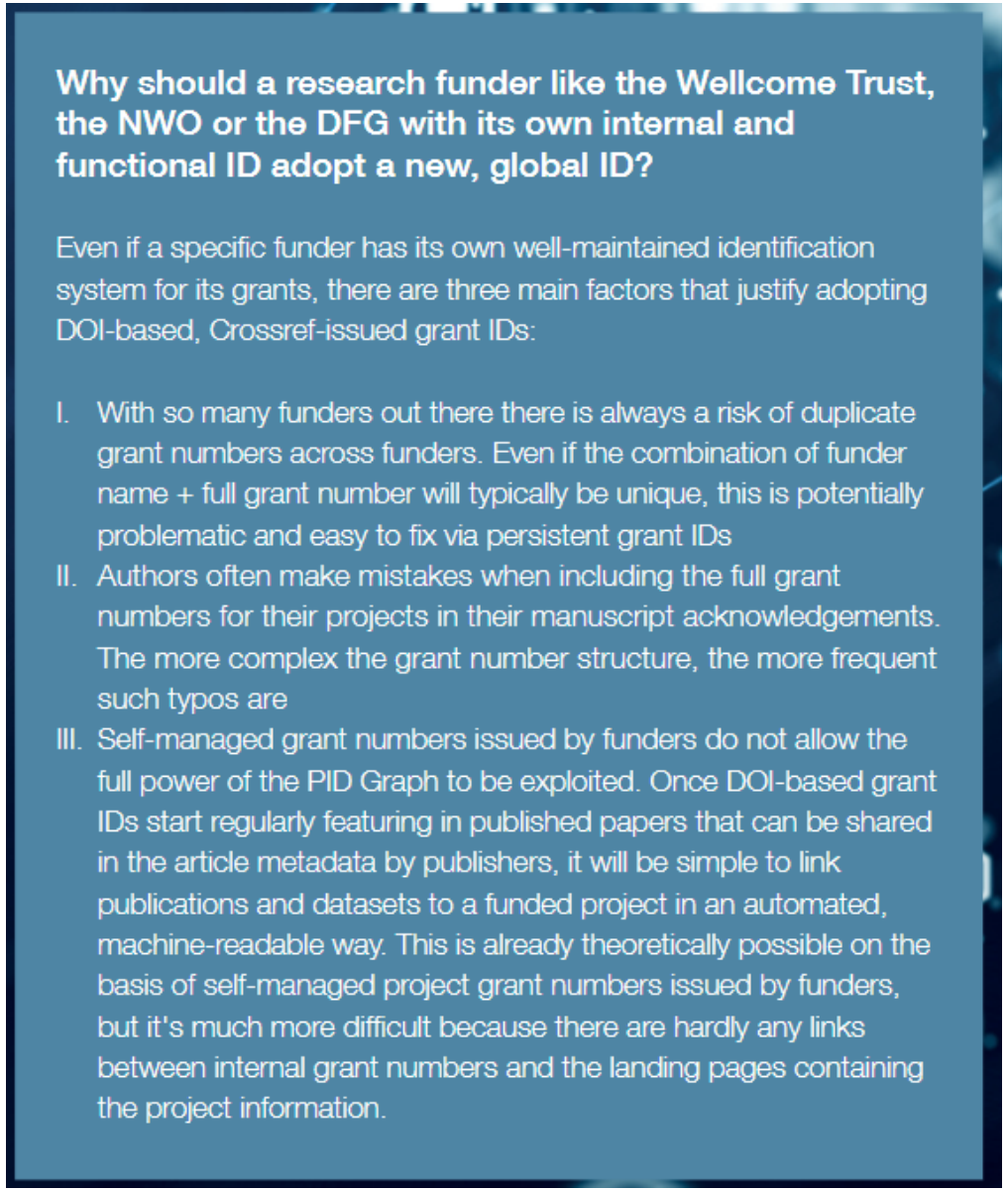
1-2 of 2 results

Deep Earth Chemistry of the Core (207467)
Dr Badro, J, INSTITUT DE PHYSIQUE DU GLOBE DE PARIS, European Research Council 2008-2013 Expired
8 publications | 6 free full text articles

Connecting causes and immune consequences of infection-induced metabolic change (207467)
Professor Dionne, Marc, Imperial College London, Wellcome Trust 2018-2023
7 publications | 6 free full text articles

1-2 of 2 results

Rationale for Grant ID adoption by funders



- Risk of duplicate full grant numbers across funders
- Typos frequently made by authors when acknowledging funding in their manuscripts
- 'Harnessing the power of the PID Graph': machine-readable grant IDs will allow easy interlinking across entities (esp btw grants *and* publications, datasets and other research results)

