Building the plane as we fly it: the promise of Persistent Identifiers

Pablo de Castro, Ulrich Herb, Laura Rothfritz, Joachim Schöpfel

HAL Id: hal-04240585
https://hal.univ-lille.fr/hal-04240585

Submitted on 13 Oct 2023

⟨KE⟩ *Knowledge Exchange*

Building the plane as we fly it: the promise of Persistent Identifiers

Publication date: February 2023

DFG · CSC · Jisc · CNRS · SURF · DeiC
DANISH e-INFRASTRUCTURE COOPERATION

Contents

Executive summary

Persistent Identifiers (PIDs) are globally unique, permanent references to potentially any sort of digital or non-digital entity. PIDs are of vital importance to modern day computerised research. PIDs ensure that all elements of research are uniquely identifiable and accessible, by humans, as well as by the rapid expansion of machines, that compute a staggering amount of complex data. A well-functioning PID infrastructure for research must be pursued such that research agendas can still grow in complexity. It must be done collectively and internationally. Unfortunately, it is no easy task; it presumes not only technical understanding but insight into the organisational, social, political, and economic aspects of PIDs. This report offers just that, as well as a set of recommendations to engage in, when continuing that pursuit.

Although PIDs are not only used in science communication, this is their primary context of application. PIDs exist to identify a variety of entities, among others, publications, data, software, physical samples, people, research performing organisations (RPOs), research funding organisations (RFOs), individual grants, project activities, conferences and instruments and facilities.

PIDs as indispensable building blocks of today's research

PIDs are essential for modern research, as they are tools to guarantee discoverability, identifiability and traceability of scientific results – if fully adopted, they describe every product and activity throughout the research cycle. PIDs contribute to the integrity of scientific communication and also to its reproducibility, where identifiability is a minimum requirement. The application of the FAIR principles as guidelines for the implementation of Open Science also depends on PIDs: making scientific activities and entities findable, accessible, interoperable and reusable is unthinkable without the features of PIDs. The permanent identification of entities and the provision of metadata describing the entities in terms of content, technical

and legal properties are therefore a core element of FAIR digital objects.

While the permanent identification of entities is crucial for the discoverability, verifiability and quality control of research results and their attribution (in the sense of impact measurement for authors and their organisation as well as research monitoring), the combined metadata describing the entities offer numerous added values. The integration and cross-linking of research objects through their metadata makes the "research graph"(see figure 1) a reality, displaying at a glance all persons and institutions involved in a research process, the results as text, data, software, the processes (e.g. funding), project events and thus giving a 360° view on linked open science.

The value of PIDs

Furthermore, using PIDs to identify authors and their publications makes the management of information much more efficient for Research-Performing Organisations (RPOs) and Research Funder Organisations (RFOs) – not only for administrative staff, but especially for time-pressed scientists. If submission workflows at funders would allow the importing of personal,

Figure 1. Open research Graph (OpenAIRE 2022, OpenAIRE website)
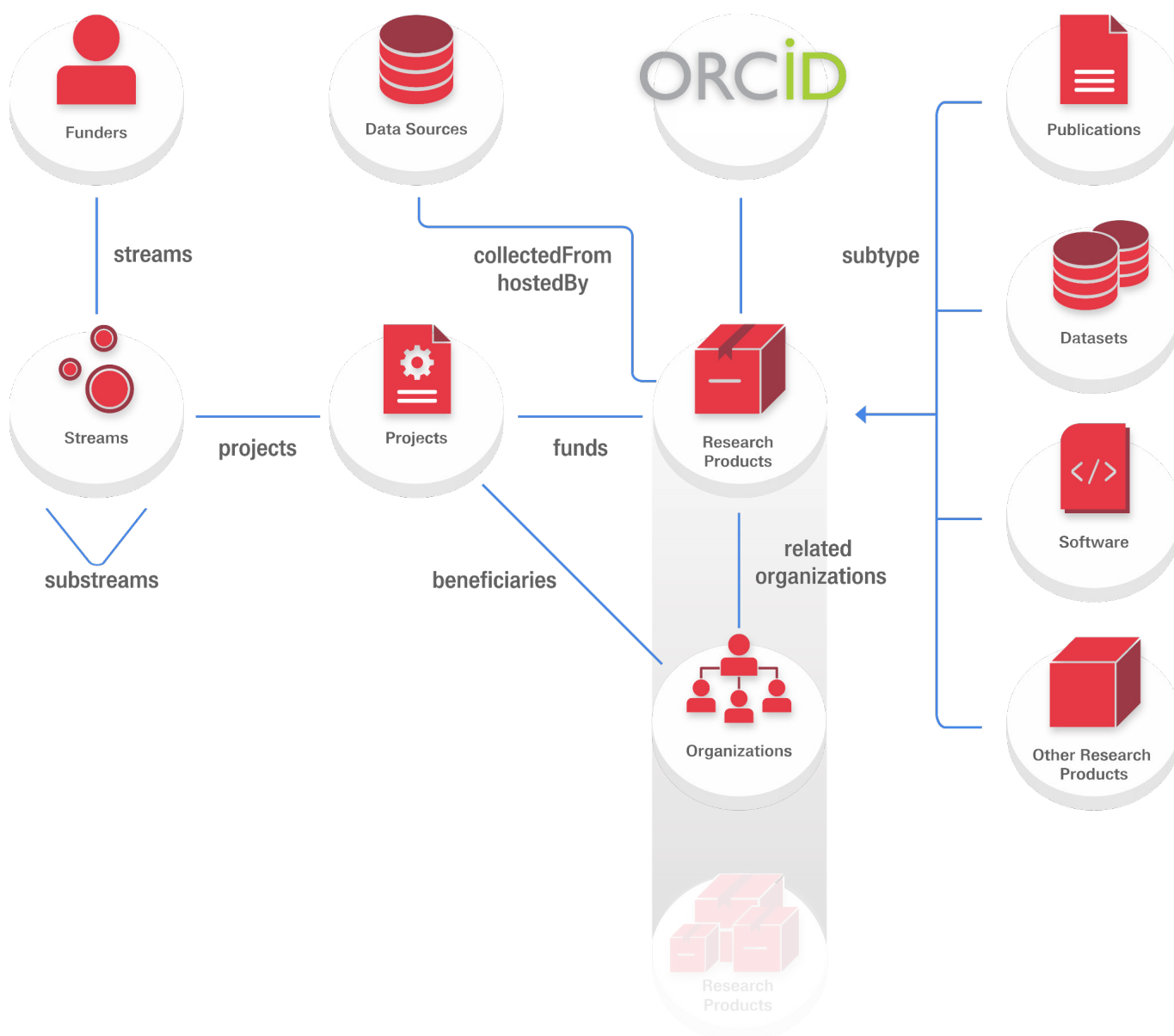
publication or project-related information via PIDs, more time would be left for the authors' actual research work and not spent on reporting. The situation is similar with regard to RPO-internal data warehouses: if publication databases were connected to the leading PID service for author data, ORCID, this information could be fed in at the click of a mouse and not through tedious multiple data entry. The potential becomes clear when one considers that in addition to the research areas for

which established PID services already exist (authors, publications), others are emerging or in the process of differentiation, above all grant IDs, with which RPO-internal project databases could be synchronised on the fly. Unsurprisingly, Brown et al. (2022) concluded that in Australia's public research sector 38,000 person days per year are wasted as the value of the metadata provided by PIDs remains underexploited. According to the authors, this corresponds to a direct financial cost

of nearly $24 million per year; they even estimate savings of as much as $84 million per year, taking into account the opportunity costs associated with technology transfer and innovation-driven growth.

The dangers of missing PIDs

However, these benefits in terms of research integrity, automation and cost savings will only materialise if the individual PID services, e.g. for identifying entities such as authors or publications, function smoothly and reliably and are integrated in the best possible way (e.g. by exchanging metadata in agreed formats via APIs). Without integration and interoperability, the promises of automation and efficiency remain unfulfilled, the same applies to monitoring or visualisation in the research graph. Without permanent identification by mapping identifiers to an entity, PIDs do not add value in ensuring scientific integrity and traceability. As a sort of worst case scenario, if a PID service that e.g. assigns a unique identifier to publications ceases to exist overnight, it leaves behind huge damage: countless unidentifiable and unverifiable documents that exist in isolation, unconnected to other research activities as well as suddenly missing nodes in the research graph that once represented links between other (now apparently unconnected) entities or activities and gaps in data repositories for research information that could only be filled with a lot of work, if at all.

The study "Risks and trust in pursuit of a well functioning Persistent Identifier infrastructure for research"

The outstanding promises of PIDs and how to harness them are the subject of this report, as well as the risks of unreliable PID services, which in the worst case can make parts of an otherwise functional infrastructure of connected PID services fragile. The report is the main outcome of a study commissioned by Knowledge Exchange (KE). This study was aimed at investigating "Risks and trust in pursuit of a well functioning Persistent Identifier infrastructure for research" analysing the current state of the Persistent Identifier (PID) landscape in the six Knowledge Exchange partner

countries and beyond, taking emerging PIDs particularly into account and examining the roles of relevant stakeholders such as PID service providers, Higher Education Institutions, researchers, publishers and national libraries.

While maintaining an emphasis on the risk and trust aspects, the report also aims to provide a sense of the current status of the PID landscape. It's been quite some time since an attempt was made to provide such an up-to-date snapshot[1], and the landscape has kept evolving rather rapidly in the meantime. Areas addressed in the study that have recently seen significant progress are – among others – grant IDs issued by research funders, Research Activity Identifiers (RAiDs) to be issued by institutions, PIDs for research instruments and facilities or the gradual emergence of organisational identifiers (OrgIDs) or PIDs for physical samples.

This multitude of PIDs for different entities may signal a high demand for PIDs in ever new scenarios, while at the same time raising questions on the business models for ever new services and on their technical and organisational justification and sustainability. One of the most prominent recommendations of the report is that this risk of fragmentation must be addressed by more coordination mechanisms, e.g. by national or even supra-national PID strategies or bodies. Other recommendations address the sustainability of the PID infrastructure, or the roles of specific stakeholders, such as researchers or funders, which should take the lead in promoting the PID infrastructure. The main recommendation is to establish a PID Federation that supports stakeholders such as service providers in achieving technical reliability or organisational sustainability. The central element of such a federation (as a strategic and technological convergence within KE countries or even beyond) should be a PID observatory providing an up-to-date and comprehensive snapshot of the PID landscape, its key players and best practice case studies in PID implementation by specific PID types, stakeholders and/or countries.

---

1   The Dec 2017-Nov 2020 H2020 FREYA project ("Connected Open Identifiers for Discovery, Access and Use of Research Resources") provided such an up-to-date snapshot of the PID landscape, https://www.project-freya.eu/en/resources/project-output

Audience & reading recommendations

The report is primarily addressed to two groups of actors: First, policymakers and persons with strategic management functions at RPOs and RFOs, who will find the quintessence for a practical exploitation of the project results in the recommendations (Chapter 3). Second, professionals in charge of providing PIDs for their own content (e.g. publishers, text/ data/ software servers) or other PID experts, e.g. from PID service providers. They will find valuable information on trust and risks in PID services and infrastructures in chapters 3-6, which present findings on the communities (or stakeholders) that populate the PID landscape and from the expert interviews and their content analysis.

Readers concerned with open infrastructures may find some reflections in chapter 7 on the contribution that PIDs make to open infrastructures, how PIDs should be understood as open infrastructures and how the PID infrastructure should benefit from openness. Chapter 8 provides a very brief summary of the report and refers

to the recommendations. Chapter 10 includes appendices. The complementary case studies produced in this study are particularly relevant in this regard, since they informed the formulation of the recommendations. These case studies illustrate new types of PIDs, community-driven PID services, PIDs superseded by later arrivals, failed or unreliable PIDs, and the role of research funders in the adoption of PIDs. The report starts with an introductory chapter, which provides an analytical characterisation of PIDs as well as the organisational, socio-technical arrangements between stakeholders in the PID system. Readers will also find information here on how risk and trust are attributed to PIDs and to the PID landscape. This chapter 1 is primarily relevant for researchers in science and technology studies, but also for PID experts who want to delve deeper into the theoretical foundation of the study. The same audiences will find valuable reading in Chapter 2, which describes the challenge of the study and the main findings of the surveys.

1. Introduction to PIDs and theoretical foundation on risk & trust

This chapter compromises an analytical characterisation of PIDs, the integration of these in a PID graph, organisational arrangements between stakeholders in the PID system, the socio-technical character of PIDs and an analysis of what specificities the perception of risk and trust in PIDs show. These explanations are not only of a definitional and illustrative nature, but also aim to make the reader understand how the challenges of the study were defined and how the analysis of risk and trust is specifically needed to formulate focused recommendations for achieving a functioning PID infrastructure.

PIDs

In this study, we consider PIDs as identifiers that can be assigned to any type of object (physical or digital). They are mostly being assigned to persons, organisations, and research output. Their purposes (besides persistent identification) are to improve research management and information retrieval. In a more formal and definitional approach we rely on the EOSC PID Policy (Hellström, Heughebaert, et al., 2020) that characterised PIDs by the following features: PIDs are …

▸ **globally unique**, including controlled syntax and namespaces governed by clearly defined authorities

▸ **persistent**, including persistent and stable link and resolver functions, persistent syntax and schemes, persistent referred objects

▸ **resolvable for both humans and machines**, including information on how a referred object can be found, accessed, or used to be found (tombstones).

Even if these features of a PID are very rarely questioned, the issue arises if they can be taken for granted. All the more so as the integration of PIDs and the use of their added values (linking, metadata re-use) as outlined in the following paragraph presuppose that PIDs actually possess the properties required here.[2]

PID system

A **PID system as a whole** is considered a mutually referenced combination of definitions, policies, services and data sources which are used for the administration and use of PIDs (Bütikofer, 2009). Even if PIDs are mainly required for research information management (Science Europe, 2016), they should also unburden the individual researcher. This added value is clearly illustrated by the concept of the PID Graph (see figure 2), which involves linkage and semantic relations between entities to which PIDs have been assigned, such as persons (researchers), organisations, research publications, research data, funding and other research results.

Figure 2. DataCite PID Graph KPI as of May 4th, 2020 (M Fenner, DataCite blog)
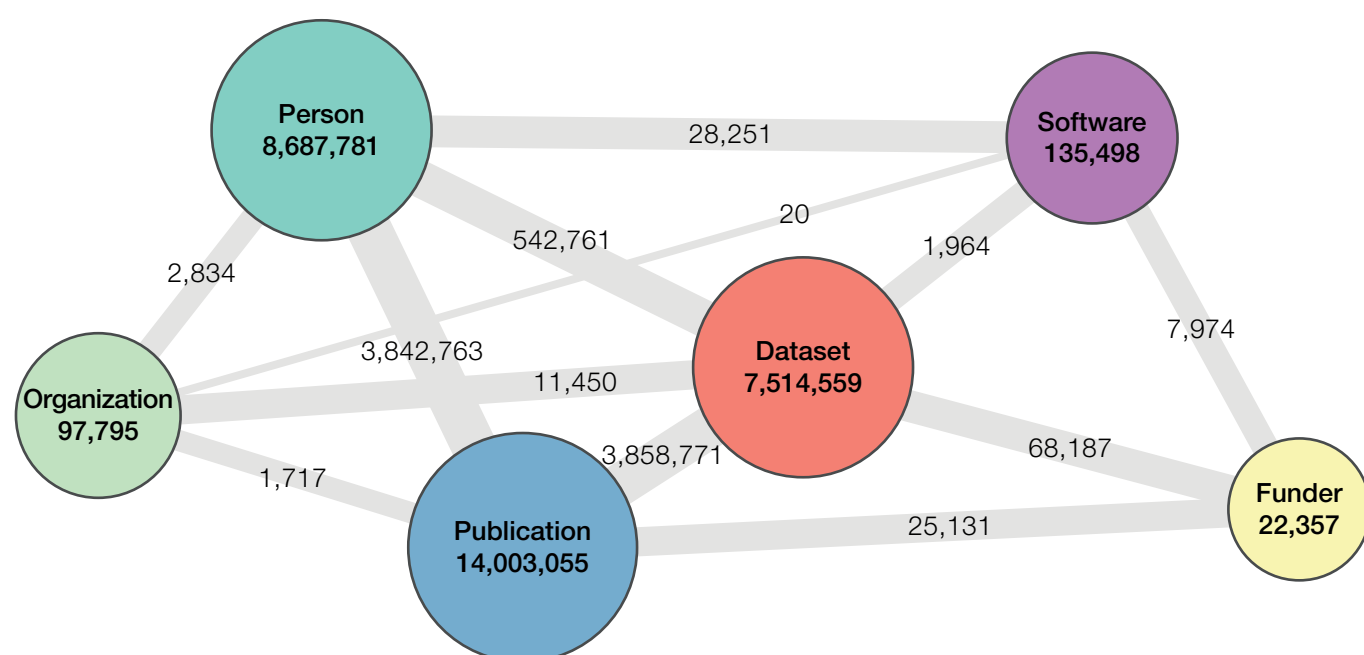
A consolidated PID landscape would involve a fully developed and interconnected PID Graph in which new entities currently associated with emerging PIDs such as grants, RAiD project IDs, instruments and facilities and samples would also be integrated. This is where the actual promise of PIDs lies: a well-functioning PID landscape should result in a machine-readable network of identifiers that would allow to easily tell what datasets stemmed from a given grant or what outputs produced by what organisations were the result of using a given research instrument. Even if we are still quite far away from such a consolidated landscape, the early results of such an interconnected PID network can already be seen in for instance the way ORCID records are able to import references for publications and datasets, organisational affiliations or funding references.

Such an ideal application and integration of PIDs is still far from becoming reality, partly because the prerequisites described above are not fulfilled, partly because not all relevant outcomes and activities are identified with PIDs yet[3]. Upstream, it should even be noted that the question arises as to which actors should/can define which outcomes and activities should be provided with PIDs. Furthermore, the question arises to what extent it makes sense to establish new PIDs, including new technical and organisational infrastructures, in order to make applications such as the PID Graph a reality. Also, we should add here, as it was sometimes mentioned in the interviews we conducted, that a well-functioning PID landscape may raise other critical issues, in particular on the political, legal and ethical levels (control, privacy, academic freedom…). After describing the desired characteristics of PIDs and their integration, we will now look at how organisational arrangements within the PID system should best be designed, both technically and socially.

Organisational commitments and arrangements within the PID system

Ideally, within the **organisational framework**, the operators of a PID system are key players with the PID system as their core task. They are committed, with legal bindings to adhere to standards, procedures and

long-term operation. Additionally, a long-term financial business plan is expected for the PID system, with the operators having an exit strategy to ensure ongoing resolution in case of discontinuments or organisational failures. Therefore, it would be necessary for these operators to possess all necessary rights, to foster transparency, and to pledge neutrality for linking data objects (Bütikofer, 2009).

Regarding **object management,** the identifier clearly indicates to which PID system it belongs. Additionally, PID issuing is considered discrete, so there is no multiplicity in PID issuing and the resolver should periodically check the validity of the associated metadata. Encoding schemes are expected to be scalable, to have a simple (enough) structure and it should be possible for users to resolve PIDs without any hindrances. Also the rules for managing objects and PIDs are expected to be transparent to the user community (Bütikofer, 2009).

As for aspects of **infrastructure and security,** the PID resolvers have to be accessible via distributed public networks and the operators take suitable and approved measures to ensure computer security, including failsafe solutions for the resolution service (Bütikofer, 2009).

Again, these are theoretical requirements that represent an ideal PID implementation. Even though the results of the interviews conducted indicate that PID managers, owners and users largely assume that PID systems meet these requirements, interviews with other PID experts (and partly also the case studies) showed that this is not completely the case.

Consequently, the trust placed in PIDs is partly not factually verifiable, as not all actors can verify all technical, organisational and social promises made by their partners in the PID system. This points to the last aspect to be highlighted in this chapter, namely that PIDs and the PID system are socio-technical in nature.

PIDs as socio-technical infrastructures

Apart from these ideal-type depictions of the nature of individual PID systems and the PID infrastructure as a

---

3   See e.g. the case study on "The role of research funders in the consolidation of the PID landscape" in chapter 10 G.

whole, it is important to emphasise its social conditions: This system or infrastructure consists of service providers, repositories, curation systems, aggregators, indexes, metadata, standards **and people** (Cousijn et al., 2021) - with every item being the product or manifestations of man-made decisions and concepts.

This study attempted to address these findings by understanding the PID landscape (and single PID services) as technical **and** social infrastructure(s) (or socio-technical infrastructures) as a PID is only as good as the services built around it, and PID services are only as good (or trustworthy) as the social adoption and sustainability they achieve. This is clearly expressed by Askitas (2010): *"The main point is that persistence is not about technology but about commitment of communities organised by knowledge domains. (…) The persistence of PIDs is a pledge and a commitment one makes. It therefore benefits if it is built on community values such as trust."*

Attribution of risk and trust in PIDs

The attribution of trust in and riskiness of PIDs relies on the assessment of technical and social characteristics. However, on a second level of analysis, this assessment itself is socially constructed and usually generalises from socio-technical indications to a generalised trustworthiness. The generalising evaluation of trust and risks is based on a reduction of complexity, perhaps even more so in the case of technical properties than in the case of social properties, since the technical ones (even if they are openly accessible and well documented) often elude factual comprehensibility and verifiability, and thus the locus of control is external (Rotter, 1966).

Looking more closely at the social components, PIDs and the PID infrastructure appear to be only as good or trustworthy as specific **trust markers** are attributed to them. These trust markers are technical and social in nature and are difficult to specify globally because they are **constructed** by individuals and different organisations. Furthermore, it must be kept in mind that the attribution of these markers does not mean that the entities to which they are attributed (PID services, PID service providers, etc.) actually justify this attribution in their social and technical characteristics. The attribution does not represent reality, but a more

or less accurate assumption based on the usage of an individual and a more or less considered set of criteria.

Of course there is a certain level of reciprocity needed in a trust relationship between e.g. PID service providers and PID users, because providers have to justify to their users that they are trustworthy (because otherwise their product would not be used). This means that trust markers are important for all stakeholders in a PID infrastructure and it is vital for a well-functioning PID landscape to find out which aspects influence trustworthiness judgments concerning PIDs and which aspects might signal risk or produce a feeling of unreliability. Since we view the greater PID infrastructure as a system entailing both social actors (people providing, managing and using PIDs) and technical actors (the underlying technologies for PIDs), trustworthiness and risk judgments are directed towards both social and technical actors. However, we must keep in mind that trust is not the only reason for usage; lack of alternatives is another reason, along with mandatory requirements and perceived ease of use. In other words, in some cases, people may (will) use services they don't really trust - which is of course as well evident for many services aside from PIDs.

2. Challenge & take-away messages

This chapter provides an overview of the research conducted for this study, the main research questions, as well as a summary of the most important outcomes. This overview provides key take-away points from the interviews and case studies, with a focus on trust markers and risks that are present in the current persistent identifier infrastructure for research.

The challenge

The mandate of the study was to analyse the impact and significance of *risk and trust in pursuit of a well-functioning persistent identifier infrastructure for research*, however, the PID system is often considered a technical entity even though there are social components such as service commitments. Risks and trust are social phenomena whose manifestation can hinder or promote the pursuit of a goal such as the realisation of a functional PID infrastructure more than the de facto existence of functional technology or sound financing. This fact requires a sharp analysis of how trust is created and risks are identified. Consequently, the challenge of the study was to identify the following elements:

a. Which technical properties of PIDs and the PID landscape establish trust or signal risk (in terms of manifest and verifiable properties)?

b. Which social properties of PIDs and the PID landscape establish trust or signal risks (in the terms of manifest and verifiable properties)?

c. Which properties produce trustworthiness or signal risks without being manifest (but rather diffuse in nature) and which largely elude factual verifiability?

This differentiation of risks and the ways in which trust in the PID landscape can be built made it possible to formulate practical recommendations to the various stakeholders. These recommendations will help to promote acceptance and uptake of PIDs in the KE countries and beyond in order to leverage the added values of PIDs for science communication.

The methodological approach was based on logically sequenced work packages that were characterised by different surveys and analyses (literature study, interviews, content analysis, design of the case studies). With each package, the focus narrowed from a broad theoretical analysis of what constitutes risk and trust to the case studies, which looked at the conditions of success and failure of concrete PIDs or PID use cases under the lenses of risk and trust. These findings fed into recommendations on how to achieve a well-functioning infrastructure for research. The most important findings from each survey respectively work

package are summarised in the following section of takeaway messages.

Take-away messages

Literature Study

In the literature study, the main characteristics of PIDs, especially in comparison with other identifiers and unique (in-system) identifiers have been identified. For a general overview of different types of PIDs a paper on the PID Graph recently published by the FREYA project (Cousijn et al., 2021) has proven valuable to our study. This study will also be helpful for us to further our investigation into specific use cases for PID systems.

Concerning the area of research on trust and risk in PID systems, three different types of trust important for this study have been identified:

▸ Interorganisational trust: perceived trustworthiness between different organisations;

▸ Trust in technology: perceived reliability, functionality, helpfulness, ease of use;

▸ Institution-based trust: perceived situational normality and structural assurance of organisations and technology.

According to our working definition of trust as the "willingness of a party to be vulnerable to the action of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other party" (Mayer et al., 1995), risks occur through the trustor's willingness to engage in risky behaviour that stems from the trustor's vulnerability to the trustee's behaviour and the trustor's inability to control the trustee's actions.

Key take-away points:

▸ (The feeling of) **Trust in general** is mainly based on the indicators competence, integrity, benevolence, predictability. (Mayer et al., 1995)

▸ (The feeling of) **Trust in technologies** is mainly based on the indicators predictability, utility, reliability,

functionality, helpfulness, performance, process and purpose. (Lippert & Michael Swiercz, 2005; Mcknight et al., 2011; Söllner et al., 2012)

▶ Regarding utility, functionality and helpfulness, **objective performance and subjective perception** should be distinguished, as perceived usefulness and perceived ease of use are significant determinants of user acceptance of technology and predictors of related behaviour (Davis, 1989); this may be a specific risk if action (e.g., curation) is required from individuals or groups that do not perceive its usefulness.

▶ Trust in technologies is by no means defined barely by technical features or properties, but also to a very large extent by **social factors** (e.g. trust in organisations or their representatives).

▶ In addition to trust (and the trust-inducing indicators), other properties (attributed and factual) also influence the acceptance of a technology, e.g. **ease of use and usefulness**.

Expert interviews & content analysis

The sample of interviewees considered all types of players in the PID infrastructure as defined by KE (Belsø, Rene et al., 2021) and EOSC (Hellström, Maggie et al., 2020): PID Authority, PID Service Provider, PID Manager, PID Owner and PID End User (see for more information appendix 10 C PID roles).

Two different questionnaires were used in this study, with a set of questions to discuss with all experts (regardless of the role they represented in the taxonomy mentioned) and a complementary set of questions that was to be respectively answered only by PID Managers, Owners, End Users and representatives from PID Authorities/ Service Providers as there was a quite different perspective on trust expected to occur during the interviews. You will find the questionnaires in appendix 10 E Interview protocol Risk & Trust in PID Systems.

There were sometimes (and topic-related) significant discrepancies in the interviews, but there was also a high degree of consensus or convergence. There was disagreement on the questions of whether top-down or bottom-up approaches to the design of PIDs and the PID infrastructure as such should be preferred. The consensus was that PIDs cannot exist on their own, but that funding is needed. The experts showed convergent opinions on the persistence of PIDs as a valuable technology (PIDs are here to stay), there was also consensus that PIDs are means and not goals in themselves (they derive their raison d'être from purpose and services) and that there is some sort of competition and coexistence in place, while coordination (e.g. to guarantee interconnection) is needed. There was also an agreement on the fact that what is especially needed is this interconnection of PIDs, e.g., as PID graphs.

Key take-away points:

▶ The interviewees mentioned predominantly **well-established PID**s 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[4].

▶ **Interoperability, value-added services, the availability and interconnectivity of rich metadata** are considered the main general benefits of PIDs. However, PIDs oscillate to a certain degree between this feature (facilitating science and their traceability) and efficient monitoring.

▶ There was broad consensus that PIDs are a **valuable technology** and that there is some sort of **competition and coexistence** in place.

▶ The PID landscape is not an offer-driven marketplace, but rather evolves following a **demand-driven logic**.

▶ Establishing and nurturing a **community** of PID users are key factors for success and trustworthiness. However, there was no consensus on what constitutes a community.

---

4   https://datatracker.ietf.org/doc/html/draft-kunze-ark-01

▸ PIDs are (in line with the outcomes of the literature study) considered **socio-technical infrastructures**.

▸ The interviewees stressed the importance of an **adequately-funded** open infrastructure for the sustainability and resilience of PIDs and the PID landscape. At the same time, there is dissent on the sources of funding.

▸ **Open source and open data** ("forkability") are a key feature for trust and reliability.

▸ In the interviews, there was dissent whether **top-down or bottom-up approaches** are to be preferred in the design of PIDs/the entire infrastructure.

▸ The implementation of PIDs requires a **strategic analysis** of a given situation (i.e., strengths and weaknesses, opportunities and threats); exchange and coordination (in some sort of forum) are considered to be of great importance.

▸ **Funding models for the development of a PID infrastructure show a dichotomy** between member-based marketing-savvy non-profit organisations and (for instance) national library networks, with each of these having their own strengths and weaknesses.

▸ Some experts stated that people trusted PIDs a) because the **risks associated with their use were too amorphous** to be evaluated and b) because they believed that PIDs could not simply fail once they reached a certain level of adoption.

▸ **Risks of PID Systems tend to be overlooked** by PID End Users, Managers and Users. The experts primarily identified these **gaps and possible risks** in infrastructure:

› danger of an infrastructure enclosure,
› dependency on a wide range of diverse actors and technologies,
› a lack of global inclusivity,
› a lack of control mechanisms,
› a lack of funding,
› a lack of knowledge and manpower needed to guarantee high quality metadata and services,
› financial sustainability (e.g. for funder or grant IDs) and affordability,
› organisational failure,
› a lack of uptake (of PIDs),
› loss of trust from the community,
› technological risks, e.g., DOI/Crossref server outage,
› general URL failures or general internet failure.

Case studies

The results of the content analysis fed into the case studies, which aimed at gaining a deeper insight into particular aspects and situations ("zoom in"), providing more details and linking to the PID literature. You will find a list of all case studies produced in appendix F from chapter 10.

The main criteria for the selection of topics as case studies were:

▸ The topics should be revealed by the content analysis of the interviews with PID experts;

▸ The topics are highly relevant for the report on risk and trust of PID infrastructures;

▸ The topics require further investigation, with complementary information sought from external sources.

The following items list some general findings that were encountered in the case studies and a few short remarks on these.

Key take-away points:

▸ **There is a dichotomy of 'technical' vs 'admin-oriented' PIDs:** Technical PIDs are promoted via (mostly) bottom-up workflows by researchers who perceive the need for the persistent identification of objects they are regularly working with, be it geo samples (IGSN), research equipment and facilities (PIDINSTs) or clinical trials (ISRCTNs). On the other hand, 'admin-oriented PIDs' tend to be implemented in a more (if not completely) top-down fashion by a range of stakeholders that do not necessarily include researchers -– these are typically institutions, publishers and research funders – in order to introduce some much-needed standardisation in the scholarly communications landscape for research information management purposes.

▸ **Community awareness or meeting the needs of a community** was reported as consistently inspiring

trust. However, the use of the term "community" turns out to be very heterogeneous and it includes researchers (of all disciplines in the case of ORCID or just one for the RePEc Author Service), infrastructural institutions (libraries for ROR), scientific organisations (ROR, ORCID, DAI), publishers (ROR, ORCID) or funders (ROR, ORCID). It might be advisable to consider aspects of risk and trust each from the perspective of these different communities.

▸ The case studies show that trust is significantly based on **brand image**. The use of this effect obviously provides PID players with a plus in trustworthiness and helps to achieve uptake.

▸ Especially for the early-stage, discipline-specific attempts at adopting PIDINSTs, a risk of **technical divergence** is presently looming over the whole domain. Several examples have been provided in the related case study on how different initiatives are using different PID standards such as DataCite DOIs, ROR IDs and DOIs for articles on research facilities published in specific journals as the basis for the persistent identification of research instruments and facilities. This risk is also identified in other emerging PIDs such as grant IDs, with a possible choice between Crossref grant IDs and RAiDs.

▸ The widespread worldwide adoption of DOIs and ORCIDs suggests a straightforward mechanism for PID implementation, but both the case studies and the interviews highlight a much more frequent status of landscape **fragmentation**. This situation is particularly severe for emerging PIDs and it applies to both the technical solutions and (especially) the community management workflows.

▸ Emerging PID domains like OrgIDs, PIDINSTs, IGSNs, grantIDs and ConfIDs[5] are likely to be simultaneously developed and matured. The fact that the underpinning PID infrastructure and the community coordination for all these initiatives mostly sit with a single actor in the community – namely DataCite – is also seen as a significant **bottleneck risk**, both regarding different

---

5   ConfIDs are supposed to provide persistent identification of scientific conferences, see also Franken et al. (2022)

requirements to curate/manage and generic upscaling. This would be about an overload for the services – especially in the area of community engagement and management – to be provided by a single organisation. This risk could result in a slowing down of the progress in the implementation of these emerging PIDs and the inability to counter the natural fragmentation that tends to arise from early-stage efforts.

▶ A widespread and harmonised uptake of a PID requires **clear use cases** shared by different actors and communities, ideally technical and admin-oriented actors. These use cases are not always sufficiently defined from the outset and such a lack of definition involves a risk of not addressing the needs of specific stakeholders. The case study on PIDs for instruments provides perhaps the best illustration for this issue.

▶ The **advent of a technically superior competitor** with a more global claim does not necessarily lead to the abandonment of a PID, as shown in the case study devoted to the RePEc Author Service. However, this is a severe risk for any PID with a limited scope (either from a disciplinary or a stakeholder-specific perspective) and should always be kept in mind. Key factors in the outcome of such a clash between competing standards include not just the admin-oriented use cases, but also factors such as awareness and use of a given PID among scientists and the attractiveness of the use cases/functionalities for them.

▶ As shown in the case study devoted to OrgIDs, **commercial and public services may have the opportunity to coexist** insofar as they offer different albeit complementary workflows, in this case ROR and Ringgold (although the latter one is not handle based). The tension between public and commercially provided PID services is always there and has frequently been highlighted in the interviews, but such coexistence is expected to persist especially in areas related to emerging PIDs.

▶ **Technical maturity** is helpful in building trust among researchers and organisations, potentially including research funders. It also facilitates PID

sustainability and maintenance. PIDs being socio-technical pieces of infrastructure, there is a balance to be achieved between technical maturity and the maturity of the community management workflows. PID landscape analyses often show the level of maturity for specific solutions and this includes both areas. However, a case study (RAS) in the series also shows a service intensively used by scientists but whose technical basis and organisational sustainability is difficult to assess.

▶ The **clear, dedicated mission** of a service and the commitment of the organisation providing it are essential requirements for a functional use and provision of PIDs. If any of these are missing, as in the case of PURL, the service becomes not sustainable – or their implementation lacks trustworthiness (as in the case of the DOIs minted by specific publishers or by a repository under investigation).

▶ In both the case studies and in the interviews it became apparent that it is advisable to **distinguish between trust and trustworthiness**. PID owners and managers trust PID service providers and authorities, PID users trust owners and managers (and implicitly providers and authorities) without necessarily having the evidence for supporting this trust. This can be witnessed in the case study of failed PID implementations. This longing for trust goes so far that - as in the case of PURL - even obvious references to the beta status of a service are ignored.

3. Recommendations

This section provides a series of recommendations for a harmonised PID implementation classified by stakeholder level. These recommendations have been structured in a way that allows the key players in the PID landscape to be identified. In such a complex framework it's difficult to unequivocally categorise stakeholders as international, national or sub-national, but an attempt has been made to list all relevant ones starting with those who may be able to provide some governance and then moving downwards on the scale of PID implementers and users. The roles listed below include 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.

The following paragraphs will list recommendations addressed at the following players: National-level stakeholders, Research funders, PID Service Providers, Institutions, Researchers, Publishers, and (not yet existing) PID Federation.

A. National-level stakeholders

There is a wide range of relevant players at a national level, all of whom should ideally be on the same page with regard to a national-level PID strategy. These include National research and education networks (NRENs), national organisations for research e-infrastructure (the likes of SURF or Jisc, sometimes overlapping with NRENs), research funders, institutions and also projects such as the Open Knowledge Base (OKB-NL) in the Netherlands or (on a wider scope) the EOSC-funded FAIR-IMPACT or FAIRCORE4EOSC.

The key recommendation for national-level stakeholders could be summarised as **"Put your house in order" - Establish a national PID roadmap**. All items listed below can loosely be considered part of this high-level advice.

There are currently strong differences across countries but many of them already have some well-established coordination mechanisms to start with – national-level ORCID and DataCite consortia being the most frequent forums. The recommendations below assume a situation where not much progress has yet been made in this area, but it's worth pointing out that a number of countries have already gone through all these recommended steps.

A1. Identify the key stakeholders at a national-level in your country and explore the feasibility of bringing them together to discuss the PID implementation strategies. These stakeholders may include – but are not limited to – research funders, national libraries, research-performing organisations and large cross-disciplinary research supporting infrastructures. Some frontrunner countries may serve as best practice examples for the purpose.

A2. Explore the feasibility to discuss national PID implementation strategies with relevant stakeholders. The purpose is to design a PID strategy with input from the various relevant stakeholders in the country. This strategy should ideally state which PIDs ought to be prioritised in the gradual development of a comprehensive PID layer at a national level.

A3. Form national-level governance instruments. As per the emerging best practice examples, these could include a PID Advisory Board with the key stakeholders represented in it. The possibility of drafting a PID policy that underpins the agreed strategy forward should also be considered

A4. Be aware of the socio-technical solutions in place for various PIDs. Although a consensus is quickly emerging, sometimes there are competing solutions in place for the same PID – such as for instance for Organisation IDs. The case studies produced in the course of this work may help in providing an up-to-date insight on the current PID landscape. This landscape is however quickly evolving and it's worth keeping up with the various ongoing international initiatives.

A5. Do not reinvent the wheel. Before even starting to implement a national PID strategy, familiarise yourself with the challenges faced by and the solutions adopted in other countries. An effective way of doing this is by joining – even as an observer – international coordination initiatives on the design of national-level PID strategies such as the dedicated WG within the RDA[6].

A6. Design an awareness-raising communication campaign highlighting the relevance of this domain for the progress of research management and administration. This effort should mainly target institutions, while keeping in mind that researchers largely remain the key end-users of PIDs. Parallel top-down and bottom-up communication strategies should be considered in this design.

---

6   The RDA National PID Strategies Working Group brings together various national agencies and initiatives with the purpose of mapping common activities and reporting on the specific PIDs adopted in the context of national PID strategies, https://www.rd-alliance.org/groups/national-pid-strategies-wg.

B. Research Funders

The national-level research funder landscape shows great variations across countries, as explored in the case study devoted to the topic. Countries where there is one or more 'hegemonic' funders such as the Netherlands, Germany, the United Kingdom, Portugal or Austria may be particularly well placed to benefit from a significant funder engagement in PID implementation initiatives, but the research funding landscape is usually quite complex. These recommendations are for all funders regardless of their size.

B1. Make sure you are represented in – or at least informed about – national-level coordination initiatives described in A1 above.

B2. Be aware of what PIDs are relevant for your activity, including for project proposal evaluation, reporting on funded research outputs and grant identification.

B3. Consider assigning grant IDs to your grants whenever possible, allocating the appropriate human and technical resources to make it possible. Best practices are already available in this domain that may be replicated.

B4. Consider requiring specific PIDs from your funded researchers, even for applicants to your funding calls.

B5. 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. Adoption of new PIDs should only be considered when they clearly add value, as it may otherwise just add to the considerable landscape fragmentation. Adoption of any new PID should first be discussed at a PID policy level.

B6. Be aware of funder-specific coordination initiatives at a national and international level, promoting and joining them whenever possible. Science Europe, Europe PubMedCentral and the Crossref Funder Advisory Group could all be suitable forums for such a policy-making and technical collaboration.

C. PID Service Providers

The current PID landscape is a very complex and fragmented one, with a wide range of stakeholders required to play a relevant role in the emergence of a coordinated approach to PID implementation. PID service providers such as – among others – Crossref, DataCite, ROR, Ringgold, the ISSN International Centre or specific national libraries in KE member countries are seen as key actors in the effort to harmonise this PID landscape[7]. This is mainly due to their mostly international nature, which significantly helps the coordination effort. Strong differences also exist between providers and frequently there are competing solutions, but the recommendations below are generic ones that should apply to them all.

C1. Ensure the sustainability of your initiative from both a technical and an economic perspective. This includes having contingency plans[8] in place for an eventual discontinuation of the PID providing services

C2. Business models underpinning PID provision must be clear and transparent.

C3. The underpinning data and process documentation should be open and allow their being taken over by the community in case of failure

C4. Explore the possibility of taking part in a PID Federation as described in the literature. While it may be too early at the time of writing for such an international, cross-PID coordination and governance body to emerge, the opportunities and the need for such a mechanism will only increase as the PID landscape consolidates.

C5. The recommendation for **putting together a PID Observatory** would most effectively suit such an international PID Federation, but the present absence of this specific coordination mechanism does not prevent the observatory from being a desirable development.

C6. Have mechanisms in place to report failures on specific implementations of the PIDs you provide (such as the DOI error report).

C7. Were the PIDapalooza series of events to be permanently discontinued, **come up with possible suitable alternatives** for sharing developments with the PID user community. As mentioned in the introduction to section A above, EOSC-funded projects such as FAIR-IMPACT and FAIRCORE4EOSC may be able to play a relevant role in this regard.

C8. Ensure transparency in the communications towards users of the PIDs you provide. This includes identifying the appropriate communication channels – blog posts and webinars being two of the most used mechanisms at the moment – and regularly reporting on any relevant progress in the implementation of a specific PID.

---

7   The PID provider landscape also includes IGSN and SciCrunch as less mainstream initiatives, and new entrants are expected to join as for instance RAiDs become more widespread

8   See as an example the strategy for coupling URNs to handle IDs in requirements PID-12 and PID-45 in Wittenburg et al. (2017), https://zenodo.org/record/1116189.

D. Institutions (Research-Performing Organisations, RPOs)

Especially in big countries, there are too many institutions of many different kinds in the research landscape for all of them to be represented in national-level governance bodies or initiatives. However, RPOs have a key role to play in the awareness-raising process towards 'their' researchers and are thus a very important player to keep in the loop. Moreover, institutions are the key stakeholder for leading the implementation of specific PIDs such as OrgIDs and DataCite DOIs.

D1. Make sure you are represented in – or at least informed about – national-level coordination initiatives described in A1.1 above.

D2. Consider the possibility of drafting an **institutional PID policy** (see an example here).

D3. Raise awareness of the existing and emerging PID landscape among institutional researchers, including prompting them to use the appropriate ones. Effective communication strategies from institutions, research funders and publishers are therefore critical for raising awareness of this area. Institutions have a particularly important role to play in terms of offering support to make specific PIDs regularly used by researchers.

D4. Be aware of your key role in the implementation of specific, admin-oriented PIDs: make sure your organisation has an up-to-date ROR (eventually a multiple-level one when this feature becomes available) or alternatively an ISNI-based OrgID[9]. Become a DataCite member if not already. Join the communities dealing with the implementation of specific PIDs.

D5. 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).

D6. Be aware of technical PIDs directly emerging from researcher communities in a bottom-up fashion such as (among others) PIDINSTs, IGSNs, CETAF or DiSSCo.

D7. Stay informed about (still to come) mechanisms to issue (and share and use) institutional PIDs such as RAiDs or PIDINSTs.

E. Researchers in their institutional context

As end-users for most PIDs, researchers are key stakeholders for PID implementation. However, due to the current complexity of the PID landscape, researchers are easily left behind[10]. This is partially due to the widespread perception that PIDs represent an additional administrative burden. Also, it is self-evident that the individual practice of researchers is closely linked to and impacted by their institutional context as described above (communication, support, strategy…).

E1. If you do not have an ORCID already, get one. Don't set all the information in your ORCID profile to private as this will render it useless. If you think you may have multiple ORCIDs, let your research support team know and they'll fix it.

E2. Follow funders', publishers' and institutional requirements in this area. This is likely to already include a requirement to use DataCite DOIs for your datasets and software, plus eventually Crossref DOI-based grant IDs in the funding acknowledgements in your funded manuscript.

E3. Stay actively informed about ongoing PID-related initiatives – such as grant IDs issued by research funders – and their relevance to your research. Your institutional research support officer will be able to provide more info if/where needed (if you're not affiliated with any RPO, you are likely to have colleagues in project consortia who are).

E4. Discipline-specific efforts to persistently identify relevant digital objects such as research

---

9   Ringgold is one of such alternative options, https://www.ringgold.com/ringgold-identifier/, but there are others as well.

10 G Macgregor, BS Lancho-Barrantes, DR Pennington (2022). Exploring the concept of PID literacy: user perceptions and understanding of persistent identifiers in support of open scholarly infrastructure (In press). https://doi.org/10.48550/arXiv.2211.07367

instruments and facilities may well be going on in your field, including under the EOSC umbrella. Consider the possibility of engaging with collaborative initiatives in this regard.

E5. Be aware of your institution's PID policy if there is one. Some admin-oriented PIDs such as OrgIDs may not seem that valuable to you, but they are important and publishers will eventually require them to be included in their manuscript submission systems.

F. Publishers (including Diamond OA publishers)

Same as research funders, publishers come in many sizes and with varied technical capabilities and business models. Publishers have in any case a key role to play in the implementation of PIDs by making opportunities available to researchers/authors to use them when submitting their manuscripts.

F1. 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.

F2. Include entries for additional PIDs in manuscript submission systems as these PIDs become more widely implemented. ORCIDs and DOIs for datasets are already being required regularly, but OrgIDs (RORs) and grant IDs for acknowledged funded projects should soon follow.

F3. Provide information snippets to researchers/authors on why PIDs are important.

F4. Be aware of the level of maturity of specific PID initiatives in order to allow references to these to be included in manuscripts.

F5. Make sure the PIDs you provide in your publications **are operational and resolve correctly**. Do include some explicit indication on the webpage when a DOI is under registration.

F6. 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.

F7. Diamond OA publishers may often lack the human and technical resources to ensure a widespread PID implementation within their journals, but they are recommended to:

▸ try to use as many PIDs as possible (with DOIs the bare minimum)

▸ be aware of the developments on the PID landscape – including emerging initiatives to provide specific support in this domain like the *Diamond OA Capacity Centre*

▸ join the appropriate coordination initiatives where best practices may be shared and replicated

G. A possible PID Federation

No such thing exists at the time of writing as a PID Federation, but the recommendation is to explore the feasibility of setting up such a body as described in the FREYA project literature. Joining a possible PID Federation is already included in the recommendations for PID service providers in C4 above, but this PID Federation is a very ambitious endeavour – which the current PID landscape is nevertheless evolving towards in areas like RAiD implementation. While this governing body with all PID providers and additional relevant stakeholders is built, there could also be opportunities for 'smaller PID Federations' to be implemented for instance by the Knowledge Exchange.

G1. Support sustainability of organisations within the PID landscape from an economic and a technical perspective. This includes making sure contingency plans are in place should any player go out of business

G2. Ensure technical resilience, openness and transparency of such organisations, including storage of data and documenting relevant operational processes

G3. Set up a PID Observatory providing an up-to-date and comprehensive snapshot of the PID landscape, its key players and best practice case studies in PID implementation by specific PID types, stakeholders and/or countries

G4. Provide a mechanism to report failures in PID implementation – this may be a collection (or aggregation) of the various help desks made available by the different organisations – and make sure the reporting is followed by some corrective action

G5. Conduct generic and specific communication activities on the value of the aggregated PID network. This may include setting up a follow-up for the PIDapalooza series of events if these are not resumed

G6. Make sure the coverage of PID initiatives and stakeholders is truly inclusive, with representation of all regions in the world, including the Global South

H. Knowledge Exchange

There is also a series of direct recommendations to the Knowledge Exchange as commissioner for the current study on PIDs. These mainly reflect the need to take advantage of already well-established collaboration mechanisms across its six member countries (Denmark, Finland, France, Germany, Netherlands and the United Kingdom) to address this additional area of activity.

H1. Keep operating the PID Working Group, ideally opening it up to like-minded members from stakeholders not currently represented in the group such as research funders. Where there is a lead for national-level PID implementation in a KE member country, that person or stakeholder should be represented in the group.

H2. Organise the appropriate communication initiatives to highlight the work undertaken by the KE in this domain and its continuity. These may include a workshop or series of workshops to discuss the results of the current study and its follow-up actions. Webinars on the topic could also be a valuable communication strategy to highlight for instance best practice case studies in the PID implementation strategy followed by specific stakeholders. These communication activities would mainly be addressed to relevant players in the KE member countries but could also occasionally feature outstanding work in the area in neighbouring countries.

H3. Maintain an up-to-date snapshot of the status of PID implementation in the six KE member countries and consider supporting specific coordination initiatives among them.

H4. Explore the connections between PIDs and the various additional areas of activity in which KE is involved, with an emphasis on Open Science. Parallel efforts going on in KE member countries like the Dutch Open Knowledge Base (OKB-NL) or the EOSC-funded projects FAIR-IMPACT and FAIRCORE4EOSC could also be relevant candidates for specific meetings and webinars.

H5. Promote a specific public-sector-driven approach to PID implementation where the advantages of PIDs are made clear to policy-making stakeholders in the KE member countries and the appropriate mechanisms to ensure sustainability are discussed.

4. Community

The following section provides some insights, ideas and questions regarding the meaning and use of the term "community". The reason is that most experts interviewed for this study have highlighted the relevance of the community effort for the consolidation of PIDs. As a socio-technical infrastructure, it is clear that in order to experience a significant uptake, PIDs – and above all services associated with them – need to be perceived as valuable and be in turn promoted by "the community".

This term is understood to have a different meaning for different stakeholders (and PIDs). If a given country has for instance a national-level PID Advisory Board in charge of designing and implementing a strategy for a nationwide PID adoption, then the term community may be taken to mean all those stakeholders represented in this Advisory Board, i.e. the likes of – among others – national agencies, research funders, national libraries, associations of research-performing organisations and perhaps publishers.

However, if we think about for instance the way the adoption of persistent identifiers for research instruments and facilities is currently progressing, it's interesting to see that practically none of the stakeholders mentioned above are part of the community of (early) adopters, which in this specific case involves researchers with a strong disciplinary alignment (geosciences), an RDA Working Group (PIDINST) and DataCite as a supporting organisation ensuring that the emerging technical standards are aligned with other PID domains[11].

This discrepancy on the meaning of community may then be influenced by the nature of a specific PID. As per the two broad groups into which the PID infrastructure is categorised in this study, 'admin-oriented' and 'technical' PIDs, both groups tend to have their specific communities. The IGSN is another example featured in the case studies where this bottom-up adoption means that it's essentially researchers who are governing the evolution and gradual adoption of this standard via the bottom-up governing body IGSN e.V.[12]. Of course this distinction is particularly relevant for emerging PIDs: once a PID becomes consolidated (such as DOIs or ORCID), these two different communities that existed at an earlier stage tend to merge into a single, all-encompassing one. But when considering the general PID landscape as a whole, we are still quite far from this consolidated stage.

Besides these differences on the meaning of community for different PIDs, there are further aspects to be kept in mind when trying to explore what the concept actually entails. There is a strong public/private dichotomy underpinning the development of and the business models behind PID implementation that has also been raised when discussing risks and trust-related issues during the interviews. From this viewpoint, the public actors would be a specific category, mostly comprising PID users and managers, while the PID service providers and some PID managers would be a mix of non-profit organisations and commercial stakeholders. Aware of the fact that it's difficult if not impossible to keep the full PID infrastructure under a public ownership/governance, the recommendations in chapter 3 make emphasis on the need for coordination mechanisms and technical interoperability across solutions and stakeholders, as well as on the need for openness and transparency.

This public/private sector divide also manifests itself in the adoption of PIDs at a very basic user level: when we speak about consolidated PIDs like ORCID, it's the publicly-funded research environment we are considering, i.e. universities, public research centres and institutes, etc – as well as the researchers working in them. It's hard to ascertain how much awareness of and appetite for PIDs there may be in the private research sector, meaning essentially industry. While research funders and publishers may both require ORCIDs when applying for funding or when submitting a manuscript, researchers in industry may not be covered by any of these workflows. This lack of clarity on the level of interest for PIDs within the private research sector is one of the limitations of this study, especially in view of the EU statistics showing that more than half (55.4%) of full-time equivalent researchers in the EU worked in business enterprises, 32.6% in higher education and 11.1% in the government sector in 2020[13].

---

11 While there are emerging initiatives to include persistent identifiers for instruments and facilities in national/regional research portals like Research.fi or the Flanders Research Information Space (FRIS), these are somewhat below the radar and hard to map – hence the recommendation for PID Observatory above.

12 https://www.igsn.org/about/

13 Eurostat (2021) Statistics Explained: R&D Personnel. Researchers. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=R%26D_personnel#Researchers

The ever increasing number of academia-industry collaborations within joint projects may offer a way forward in order to expand the PID outreach, but whether the PID infrastructure can be considered to be consolidated while largely addressing public-sector researchers remains a little-explored open question.

The stakeholder categorisation by PID roles – PID authorities, service providers, managers, owners, users, etc. – is another way of examining the community of actors involved in the PID implementation. The provenance for this specific classification is the previous landscape analysis document produced by the Knowledge Exchange (Belsø et al., 2021). A summary of the classification that is based on the PID policy for EOSC is provided in appendix 10 C.

The recommendations are classified by stakeholder, and it's this classification we will use as a starting point to analyse opportunities and challenges from the perspective of each actor. First, in order to illustrate the discussion on the various stakeholders and PID roles within the community, it may be useful to examine a few example workflows for the issuing of PIDs and what roles different stakeholders play in them. This is mainly aimed to show the variations across PIDs and the complexity involved in coordinating these different actors. Some of the examples shown below have deliberately targeted emerging workflows in order to provide a sense of the missing actions and interactions at the time of writing.

1. DOIs for datasets minted by Higher Education Institutions (HEIs)[14]

Stakeholder	Role
DataCite	PID service provider: DataCite assigns a stack of DOIs (via a prefix) to institutions and guarantees persistence and the correct resolving
HEI	PID manager: institutions mint dataset DOIs for their researchers on the basis of their DataCite subscription. Other third-party services like CERN may also play this role
Researcher (at HEI)	PID user: researchers include the URLs in the data statements on their manuscripts. Datasets with their PIDs are listed among the researchers' outputs
Publisher	PID user: publishers use these dataset DOIs as provided by researchers and include it in the metadata set exported to Crossref – allowing the references to be picked in for instance individual ORCID profiles
Research Funders	PID user: funders use the dataset DOIs to identify outputs arising from their funded projects and link them to these for project assessment purposes

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***

*The correct resolving of PIDs is a joint task shared by the PID manager (the research funder) and the PID service provider (Crossref). It's worth mentioning here that the first grant ID minted for a Wellcome-funded grant, https://doi.org/10.35802/207522, a snapshot of whose landing page in Europe PMC is still shown in Kiley et al. (2018) at the time of writing, does not resolve due to a URL update not reflected in the corresponding DOI. This is particularly revealing in a study examining the risks associated with PIDs and highlights the need to be extremely careful when updating any metadata element in a third-party-hosted grant ID landing page at which a PID is pointing.** Only a few funders – such as the Wellcome Trust or the European Commission – have started issuing DOIs for their grants at the time of writing. *** These integration processes haven't yet started.*

3. OrgIDs issued by ROR or Ringgold

14 A Lahtinen et al (2020). Choosing and implementing persistent identifiers: Guide for research organisations. https://doi.org/10.5281/zenodo.4395767

Stakeholder	Role
Service provider (ROR or Ringgold)	PID service provider: OrgIDs assigned to batches of organisations (ideally on a multiple-level basis). Available workflows for maintaining up-to-date-records with guaranteed PID resolving*
National office	PID manager: national offices with an up-to-date database of relevant research-performing organisations in the country arrange the issuing of OrgIDs for all those organisations (usually via Ringgold)*
HEI	PID user: institutions ensure they have an OrgID, individually registering where there is no national-level initiative, and disseminate it to their researchers. They also use their OrgIDs in workflows around transformative agreement management. OrgIDs are also included as a part of the RAiDs HEIs mint*
Researchers	PID user: prompted by their funders and HEIs, researchers include their OrgIDs in the metadata provided upon manuscript submission to ensure harmonised institutional affiliation and smooth management of transformative agreements*
Publisher	PID user: publishers allow OrgIDs to be provided on affiliation fields and use them for transformative agreement management (including notification services to HEIs)*

All these processes are only being started at the time of writing. It is expected that frontrunners will share their best practices

4. URN:NBN (National Bibliography Number) issued by various national libraries[15]

Stakeholder	Role
National Library	PID service provider: manages the urn:nbn namespace PID manager: issues PIDs for the appropriate items (mainly publications, but potentially also authors or organisations)
HEIs/Researchers	PID users: urn:nbns issued by national libraries may be used as an alternative persistent identification system (or "fallback mechanism") to DOIs or handle IDs
Other organisations	PID manager: organisations interested in setting up a new URN namespace can manage the issuing of PIDs once they've been authorised to do so (see for instance the urn:uvci namespace established by the EU eHealth Network for a Unique Vaccination Certificate/assertion Identifier for Covid-19) Open Access repositories may also operate as PID managers by using urn:nbns to persistently identify the records they store (sometimes alongside DOIs), see for instance https://edoc.ub.uni-muenchen.de/29298/

Some of the example workflows described above show how complex it is for a default stakeholder classification – such as that based on PID roles shown in appendix C – to provide a good fit for all the very diverse case studies for their application in the PID domain. For analysing the opportunities and challenges associated with the various stakeholders

15 "[The] National Bibliography Number (NBN) is a group of publication identifier systems used by national libraries in countries such as Germany, Italy, Finland, Norway, The Netherlands and Sweden", https://en.wikipedia.org/wiki/National_Bibliography_Number

this chapter loosely follows the stakeholder classification used in the recommendations in chapter 3, adding an additional level (international, national or sub-national) to categorise these actors and the scope of their interactions.

PID community stakeholders: international level

Governing bodies

These stakeholders, termed PID authorities in the PID role classification in Appendix C, are the bodies sitting at the top of the PID stakeholder hierarchy. Their role is to establish and enforce harmonised processes for creating, approving, maintaining and terminating PID standards that PID service providers should follow. There may be different hierarchical levels of governing bodies such as the DONA Foundation in charge of the Global Handle Registry, of which the International DOI Foundation is a member or MPA (Multi-Primary Administrator).

PID Service Providers

These are actors such as Crossref, DataCite or others, that provide PID services in conformance to a PID Scheme, subject to its PID Authority. PID Service Providers are responsible for the provision, integrity, reliability and scalability of PID Services, in particular the issuing and resolution of PIDs, but also lookup and search services.

PID Federation

The possibility and potential usefulness of setting up an international PID Federation was explored by the FREYA project (Brown, 2020b) in view of the ever expanding range of entities for which persistent identification was being considered. At the time of writing such a body has not started being built yet since the PID landscape hasn't yet achieved the required level of consolidation, but gradual progress in PID implementation in forthcoming years could well lead to its being set up. A number of recommendations are provided in chapter 3 for such a body should it finally emerge.

A PID Federation would also make it possible for the leading initiatives in PID implementation that are currently mostly taking place in Europe, the United States and Australia to rely on some degree of coordination. Right now, while some mechanisms allow the information to flow across geographies, it is remarkably difficult to collect a wide snapshot of who is doing what and where.

RDA Working Groups

A number of international working groups operating within the Research Data Alliance (RDA) – such as the PIDINST WG on PIDs for instruments and facilities or the National PID Strategies WG – are conducting a very effective international, bottom-up coordination activity, especially in the area of emerging identifiers or PID initiatives. While several relevant stakeholders in the landscape tend not to be represented in these WGs such as research funders or HEIs, national bodies are usually aware of the work being done and will be able to bring it into their discussions.

PID-related projects and initiatives

There are a good number of international projects dealing with PID implementation and offering valuable opportunities for coordination, often operating within the framework of the European Open Science Cloud (EOSC). A non-comprehensive list of these includes EUDAT, a project turned into a legal entity that provides its own PID services[16], the DICE (Data Infrastructure Capacity for EOSC) H2020 project or the recently kicked-off FAIRCORE4EOSC and FAIR IMPACT projects respectively led by the CSC in Finland and DANS in the Netherlands. The EOSC Association also operates a Task Force for PID Policy and Implementation whose activity should foster international collaboration to define and apply harmonised PID implementation workflows.

International coordination bodies

Stakeholders such as research funders are listed in the national-level group of actors below, but there are organisations such as Science Europe, the Global Research Council and others which could play a key role in promoting a coordinated PID implementation

---

16 See for instance J Nordling, M van de Sanden (2022). Developing the B2INST service for registering and persistently identifying instruments. FAIRsFAIR Implementation Story. https://doi.org/10.5281/zenodo.6411786

agenda among their members at an international level. The bodies could also provide critical support to ensure that a wide-range PID implementation is not limited to the most advanced geographies but that it gradually trickles down to the whole world in an inclusive way.

Publishers

Publishers can be international or national-level players, but their role for the adoption of specific PIDs is critical. Publishers are usually Crossref members and this means an opportunity to harmonise the workflows for PID use by and collection from researchers via manuscript submission systems. The adoption of DOIs for publications meant a first pioneering PID implementation and was very much driven by publishers and Crossref. These days discussions are already taking place among publishers on how to best collect OrgIDs, grant IDs and other emerging PIDs and how to share the identifiers in the article metadata they provide via Crossref, allowing a further deepening in the realisation of the PID Graph.

National-level stakeholders in the PID landscape

National offices

In a study carried out for the Knowledge Exchange, these are mainly meant to be organisations like SURF in the Netherlands, CSC in Finland or Jisc in the United Kingdom that are on the one hand KE national representatives and on the other hand able to lead the national-level discussion on a PID implementation strategy. It's fair to mention though that KE national reps may also be research funders such as the German Research Foundation (DFG) which do not fall under the "national office" category.

National offices often play a leading role in the ORCID Consortia that are currently operating in many countries (and specifically in all six KE member countries). These ORCID Consortia are worth mentioning as a separate stakeholder since while they were originally designed to govern, promote and oversee the national-level implementation of ORCID, their scope has regularly expanded in the course of time into additional PIDs such as OrgIDs. Because these consortia typically include HEIs, the well-established coordination networks are very useful for its members to jointly

explore the most suitable mechanisms for the implementation of other PIDs.

NRENs

National research and education network (NRENs) are specialised internet service providers dedicated to supporting the needs of the research and education communities within a country. NRENs may also play a relevant role in defining the PID strategy for a given country, especially in the absence of other national offices with specific competences in the matter. The Danish e-Infrastructure Cooperation (DeiC) is for instance the national representative from Denmark in the Knowledge Exchange network. Jisc is another example.

A significant number of NRENs were represented in the 1st IRISC workshop for Identity in research infrastructure and scientific communication held at CSC in Helsinki in Sep 2011. This conference predated the launch of ORCID and gave rise to the PIDapalooza series of events that started in 2016 in Reykjavik, Iceland.

NRENs and national offices are often the same organisations, as it makes sense to bring the internet provision services for research and education communities together with the management of research information in a very generic sense.

Research funders

Research funders are critical players in the PID implementation landscape for their close connection to researchers and their opportunities for defining a policy whereby a specific PID is required from researchers they fund and/or the institutions they are affiliated with. This has so far been the case for ORCID, which funders in many countries are already requiring from researchers when submitting project proposals. OrgIDs may be expected to follow suit in due time, but at the time of writing the key development certain funders are starting to undertake in the PID landscape is the issuing of grant IDs for the projects they fund. The case study "The role of research funders in the consolidation of the PID landscape" examines this recent development in more detail.

HEIs/research centres/institutes

Publicly-funded research-performing organisations such as universities or research centres also play a key

role in PID implementation. Not only they mediate between 'their' researchers and all the external bodies whose PID policies may affect them, but they have the necessary communication mechanisms towards researchers to ensure they meet the requirements posed by research funders or publishers and to keep them informed about recent developments in the PID landscape. HEIs could also play a critical role in the adoption of emerging PIDs like identifiers for research instruments and facilities. These institutions have traditionally had strong collaboration mechanisms across them – for instance for Open Access and Open Science implementation – and PIDs could easily fall within this joint activity.

HEIs institutions and research structures often operate platforms in the form of publication repositories, data repositories or CRIS systems. All these platform types use PIDs to identify scientific content and are therefore relevant for the study: CRIS systems integrate information on the scientific activities of the institutions and rely on identification of outputs, but also on identification of funding or equipment and facilities in the future. While repositories for text publications have largely implemented specific PIDs, this only applies to a limited extent to data repositories. As of 5th October 2022, the re3data directory lists 2,961 repositories. For 2,757, the registry provides information on PID implementation, with 1,342 (49 %) of these 2,757 servers not assigning PIDs to their repositories.[17] Certification processes can make an important contribution to quality assurance, for example, the current version of the CoreTrustSeal[18] requires the assignment of PIDs. All data accredited by the German Data Forum

RatSWD (Rat für Sozial- und Wirtschaftsdaten, Germany) also uses PIDs, even though their issuing is not a certification criterion.[19]

As far as text repositories are concerned, the DINI Certificate 2022 for Open Access publishing services[20] requires the assignment of PIDs to deposits. In practice, however, the issuing of PIDs can lead to the allocation of a confusing number of PIDs of the same type for one and the same content: if a version of a record is published in a repository in parallel with the journal, it usually receives a new DOI within the prefix range of the repository. If this DOI is also displayed on the splash page of the record (and possibly given as a citation recommendation), confusingly a different PID issued by the publisher is found in the full-text.[21] This issue even multiplies if the document is also published on other repositories, e.g. in case of multi-authorships. It would be helpful in such scenarios to have some kind of recommendation regarding the multiple assignment of PIDs to the same content, which would have to clarify whether the PID refers to a bitstream or an intellectual content in a specific manifestation.[22] Should the PID refer to the intellectual content, it would stand to reason that the final author manuscript and version of record would also have the same PID.

On a global level, CRIS systems and institutional repositories are converging. Since more and more repositories carry out CRIS functions by performing reporting functions for publication volume and expenditure[23], their enrichment with PIDs (in addition to PIDs for texts published in the repository itself, also PIDs for authors, organisations, funding, related objects

---

17 Following https://www.re3data.org/search, the vast majority assigns DOIs (34%), with only a few assigning handles (9%), URNs (2%), ARK (1%), PURL (1%) or other identifiers (4%, e.g. Research Resource Identifiers RRIDs).

18 https://www.coretrustseal.org/why-certification/requirements/

19 Although the use of persistent identifiers (PIDs) is not a criterion for the accreditation of a research data centre by the RatSWD, the use of PIDs is nevertheless queried during accreditation. In addition, the reasons why PIDs are not used are captured.

20 https://dini.de/dienste-projekte/dini-zertifikat/

21 See for instance: http://doi.org/10.25358/openscience-5863

22 ISO 26324:2012 gives room for both interpretations: It "defines the syntax for a DOI name, which is used for the identification of an object of any material form (digital or physical) or an abstraction (such as a textual work) where there is a functional need to distinguish it from other objects."

23 See as an example, https://epub.uni-regensburg.de/52447/

such as research data) seems more than reasonable. Enrichment is even more fruitful if these PIDs enable referencing (e.g. in human resource systems, e.g. for the assignment of APCs to persons) or integrate authentication (and thus allow automatic assignment and synchronisation with ORCIDs when uploading to the repository). Of course, the above applies to CRIS systems as well. The current version of the COAR Community Framework for Good Practices in Repositories[24] therefore not only requires the mandatory assignment of PIDs for published items, but also recommends to include a link in the metadata record of published items to related contents such as preprints, published articles, data, and software and integrating PIDs for authors, funders, institutions, funding programmes and grants, and other relevant entities.

Researchers

Researchers are the ultimate PID end-users. As such, they should keep up to date with developments in the PID domain. However, PIDs – especially those termed 'admin-oriented PIDs' in this study – are often implemented without a direct researcher involvement. It's then for institutions to raise awareness of PID developments among researchers and for research funders to reach out to their funded researchers to explain the rationale behind their PID policies and requirements. Publishers also have a role in this dissemination effort by clearly stating in their manuscript submission systems what identifiers researchers are expected to provide. All these stakeholders (HEIs, funders, publishers) may also wish to pay attention to the bottom-up PID initiatives 'their' researchers may be taking part in when considering whether specific emerging PIDs might be worth adopting in a top-down approach.

Start-ups

Because PIDs are ultimately a matter of assigning identifiers to digital objects, there are opportunities for start-up companies to play a relevant role in the PID landscape too. These actors are difficult to coordinate

under a publicly-led drive, but due to their independent character they may be able to integrate workflows in a way that may offer value. See for instance the work of SciCrunch in the United States and their use of Research Resource Identifiers (RRIDs).

5. Risks

This chapter presents an in-depth exploration of risks for a well-functioning PID infrastructure. Risks were identified from the expert interviews and analyzed for the arenas suggested by the KE OS framework. There are political, economic, social and technological risks associated with PIDs.

Risks and trust in a well-functioning PID infrastructure are closely connected, since trust is theorised to only be relevant in risky situations. Risk is present in a situation where the possible damage is greater than the advantage that is sought (Luhmann, 1988). In fact, scholars have argued that risk, consisting of uncertainty (e.g. Hardin, 2001: Gambetta 1998) and vulnerability (e.g. Rousseau et al., 1998) is a precondition to the trust development. The conscious acknowledgement of risk actually distinguishes trust from related concepts such as confidence (Kelton et al., 2008).

A classical approach to defining risk is the probability of an event multiplied by the magnitude of its consequences (Leveson et al., 2009). Nevertheless, risk is also socially constructed, meaning that social factors influence how people perceive, understand and experience risk (Frank, 2020). By trusting the PID infrastructure, PID managers and users make themselves vulnerable to a number of risks, because they depend on the functionality of a PID service without being able to control it themselves.

The literature study revealed that risks in PID systems had been discussed in some cases before, most notably in Car et al., 2017.

PID-related risks reported in the literature were mostly *technical failures* such as …

▸ loss of link between object and identifier,

▸ assignment of multiple identifier for one object or multiple objects to one identifier,

▸ metadata for object is wrong/wrong object linked/ metadata changed by unauthorized source,

▸ object has been changed, access to object is lost, no tombstone page.

… or *environmental risks* (as causes for technical risks) as a loss of database, metadata etc. through environmental threats (fire, storm, flood) and *organisational risks* such as

▸ loss of financial stability or sustainable business model,

▸ vast changes in government, leadership, responsibilities,

▸ loss of interest/priority of PID system, lack of maintenance and governance,

▸ lack of community involvement or community support,

▸ abandonment of PID services, "Zombie PIDs".

These risks described in the literature can be mapped to the arenas of the Open Scholarship Framework as follows:

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.

During our interviews with PID experts, it quickly became clear that all participants shared the opinion that thinking and communicating about risks around PID systems and infrastructures is very important and sometimes overlooked.

PIDs should not just be seen as a solution to everything: they are not a "holy grail" or "trust markers", they don't solve all problems. On the other hand, adoption and uptake need to be supported and should be relatively high to make them be useful at all. This is a fine balance and it's important to be aware of both sides to this. Almost all experts agreed that talking

about risks associated with a technology that is designed to guarantee the sustainability of scholarly communications as an underlying infrastructure is very important. There is a real risk of over-estimating the "power" of PIDs as a solution for problems such as discoverability, access, preservation, etc. by the scholarly community at large.

"Well, it's a little bit of a delicate point, because if I start talking to the major stakeholders in the Netherlands about all the possible risks of PIDs, while I am still trying to get them to adopt the idea of PIDs in the first place, that might discourage them from stepping in. And the other hand, you don't want to be like hypocritical in the sense that you say, you promise them the world, because PIDs are the basic elements of FAIR and are the basic element of, you know, digital sovereignty to me, without talking to them about the fact that there's risks involved as well. So I think it's a balance you should strike between being enthusiastic about potential and the use of PIDs for those themes that we talked about, whilst also not presenting them as the Holy Grail."

People should not have inflated expectations towards PIDs but stay aware of requirements and risks.

"At the same time, what I really do not like is the magic power that is being associated in a lot of communications and actually, most of the time the persistent identifier people themselves, the experts like Geoffrey Bilder and all, they don't do that. They understand that there's risks there also, and you know that, but in a lot of stories, a lot of these European Union projects, 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."

In terms of the Gartner hype cycle, the PID technology is useful but more realism about the implications and risks is needed, especially in the long term.

"Well, yeah, I don't think it's a technology hype, in the sense of it's totally not useful. I don't think so. I mean, it is, again, I emphasise, I think it is a useful technology. I just think that we too rarely hear the "but" part of the story. And as I said earlier, really the understanding that

we've signed up on this, and we are now going to have to deal with this for the next 100 years or so or for as long, let's say as we assign DOIs plus 50 years, I guess we'll have to deal with it."

At the same time, when asked about fallback plans or emergency plans in case PIDs such as DOIs would suddenly stop working, it was interesting to note that most interviewees from the user side did not have specific plans in place. In some cases, during the conversation it appeared that this was a question that had not been considered before.

"Really good question, because no, we don't. We don't really have a fallback plan, I think. Especially with DOI, which is so crucial to our work. We just sort of expect them to always be there.. They are just such an established, I think, part of the research infrastructure at the moment, so. But I think it's a good point. Because we make risk assessments for all other kinds of things, actually, I'd say. But this particular risk, we haven't. So I think it's something for me to take back at least to discuss how, yeah..... And I think also, because we get, well, most DOIs come from Crossref and DataCite and I think the way we perceive Crossref and DataCite is that it's a well established organisation. We have trust in them. So....Maybe that gives us the sufficient sort of reassurance, let's say, that we don't have to think about risks. In fact, that also can be a bit fragile, I suppose."

We analysed risks in PID systems in accordance with the Open Scholarship Framework. Four types of risks can be distinguished: social, political, economic and technological risks.

Political

The political aspect focuses on the areas in which government or institutional policy (decisions) and/or regulations (rules) affect the PID landscape as a whole or the functioning of a specific PID. Some issues can be described as political risks, with legal, commercial and financial dimensions, including governance. The first is the **risk of discontinued service due to organisational change, such as a takeover by commercial companies or a merger.** Political and commercial independence should protect the research infrastructures against financial and commercial interests.

Another risk is the **non-conformity with European privacy laws (GDPR)**. Who is in control of the data? What about personal data transfer? The risk is not only "objective" but also "subjective" because some people may be aware of the risk and refuse to release their data. Legal advice is required, as well as communication.

"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. Perhaps another term that I should mention, which has been an important theme for the university recently, is that of digital sovereignty. I think it's also an important theme for the European Union. So the whole idea that you want to continue to be in charge of what happens to the data that you work with, and that you know what happens to all the data that you collect and that you create, I think this has been an important concern also, when the university began to adopt ORCIDs. Of course, we stimulate our researchers to enter all the data. Because we know that ORCID has databases situated in the United States, so there's a transfer of personal data."

Leadership, governance and participation have been mentioned as a third political risk, especially because of the **role of commercial stakeholders**. Are the needs and the requirements of all stakeholders proportionally represented? How to draw a consensus about the future of a given identifier? How to prevent increasing control from publishers? How to maintain transparent leadership, and how to guarantee a multilateral, participative governance?

"Another problem is the political problem: who is leading every identifier infrastructure, where is the governance and how to participate in it ? (...) In the "pre-Berlin-wall-fall-world" there was a general governance system based on the concept of multilateralism, illustrated by the United Nations, UNESCO, ISSN, ISO, or even AFNOR in France and DIN in Germany. Things have changed and today,

we have a multi-centric world where the governance of the PIDs is disseminated among many, many bodies and organisations. Before it was centralised in multilateral institutions, which are not perfect of course. But when I want to know who is the owner of the ISSN, who are the stakeholders of ISSN, it's clear, I can know it, it's transparent, it's public, even if the economic model is totally obsolete. But when I see ORCID, when I see ROR, CrossRef: who is it? Who pays? Who has got the power? Is it democratic? Is it here for a long time? No. So this is a problem. And this is a problem we must address collectively."[25]

Many PID systems are **very western-centred**. It is difficult for countries from other parts of the world to be involved in PID developments and governance (on the adoption of industry standards see Chan, 2018). If we want to build a reliable infrastructure, organisations need to figure out how to best involve all countries.

"I mean there was a big thing at that point in time, that hey, you know what, DOI means scientific quality. No, it doesn't. It's an identifier. 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. How do we get the Third World countries into all of this? I mean, there's things happening out there and they need to be more able to participate in it, otherwise we're just again building inequality into the thing. That's not good. I mean, the Internet kind of tore down a lot of these barriers. Yeah, well, we're just building some of them back up."

---

25 The business model of the ISSN Centre is based on contributions from the French government, contributions from 92 member countries and sales of services, 1/3 each of its operational budget. Contributions of member countries are based on the GNP and adapted to their national budget.

It can be hard to argue for the importance of PIDs at a policy level. Communication and promotion is required, not only about the technical side of things. **Lack of strategic communication** is a risk.

Centralised solutions represent a particular risk for the PID ecosystem, in two ways: they can prevent easy solutions, and they can cost more money than other solutions. Yet, a centralised solution may not be the worst option. A realistic assessment is required.

"Instead of making that immediate leap to go for a centralised solution, why not also consider decentralised solutions and then compare: If you do it in a centralised way, that's the net result, if we do it in a decentralised way, that's how it works."

Economic

The economic aspect focuses on the business model, on revenue streams, cost structure etc. that affect the PID landscape as a whole or specific PID stakeholders. According to the interviews, **lack of funding** is the greatest risk for the PID ecosystem, in particular no stable funding to ensure sustainability.

"The strength of the current system is that there is a new PID for any use case. So it's a strength and also it's a weakness, because who funds any different PID? What is the economic model of every PID? And who should take care of it for the long term? So we see that the weakness of the system today is that every PID has got its own economic model and no clear governing body, no clear funding, and no clear future, no clear strategy. We see with ORCID, for example, that their economic model is very weak, very fragile. And they have decided to increase the price for the consortial members, the group of institutions paying an annual fee, by 10% next year, in order to finance their budget of 6 millions US$ per year to maintain their activity. It's very expensive. It's too expensive. So if they need money to manage ORCID, how much will cost ROR, how much will cost RAiD? The problem today is who pays for the identifiers."

Experts highlight the **necessity of contingency funding**, i.e. paying to sustain the PID infrastructures.

"And one problem then and I think it's not solved now is the idea that scholarly infrastructure in the academic context should basically be free. It's fine if it's free to use for academics, but the idea that these things cost money... There's a lot of grant funding to get things started, but to think, what's the business model, I think that's still not really addressed for infrastructure in general, not necessarily PID specifically. So I think it's a problem that's not really solved. There are always these sort of special cases like Wikipedia, which have a very unique and very different business model, but that's not probably something that can easily be transferred to other kinds of infrastructure. So it is still a problem."

Membership organisation is an option, but then the question is about **calculating the benefits for membership** (e.g. DataCite).

"I don't know. Yes, it's maybe another weakness or risk, it's the cost of some ID and also the governments of PID providers. Some IDs like DOIs have fees, and this may be problematic in the medium- to the long-term for an infrastructure like HAL that manages an increasing amount of data. And we are in the process of stabilising it all on a funding model. So if I have to buy DOIs, but also maybe ORCIDs, I don't know, and also maybe ID Structures because we would like also to have a description of the structure at the French level, but it will be interesting to be linked with the ROR at the international level, so it's not clear but maybe there will be also fees to buy. I think it could be a real problem for infrastructures as HAL because in the open science ecosystem, the funding of infrastructures, as HAL, you know, is still in a transition period. It's not so easy for us."

"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."

Social

The social factors that may affect the PID landscape are cultural (scientific, professional) attitudes and values, professional skills, workplace conditions (resources), workforce etc. Social risks to PID systems are discussed by all experts and mentioned most frequently during all interviews. This is in part because most experts agree that PIDs are a social-technological system, where the social part is most important. In other words, for the experts, the most vulnerable part of PID systems is made up of people - pointing both towards the people that provide PID services and the people meant to use these services. Again, we can identify different levels of social risks:

▸ social risks concerning PID infrastructure sustainability and functionality,

▸ social risks concerning PID users,

▸ general system-based risks.

Risks around sustainability and functionality were most commonly mentioned in conjunction with a **lack of institutional commitment and of human resources** (workforce) needed for supporting and maintaining PIDs. This includes staff for helpdesk and user assistance.

"You know, and the other thing was sort of a lot of things… We have, you know…. we would like a new PID for this, and we'd like a new PID for that. These are all great, but they're not the substantial issues of running infrastructure, the substantial issues of running infrastructure, or getting communities to agree to a set of norms, getting sustainability, ensuring that the thing can… if everything else goes down, that that data is open, and so on and so forth. And then I hate to say horrible things like support and maintenance, which is, of course, terribly, terribly un-, you know, -sexy to talk about. It's the fact that you have to have the help desk, and the fact that you have to keep servers running and that you've got a person who wakes up at three in the morning, because, you know, because the server's gone down. You know, it's not just about the superficial stuff."

Not being able to show examples and value is another social risk. Starting new services is a leap of faith until you reach a critical mass. Examples are needed to show value, but examples can only be shown if people are using an identifier, meaning there is a direct link with the **risk of lacking uptake**. Even the biggest investment in PIDs would not equate to its actual value, if the system is not used.

"Building PID infrastructure is not a sort of multi billion dollar sort of high risk game, where you just need a lot of money to get anything started. But it's really more about having enough people that support both the organisation but also of course, the use cases etc."

Functioning PID services need a community that supports them. Beyond the technological infrastructure, there is a people infrastructure. Identifiers are about communities agreeing to common practice, and implementing practices that are not common will make them go away. Working without community will fail. The **lack of community engagement** is another social risk.

"I think, maybe one thing that I'll mention, is that all of this, that we talk about the services, the trust the risk, involves people. And, for me, this is the most important thing. We can build the most fancy, exciting, different system, address all the use cases, put in all the risk mitigation strategies we want, but if we neglect to actually genuinely work with the community and do it as a collective, we will fail. And we may not fail in the next five years, but in 20 years time, I can guarantee you it will be a failure. For me, at its core, is people. And it's working through conversations, talking to people, working through disagreements, we don't agree always, understanding what the shared path is forward. That's really important in trying to address risk, understand how we build trust and really, really important. Without that I think all of these other things, you know, become obsolete, because we will just fail and because we won't have that connective community buy in."

*The lack of uptake and actual usage of PIDs in different systems poses a challenge. Without uptake and usage, the whole thing is a "pointless exercise" because infrastructure is nothing without the people:
"....Infrastructure is nothing if people are not using this, and if it's not solving problems."*

The problem cannot be reduced to disciplinary differences but is also conditioned by the **plurality and the diversity (inconsistency) of stakeholders in general.** The reliance on PID providers in terms of sustainability is difficult, because it encompasses all aspects of socio-technical systems (technical, financial, political, organisation, relational aspects). It is not a stable state but a multidimensional process. The business model is crucial but it is definitely not the only key factor. Open data is another key factor of sustainability.

"And ultimately, a community needs to be able to say, "You know, you're not doing what we want you to do anymore. We're going to take our stuff, and we're going to go and do it elsewhere." Right? And so the… to my mind, the thing that an organisation that's running infrastructure, or any service can do, that's most useful and the best insurance that they can't be co-opted, is to make the data as forkable as possible. Now, clearly, the data and the service, not just the data. So that, you know, if a sufficiently large part of the community is unhappy with what the organisation is doing, they can go and start another way. Right? And there are no artificial, technical, or data barriers to them doing that. Clearly, there's always the barrier of bringing the rest of the community with you. But if you can't do that, then the question is, are you actually an outlier? I mean, so it's sort of a self reinforcing thing."

Centralization of PID infrastructures and their **dependency on the people responsible** for them are problematic, because "(..) things run on passion."

"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."

A lot of the **burden is put on researchers and/or research communities**; more support is needed from governments and publishers, and there should be benefits for the individual researcher. If not, there is a social risk of non-adoption and unsuccessful systems. Insofar institutional IT systems and other scholarly infrastructures are increasingly dependent on PIDs, this reliance on multilateral, close-coupled systems becomes a particular risk, a vulnerability and a challenge. **Unreliable data, i.e., the lack of control and curation**, can undermine the value of the entire system:

"(…) every PID will fail if it is not managed. And administration is always an issue no matter how good your technical platform is. If you don't know how to administer your DOIs or URNs or whatever, they will fail, so that is, I think, a major risk even in systems that are well established. And you only need to have a few rotten eggs to undermine the value of the entire system. "

"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. And I think there's supposed to be these types of dependencies. So I talked about this whole scholarly infrastructure, and it's, of course, based on the foundation that's put in place by these persistent identifiers. Yeah, but I think that's also a vulnerability. I think that's probably the main cause for concern, this produces challenges."

PIDs are often seen as trust markers, even though they are clearly not. They are not a quality stamp, and the official requirement to assign DOIs to publications and to evaluate only publications with DOIs is not an indicator of quality.

"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."

Technological

The technological aspect puts the focus on the specific role and development of technologies within the PID landscape, as well as the wider uses, trends, and changes in technology that may affect the functioning of PIDs. Even though technological risks play a big role for PID systems, because they naturally are dependent on functioning technologies, during the interviews technological risks were not seen as the most important ones. In general, the technology of PID systems is well established and runs quite smoothly. As one expert put it:

"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 (...)."

Mostly, concerns about technological risks can be subdivided into two different groups: Risks concerning the metadata associated with a PID, such as quality, richness, completeness and risks around PID systems, especially in terms of their interoperability and scalability.

The lack of quality, richness and completeness of metadata is a risk for the whole system.

"Another issue is around quality of the metadata in the business of identifiers. This happens in all facets, you know, we can go to ORCID where you can find auto dealerships as ORCID iDs. But you can also find in DataCite, the same notes, where there's a required field that says "not applicable", because it needed to be registered and somebody didn't have that information and they just filled in "not applicable". We are also working on various sorts of tools and things to help support members in this. Because one, the resolution in trying to pick these up and identify these through automated processes and then this triggers various processes that happened."

Other identified risks are more general risks at the system level, related to **scalability and interoperability**.

"And maybe another weakness is that for all these identifiers, we have to use a register to keep the link between the identifier and the object, the entity. And so registry is a fundamental component. And what happens if this registry were lost, corrupted, this kind of thing? So I think yes, maybe it can be considered a weakness."

6. Trust

This chapter shows how interviewees conceptualized trust in the PID infrastructure. The analysis was done according to a framework developed from existing trust research. It is not easy to discern, which part of the PID infrastructure trust is actually directed towards, but the perceived trustworthiness of PID providing organizations seems to be most important for PID users.

We understand trust as the "willingness of a party to be vulnerable to the action of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other party" (Mayer et al. 1995) with the contextual factor that "A trusts B to do X" (Hardin, 2004). From this view, there can be trust in PID systems, in organisations that manage these systems and in their technological function. Trust can be seen as a cognitive state (a belief) that has intentional (e.g. the willingness to accept risks contained in PID systems) and behavioural (e.g. the prolonged reliance on and usage of PIDs) consequences. In order to decide whether a system can be trusted, its trustworthiness has to be checked. Trustworthiness refers to a set of qualities that (should) signify to the trustor whether the trustee can be trusted. A trustworthiness judgement predates the actual appearance of trust. Trustworthiness is something that can be observed by the trustor and demonstrated by the trustee, meaning that the trustee also has a certain amount of control over how trustworthy they appear. For example, a PID provider can provide evidence to PID users that they are trustworthy and therefore should be trusted. They can hardly control, however, if trust in them actually develops, especially with researchers who might not immediately be affected by the failure of a PID system. Trustworthiness and trust are closely related but should not be viewed as the same thing.

The results of the literature study show that trustworthy *PID services* can be characterised by different properties (Weigel et al., 2018). Trustworthy PID systems are:

▶ maintained by dedicated and reliable team,

▶ based on a transparent and sustainable business model,

▶ provided by a non-profit organisation,

▶ subject to regular quality assessments by external parties,

▶ governed by international boards,

▶ based upon open standards,

▶ based on a redundant and secure architecture,

▶ support a huge address space,

▶ supported and openly documented API optimally supporting accepted data models.

Trustworthy *PID service providers* in turn feature these characteristics (Hellström, Johnsson, et al., 2020):

▶ PID registration and resolution has no costs to end users,

▶ PID Services should have Technology Readiness Level 8 (system complete and qualified) or 9 (actual system proven in operational environment),

▶ 24/7 availability is ensured, responsibilities for service maintenance are documented clearly,

▶ there is a clear sustainability and succession plan with an exit strategy in place,

▶ PID Service providers and Authorities are regularly certified based on agreed standards,

▶ an accessible API is in place for the development of a generic, global resolution system across all systems and providers.

If these requirements are complemented by others that Car et al. (2017) impose on PID systems, then stakeholders and trust indicators can be presented as shown below in figure 2.

Trust always involves at least two parties: a trustor (a person who trusts) and a trustee (a person or object that is trusted). Since PID systems can be viewed as an infrastructure consisting of technology, people and organisations, it is not easy to distinguish who or what exactly the trustees are in a trusting relationship between PID users and managers and PID systems or providers.

The interview questionnaire and subsequent analysis of the interviews regarding trust in PID systems was based on a theoretical framework deducted from the literature study. The analysis focused on factors influencing the perceived trustworthiness of an organisation and the perceived trustworthiness of technology. Trustworthiness of an organisation was analysed using the foundational aspects in trust research by Mayer et al. (1995): *ability/competence*, *benevolence* and *integrity*. Added to these from the literature on trust in digital repositories (e.g. Yoon, 2014) were the factors of *transparency* and *reputation*.
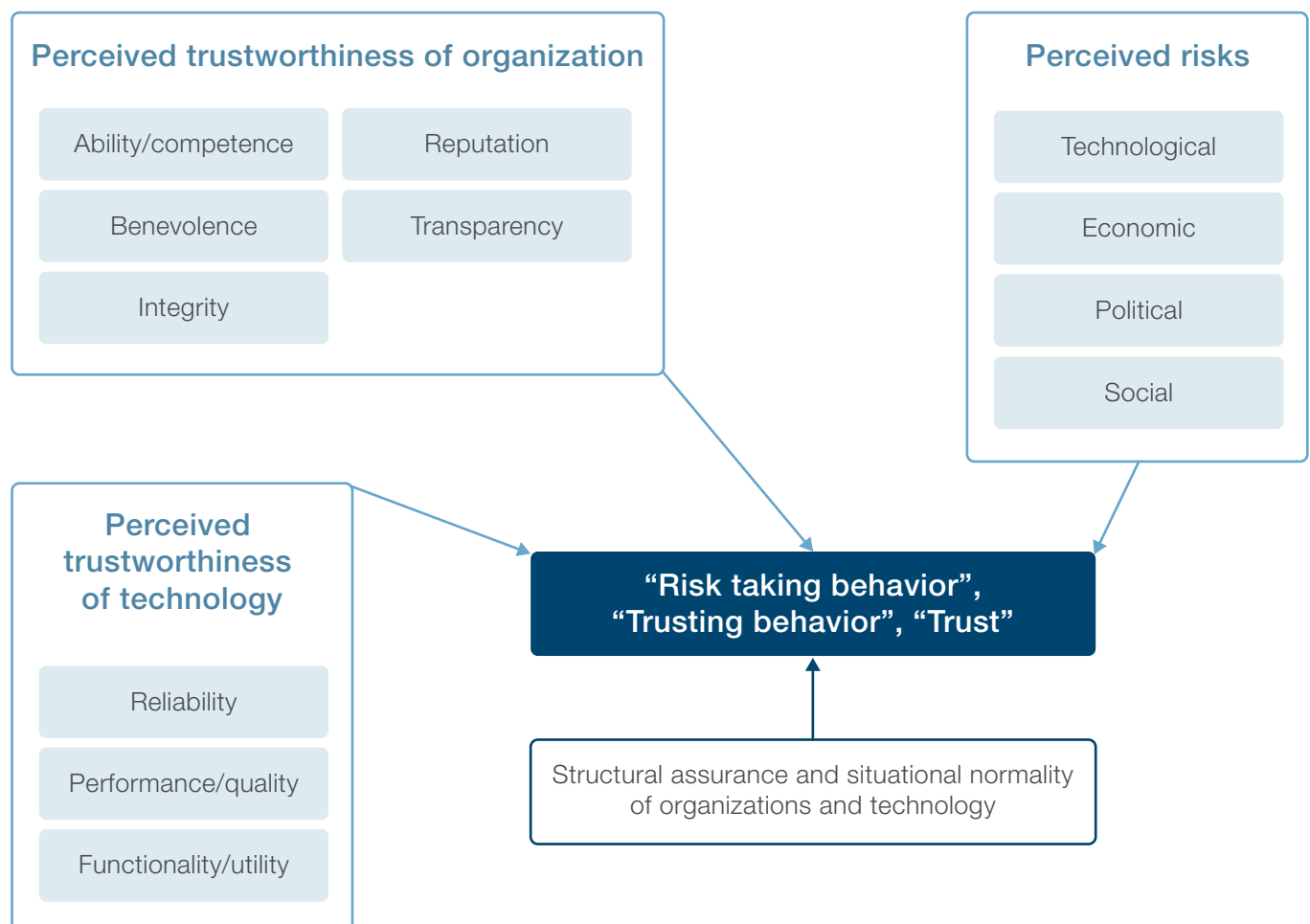
Figure 3. Framework of the analysed risk and trust variables
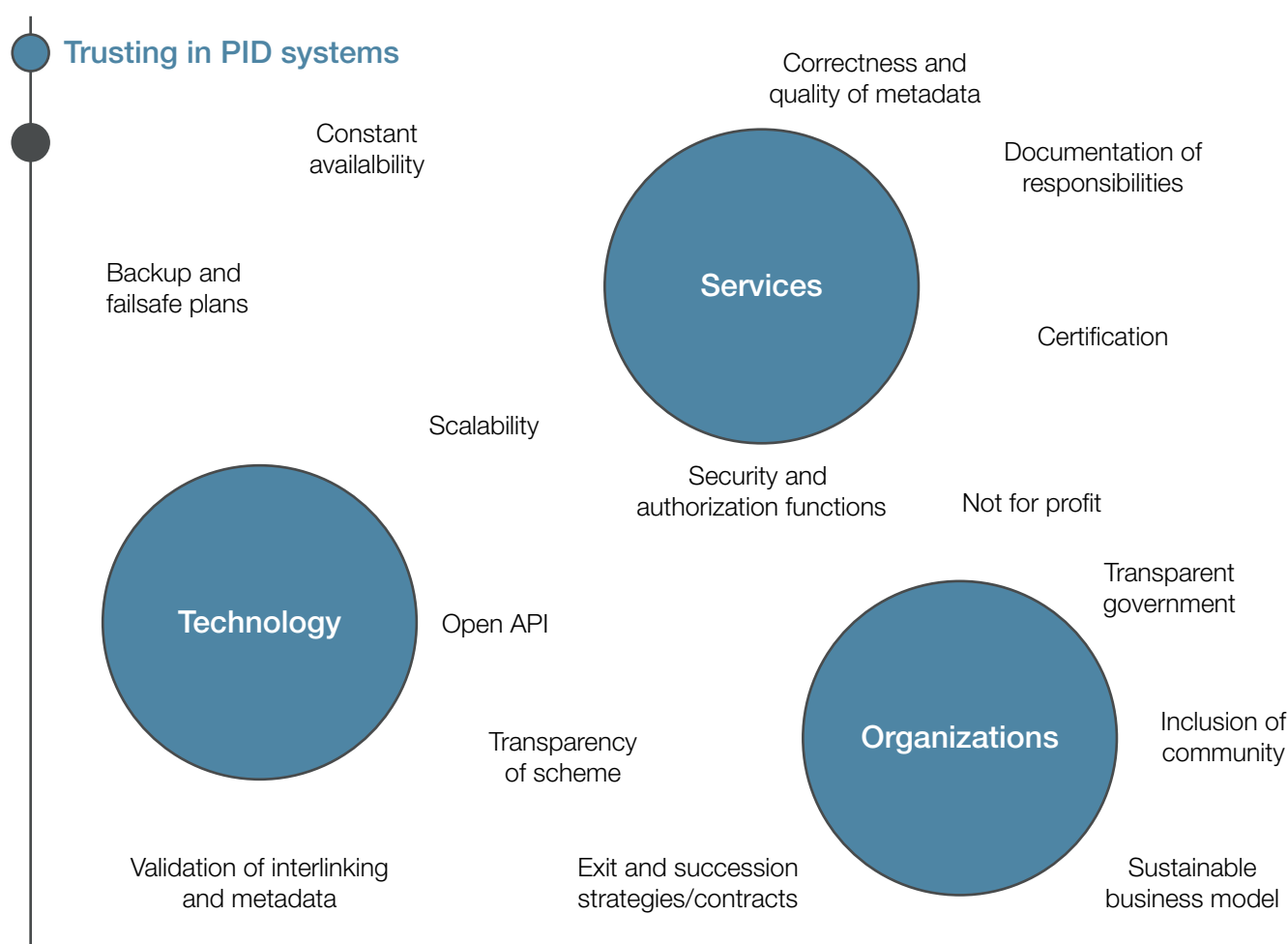
Trustworthiness of technology was analysed in accordance with the Technology Acceptance Model (Gefen et al., 2003) by using the factors *reliability*, *performance/quality* and *functionality/utility*.

Trust research has also stressed the importance of structural assurance and situational normality, especially in cases where trust is not interpersonal but directed towards systems or institutions (e.g. Gefen et al., 2003). These factors were also analysed.

Figure 4. Framework of the analysed risk variables

Perceived trustworthiness of organization

Ability/competence

Reputation

Benevolence

Transparency

Integrity

Perceived risks

Technological

Economic

Political

Social

Perceived trustworthiness of technology

Reliability

Performance/quality

Functionality/utility

"Risk taking behavior", "Trusting behavior", "Trust"

Structural assurance and situational normality of organizations and technology

Situational normality

When entering a trust relationship, one of the factors a trustor pays attention to is the fact if the relationship is similar to other successful trust relationships and that a sense of reliable circumstances can be perceived. Situational normality in our interviews was not talked about a lot, perhaps because it is an unconscious feeling that exists without much reflection by either party. However, when judging trustworthiness of PID organisations, three aspects came up that are connected to other successful trust relationships and the reliability of the situation of PID infrastructures as they exists right now:

Trusting relationships between PID users and PID providers are reciprocal and reinforce each other:

"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."

Opening and duplicating data helps to insure that a 'normal' situation can at least be technically recreated if failure would occur:

"One of the things is that, by exchanging this content with other other systems, then hopefully we have a copy of these things, and by having had PIDs, persistent identifier, on it, we might be able to duplicate the things in the future, to know that this was actually the same thing, right. But, I mean, then we go into that, I mean, if a bomb pins down in the National Library you're also screwed, right? There's always some level of failure, where things… they go wrong, right, and I think the one that's the most fragile at the moment is the people infrastructure."

Established relations between traditional organisations and their users are already built on trust:

"Yes, the contract is maybe one aspect, the other is some traditional relations, I think. So we are already working with a lot of these institutions in other contexts as a big national network of libraries in Germany and so these partners used to work with TIB and they know that we understand their workflows and what is that important task in each of our fields. So that we can answer their questions."

Structural assurance

In trust theory, structural assurance refers to guarantee that structures are built into the relationship between a trustor and trustee (here: mostly institutions/organisations). These guarantee structures are mostly formalised in a way that protects the trustor party from succumbing to risky situations and increases the perceived reliability of the trustor. Structures, formalised guarantees or warranties are normally provided by the trustee and made more or less available to the trustor.

In PID infrastructures, our interviews showed that structural assurance does play a role in the trust relationship between PID users and PID providing organisations, but are perceived as more important to supply by PID providers than actually looked for by PID users. This may be the case because other trust factors (see below) were emphasised more strongly by PID users and since in most cases there was a feeling of general trust into PID providers, a need for concrete evidence of structural assurance factors was not sought after as much. This argument is also supported by the fact that fallback plans were something that PID users mostly did not have in place and did not put much thought into (see above).

However, structural assurance measures such as **contracts, policies, risk management workflows** etc. did come up in our interviews, especially during interviews with PID providers:

"I think that should be part of what kind of service levels you provide, but also how it is defined using your own policy structure, and this kind of information should be provided to the organisations who will make use of your

service. It depends on what kind of sustainability you plan. For example, I know from DataCite that when you register a DOI, within their policies they describe that okay, "we will maintain the PID and metadata which you submit with the DOI, it will be maintained forever even if you do not pay for the DOI or at least the DataCite contract, so, even if you stop being a member". And that will be confined within the agreement which you have with the providers."

Escrow setups and contingency plans (living wills) are elemental factors to structural assurance.

"And obviously, that's a challenge when you're talking about data that's sensitive, that's private, and so on and so forth. Not every organisation could just say we're going to open our data, right? But they can make provisions with other organisations, escrow setups with other organisations and things like that. And we have to start exploring those kinds of things, right? How do you set up a living will that says that if people need to get this data, they can somehow get it and they can move it on? Those are the kinds of things that I think we have to start focusing on."

Risk management plans should be well communicated to the users. This increases perceived reliability.

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."

On the other hand, from the perspective of PID users, structural assurance mechanisms were paid a lot of attention to.

"But when they get to a point where they are.... the service itself is taking for granted and I think you know, Crossref, DataCite and ORCID are at that level, there

isn't too much talk about risks or other possible problems that you could run into anymore. I'm trying to figure out is that because I don't want to read it, because I'm a PID proponent, or is that because it isn't there. And I'm not entirely sure. I haven't really specifically looked at their contracts or at their websites to see how they present the risks of their own services. I do know in terms of business models that ORCID has, you know, always been very transparent about: This is our business model. We're going to change it now. So there was transparency there. But I'm not entirely sure if I would find anything at another risk level, like, you know, what if we're hacked? Or what if the PIDs start falling apart for some reason? Or what if, you know... I'm not sure if I would find anything? I haven't looked, I must admit."

Trust in technology

Because PID infrastructures are socio-technical systems, it is important to distinguish between human and technological factors when studying a trust relationship. While the trustor in this relationship is most likely always human (i.e. actual PID users), the trustee is not always easy to distinguish. When people talk about their trust in DOIs for example, do they mean the underlying technical system or are they actually talking about a DOI providing organisations, such as DataCite or Crossref?

In our interviews, it quickly became clear that even though the PID infrastructure is built on technology, this part is seen as mostly reliable and trustworthy, or at least as easy to fix if something went wrong. In trust theory, trust in technology can be studied based on several different factors. In our interviews, we mainly focused on three different aspects: Functionality and utility, quality of the performance and the reliability of the technology.

Functionality/utility

Functionality and utility refers to the fact that a PID system is actually usable and integretable into a PID manager's system. It also refers to the specific instances PIDs are actually used for on a technological level in an organisation. Perhaps not unsurprisingly, this aspect came up most often in our interviews. Of course, **PIDs have to be technically usable**, otherwise their adoption and usage would not make sense at all.

"Yeah, I think the ability to create the visibility of connections between the different elements of research outputs. So, the output themselves, but then the creators and then the supporters of those creators. That kind of line of connection between researcher, institution, funder, government. And then actually, we should go beyond that and talk about impact, because that's a piece that kind of gets inferred in there. But isn't actually demonstrated at the moment as much. But that line of connection, the visibility there is poor, generally. And the benefits that a strong consistent PID landscape provides is to understand what the interconnections are and the results of those are and we get some of that."

However, **usability has to be demonstrated, ideally through thorough use cases, so that the end-users understand the utility of PID**. There is still a debate about how much of the technical functionality of PIDs e.g. researchers should need to understand.

"So for the trust, I think one important question you have to answer is, what is your expectation from a researcher? Basically, you should have a very high level of understanding, because that's probably needed. Otherwise, you would never sign up for ORCID or use DOIs. But is that all that's needed, and everything else is just a distraction for busy researchers? But it makes it very hard....certain things if you basically leave out all the details, when you talk to researchers or you never talk to them."

From our interviews it seems that functionality, even though mostly invisible for PID users, may be something that users actually don't have to fully understand to rely on. While PID managers and providers are much more dependent on the technical function of PIDs as trust factors, users actually could be more reliant on the services built on top of functional PID systems.

Performance/quality

Another important aspect of trust into the PID system and especially the technology is the perceived quality of performance the technology offers. In most interviews, **quality was a big topic with regards to metadata quality and standards and specifications from different PID systems**.

"Generally, I think if there is a preference, and that is probably based on the question whether or not they meet certain requirements, of course, in general, we're interested in the adoption of the persistent identifiers among stakeholders in this infrastructure for scholarly communication, and assume that we want to work with persistent identifiers that have reached a certain technical maturity. So, at least the PID system should meet some basic technical requirements such as 24/7 availability and stability. And of course, the uptime and response time should meet all expectations. So it should perform well, definitely. So I think that are all kinds of very evident requirements, basic technical performance, resistance."

In general, it was easier for interviewees to talk about problems with quality when asked about trusting the technology than about most other technology trust factors. This is probably the case, because problems show up with high usage and evolving use cases.

Reliability

In terms of reliability, there was consensus across most interviewees, that a PID technology should be as robust as possible. This aspect of technology reliability was often referred to in **connection with long-term preservation of digital content**.

"For me, using PIDs is a way to generate trust because we are able to give access to qualitative metadata and data. We have a strong commitment to long-term preservation, so it's important for us. Using PIDs also is a way to ensure interoperability, so it's important for us as a national infrastructure."

Trust in an organisation

Reputation

Reputation can be earned by the providing organisations and influences their perception by their users over a long time. The interviewees discussed several points relating to reputation. One of them is connected to the nature of an organisation: is it better to supply long-lasting and high quality identifiers through **state-funded, more traditional information management organisations, such as libraries**? Many interviewees agreed to that, but on the other hand discussed, that these organisations are not

membership-driven and that PIDs are not their only priority, even though they were expected to deliver high quality products because of their extensive experience. Additionally, these kinds of organisations tend to be more reliable in terms of **sustainability**, because they have been in existence for a long time and most likely will never go "out of business". Long existence also influences their **relationship and knowledge of their users** - the level of trust earned has most likely been pretty high because of a much longer partnership.

On the other hand, reputation also refers to the kind of **"branding" or "image"** organisations have. Interviewees mostly agreed that newer organisations, such as DataCite and Crossref were much better suited in terms of **agility and marketing experience** than traditional institutions. These aspects can be very important for PID uptake in many communities. These organisations have accumulated a so-called "record of trust" over time, because they have constantly been performing to a level that the communities need. This is also in stark contrast to some commercial suppliers, e.g. some publishers, who have a bad reputation in terms of community orientation.

"Maybe… now okay, so it's mainly the experience we already have in this case, because TIB is doing that for more than 15 years now, which is quite a long span of time in this context. Yes. We have… so we have experiences with PIDs we have experiences in offering infrastructure for research organisations and maintaining them. Although, in this case we don't even offer the infrastructure, because it's the DataCite infrastructure and the infrastructure of the repositories. We have good knowledge about the German landscape of research. Which places have libraries, for example, in the research landscape and the universities, and how do they work together and which really have research data repositories."

"Well, if you look at it from the point of view of organisations, then the libraries and especially national libraries that have a legal deposit, are the most solid basis that we have for long term preservation. It has already worked [for] centuries for printed documents and we are hoping that it will be the same for digital documents, but of course nobody knows if we will be able to preserve anything digital for a very long term. Part of my concern is because, for most IT people, the

long term means a couple of decades at most. I had a great time reading what W3C was writing about URIs. They said URIs must be persistent, so think about it in terms of years or even decades [laughs]. At that point I started laughing because a few years or a few decades is not at all persistent from the National Library point of view, our oldest documents are a thousand years old."

Views on whether traditional knowledge management and cultural heritage institutions would be better suited for managing PID systems differed between interviews. On the one hand, organisations such as libraries have a reputation of handling metadata very well and providing better defined and concise information. On the other hand, these kinds of organisations have a reputation of maybe being "too old fashioned" to be able to supply the flexibility and agility that makes independent organisations such as DataCite or Crossref valuable to the community.

One argument in this area of discussion that was brought up during the interviews can be summarised in terms of reputation: independent organisations have a better marketing strategy, whereas cultural heritage organisations have a better view of sustainability:

"Again, for me, it's way more the long-term sustainability. And basically, the budget that comes with all that, and especially this notion of really long term, I mean, we're talking not five years, 10 years, we're talking really long term here. And when I think about really long term, I keep coming back to the cultural heritage organisations, the national libraries, and so on, that have been with us for 100, 200 years as kind of good places for that kind of infrastructure. Then on the other hand, just for the sustainability of it, I cannot imagine those organisations to be equally innovative, and hip, and all, you know, in the way that these PID organisations are. So from a marketing perspective, I think it would not be a great idea. From a long term sustainability perspective, I think it's way more logical to have this kind of very significant long term responsibility in the hands of multiple cultural, long-term cultural heritage organisations."

Independent organisations are also viewed as **more in touch with their users and focused on a more collaborative and democratic approach**, which was seen as favourable for PID management:

"Yeah, I think they have a good ability to be independent, while collaborative. And so without having to be individually or specifically beholden to one group or one approach, but are actually very, very good at collaborating with the kind of stakeholders and the end users of their outputs. I think that gives them a better position then, say, a consortium of publishers that then decided to.... which there has been other attempts in the past to do similar activities. So I think that makes them a kind of trusting, trusted partner, that they definitely have connections with their stakeholders, but they don't...they're not beholden to any one group."

Very important to reputation is the 'branding' and public image of an organisation providing PIDs. During the interviews, it became clear that PID users rely on their own image of a PID provider, which has been formed through their experience of their behaviour. Trust judgement is therefore dependent on **the organisations' track record of favourable behaviour**.

"Hmm…From…. as a customer, I care about the service that's working and providing useful stuff. So it's….I would like a little bit of information on it, but I don't need a lot, because I have a track history of how they have been behaving in the past. That there wasn't things that weren't okay. If we take other examples, and Open Source software nowadays, right, you have Elastic as a company changing license on the thing, you have Docker changing license on Open Source stuff that have massive impact on license and on people. Because of a change, they make you lose trust in the company that they're going to manage this in the future, and it makes you think "What should I switch off?" And I don't have the same kind of thoughts with DataCite, because they have a track history of operating according to best practices and principles, and being useful for the needs, and being attentive to the needs of evolving metadata schema and things, right? So that trust is not easily…..is very easily broken. Right, so of course I think that's the main thing they have to keep first, but of course talking about the risk and telling "This is how we're addressing them" is something that helped

build that trust, but as soon as you lose it it's very extremely hard for them to regain."

Unfavourable behaviour in the past in terms of e.g. Open Science practices weakens the ability to trust an organisation to deliver good PID services. The following exchange serves as an example for this case, in which a company had shown intransparent and somewhat "egotistical" (for profit) behaviour in the past. Interviewer:

"Well, this is kind of a weird question but, for example, if XX (company) now would offer some sort of really well functioning, easy persistent identifier system that would be kind of comparable to DOI, which is governed more openly, would you at Zenodo consider using, let's say, XXX (company) PIDs?

Interviewee: I wouldn't trust them.

Interviewer: Okay.

Interviewee: Because they don't have a track record that builds trust.

Interviewer: Okay.

Interviewee: They don't have a track record…. for instance, they were holding off in opening up their citations metadata for so long that it was not until there was really humongous pressure on them, that they actually opened. They've never been leading in these kinds of things. They're doing things for profits. Which… they're a commercial company, I shouldn't…. They can do what they want, right? But they don't have a track record of trust, and that means you can't, you can't be based on and that's….. it's as simple as that. They don't have trust. It doesn't matter how much they do, it's like, to compare with software, it's like YY and ZZ now taking actions that may…... I understand, again, perhaps their reasoning behind taking this action, because anything costs money, and they need to get money somehow, right? But they have done it in ways that make me lose complete trust in these companies, of how they're going to manage it in the future. The question is not the change they did, the question is, what is the next change? And that's the same with XX. If they came and said, we have this thing with open

governance. Then the question is, what is the next thing you're going to use this for? How are you going to monetize it? How are you going to make sure that we pay you money?"

Reputation is not only seen as a factor for a favourable trustworthiness judgement, but a good reputation also minimises the perception of risk, or even the amount of interest in which eventual risks there might be.

"And I think also, because we get, well, most DOIs come from Crossref, and DataCite and I think the way we perceive Crossref and DataCite is that it's well established organisation. We have trust in them. So…. Maybe that gives us the sufficient sort of reassurance, let's say, that we don't have to think about risks. In fact, that also can be a bit fragile, I suppose. (...) But it's also dangerous to see, Oh, everyone else thinks that these are good enough. So do we. But I think actually, that's for me, if well established organisations and initiatives, etc, use them..... I'd say that's something that for me, would then indicate that .... yeah, I could have trust in them actually. That they are doing a good job."

Transparency

Transparency was one of the most cited factors that established trust in PID service providers. Both from the user point of view as well as from the provider side of view, establishing transparency about the processes surrounding the PIDs and being able to provide a feeling of inclusion in decisions and workflows was important. I**n contrast to less transparent commercial identifiers, transparency provides a feeling of control over how a system is managed and run. This can be achieved by open documentation, close communication, boards and democratic decision-making**.

Providers

From the providers' point of view, it was emphasised that transparency played a huge role in their business operations. In their opinion, being as transparent as possible about every aspect of their organisation demonstrates trustworthiness to their users. All interviewees from PID providers stressed that this is one of the most important aspects in their government.

Managers, users

PID suppliers see transparency as very important and our interviews also confirmed that transparency was one of the most important aspects from the PID users' and managers' point of view that establishes a trusting relationship with providing organisations.

Users and managers named a number of different factors that contribute to transparency and therefore perceived trustworthiness of organisations. However, the two most important aspects that appeared in a lot of interviews were related to a **general feeling of being able to understand how decisions are made inside the organisations, how business models and funding works and being able to actually have a say in those decisions through e.g. sitting on the board of an organisation**. In general, transparency provides users and managers with a feeling of being in control of the organisations and services they are reliant on.

Integrity

Integrity refers to the feeling of users who rely on a certain service provided by an organisation, that the organisation does exactly what it promises to do and has no secrets from the community. This part of trust development is closely connected to structural assurance, because organisations seem most integer when some kind of (contractual) assurance is in place.

First of all, integrity may show through **transition plans and protections against commercial takeovers** being in place. These kinds of plans can assure the community that the organisation is clear and willing to invest in their future as a (nonprofit) company and that they are planning for their PID to be available *"in the long run"*.

"They should also have a plan or a vision to ensure the sustainability and the longevity of the organisation in the longer run. And ideally, there should also be some transition plans explaining what should happen when the system needs to be terminated unexpectedly. Also some protection, I guess, against the organisation's takeover by commercial companies. We want to make sure that these crucial building blocks of our scholarly infrastructure do not depend on commercial interests, that it's not subjected to economic rules of the commercial market. Of course, I guess we also want to

avoid a lock in with a specific commercial vendor. Though I realise that this is difficult to avoid, takeovers and management mergers probably happen quite frequently. But I think it's important to ensure that there are measures in place that PIDs continue to resolve reliably, even in the case of organisational changes or changing financial conditions. So organisations should minimise the risk of non interoperability and avoid all kinds of legal obstacles.

Another way for an organisation to show integrity is to have **clear values that align with their user community and that are communicated and demonstrated frequently**.

"And maybe related slightly to trust is our mission, you know, it's a global mission and it's not about financials it's really about providing value to a global community. And so, providing equitable access and trying to help support the community and demonstrating on that, is something that really shows that we have trust as a community in what we're doing. These things are, you know, not always easier said than done, and so we still have a lot of work to do in various areas. But yeah, so I would say a very sort of principle to organisations is that they have very clear values and transparent processes is really important in building trust."

Lastly, interviewees talked about being able to tell if an organisation was committed or not if they were **willing to invest** into the service they are offering.

"So one important measure for me whether they trust in the system is whether they invest in it. Whether they put resources to it. This is in particular, with government agencies, when they make a commitment. Once they make this a line item, they stick with it. But to get to that point is hard. So they have to be convinced it's worthwhile going through all this. So that's from an infrastructure perspective."

Benevolence

Benevolence describes a feeling towards an organisation by its users, that the organisation has the community's best interests in mind during their operations. This factor is closely connected to reputation, which forms a track record of benevolent or non-benevolent behaviour by an organisation.

"But I would also stress the supporting site so that you engage with all the different kind of stakeholders to integrate or make use of the PIDs for their domain and that is collaborating with the higher education organisations, universities, research organisations, with different research communities so that they make use of the different kinds of PIDs. But you also need to provide support to them, but that is labour intensive to do. You need to find the right channels, you need to develop documentation, training, but also collaborate with the different stakeholders to make them aware, but also to support them to make use of those PIDs."

Ability/competence

Lastly, trust in the PID system is established through the general feeling that the organisations providing PIDs are able and competent enough to actually do what is required from them. In trust theory, ability and competence are measured from what is known about the trustee and how they present themselves. Therefore this aspect is reflected in most other factors included in trust judgements, but especially closely connected to reputation and benevolence.

Our interviews showed that there were **very few doubts about the competence and ability of PID providers**. This is probably also the case, because interviewees had been relying on them for quite a long while at the time of the interviews and the organisations had already proven their competence. In cases where there had been issues of doubt in the past, the general judgement was that the organisations were improving and evolving according to the communities needs.

"So there is leadership and getting people to collaborate and get them moving in the same direction. So, from my work with InvenioRDM I know that's extremely difficult and it takes a lot of effort. And that's what an organisation like DataCite does and it's extremely difficult and requires a lot of time. Then, leadership is, I mean it is getting this thing, everything running in the same direction, but it doesn't… that doesn't fix this huge chain link system. Okay, it helps move it a bit from forward but without it, we would be nowhere without the leadership. But you still have a humongous chain system where, you know, people are just not fully subscribed to this thing because they don't know DataCite. But at least without having one thing there

that can move ahead than where we have nothing at all. So again, being pragmatic about things, it's better than most of the things we had, and most of the things I can think of. I'm happy with most of what DataCite does. Everything can be improved, yes, right."

However, for emerging PID providers ability and competence cannot easily be judged yet. This means that for emerging PIDs, **early and ideally favourable experiences with the PID are very important to establish trust**. Additionally, for emerging PIDs utility and performance of the product (not the organisation) might be more important aspects for trustworthiness judgements.

7. Open infrastructures

As mentioned before, openness of PID infrastructures is highly relevant for their trustworthiness and sustainability, insofar it may safeguard the persistence of the infrastructures and the data. The term of openness covers two aspects, technical openness in the sense of connectivity and interoperability but also, open availability of data and metadata. Open source and open data ("forkability") are considered a key feature for trust and reliability. Some experts consider PID infrastructures as common goods, in the wider context of open science. As one service provider puts it: *"For us it's really important that both our services and everything that we build is open source, so it can be used by the global community".*

This user-centred attitude generally goes along with a refusal of closed, commercial systems. PID Providers, so the common opinion, should work according to a philosophy based on open science principles, i.e., transparency, reliability and sustainability; and of course, openness. Non-profit may not be a guarantee of sustainability (and quality), but commercial systems are criticised as risky or simply bad, as they do not prioritise the user (community) needs and as they are generally closed, without any option to recreate data and metadata if the system is down. *"Metadata is the basic info that we need to be open at all times. So we don't want to be dependent on commercial parties…"* Implicit in this view is a criticism of the political and economic power of certain organisations; or rather, of its potential abuse.

The discussion on open PID infrastructures and the requirement of open source and open data is part of a broader initiative in favour of open scholarly infrastructures which includes governance not co-opted by particular interest groups, financial sustainability and

forkability, i.e., the ability of the community to recreate the infrastructure, software systems and data (Bilder et al., 2015).

However, openness is not only a required feature of PID infrastructures. Moreover, PID infrastructures can (and should) contribute to the openness of research output. In fact, in functional terms, PIDs can be used to produce narrow criteria and indicators of research quality and impact, without any link to open science, as well as they can be applied to innovative and support responsible research assessment, in line with the principles of open science. A recent report of Knowledge Exchange on "Openness Profile" draws attention to the limits but, above all, to the potential of PIDs in relation to research evaluation and open science (Jones & Murphy, 2021). PID Providers (ORCID, Australian Research Data Commons RAiD, Crossref, DataCite) were, along with funders, part of the interviewees and focus groups of this study which highlights the role of PIDs for the assessment of "structured content" and for the automation of workflows.

Figure 5. A representation of the Openness Profile as a user-curated portfolio of contributions to open scholarship (source: Jones & Murphy, 2021)
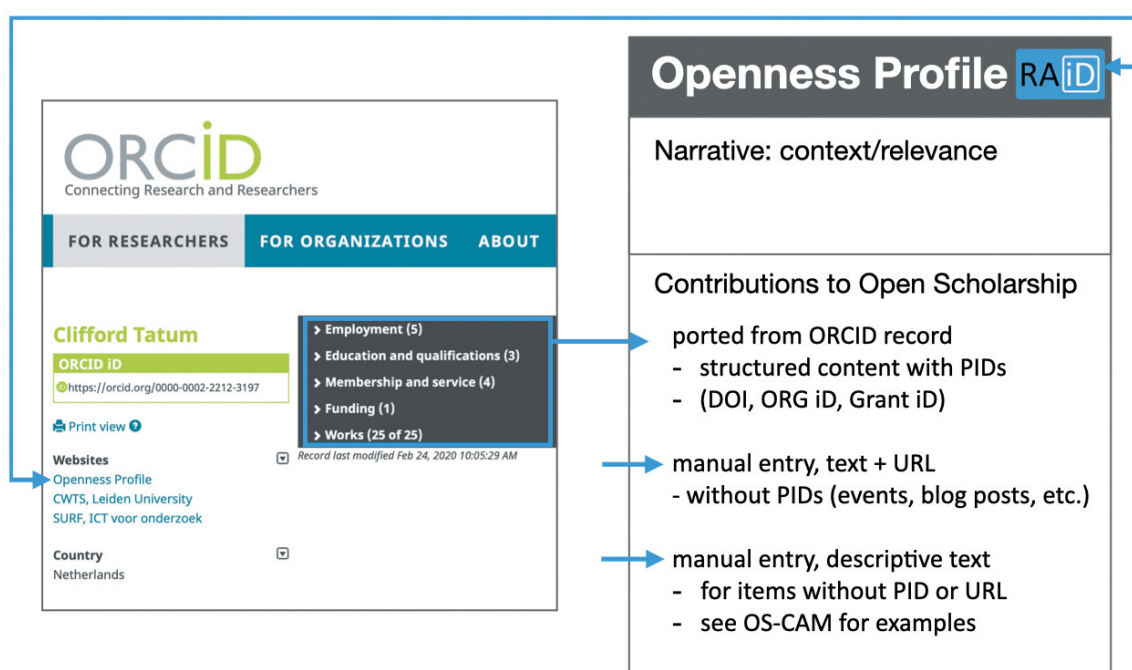
Following the Jisc roadmap for open access to UK research (Brown, 2020a), the KE report identifies five priority PIDs (people, institutions, funded grants, projects, outputs) and three key systems or infrastructures that "should be prioritised in order to more efficiently gather and curate open scholarship contributions: funding systems, CRISs, and institutional repositories". Also, the report describes a few "key identifiers", in particular, DOI (Crossref, DataCite), ORCID, RAiD, OrgID and GrantID, but leaves it open how (and if) PIDs can be useful for qualitative assessment, for narratives, events, blog posts and so on (here, the report only mentions URLs) and, more generally, for criteria based on the Open Science Career Assessment Matrix (OS-CAM) (European Commission Directorate General for Research and Innovation., 2017). Should existing identifiers be extended to those criteria? Should new identifiers be developed? Or should there be no such identifiers because of the risk of "diversion of policy and managerial attention towards things that can be measured" (Jones & Murphy, 2021)?

In all cases and fundamentally, the KE report on Openness Profile considers PIDs, "their associated metadata, and modern IT integrations through APIs (as) necessary to improve the flow of information between funders, national research organisations, assessors, institutions, publishers, and individual research contributors", especially for funder systems where the report observes "poor adoption of PIDs and little to no interoperability with downstream stakeholders" (Jones & Murphy, 2021).

The KE report's recommendations put forward the need for collective action to achieve the "Openness Profile", including a large variety of stakeholders, and unsurprisingly, one part of them are similar or consistent with our own analysis, namely the primary recommendation on the macro-level which is a call for action to the main players in the research community, to ensure continued consensus-building by investing in productive exchange and collaboration ("facilitate a stakeholder summit"). Another recommendation on the macro-level is to establish an ongoing working group with a focus on two PID-relevant areas (see Jones & Murphy, 2021):

▶ Technical facilitation of research management workflows ("developing standards for information interchange and interoperability as well as identifying key points for integration").

▶ Infrastructure survey and gap analysis (including coordination with related initiatives such as the "workflow and PID metadata recommendations from the Jisc PID roadmap").

The KE report also recommends the identification and recruitment of one or more "sponsor(s)" which in fact would play a kind of operating agent role and be responsible for a number of programme management tasks, including engineering of middleware to connect information systems using PID metadata.

On the meso-level, the report recommends

▶ For funders: Implementation of PIDs in grant information systems, beginning with Crossref's DOIs for grants and then extending to ORCID, RoR, and RAiD.

▶ For national research organisations: Promotion of community governance within the scholarly infrastructure, and setting up consortia to support PID subscriptions (membership) and development, with local registration agency services where necessary.

▶ For infrastructure providers: Development and support of greater interoperability between research systems and in particular, promotion and adoption of ORCID in CRIS and institutional repositories.

▶ For institutions: They should make use of PID- and metadata-enabled workflows, beginning with mandating ORCID for all research contributors, including technicians, engineers and support staff, and expand to DOI for outputs and awarded grants, RoR, and RAiD.

Looking more closely at these recommendations, only one of them addresses the requirement of open infrastructures, i.e., the promotion of community governance within the scholarly infrastructure whereas the other recommendations are more about efficiency

Figure 6. A representation of the KE Open Scholarship framework (source: Jones & Murphy, 2021)
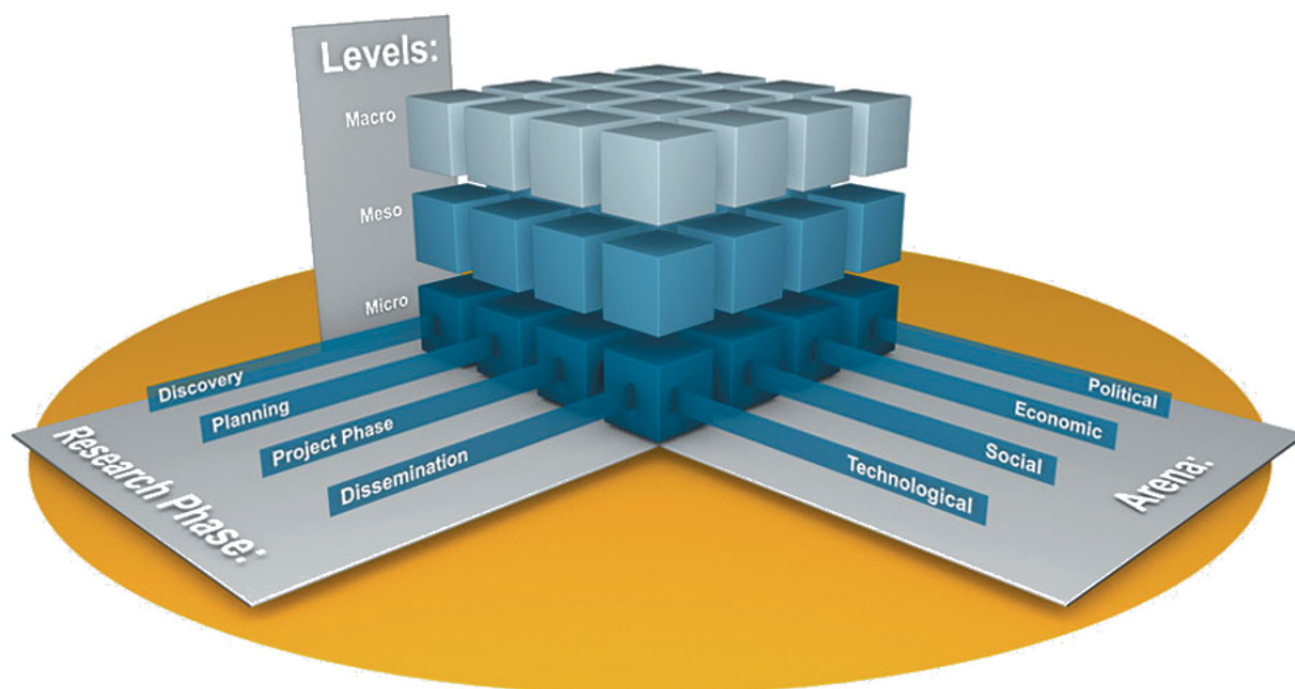
and performance of existing research infrastructures (funder systems, CRIS, repositories), without taking account of their openness.

Thus, the KE report on Openness Profile covers essentially the macro- and meso-level of the KE Open Scholarship Framework. The approach we take here, adds recommendations on the micro-level, i.e., regarding individual researchers.

Most of the PID-relevant recommendations clearly refer to the political and technological arenas of interest, such as coordination of actions, promotion of solutions, development of standards and engineering of systems, covering either the project phase (funding) or the whole research cycle. The economic arena is addressed only on the macro-level, by the recommendation to all stakeholders to identify and recruit one or more sponsors for PIDs.

Based on the expert interviews, on the case studies and on the discussions with KE experts, our conviction is that the recommendations should cover the whole

spectrum of the KE Open Scholarship Framework, i.e., all levels (macro, meso, micro) and all arenas of interests (PEST: political, economical, social, technological), and the whole research cycle, with a focus on the project phase and the dissemination of results.

Also, we include recommendations on the micro-level for individual researchers, concerning some political and socio-cultural factors (awareness, engagement in initiatives, following requirements, ORCID registration…). However, and similar to the KE report on Openness Profiles (Jones & Murphy, 2021), most of our recommendations (chapter 3) are positioned on the macro- and meso-levels, in the political and technological arenas. This is compliant with the general opinion and assessment from experts: the issue with PIDs is technological (of course) but above all, political, and it should be addressed via a global approach (national roadmap…) at the same time as by institutions, organisations, networks, communities and so on. The major part of the recommendations covers the whole research cycle while others (less) refer only to the project phase or to the dissemination of results (publications, data).

Looking more closely at the elements of the PEST scheme, the following risk and trust signals can be identified with regard to PIDs:

Dimension	Risk signals	Trust signals
Political	▶ Unstable or unclear governance of PID authorities or service providers ▶ Lack of commitment to a PID	▶ Historical record of the performance of their function ▶ Continuous support ▶ Participative governance (Openness)
Economic	▶ Lack or unclear funding ▶ Lack of investment	▶ Transparent reports on financial stability of a PID Service ▶ Invest in functionality and human resources ▶ Financial stability can be shown over a long period of time prospectively or retrospectively (e.g. in the case of national libraries)
Social	▶ Little or no communication with the community that uses or should use the service ▶ Unclear intention of the providers of the service, e.g. in terms of commercialisation	▶ Proven or assumed competence and resilience of the organisation providing the service ▶ Proven or assumed competence of persons in charge of the service ▶ Presumption, the services, the organisation and those responsible pursue goals that are in line with the goals and needs of the community
Technological	▶ Downtimes ▶ Non resolving PIDs ▶ Features are inferior to a competitor	▶ Highest possible availability of the service ▶ Transparent reporting on downtimes ▶ Correct addressing of objects and provision of metadata ▶ Risk mitigation ▶ Openness (open data)

8. The plane has taken off

Research funders, higher education institutions, research institutions, infrastructure organisations, publishers, communities and scientists use a portfolio of PIDs that has reached a high level of acceptance in some sectors, e.g. PIDs for publications, data, software or PIDs for persons and organisations. However, even in these consolidated areas, competing PID systems exist (DOI vs. URN, the global ORCID ID vs. disciplinary or national person PIDs, ROR vs. Ringgold), which need not be harmful per se, but can spur innovation or serve complementary needs.

Final considerations

Fragmentation is even more noticeable with emerging PIDs, e.g. those for instruments, facilities, conferences or grants. For example, different types of PIDs are already being awarded for grants without, for example, satisfying complementary needs. The dangers associated with fragmentation are manifold: wasting resources on solutions that prove technically, organisationally or economically unsustainable, plus resources for transferring PIDs of abandoned systems to those of superior competitors. Or worse, simply shutting down a service that the provider deems no longer financially viable or useful to operate - thwarting the PIDs' promise of permanent content identification and damaging the PIDs landscape as a whole by eliminating nodes in a network of interlinked PIDs.

The decision on which PIDs to use needs to be well-considered. Also, our recommendations propose to establish an overview of PID systems in the form of an observatory and, at the same time, to set up a PID Federation that accompanies the implementation of PIDs and the formation of new PIDs, e.g., by formulating best practices with regard to sustainability, technology and community work. This federation must bear in mind that the success of PIDs depends to a large extent on the needs of the communities and is a social as well as a technical challenge, because no matter how good the technology is, what counts in the end is trust in its provider. Rolling out these PIDs into infrastructures or processes depends largely on PID policies, e.g., of funders or HEIs, but it requires informed decisions, which in turn are enabled by the PID Federation.

As provocative as it may sound, to speak of a well-functioning PID infrastructure can be metaphorically summarised as "building the plane as we fly it". The PID plane has long since taken off, essential parts of it are in place (PIDs for people, organisations, outputs), the identification of other necessary parts is still in progress (grants, instruments, conferences, facilities), in part it has not yet been specified how they are to be identified. Even more, it is not foreseeable which parts will eventually make up the plane or whether the design process will ever end. Nevertheless, the plane is already flying. Achieving its maximum functionality, however, requires coordination that involves extensive exchange with designers of the individual parts, gives them enough space to develop solutions that precisely meet their requirements and yet harmonise with the technical and social conventions of the plane as a whole. The better they succeed, the more passengers dare to board and the more comfortable the journey.

9. References

Askitas, N. (2010). *What Makes Persistent Identifiers Persistent?* (RatSWD Working Paper Nr. 147). German Data Forum (RatSWD). https://doi.org/10.2139/ssrn.1639996

Belsø, