

# The Role of Universities in the Implementation of Persistent Identifiers (PIDs)

Pablo de Castro<sup>1,5,6\*</sup>, Ulrich Herb<sup>2,6</sup>, Laura Rothfritz<sup>3,6</sup> and Joachim Schöpfel<sup>4,5</sup>

<sup>1</sup> University of Strathclyde, United Kingdom

<sup>2</sup> Saarland University, Germany

<sup>3</sup> Humboldt University Berlin, Germany

<sup>4</sup> University of Lille, France

<sup>5</sup> euroCRIS, The Netherlands

<sup>6</sup> scidecode science consulting, Germany

[pablo.de-castro@strath.ac.uk](mailto:pablo.de-castro@strath.ac.uk), [u.herb@sulb.uni-saarland.de](mailto:u.herb@sulb.uni-saarland.de), [laura.rothfritz@hu-berlin.de](mailto:laura.rothfritz@hu-berlin.de),  
[joachim.schoepfel@univ-lille.fr](mailto:joachim.schoepfel@univ-lille.fr)


## Abstract

The report “Building the plane as we fly it: the promise of persistent identifiers” was published last February by the Knowledge Exchange. The report explores the challenges, opportunities, risks and trust-related issues associated with the quickly-developing PID landscape with an emphasis on the six KE member countries (Denmark, Finland, France, Germany, Netherlands and the United Kingdom). A series of seven complementary case studies were published alongside this report examining more in depth the current PID landscape for specific entities such as authors, organisations, instruments and facilities, grants and projects or physical samples. The report summarises the findings of a series of interviews with PID experts that provided the basis for the study. Sets of recommendations are also provided for a range of relevant stakeholders in the PID implementation domain. The current fragmentation of the PID landscape is one of the main challenges highlighted in the report and the case studies. This contribution provides a summary of the findings of this study and analyses this fragmented PID landscape in some detail, specifically exploring the meaning of “community” in expressions like “a community-driven PID landscape”. The role universities are expected to play in the gradual, unstoppable adoption of a wide range of PIDs is laid out, together with some early best practices in the domain.

---

\*  Pablo de Castro, <http://orcid.org/0000-0001-6300-1033>

 Ulrich Herb, <https://orcid.org/0000-0002-3500-3119>

 Laura Rothfritz, <https://orcid.org/0000-0001-7525-0635>

 Joachim Schöpfel, <https://orcid.org/0000-0002-4000-807X>

# 1 Introduction

A working paper called “Risks and trust in pursuit of a well-functioning Persistent Identifier infrastructure for research” was published in 2021 by the so-called Task & Finish Group for PIDs Risk and Trust of the Knowledge Exchange (KE) [Belsø, R., 2021]. This T&F Group is formed by a set of experts in scholarly communications and research information management [Knowledge Exchange, 2021] from the six KE member countries, including the six organisations that constitute the Knowledge Exchange network: DeIC in Denmark, CNRS in France, CSC in Finland, the German Research Foundation (DFG) in Germany, SURF in the Netherlands and Jisc in the UK. The working paper was aimed at identifying, through investigation, analysis and recommendations, the best possible strategic and operational paths to achieve a well-functioning PID infrastructure for Knowledge Exchange (KE) member states and beyond. The paper defines persistent identifiers (PIDs) as “a sequence of characters that uniquely denotes a referent. This sequence is deemed persistent when the identifier, its binding to the referent and the related metadata survives over time and technical evolutions”.

A call for proposals followed in June 2021 for a team of consultants to more deeply explore these risks and trust-related issues against the current, swiftly-evolving PID landscape. The study was awarded to the four authors of this paper, who worked as a team under the banner of scidecode science consulting. 18 months after the start of the project in Sep 2021 the report “Building the plane as we fly it: the promise of persistent identifiers” [De Castro, P. et al, 2023] was published together with seven case studies describing the PID landscape for a number of entities (titles and URLs for the case studies and the report are shown in the text box below). This work built on a number of interviews with PID experts in the various PID roles identified by the original T&F Group working paper – such as PID end-users, owners, managers, service providers and authorities – and a comprehensive literature study to identify the current status of the many parallel PID initiatives. Regular discussions were also held with the T&F Group members, including a 2-day in-person workshop held at SURF in Utrecht in Oct 2022.