The (Currently) Fragmented PID Landscape

1. ‘Technical’ vs Admin-Oriented PIDs

- Technical PIDs: Researcher-led, bottom-up implementation, little if any direct involvement from research funders, research-perfoming organisations (RPOs) or national offices
  Examples: PIDs for instruments and facilities, IGSNs, ISRCTNs (International Standard Randomised Controlled Trial Numbers), Biomedical IDs (GenBank etc)

- Admin-Oriented PIDs: led (generally in a top-down fashion) by national offices, RPOs, publishers and (some) funders, all of whom directly reap the rewards. Little researcher involvement*, even awareness (as a rule)
  Examples: DOIs for publications/datasets, ORCIDs, RORs, Grant IDs, RAiDs, etc

** Researchers often see these as unwanted additional bureaucracy*

HORIZON 2020

Stardust Reloaded

Project description

DE EN ES FR IT PL

Training young researchers in cutting-edge technologies for a safer space

As satellites get smaller and easier to build, companies around the world are launching vast numbers of them into low Earth orbit. However, the amount of space debris becomes a crucial problem for maintaining space safety. The implications are significant: would an asteroid impact have devastating consequences for our planet? Funded by the Marie Skłodowska-Curie Actions programme, the Stardust-R project will train a new generation of scientists and engineers in enabling technologies and effective solutions to critical problems in planetary safety, minor body exploration and the sustainable use of space. Fifteen early-stage researchers will receive interdisciplinary training, which will cover mathematics, physics, computer science and aerospace engineering.

Show the project objective

Project Information

Stardust-R
Grant agreement ID: 813644

DOI
10.3030/813644

Start date 1 January 2019
End date 30 June 2023

Funded under
EXCELLENT SCIENCE - Marie Skłodowska-Curie Actions

Total cost
€ 3 867 284,52

EU contribution
€ 3 867 284,52

https://scidecode.com/2023/05/22/at-present-wikidata-id-is-the-most-suitable-id-for-organisations-in-terms-of-coverage/

Pavel Zbornik: „At present Wikidata ID is the most suitable ID for organisations in terms of coverage“

(interview by Pablo de Castro)

[with additional input by Stephane Ndong at the EC RTD]

What led the EC to take this step? What main use cases do you see for grant IDs?

There was an internal discussion on the use of grant IDs which resulted in an agreement that the EC should issue grant IDs for Horizon Europe. The overall purpose for the Commission is to contribute to the general development of PIDs in a mid- and long-term timescale. The idea underpinning this effort is that the availability of PIDs will become an important factor in the steering and evaluation of R&I Policies.

One major use case for grant IDs is linked to the evaluation and monitoring of the impact of the framework programme. The expectation is that better links between grants and related publications – plus other research outputs such as patents, prototypes, software components, etc – will simplify the analysis when the link is declared via a PID instead of as free-form text.

https://scidecode.com/2023/05/22/at-present-wikidata-id-is-the-most-suitable-id-for-organisations-in-terms-of-coverage/

The (Currently) Fragmented PID Landscape

2. Competing Technical Solutions

- OrgIDs: ROR vs Ringgold, record maintenance and the issue of multiple-level Org IDs

Copyright Clearance Center

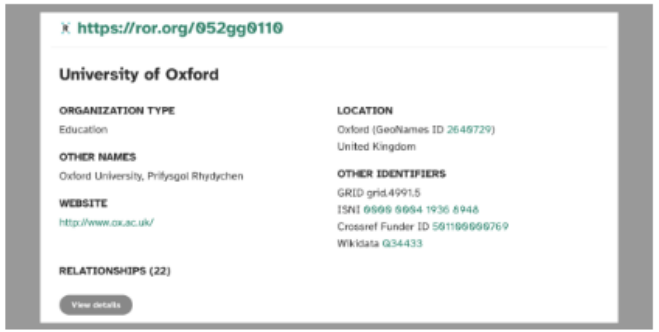
Home › Press Releases › CCC Announces Acquisition of Ringgold, Leading Provider of Organization Identifiers in Scholarly Communications

Press Releases

CCC Announces Acquisition of Ringgold, Leading Provider of Organization Identifiers in Scholarly Communications

3 May 2022 – Danvers, Mass.

Acquisition Reflects CCC's Ongoing Commitment to Promoting Interoperability, Addressing Market Friction, and Collaborating with Stakeholders

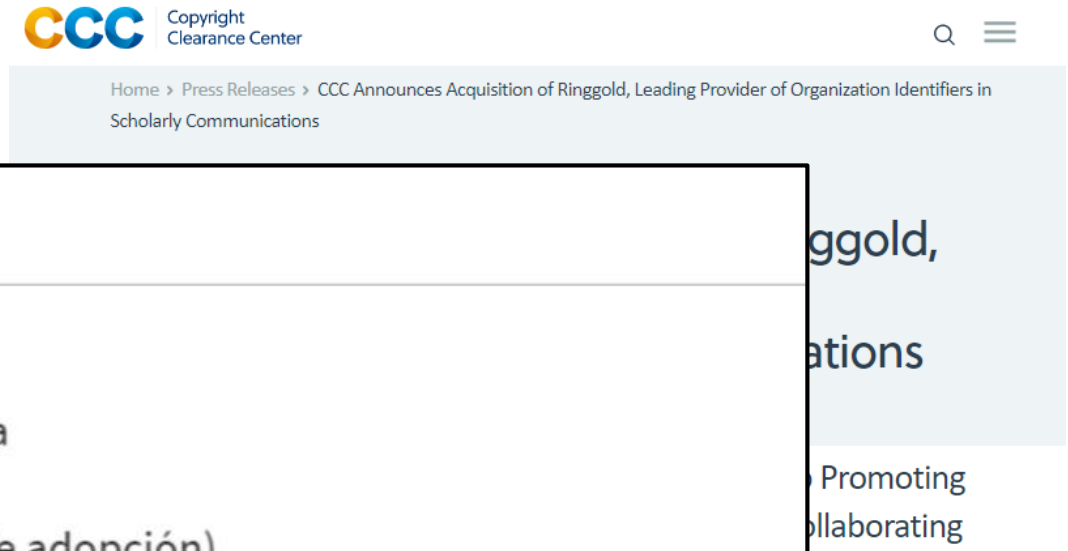


Use the registry

The Research Organization Registry (ROR) includes IDs and metadata for more than 102,000 organizations and counting. Registry data is CC0 and openly available via a search interface, REST API, and data dump. Registry updates are curated through a community process and released on a rolling basis.

The (Currently) Fragmented PID Landscape

2. Competing Technical Solutions



¿CÓMO ADOPTAR ARK-CAICYT?

Etapa inicial: primeros adoptantes

- Disponer de un ISSN en línea provisto por la agencia argentina
- Publicaciones con URL únicas por recurso o parte
- Dos años de antigüedad en su versión en línea (etapa inicial de adopción)
- Adhesión a políticas de Acceso Abierto (Acceso inmediato, sin requerimientos de registro, suscripción o pago a los recursos)
- Nombrar un editor responsable para la gestión de su ARK-CAICYT

metadata for more than 102,000 organizations and counting. Registry data is CC0 and openly available via a search interface, REST API, and data dump. Registry updates are curated through a community process and released on a rolling basis.

The (Currently) Fragmented PID Landscape

3. ‘Community’: an ambiguous concept

- PIDs and services associated with them need to be perceived as valuable and be in turn promoted by "the community"

2. Crossref grant IDs minted by research funders

Stakeholder	Role
Crossref	PID service provider: Crossref assigns a stack of DOIs (via a funder ID praefix) to research funders and guarantees PID persistence and the correct resolving*
Research funder	PID manager: funders join the Crossref funder advisory group and gather the expertise to start minting grant IDs for their funded projects**
HEI	PID user: institutions store the grant IDs in the metadata set for funded projects they keep in their CRIS systems*** Grant IDs are included as a part of the RAiDs HEIs mint***
Researchers	PID user: prompted by their funders and HEIs, researchers include the grant IDs in the acknowledgements section of their manuscripts***
Publisher	PID user: publishers allow these grant IDs to be provided on the manuscript submission systems and include them in the metadata set exported to Crossref – allowing the references to be picked in for instance individual ORCID profiles***

PID community stakeholders

- **Governing bodies**
- **PID Service Providers**
- **(Possible) PID Federation**
- **RDA Working Groups**
- **PID-related projects and initiatives**
- **International coordination bodies**
- **Publishers**
- **National offices**
- **National research and education network (NRENs)**
- **Research funders**
- **HEIs/research centres/institutes**
- **Researchers**
- **Start-ups**

Some recommendations for RPOs

1. **Make sure you are represented in** – or at least informed about – national-level coordination initiatives

2. **Consider** the possibility of drafting an **institutional PID policy**

3. **Raise awareness** of the existing and emerging PID landscape among institutional researchers, including prompting them to use the appropriate ones

4. **Be aware of your key role** in the implementation of specific, admin-oriented PIDs

5. **Include as many PIDs as possible** in your research information management systems such as institutional repositories and CRIS systems (plus any other institutional system that feeds these)

6. **Be aware of technical PIDs** directly emerging from researcher communities in a bottom-up fashion

7. **Stay informed about (still to come) mechanisms** to issue (and share and use) institutional PIDs such as RAiDs or PIDINSTs

EUNIS VIGO 2023

Thanks!

Questions?

University of Strathclyde Glasgow

Pablo de Castro
euroCRIS Secretary

Open Access Advocacy Librarian
University of Strathclyde
pablo.de-castro@strath.ac.uk
http://orcid.org/0000-0001-6300-1033
@pcastromartin