

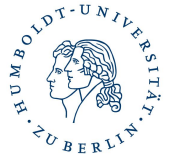
# Risk and Trust in PID infrastructures

Open Science Conference 2023

Laura Rothfritz  
School of Library and Information Science,  
Humboldt-Universität zu Berlin &  
scidecode science consulting

<https://scidecode.com>

 <https://ror.org/01hcx6992>



 <https://ror.org/02c0bjd31>



scidecode science consulting  
De Castro, Herb, Rothfritz GbR

## ● The mission

● Knowledge Exchange (KE) is a collaboration between six national research supporting organisations - CSC (Finland), CNRS (France), DeiC (Denmark), DFG (Germany), Jisc (UK) and SURF (the Netherlands) - working together to support the use and development of ICT infrastructures for higher education and research.

“... to identify, through investigation, analysis and recommendations, what could be the best possible strategic and operational paths to achieve a well-functioning PID infrastructure for Knowledge Exchange (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. Equally important is an analysis that reveals how outcomes and knowledge emerging from this investigation can be transformed into stakeholder recommendations.”

<https://www.knowledge-exchange.info/news/articles/24-06-2021>



## The study

**Analysis** of the current state of the Persistent Identifier (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, etc.) and *new PIDs* (eg conferences, research equipment, facilities).

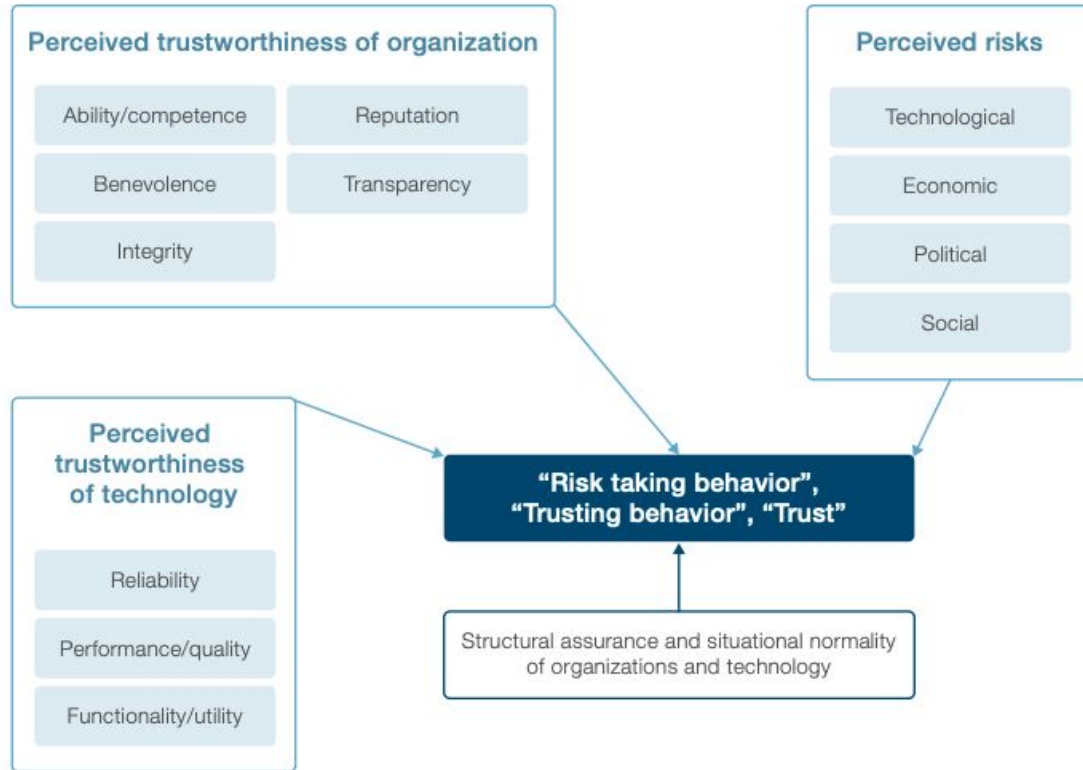
Data collection by **literature study** & **expert interviews**

These fed into

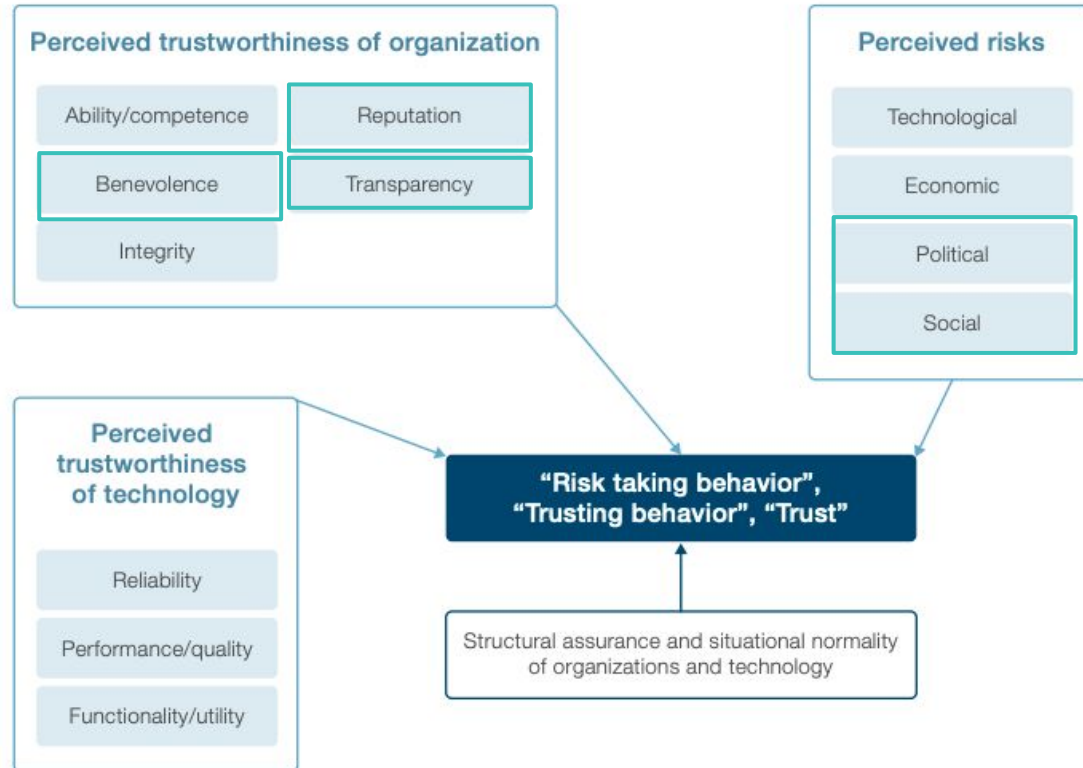
- the construction of **seven case studies** highlighting issues of risk and trust in the PID infrastructure and
- the formulation of **recommendations** for good practice and on the best possible strategic and operational paths to achieve a well-functioning PID infrastructure.



# Findings



# Findings



## Findings: Risk

- Fragmentation of landscape and resulting lack of resources
- ALSO: Risk of centralization
- Dependency on organisations, sometimes people
- Lack of community uptake
- Lack of intra-organisational contingency plans
- Lack of financial sustainability

*“And then the other risk, if I’m allowed two risks, is that I think we must be careful of starting an organisation for each and every persistent identifier, I think. We run the risk, in doing that, in adding more costs to the community. That’s where we must be very mindful about, if there’s something that needs to scale up, how we can do that as a community with existing services and infrastructure.”*

Arena	Possible event
 Political	PID owners decide to stop maintaining metadata, loss of organisational government
 Economic	Financial sustainability is no longer given, financial support is lacking
 Social	Key players in the PID system change or end their involvement, lack of community uptake
 Technological	Technology the PID relies on is changed for any reason (e.g., vendor lock-in), or ceases to support new requirements.

## Findings: Risk

Risk of overestimating the power of PIDs:

- As “trust markers”
- As “quality stamps”
- As “the holy grail”
- Solvers of all problems in scholarly communications

*“At the same time, what I really do not like is the magic power that is being associated in a lot of communications (...). In a lot of stories (...), **people talk about PIDs like they’re absolute magic.** Well, no, you’ve got to really love them, take care of them, you know, make sure that the infrastructure exists for a long time or you’ve basically bought into nothing.”*

*“I think within the scholarly community, the PID story has been sold so well that the researchers, the authors themselves, will start for their own work using them of course, because otherwise their publications don’t even count in their evaluation and so on. So they’ve even penetrated in that. So now, even if your little three pager has a DOI, you can now put it on your list of accomplishments, right? So it’s almost like it’s a quality stamp also, which we all know it’s not. **Again, you know, the people that know, it’s not the quality stamp it is just a freaking identifier.**”*

## Findings: Trust

- PIDs are considered **socio-technical** infrastructures. It seems that **trust in organisations or individuals** is more important for the acceptance of PIDs than the technology used, as the risks associated with the technology are considered amorphous.
- **Transparency** provides a **feeling of control** over how a system is managed and run. This can be achieved by open data, open documentation, close communication, boards and democratic decision-making
- **Structural assurances** like contracts, policies, risk management workflows can build trust

*“But I think, probably collaboration is really important. We’ve talked about this before. If the systems begin to occupy a really central position in the scholarly infrastructure, there are many systems based on it, and these systems probably also have a good motivation to keep all these systems in place. So I think broad adoption and a community-wide use of all of these systems also helps to promote the trust in the systems, and it creates additional motivation for the people behind these systems to keep all the building blocks of the infrastructure in place. So broadly, it’s a two directional movement.”*



## Findings: Trust

Trustworthiness of PID providers is judged by

- Favorable behavior in the past (**reputation**)
- **Relationship** with and knowledge of user community
- Plans for **sustainability, transparency** of operations
- Clear **values** that align with their user community and that are communicated and demonstrated frequently
- Willingness to **invest and commit**

*"I think, if you want to be realistic, then you probably should accept that whenever you adopt something that's not fully under your own control, that always implies a degree of risk. I guess you want to develop a degree of trust in the system that you work with, to develop understanding of the type of procedures that these other organisations have put in place while to mitigate these risks and ultimately to deserve the trust. So we need to do a study on the things that I've mentioned, [the political](#), [the commercial independence](#), [the governance system](#), [the measures they've taken to avoid commercial takeovers and independence](#). Trust and reliability is really useful."*

# Take-away messages



## **PIDs are not magic!**

Value for scholarly ecosystem is very high, but investment and maintenance costs are as well. They are not trust marker or quality stamps.

## **“It takes a village...”**

PIDs are socio-technical, taking care of the social part is equally or even more important than functioning technology.

## **Trust, but verify!**

Caring and talking about risks is important, blind trust is not a good idea.

## **“Trustworthy” ≠ “Trusted”**

PID Providers who describe themselves as trustworthy: how do you know you are actually trusted?

## Case studies on risk and trust

**KE**  
Knowledge Exchange

Case study

### Failed PIDs and unreliable PID implementations

February 2023



This case study is part of a series that has been produced within the study on "Risks and Trust in pursuit of a well-functioning PID infrastructure for research" commissioned by the Knowledge Exchange in July 2021. The main outcome of this study is a report examining the current PID landscape with an emphasis on its risks and trust-related issues.

This complementary series of case studies aims to provide a deeper insight into specific areas of activity, workflows and stakeholders within this wider PID landscape.

DFG CSC Jisc CNRS SURF DeIC

Pablo de Castro, Ulrich Herb, Laura Rothfritz, & Joachim Schöpfel. (2023). Failed PIDs and unreliable PID implementations. Zenodo. <https://doi.org/10.5281/zenodo.7330527>

**KE**  
Knowledge Exchange

Case study

### IGSN - building and expanding a community-driven PID system

February 2023



This case study is part of a series that has been produced within the study on "Risks and Trust in pursuit of a well-functioning PID infrastructure for research" commissioned by the Knowledge Exchange in July 2021. The main outcome of this study is a report examining the current PID landscape with an emphasis on its risks and trust-related issues.

This complementary series of case studies aims to provide a deeper insight into specific areas of activity, workflows and stakeholders within this wider PID landscape.

DFG CSC Jisc CNRS SURF DeIC

Pablo de Castro, Ulrich Herb, Laura Rothfritz, & Joachim Schöpfel. (2023). IGSN - building and expanding a community-driven PID system. Zenodo. <https://doi.org/10.5281/zenodo.7330498>

Read the report at:  
de Castro, Pablo, Herb, Ulrich, Rothfritz, Laura, & Schöpfel, Joachim.  
(2023). Building the plane as we fly it: the promise of Persistent  
Identifiers. Zenodo. <https://doi.org/10.5281/zenodo.7258286>



scidecode science consulting

De Castro, Herb, Rothfritz GbR



## Keep in touch if you like!



### Laura Rothfritz



[laura.rothfritz@hu-berlin.de](mailto:laura.rothfritz@hu-berlin.de)



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



@ztirfhtor

<https://scidecode.com>

<https://twitter.com/scidecode>

<https://openbiblio.social/@scidecode>

<https://www.linkedin.com/company/scidecode>



# BACKUP



## ● The team

- 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 associate 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.



## Recommendations



... addressing a wide range of stakeholders

- **National-level stakeholders**
- **Research funders**
- **PID Service Providers**
- **Institutions/ Research Performing Organisations (RPOs)**
- **Researchers**
- **Publishers**
- **A (possible) PID Federation**
- **Knowledge Exchange**

## ● The recommendations

### ● Research Funders

1. **Make sure you are represented in – or at least informed about – national-level coordination initiatives.**
2. **Be aware of what PIDs are relevant for your activity**, including for project proposal evaluation, reporting on funded research outputs and grant identification.
3. **Consider assigning grant IDs to your grants** whenever possible, allocating the appropriate human and technical resources to make it possible.
4. **Consider requiring specific PIDs from your funded researchers**, even for applicants to your funding calls.
5. **Be aware of the developments around emerging PIDs** that may be relevant to your area of activity including PIDs for instruments and facilities and PIDs for geo samples.
6. **Be aware of funder-specific coordination initiatives** at a national and international level, promoting and joining them whenever possible.



## ● The recommendations

### ● Institutions (Research-Performing Organisations, 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.

## ● The recommendations

### ● Publishers

1. **Ensure long-term availability** of publications with a PID through agreements with long-term archiving agencies or national libraries. Have exit policies in place stating you will notify the PID provider about the findability of publications in case of journal discontinuation so that resolving is maintained.
2. **Include entries for additional PIDs** in manuscript submission systems as these PIDs become more widely implemented.
3. **Provide information snippets** to researchers/authors on why PIDs are important.
4. **Be aware of the level of maturity of specific PID initiatives** in order to allow references to these to be included in manuscripts.
5. Make sure the PIDs you provide in your publications **are operational and resolve correctly**.
6. Where these are available, **consider including pre-existing PIDs for pre-prints in the final research publication webpage** alongside the PID for the Version of Record.
7. **Diamond OA publishers:** implement DOIs as the bare minimum, make use of the *Diamond OA Capacity Centre's* support, join initiatives where best practices may be shared.

## Findings: Risk

### Political

- Risk of discontinued service due to organisational change, such as a takeover by commercial companies or a merger
- Risk of privilege: many PID systems are very western-centred

### Economic

- Fragmentation and lack of funding
- Lack of contingency funding

*“I think that’s an important thing to split apart, that identifiers should not have any meaning of what is good quality, because they don’t... there’s two things that don’t fit together. I mean, you should not use a DOI as a statement that this is the thing because, again, then, I mean, what about Third World countries who can’t assign DOIs? That doesn’t mean that the thing that they come with is actually bad quality, or it’s bad science, it’s just better than that monetary registered DOI. Right? Again issues, going back to this thing about like Third World countries, like, how do we involve them? I mean this, again, we’re building things because we have the money to build it. That’s exclusively for the western world.”*

## Findings: Risk

### Political

- Risk of discontinued service due to organisational change, such as a takeover by commercial companies or a merger
- Risk of privilege: many PID systems are very western-centred

### Economic

- Fragmentation and lack of funding
- Lack of contingency funding

*“And then the other risk, if I’m allowed two risks, is that I think we must be careful of starting an organisation for each and every persistent identifier, I think. We run the risk, in doing that, in adding more costs to the community. That’s where we must be very mindful about, if there’s something that needs to scale up, how we can do that as a community with existing services and infrastructure.”*

## Findings: Risk

### Social

- Lack of institutional commitment and of human resources
- Not being able to show examples and value
- Lack of uptake
- Dependency on the people responsible for the system
- PIDs are often seen as trust markers, even though they are clearly not

### Technological

- Lack of quality, richness and completeness of metadata
- Risks concerning scalability and interoperability

*“I think a lot of identifiers, let’s take another example, like the PIC code from ADS. So ADS is a system that’s been running for a long time. They have their own identifier, that PIC code thing, that they use internally in the system, that has some other people using it from time to time, right. But then you have a guy who’s been sitting there for 20 years and knows everything about the things, but he’s going to retire at some point. Right, I think that the people factor in the people running these things, because, I mean, I was a scientist running on a shoestring a lot of the time. These things run on passion. I mean, they’re running because a lot of people put the effort into it beyond what they expect it to do.”*

## Findings: Risk

### Social

- Lack of institutional commitment and of human resources
- Not being able to show examples and value
- Lack of uptake
- Dependency on the people responsible for the system
- PIDs are often seen as trust markers, even though they are clearly not

### Technological

- Lack of quality, richness and completeness of metadata
- Risks concerning scalability and interoperability

*“Because at the end, you have to have people that run it. It’s a socio-technical system, where the easier part is running the technology, the more difficult part is the social part (...).”*