## **Study: Risks and trust in pursuit of a well-functioning PID infrastructure for research**

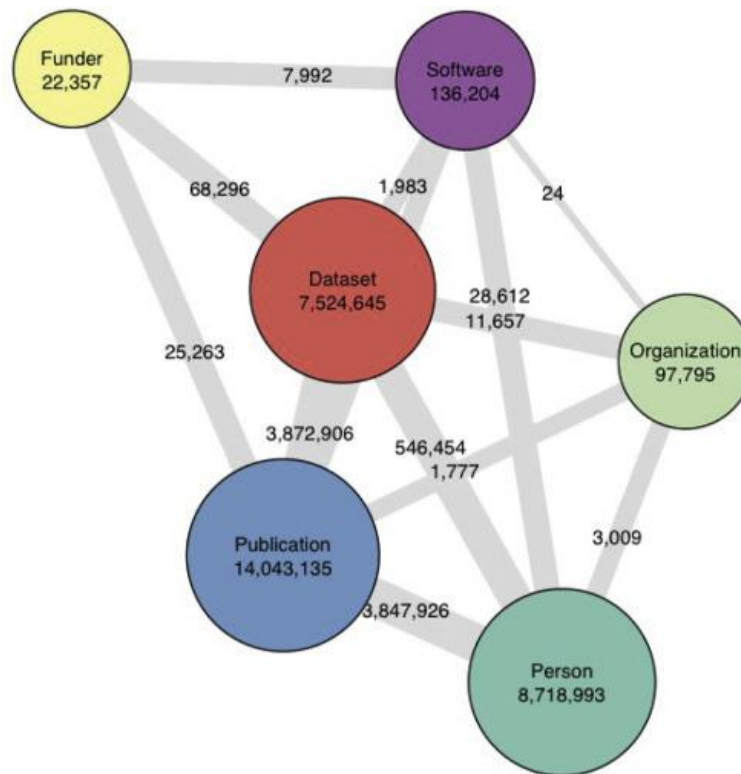
### **Report:**

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>
- Persistent identifiers 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 “Building the plane as we fly it” report emphasises the value of PIDs for an optimal management of research information in the scholarly communications domain and specifically for the implementation of Open Science [Bécquet G., 2022]. The analysis of the numerous risk- and trust-related issues identified in the way the PID infrastructure is currently being developed is conducted against an examination of the present state of the PID landscape. In this sense, the report and its associated case studies aim to provide an update on the detailed analysis of the evolving PID landscape carried out by the Dec 2017-Nov 2020 EU-funded FREYA project [Cousijn, H. et al, 2021]. A key element in this regard is the PID Graph concept, which displays a network of interconnected entities – such as authors, organisations, research outputs (publications, datasets, software), projects, grants, instruments, etc – which are increasingly being assigned machine-readable persistent identifiers by different initiatives and organisations. The PID Graph has kept expanding and deepening in the past few years and although it’s fair to say that it still remains in its early stages, it is already possible to navigate it and to grasp its enormous potential as it slowly consolidates.



**Figure 1:** PID graph as described by FREYA in Aug 2020 – with a strong focus on research outputs

Some examples are already available out there for what the report calls “the promise of PIDs”. One of these is the PID-optimised research cycle released by the MoreBrains cooperative [MoreBrains, 2022] with its analysis of 10 specific workflows and stakeholders for PID application within the whole project life cycle. Another very recent example is the analysis conducted by the BMBF-funded TAPIR project led by the TIB Hannover in Germany to explore various partially automated PID-based reporting workflows [TIB Hannover, 2022].

## 2 Some key findings from the PID study: a summary

The fact-finding exercise that provided the basis for the KE report and case studies was based on sixteen interviews conducted by the scidecode team of consultants with PID experts from a wide range of positions, organisations and countries. These included among others research funders, universities, national libraries, repository managers and publishers besides representatives from PID service providers like Crossref, DataCite and others. A list of interviewees is available in an annex to the report.

Some key findings arising from these interviews are listed below:

- Well-established PIDs such as DOI, ORCID and ROR were predominantly mentioned by interviewees, although many other emerging PIDs came up in the discussions – including among others the likes of funder and grant IDs, RAiDs, PIDINSTs, ConfIDs, standards like URNs and schemes like ARK;
- The most frequently mentioned main benefits of PIDs included interoperability, value-added services built on top of them and availability and interconnectedness of rich metadata [Car, NJ., Golodoniuc, P., Klump, J., 2017];
- A dichotomy was identified between ‘technical’ (mostly bottom-up, researcher-driven) and ‘admin-oriented’ PIDs (top-down implementation, uptake driven by institutions, publishers and research funders);
- Open source solutions and open data (“forkability”) were highlighted as a key feature for trust and reliability [Bilder, G., Lin, J., Neylon, C., 2015];
- PIDs are a socio-technical infrastructure and establishing a community of PID users (the ‘social’ aspect) is a key factor for success and trustworthiness;
- Trust in organisations and/or individuals is generally perceived by experts as more important for the acceptance of PIDs than the specific technological solutions used, as the risks associated with the technology are considered to be loosely defined;
- Experts also consider there is a general perception that PIDs can simply not fail once they have reached a certain level of adoption
- The implementation of PIDs requires a strategic analysis: while (for instance) funders are increasingly requiring the use of specific PIDs within their research information management workflows, new PID initiatives need to reach a degree of technical and social maturity before they join the list of PIDs worth being endorsed in a widespread fashion

## 3 Recommendations by stakeholders and the role of universities

One of the key outcomes of the KE report is a series of recommendations for a harmonised PID implementation classified by stakeholder. The recommendations have been structured in such a way that allows the key players in the PID landscape to be identified, including national-level stakeholders (such as the six national organisations that make up the Knowledge Exchange), research funders, PID providers, institutions, researchers, publishers (including Diamond OA publishers), a possible PID Federation and the Knowledge Exchange itself. An attempt has been made to identify all relevant

stakeholders starting with those who may be able to provide some governance and then moving downwards on the scale of PID implementers and users.

Because the EUNIS annual congress is expected to offer an opportunity to reach out to universities, this contribution focuses on the recommendations to research-performing organisations (RPOs) listed below. While pilots are being designed in some KE member countries for universities to explore the use cases and implementation workflows for ‘new’ PIDs like Research Activity IDs (RAiDs) or PIDs for research instruments and facilities, RPOs have traditionally had a low presence in events specifically focused on PIDs such as the – unfortunately discontinued – PIDapalooza series [Meadows, A., 2021]. It is encouraging in this regard that universities in all six KE member countries are well represented in the national-level ORCID consortia that are frequently being used as a venue for expanding the discussion on PID adoption beyond persistent identifiers for authors.

#### **Recommendations for institutions/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.

Some early requirements are emerging for institutions to assign PIDs to the entities they are in charge of – such as researchers, publications and datasets. One of these is the updated, Plan S-aligned Open Access policy by the UK Research and Innovation (UKRI) that was released in Aug 2021 [UKRI, 2021]. Several references to PIDs are included among the technical requirements in the policy for (i) journals and publishing platforms and for (ii) institutional and subject repositories. The wording for the latter clauses is shown below.

[UKRI Open Access policy] technical requirements for institutional and subject repositories

*a. PIDs for research outputs must be implemented according to international recognised standards, examples of international standards include DOI, URN or Handle.*

*d. Common unique PIDs for research management information (for example identifiers for funders and/or organisations) are strongly encouraged; ORCID, the researcher identifier, must be supported.*

A particularly interesting development is the issuing of PIDs – either DOIs, URNs or handle IDs – the research funder requires for Accepted Author Manuscripts or AAMs. Bearing in mind that these AAMs will eventually become Version of Record (VoR) papers with their own DOI issued by publishers, it becomes necessary to figure out a technical workflow that allows the PIDs issued for the various versions of the same publication (including pre-prints where available) to be automatically linked to each other. Although the PID graph may not yet be sufficiently developed at this point to allow this automated interconnection to happen, this is part of the early scoping work the above-mentioned TAPIR project has been looking into, and it's already clear that we will be able to get there eventually.

