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

Title: Failed PIDs and unreliable PID implementations

Authored by: Pablo de Castro, Ulrich Herb, Laura Rothfritz and Joachim Schöpfel (scidecode science consulting).

Email: **office@knowledge-exchange.info**

DOI: **10.5281/zenodo.7330527**

Contents

1. Rationale

PIDs serve as a building block of the scholarly infrastructure and might be described as the "glue" that holds together all the different parts of the scholarly record.

This includes scholarly outputs, digital objects, contributors, institutions etc. and PIDs aim to establish some sort of meaningful relation between these items by describing how they are related.

It is, however, important to not view PIDs in isolation - rather, PIDs are one part of a larger system that might be described as a "tightly coupled system". The term tightly coupled system originates from computer architecture and describes a system, which consists of several parts all dependent upon each other. "Hardware and software components that are linked together and dependent upon each other. For example, in a multiprocessing environment, where several computers share the workload, a tightly-coupled system might have to be shut down in order to add or replace one of the machines." (McGraw-Hill Dictionary of Scientific & Technical Terms, 6E., 2003)

In the case of the PID infrastructure, PIDs are a piece of technology (software, in this analogy) that is directly linked and dependent upon organizations (hardware) that both use and support it. The PID infrastructure is a multilateral system, which needs all parts of it to function properly.

> There's also, of course, a situation that as an institution, you begin to rely at some point on certain persistent identifiers. So these systems become embedded within the systems that you work with so that the functioning of the system ultimately also depends on the rate of reliability of the persistent identifier. So, you create a degree of dependence on these systems.

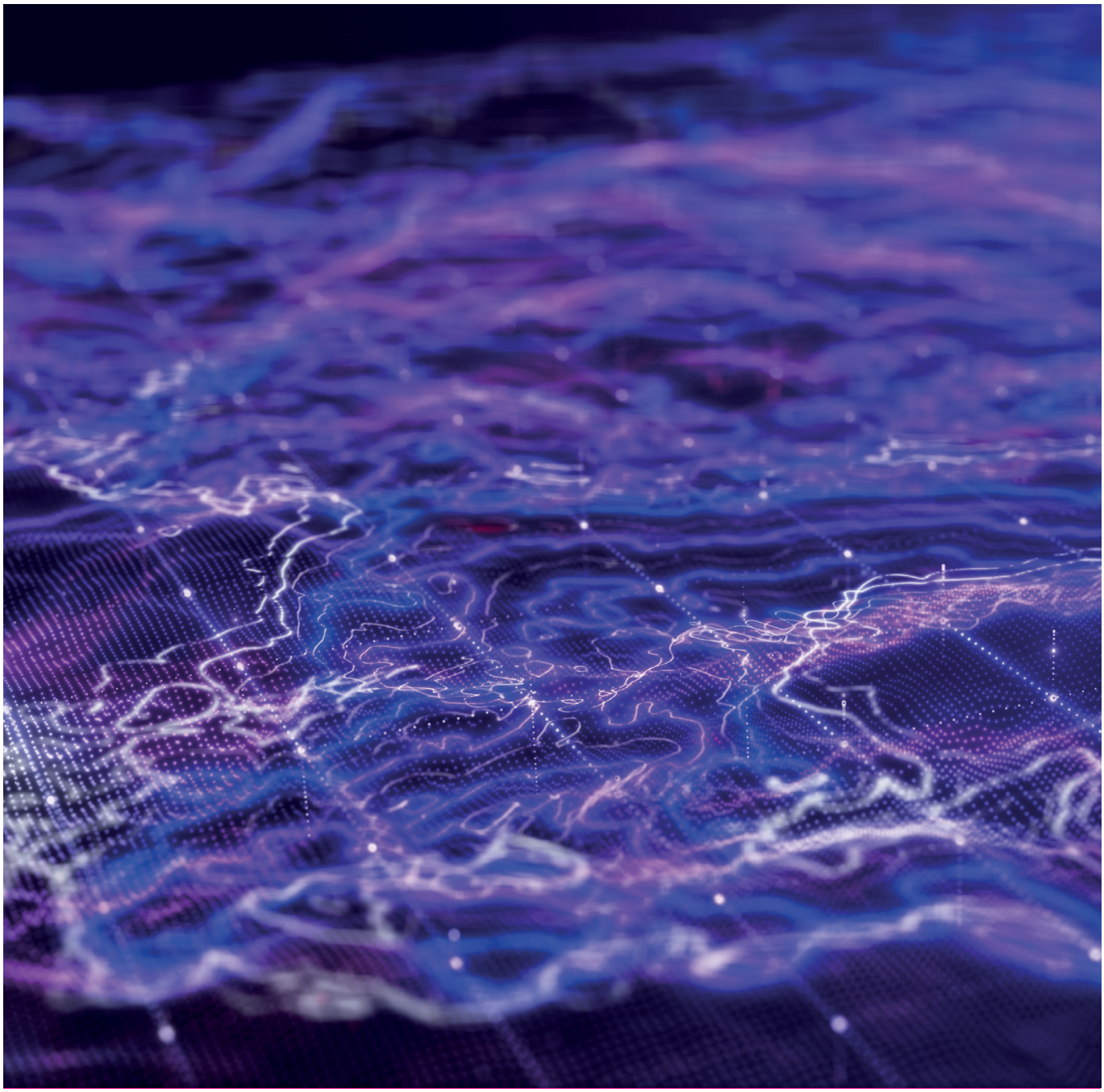
Of course, technical failure of PIDs can occur at any time. For example, Crossref experienced several longer outages, the most recent one was a 17-hour-long outage on March 24th, 2022, due to a large data center outage (similar to the one responsible for another outage in October 2021), leaving the website, content registration system, APIs and reporting function defunct.[1]

However, this case study is about another risk for PID failures: What happens if organizations providing and managing PIDs are unreliable? Does the system in fact need to be shut down in order to replace one of its parts?

1. https://www.crossref.org/blog/outage-of-march-24-2022/

It illustrates risks of failure through lack of organizational support from two perspectives: PURL serves as an example, where a PID provider ceased support for a system. The other perspective shows examples, in which PID-managing organizations fail to implement otherwise well-working PIDs in their systems, and the failed organizational implementation of the International Standard Report Number Identifier (ISRN).

These examples are tightly connected to a number of social risks that were illustrated in our study. These risks are of concern, especially for PIDs that are so embedded in the scholarly communication system that they almost function as “invisible infrastructures” that may only become apparent upon breakdown. (Star & Ruhleder, 1996)



2. Insights in failed PIDs and unreliable PID implementations

2.1 Unreliable PID providing organizations: PURL's near-death experience

The Persistent Uniform Resource Locator system (PURL) is a URI-based identification scheme that consists of a http scheme, a name authority (such as a domain name) and a path. In contrast to normal URLs, PURLs do not directly target the resource they identify, but use a resolving function (similar to other PID systems such as DOIs) that redirects to the current location of the identified object. (Hakala, 2010), However, PURLs are not fully location independent and the fact that they need a naming authority makes them acceptable as Unique Locators, but not as reliable Unique Identifiers, as described by Duerr et al. (2011).

A prominent use of PURLs can be found in Dublin Core: All DC elements are identified via PURLs. That means, all systems employing DC metadata, such as repositories, are also directly connected to PURLs. PURLs are also widely used in other digitally published ontologies for the semantic web in order to define URIs for each term without having to register new domain names for every ontology. Examples include the Human Disease Ontology[2], FaBIO, a FRBR-aligned Bibliographic Ontology[3], or the German GBV Ontology[4].

purl.org, as the major name authority, was founded, run and managed by the Online Computer Library Center (OCLC) from 1995 until 2016. It was free to use and provided a low-maintenance solution, especially for assigning persistent URIs in ontologies, as noted above. According to Car et al. (2017), over 100,000 identifiers were generated. However, the system lacked social infrastructure and formal governance (Klump & Huber, 2017). Most PURLs relied on OCLCs name authority, so when OCLC withdrew institutional support in 2014, a lack of maintenance led to a slew of technical problems, which made a lot of PURL identifiers almost unusable.

> At one point the administrative interface of PURL collapsed, so that it wasn't possible to mint new PURLs anymore or change the resolution of existing ones. OCLC didn't inform the users about the state of the system, so people were getting quite worried about it (...)

The system was eventually put into "read only" mode, making it impossible for PURL owners to manage their identifiers.

> And then OCLC lost interest in it because it wasn't really able to compete with DOI or Handle. It was sidetracked, but not before many, many PURLs had been assigned. So it was important to keep the system alive, but at the same time OCLC didn't seem to be willing any more to invest in the system.

2. https://bioportal.bioontology.org/ontologies/DOID?p=classes&conceptid=root
3. https://sparontologies.github.io/fabio/current/fabio.html
4. https://gbv.github.io/gbvontology/gbvontology.html

Even before being dropped by OCLC, trust in PURL's reliability was sometimes quite shaken, as illustrated by a Tweet by Andrew Treloar in 2011: "Persistence of object good, but persistence of cited identifier crucial. (Context? Purl.org is currently failing to resolve…)"[5].

According to a conversation in the PURL google group from 2015, PURL actually never made it past the state of a Research & Development project at OCLC during its 18 years of existence. It basically existed as a "single machine with a single, part-time administrator"[6].

However, the discontinuation of OCLC's support nearly "killed" the whole ID system. In 2016, The Internet Archive and OCLC announced that the Internet Archive would take over the management of PURLs as a "new sustainable service"[7], including the transfer of all previous PURL definitions and a new web interface to manage PURLs at https://purl.prod.archive.org/.

PURL is therefore still alive – even though it needed serious life support and a reanimation by the Internet Archive. But consistent problems like these do not make PURLs very trustworthy:

> OCLC somehow got tired of it and they went over to the Internet Archive, and they're doing a poor job running it now. Which is really dramatic, because a lot of ontologies depend on PURL URIs for their terminology. You know, I mean, Dublin Core even uses PURLs. So this is basic infrastructure that somehow ... Well, we didn't lose it yet, but it's definitely not maintained at the technical level that encourages further use.

For example, DCMI PURLs were resolving to a wrong target[8] and the SSL certificate expired without immediate action or response by the Internet Archive[9]. Dublin Core, which is heavily dependent on PURLs as described above, seems to have kept a close eye on its reliability after the move to the Internet Archive[10].

> So it looks like PURL is okay at the moment but I think there is still a risk because it's not a strategic product to the Internet Archive. So if anyone is using PURLs I would advise them to move over to some PID system, because even if the collapse was avoided this time, it might not happen next time.

The question remains: what would have happened if the Internet Archive had not taken over PURL as a service? Coming back to the analogy of tightly coupled systems, the death of PURL because of organizational "hardware" failure would have led to a huge amount of problems, especially in the semantic web. Here, an organization solely responsible for an important identifier just ceased support or "lost interest", as some of the interviewees in our study have described it.

5. https://twitter.com/atreloar/status/41827813803696129
6. https://groups.google.com/g/persistenturls/c/Zpd4BHQxxIM/m/7MBio5FsAgAJ?pli=1
7. https://cdm15003.contentdm.oclc.org/digital/collection/p15003coll6/id/649
8. https://twitter.com/paulwalk/status/1146000863690993664
9. https://twitter.com/DublinCore/status/1336723571142504451
10. https://twitter.com/DublinCore/status/1336975185837252608

2.2 Unreliable PID implementations

The handle implementation of the e-LIS repository
e-LIS[11] (eprints in Library and Information Science) is considered the largest Open Access repository in Library and Information Science.

According to information publicly available via the e-LIS website the operation of the repository is currently supported by the Agricultural Information Management Standards Portal (AIMS) which in turn is financed by the Food and Agriculture Organization of the United Nations (FAO), deplegal (a law firm specialized in legal assistance in the field of Intellectual Property, Media, Entertainment, Information Technology and Data Protection) and the Università degli Studi di Napoli Federico II. Both nature and volume of support by these stakeholders is not revealed or described on the website. It was launched in 2003 and its internal structure comprises an administrative board of three persons, an executive board with five members and an editorial board with about sixty members being responsible for the quality control of submitted items.

In Spring 2017 there were some discussions going on about e-LIS failing to keep the handle resolving or assignment alive. As one of the regional editors stated in a blog comment[12], this was “caused by the temporary change of the used repository software”. One of the effects was that handles that were already assigned to documents now referred to other documents.[13] Obviously, these complications are a consequence of the server’s uncertain funding. Organizations offer hosting as sponsorship/ funding for e-LIS, which results in regular server moves - meaning also that technical decisions around e-LIS are not in the sovereignty of its community. e-LIS migrated from Eprints to Dspace between December 2010 and January 2011.[14] After e-LIS had been running under DSpace for about a year, it was decided to switch back to Eprints. Apparently, in the course of this platform change, the mapping between handles and URLs was lost. At least from spring to autumn 2022, the e-LIS handles are not functional, the server also does not use a SSL certificate. The dependence of e-LIS on hosting by external organization reveals a serious weakness: organizational and technical decisions do not lie with e-LIS.

11. http://eprints.rclis.org/
12. https://web.archive.org/web/20200927014145/https://scidecode.com/2017/03/16/welches-open-access-repository-sollte-man-in-bibliotheks-informationswissenschaft-nutzen/
13. e.g. the handle ID 17094 once pointed via the URL http://eprints.rclis.org/17094/ to this document: Mittelsdorf, B., & Herb, U. (2009). Access Data Mining: A new foundation for Added-value services in full text repositories. (https://web.archive.org/web/20121031035059/http://eprints.rclis.org/handle/10760/13711) whereas it now points to a different one: Falcato, Pedro and López, Alicia Isabel and Araujo, Juan Facundo Directrices para las transacciones de referencia en la Biblioteca del Instituto Nacional de Tecnología Industrial. Información, Cultura y Sociedad, 2004, n. 10, pp. 67-82. Obviously, even back in 2012 there was already a re-direct active as the ID of the internet archive’s mirrored version is not the same as the one originally assigned to paper be e-LIS. This can be checked by pasting the URL (including the handle, http://eprints.rclis.org/17094/) into the internet archive’s seach slot under https://archive.org/.
14. https://web.archive.org/web/20111109103717/http://eprints.rclis.org/cms/about

Ongoing problems with DOI resolving/ registration

When on 23rd March 2022 the German publisher De Gruyter posted via Twitter[15] that they won the "Best Publisher UX Award 2022" sponsored by OpenAthens this was harshly criticized by the user Stefan Müller (@LingMuelller) stating (directed at De Gruyter) that he "found many non-working DOIs, entire volumes are missing. The database seems to be a mess."[16] He also posted an example of an DOI (10.1515/97831101 98621.185) that did not resolve at that date (and that still doesn't as of 20th September 2022). In the discussion following this posting, Stefan Müller mentioned to have already in 2021 reported a severe problem with the DOI-resolving by a range of reputed publishers.[17]

In a communication with the consultants, Stefan Müller reported that he did not investigate non-resolving DOI systematically. He simply stumbled upon many broken DOIs while editing a handbook and collecting relevant literature for seminars.[18] He provided the consultants a random sample of eight DOIs that could not be resolved Spring 2021[19]; four of them still did not work in September 2022.[20] This issue addresses DOIs that were obviously not even registered with Crossref, so they can not be resolved to any URL.

The case of discontinued journals

We analysed data provided by Laakso, Matthias & Jahn (2021) with their publication on vanished OA journals in order to find out whether DOI records are kept up to date or curated, so they link to archived articles in the case a journal ceases to exist. From the 174 journals reported, we selected only journals that were online and published after 2015, as the analysis on journals that ceased to exist even earlier proved that these mostly did not have any PID implemented. Applying the aforementioned selection criteria and after elimination of journals for which it is no longer possible to find out whether they issued DOIs, two journals remain. For both journals obviously DOIs were registered, but after the journals ceased to exist, no contingency plan became effective, so there is no archived version online to which the DOI points. In fact, the registration agency either was not informed about the discontinuation of the journals or (lacking new locations it might direct the resolving to) simply did not change their DOI-URL-matching after the journal ceased to exist.

2.3 PIDs & lacking perceived usefulness

The Technology Acceptance Model (TAM) describes perceived usefulness as one of the main variables that determine the acceptance of new technologies. In the PID ecosystem, if a new identifier is not perceived as useful by the target user community, there will be a real problem of acceptance and uptake, even if the same identifier is considered as necessary by PID service providers. This may be illustrated by the failure of the ISO ISRN identifier.

30 years ago, the International Standard Report Number (ISRN) was required by major national libraries and STI centers for the processing of scientific, technical and administrative reports, especially for the production of the European SIGLE bibliographic database for grey literature.

Created in 1980, the launch and production of SIGLE was a political European decision, coordinated by the EC and organized as a network of main STI stakeholders in different EU member states, with an operating agent; among the participants of the network were CNRS-INIST (France), Fachinformationszentrum Karlsruhe GmbH (Germany), Koninklijke Bibliotheek (The Netherlands) and The British Library Document Supply Centre (UK). The role of these institutions was the identification, collection (acquisition), cataloging, preservation and supplying of reports on a national level.

15. https://twitter.com/degruyter_lib/status/1506593201880276998
16. https://twitter.com/LingMuelller/status/1506644563984797703 see similar https://twitter.com/scinoptica/status/1368127705980559362
17. https://twitter.com/LingMuelller/status/1367501731022651398
18. see also his Twitter post: https://twitter.com/LingMuelller/status/1506682887227293701
19. five assigned by De Gruyter, two assigned by Taylor & Francis, one by Cambridge University Press
20. one assigned by De Gruyter, two assigned by Taylor & Francis, one by Cambridge University Press

After initial funding by the EC, it was currently produced by EAGLE (European Association for Grey Literature Exploitation). Its members, major European STI libraries and centres, expressed explicitly the need of an persistent identifier for reports. “The recent proposal of an ISRN (International Standard Report Number) will make possible the bibliographic identification of reports identified on the national level.” (Debachere, 1995)

The structure, metadata, provision and governance are described by Stock & Schöpfel (2000). In France, INIST has been designated as the national agency responsible for assigning these numbers, while FIZ Karlsruhe became the international agency. In fact, the role of the national ISRN agencies has been the development and maintenance of a national ISRN register and the promotion of the new identifier among the report producers who had to assign themselves an ISRN to each published report, similar to the functioning of the ISBN identifier for books.

“With the implementation of the SIGLE (System for Information on Grey Literature in Europe) system (the bibliographic database for GL described below), a set of regulations for the standardisation of data to be entered in the system, was issued. This has inevitably influenced the presentation of this documentation. Among these regulations, the description of the type of registration and the related bibliographic level introduced the ISBN and ISSN codes which, if they appear on documents, are accepted by the system. The elaboration of the standard ISO 10444 ‘International standard technical report number (ISRN)’, published in 1994 can be partly considered a realisation of the decisions made in York, even if it is based on the ANSI standard Z 39.23-1983 (first edition:1974). The application of this standard foresees the establishment of an international agency supported by national centres for the assignment of report numbers and their control.” (Alberani & De Castro, 2001)

In reality, the ISRN was not perceived as useful by the report producers who should have assigned them, either as a complement or as a substitute to their own internal reference number. The uptake was extremely low; the additional workload and the new development was not considered as necessary or facilitating the internal handling or the external dissemination.

The ISRN was perceived as useful and even essential by a limited number of stakeholders for the production of one service, i.e. the SIGLE database (at that time the only large international bibliographic database for grey literature). But as the ISRN had to be assigned by the report producers and not by the national libraries and STI centres, there was immediately a problem which obviously had not been anticipated as a risk. Communication and promotion were not enough to change the situation. The issue was not awareness , but perceived usefulness.

However, the definitive failure of the ISRN identifier was determined by another factor, i.e. the decision of the same stakeholders to end the European cooperation and to stop the acquisition of grey literature including reports and the production of the SIGLE database. The reasons have been described in Schöpfel, Stock & Henrot (2006) Today, the more than 1m former SIGLE records are freely available as a dataset on the Dutch EASY data repository DOI: 10.17026/dans-xtf-47w5, without the ISRN identifier.

3. Issues around risk and trust

3.1 Overview

The different issues around PID implementation raised in the interviews and in desk research of our study show a great diversity both in their phenomena and causes.

On the one hand, with PURL we encountered a PID that, despite the beta status stated by its provider, was used by organizations to reference objects that are highly relevant. To some extent, this can be seen as an act of naivety on the part of users. However, it also points to the problem of setting up a service and enabling its use without having an exact idea of the purpose, seriousness, and endowment with which one intends to offer it.

Next, with e-LIS, we found a repository whose lack of funding leads to dependencies on supporting institutions, which take over the hosting of the server and thus also sovereignty over technical actions such as updates or migrations and – at least the case of e-LIS suggests this – exercise these probably with less care than with in-house services. With e-LIS this caused reassignment of already assigned handle IDs, which annulled their persistent identification. Unlike PURL, however, it was not the PID as such that was at stake here, but rather a technical and organizational error, primarily caused by the lack of own resources.

Problematic PID implementations by publishers in turn are probably not caused by a lack of financial resources, they are also different in nature: we found examples of publishers that obviously even failed to register DOIs (but nevertheless displayed these on landing pages and full text PDFs). Findings[21] by Klein & Balakireva (2021) also suggest that publishers do not necessarily facilitate unique identification of documents. Especially, the utilization and implementation of PIDs by publishers were reflected critically by two of the experts interviewed in our study.

> And the reason [for the for hiding items behind re-directs and HTTP layers] is (...) because they don't want to make it easy for machines to navigate their environment, of course. So there's a weird thing going on there.

These results highlight the opinion of an expert, who portrayed publishers as ***"the weakest link"*** (PID_14) in the PID system.

The examination of the Laakso, Matthias & Jahn (2021) sample revealed journals whose editors did not inform the registration agency that the publication had been discontinued, so the DOIs assigned pointed to invalid content. This last issue reflects neither organisational nor technical issues, but simply a lacking sense of responsibility possibly combined with a lack of resources.

A more detailed investigation of the extent to which publishers miss registering PIDs for published content could provide very interesting insight into the reliability of the PID system. This issue, unlike HTTP redirects, CMS configurations, or journals that cease to exist without fallback plans for assigned PIDs, was not mentioned by any expert, but seems in no way of lesser importance than the aforementioned issues, as registration precedes them in time and missing it may be considered even more serious.

21. Based on an analysis of HTTP answer codes.

3.2 The importance of a committed organization

PID systems need commitment by an organization to run them – both in terms of financial and human resources. If organizations do not want to be responsible any more, and there are no backup plans in place, a PID system will fail.

> ...once you start a PID, there's sort of the understanding that there is somebody who will be responsible for maintaining infrastructure, etc. And that's just more wishful thinking than some law or some requirement for somebody to sign papers: yes, I will be responsible. The reality of life is, for example, that PID infrastructure costs money to maintain. So that… there's this risk.

One of the solutions for this was the commitment of a non-profit organization to PID maintenance (in the case of PURL). The majority of the interviews in this study felt that governance by a non-profit organization would vastly increase the sustainability of PID systems.

There is still a valid reason for doubt, though: “A thing that keeps bothering me is @OCLC having ended PURL support. They were nice enough to find an alternative (@internetarchive) to somehow sustain it. But, if a not for profit org with global membership and support can’t keep a PID service alive, who can?”[22]

One of the interesting points to raise here is that the tax status (e.g. non-profit) is not directly connected to an actual sustainable management of PID services. This can clearly be seen through the case of PURL.

> I would love people to stop putting as much faith in nonprofit status as they do. I just don't think it's relevant. Other people have pointed out that Harvard's nonprofit. It's got a high revenue but it's… a lot of people say it's a hedge fund with a college attached to it, right? There's nothing inherently noble about being a nonprofit. It's just the tax status.”

Commitment to sustain PID infrastructures is crucially dependent on the willingness of the organization to actually run the system in the long term, no matter if they are for profit or not.

Another problem with PURL was that OCLC was solely responsible for it and that the data was not made openly available. Ceasing support meant that the system could not be recreated by the user community; they were directly dependent on OCLC to provide the service. Communication around the maintenance and status of PURL also was not open, which left the user community in the dark about the future of the PID.

3.3 Contingency plans

How can the risk of organizational failure to commit to a PID system be countered? One of the most important aspects here are contingency plans – contractual agreements between organizations stating that the responsibility for a PID will be taken over by another organization, ideally with clear plans on how this might work and openly shared to the user community.

> And I think if you look at this, depending on how your organization is set up, then it is also the question, what are the policies you have to find to maintain the availability of all the PIDs. Maybe it is more difficult if you have a single organization because if the single organization is not available any more or cannot find the funds to sustain itself then you already have a challenge to ensure sustainability, but also the persistence of your service, but also the PIDs itself.

> You know, pretending that you can't go out of business is not… no, it's not an option. Everybody can go out of business. We've seen huge organizations go out of business, we see countries disappear, from, you know, all sorts of things. This is not a strategy. A responsible thing to do is say, Okay, if this happened, what would it take to do it?

22. https://twitter.com/hvdsomp/status/1308122184402055171

Funding is a key topic here, in this case a contingency fund as a back-up if PID providers go out of business. As seen by the time it took to work out a plan between OCLC and the Internet Archive (over a year), it is most likely that bigger PID providers such as DataCite or Crossref would also need some time to wind down an organization:

> And you're talking about actually winding down an organization, an infrastructure organization and trying to do it in a smooth way, you're going to need at least a year to do that. Let's just face that. And obviously, you've got to have the funds do that. It's easier if the data is open. Because then at least you're not having to deal with escrow and all these other weird things. I think that having something documented helps – and in the case of (organization), we've got a few things.

3.4 Other issues of risk and trust

PURL's near-death experience serves as a good example of how NOT to handle a PID system that is at risk. Quite a few things went wrong in the winding down of PURL at OCLC that might damage the trust relationship between PURL users and OCLC in the long run. OCLC did not openly communicate about the state of PURL. It was clear to users that something was not working well, but it was not clear what exactly was happening. Without honest communication, the user community had no way of knowing if their reliance on PURL was in fact too risky. Those organizations, such as the DublinCore Initiative, that did (or could) not step back from using PURLs had to reach out to OCLC several times.

This lack of communication was interpreted as a loss of interest by OCLC. There was no information on funding or development plans for PURLs and even if internally a discussion about its future might have taken place, the user community was not informed about it.

Secondly, it took almost two years for OCLC and the Internet Archive to come to terms about a takeover of PURL. There was clearly no contingency plan in place; it is in fact quite unclear if OCLC had any plans of keeping PURLs alive and stable. This fact has a big influence on the extent to which a user community can trust the sustainability of an infrastructure it depends on. With no fallback or backup planning in place, a PID system can "die" a pretty quick death or live in a "zombie" stage for a long period of time (as seen with PURLs).

Lastly, this case study has made clear that centralizing PID systems to just one responsible organization, without any sort of open data backup, is not a good idea at all. Even if the organization in question is defined as "non-profit", that does not mean that a lack of commitment to keeping an infrastructure alive will mean the end of this infrastructure.

The case of the IRSN proves that a lack of use cases or perceived usefulness by potential applicants of a PID is a severe risk, this might especially affect admin-oriented PIDs.

On the level of content providers (publishers, journals, repositories) apparently users (as authors or readers) take the reliability of these venues for granted without reflecting on possible risks. This is unsurprising, as no user undertakes a deeper analysis of CMS/web server configurations, a service's financial viability or checks whether they perform correct registration of PIDs. They usually only notice these errors when a PID cannot be resolved or their reference management software cannot import a full text, as the HTTP redirects prohibit this. This can lead to a general mistrust in PIDs as such, as well as a mistrust in the implementation at individual providers. Mistrust in PIDs as a whole could have a negative impact on their use, e.g. in bibliographies, and affect different types of PIDs (e.g. reduce the acceptance of ROR if people had poor experiences with DOIs).

4. Authorship

This case study has mainly been written by Ulrich Herb (Saarland University, ORCID https://orcid.org/0000-0002-3500-3119), Laura Rothfritz (Humboldt University Berlin, ORCID https://orcid.org/0000-0001-7525-0635) and Joachim Schöpfel (University of Lille and euroCRIS, ORCID https://orcid.org/0000-0002-4000-807X) within a team of consultants including also Pablo de Castro (University of Strathclyde and euroCRIS, ORCID https://orcid.org/0000-0001-6300-1033) under the umbrella of scidecode science consulting (ROR https://ror.org/02c0bjd31). The work has been overseen by the Knowledge Exchange Task & Finish Group whose composition is listed at https://www.knowledge-exchange.info/event/pids-risk-and-trust.

5. Literature

Alberani, V., & De Castro, P. (2001). Grey Literature from the York Seminar (UK) of 1978 to the year. *Inspel*, *35*, 236-247. https://archive.ifla.org/VII/d2/inspel/01-4alvi.pdf

Car, N., Golodoniuc, P., & Klump, J. (2017). The Challenge of Ensuring Persistency of Identifier Systems in the World of Ever-Changing Technology. *Data Science Journal*, *16*(0), 13. https://doi.org/10.5334/dsj-2017-013

Debachere, M. C. (1995). Problems in obtaining grey literature. *IFLA journal*, *21*(2), 94-98. https://journals.sagepub.com/doi/pdf/10.1177/034003529502100205

Duerr, R. E., Downs, R. R., Tilmes, C., Barkstrom, B., Lenhardt, W. C., Glassy, J., Bermudez, L. E., & Slaughter, P. (2011). On the utility of identification schemes for digital earth science data: An assessment and recommendations. *Earth Science Informatics, 4*(3), 139. https://doi.org/10.1007/s12145-011-0083-6

Hakala, J. (2010). Persistent identifiers-an overview. http://www.persid.org/downloads/PI-intro-2010-09-22.pdf

Klein, M., & Balakireva, L. (2022). An extended analysis of the persistence of persistent identifiers of the scholarly web. *International Journal on Digital Libraries, 23*(1), 5–17. https://doi.org/10.1007/s00799-021-00315-w

Klump, J., & Huber, R. (2017). 20 Years of Persistent Identifiers – Which Systems are Here to Stay? *Data Science Journal, 16*(0), 9. https://doi.org/10.5334/dsj-2017-00923.

Laakso, M., Matthias, L., & Jahn, N. (2021). Open is not forever: A study of vanished open access journals. Journal of the Association for Information Science and Technology, 72(9), 1099–1112. https://doi.org/10.1002/ASI.24460.

McGraw-Hill Dictionary of Scientific & Technical Terms, 6E. (2003). Retrieved April 3 2022 from https://encyclopedia2.thefreedictionary.com/tightly+coupled

Schöpfel, J., Stock, C., & Henrot, N. (2006, December). From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource Migration in the European Context. In *Eighth International Conference on Grey Literature: Harnessing the Power of Grey. New Orleans, 4-5 December 2006*. https://archivesic.ccsd.cnrs.fr/sic_00181592/

Star, S. L. & Ruhleder, K. (1996) Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. Information Systems Research 7(1):111-134.

Stock, C., & Schöpfel, J. (2000). *Mise en place de l'agence nationale ISRN* (Report, Institut de l'Information Scientifique et Technique (INIST-CNRS)) https://hal.inria.fr/hal-01459000/

Knowledge Exchange
C/O Jisc
4 Portwall Lane,
Bristol, BS1 6NB
United Kingdom

T +44 203 697 5804
E office@knowledge-exchange.info