In the meantime, internal coordination and trouble-shooting dissemination activities are already being held in the UK – where this Plan S-aligned UKRI policy for the deposit of AAMs under no embargo period and a CC BY licence has been implemented for some time already – so that universities are able to discuss the most appropriate technical ways forward [UKCoRR, 2023]. The ultimate goal is for all these institutions affected by the Open Access policy to apply harmonised workflows for its implementation in the area of PIDs – and these will eventually also involve the use of RORs for organisations and funders and RAiDs for projects.

The drafting and issuing of institutional PID policies by institutions is another area – as reflected in recommendation no 2 for RPOs above – where significant progress is expected to happen in a short- and mid-term horizon. These policies should describe the relevant PIDs that are expected to be applied by institutional researchers and professional services together with the associated workflows for their implementation and use. At the time the 'Building the plane' report was written there was only one such policy available at the British Library [Madden, 2021]. This was mainly an outcome of the BL being a partner in the FREYA project and the national DataCite node for the United Kingdom. At universities, elements of a possible PID policy such as the guidance around the use of ORCID by academics currently tend to be scattered across other policy documents but as the number of PIDs in use steadily increases it would eventually make sense to bring together all the guidance into a single, specific PID policy document.

## 4 A fragmented PID landscape

One of the most evident risks in the current development of the PID implementation landscape identified in the report is the fragmentation of such landscape. There are many parallel initiatives being carried out by different organisations in different countries that are not always talking to each other, and the report identifies an acute need to define coordination mechanisms and forums for a global conversation to happen. In the current absence of events specifically devoted to hosting a wide-range discussion on PIDs, the conversation is taking place in various forums not necessarily connected to each other. Fortunately there are many of these forums – the Research Data Alliance (RDA) events and EOSC-funded projects like FAIRCORE4EOSC may be mentioned as examples for relatively recent entrants hosting specific discussions on PIDs – and there is also a certain degree of overlap among the stakeholders taking part in them.

Three main aspects may be highlighted in which this PID landscape fragmentation most evidently manifests itself. The first one is the already mentioned dichotomy between 'technical' and 'admin-oriented' PIDs. A second one is the wide range of competing technical solutions and standards. Finally, another key aspect of this fragmentation is the long list of stakeholders that are part of the concept of community in terms like "a community-driven PID implementation process".

## 4.1 ‘Technical’ vs ‘admin-oriented’ PIDs

Two broad PID categories have been defined in the KE report to address the similarities in the development of specific PIDs: the ‘technical’ and the ‘admin-oriented’ PIDs. Technical PIDs include identifiers like PIDs for instruments and facilities, IGSNs for physical samples, ISRCTNs for clinical trials or a wide range of ‘biomedical IDs’ such as GenBank IDs. The implementation of all these ‘technical’ identifiers tends to be directly led by researchers mostly through bottom-up workflows with little – if any – involvement from the likes of research funders, research performing organisations (including universities), national offices or publishers.

On the other hand, the ‘admin-oriented’ PIDs include the most consolidated identifiers such as DOIs for publications and datasets, ORCIDs, RORs plus emerging ones like DOI-based grant IDs, RAiDs, ConfIDs for conferences and events, etc. These are usually led in a top-down fashion by national offices, RPOs, publishers and some research funders, all of whom are able to directly benefit from the adoption of these PIDs for their own research information management workflows. Researchers on the contrary tend to have little involvement with and awareness of admin-oriented PIDs (with the sole possible exception of ORCID, which is well-established at least among academic authors in the public sector). Moreover, researchers may often see these ‘admin-oriented’ PIDs as pieces of an unwanted additional bureaucracy and will require some significant degree of advocacy to understand their associated benefits and to regularly use them.

It could be that the technical PIDs are simply showing a less evolved implementation stage and that they will eventually become as relevant for research funders, universities, publishers etc as the admin-oriented PIDs already are for them. In fact, the growing involvement of PID service providers like DataCite and Crossref in the management of the nascent PID initiatives guarantees some common ground across PID categories. In the meantime however, the perception is that developments in these two broad PID areas are taking place with little awareness of the key stakeholders from the other area. Hence the recommendation to RPOs to keep abreast of developments in PID areas where their own researchers may already be involved.

## 4.2 Competing technical solutions

This is the context where the PID landscape fragmentation is most evident right now. The best example for this is the area of PIDs for organisations or OrgIDs, for which there are currently two different technical solutions that are not mapped against each other, the Research Organisation Registry (ROR) and Ringgold – recently acquired by the Copyright Clearing Centre (CCC). ROR currently displays OrgIDs for over a hundred thousand organisations, while Ringgold has over 600,000 entries in their database. The current disconnection between competing solutions could perhaps be addressed by using ISNIs, but it will be difficult for the intense ongoing discussion on the feasibility of having multiple-level OrgIDs to reach any sort of conclusions if there is no coordination between the actors that could make it possible.

While OrgIDs may be the clearest, they are far from being the sole example for this sort of fragmented technical solutions. A specific case study within the KE work is devoted to the RePEc Author Services (RAS). This economics/business-discipline-specific Research Papers in Economics database is currently – and very successfully among researchers in those disciplines – offering its own author and organisation IDs with no default linking to ORCIDs or RORs or Ringgold. It may be just a question of time until the different bits and pieces of the PID landscape start to come together, but it’s worth highlighting the current fragmentation issues while this process unfolds. This is particularly

relevant for universities, who have traditionally taken on the responsibility of training their researchers in the adoption and use of PIDs like DOIs for publications and datasets and ORCIDs. If researchers feel a bit reluctant to follow the developments around ‘admin-oriented’ IDs like OrgIDs, their attitude is unlikely to improve if they find out there are two different, unconnected sets of OrgIDs available.

### 4.3 The ambiguous concept of ‘community’

The concept of community in expressions like “PIDs and services associated with them need to be perceived as valuable and be in turn promoted by the community” is far from clear at present. A wide range of PID community stakeholders are identified in the report, including governing bodies, PID service providers, a possible PID Federation, the RDA working groups, PID-related projects and initiatives such as those related to the European Open Science Cloud, international coordination bodies like the KE, publishers, national offices, NREs, research funders, universities and research centres, researchers and startups. The fact that many of these have little common ground with each other in terms of sharing a unified conversation means a significant challenge for a harmonised development of the PID landscape.

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 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
Universities	PID user: institutions store the grant IDs in the metadata set for funded projects they keep in their CRIS systems. Grant IDs are included in the RAIDs issued by institutions
Researchers	PID user: prompted by their funders and institutions, researchers include the grant IDs in the acknowledgements section of their manuscripts
Publishers	PID user: publishers allow these grant IDs to be provided into the manuscript submission systems and include them in the metadata sets exported to Crossref – allowing the references to be included for instance in individual ORCID profiles

**Table 1:** Stack of stakeholders involved in the issuing, adoption and use of DOI-based grant IDs

In its section 4 devoted to “Community”, the KE report explores some examples of how different stakeholders participate in the development of specific PIDs. Table 1 above shows for instance the stack of stakeholders involved in the issuing of Crossref DOI-based grant IDs by research funders [Kiley, R., Fentrop, N., Hendricks, G., 2018]. It is very useful for these various stakeholders to be aware of the role each of them is expected to play in the consolidation of the PID landscape and to be able to interact with each other in a proactive way.



## 5 Copyright notice

The author of papers, abstracts, presentations, etc. for the EUNIS Congress retains the copyright of such material. EUNIS may publish such papers, abstracts, and enclosures (in full or in part) on websites, in print and on other media for non-commercial purposes. All papers and extended abstracts are reviewed by members of the Scientific Committee. However, the responsibility for the contents of the papers and extended abstracts rests solely upon the authors.

## 6 References / Citations

Belsø, R. & KE Task & Finish Group for PID Risk & Trust. (2021). *Risks and Trust in Pursuit of a Well-functioning Persistent Identifier Infrastructure for Research*. Retrieved March 5, 2023 from: <https://doi.org/10.5281/zenodo.5018216>

Bécquet, G. (2022). *Identifiers and Open Science: Elements for a socio-technical analysis of ORCID*. International Journal of Knowledge Engineering and Management (IJKEM) 11(29): 125-160. <https://doi.org/10.29327/265007.11.29-8>

Bilder, G., Lin, J., Neylon, C. (2015). *Principles for Open Scholarly Infrastructures-v1*. Retrieved March 2, 2023 from: <https://doi.org/10.6084/m9.figshare.1314859.v1>

Car, NJ., Golodoniuc, P., Klump, J. (2017). *The Challenge of Ensuring Persistency of Identifier Systems in the World of Ever-Changing Technology*. Data Science Journal, 16, p. 13. <https://doi.org/10.5334/dsj-2017-013>

Cousijn, H. et al (2021). *Connected Research: The Potential of the PID Graph*. Patterns 2(1): 100180. <https://doi.org/10.1016/j.patter.2020.100180>

De Castro P. et al. (2023). *Building the plane as we fly it: the promise of Persistent Identifiers*. Knowledge Exchange. Retrieved March 2, 2023 from: <https://doi.org/10.5281/zenodo.7258286>

Kiley, R., Fentrop, N., Hendricks, G. (2018). *Wellcome explains the benefits of developing an open and global grant identifier*. CrossRef blog. Retrieved March 2, 2023 from: <https://www.crossref.org/blog/wellcome-explains-the-benefits-of-developing-an-open-and-global-grant-identifier/>

Knowledge Exchange. (2021). *Knowledge Exchange Task & Finish Group for PID Risks and Trust*. Retrieved March 2, 2023 from: <https://www.knowledge-exchange.info/event/pids-risk-and-trust>

Madden, F. (2021). *The British Library Adopts a New Persistent Identifier Policy*. British Library Digital scholarship blog. Retrieved Jun 1, 2023 from: <https://blogs.bl.uk/digital-scholarship/2021/11/the-british-library-adopts-a-new-persistent-identifier-policy.html>

Meadows, A. (2021). *PIDapalooza is taking a break*. The PID Forum. Retrieved March 5, 2023 from: <https://pidforum.org/t/pidapalooza-is-taking-a-break/1858>

MoreBrains. (2022). *The PID-optimised research cycle*. Retrieved March 2, 2023 from: <https://resources.morebrains.coop/pidcycle/>

TIB Hannover (2022). *TeilAutomatisiertes Persistent-Identifer-basiertes Reporting (TAPIR) project*. Retrieved March 2, 2023 from: <https://projects.tib.eu/tapir/en/project/description/>

UKCoRR – United Kingdom Council of Open Research and Repositories (2023). *PIDs and repositories [webinar]: experiences of implementing DOIs for content*. Retrieved June 14, 2023 from: <https://www.eventbrite.co.uk/e/pids-and-repositories-experiences-of-implementing-dois-for-content-tickets-643690574837>

UK Research and Innovation (2021). *UKRI open access policy*. Retrieved March 2, 2023 from: <https://www.ukri.org/publications/ukri-open-access-policy/>

## 7 Author biographies



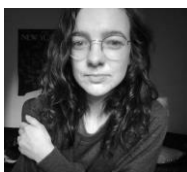
**Pablo de Castro.** Physicist. Open Access Advocacy Librarian at the University of Strathclyde in Glasgow since Jan 2017. Technical Secretary of the Dutch non-profit association euroCRIS since Jan 2018. Former OpenAIRE project officer. Member of the EOSC Association Task Force for PID Policy and Implementation.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

<https://www.linkedin.com/in/pablo-de-castro-018905a/>



**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.

<https://www.linkedin.com/in/ulrichherb/>



**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.

<https://www.linkedin.com/in/laura-rothfritz-22460b238/>



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

<https://www.linkedin.com/in/schopfel/>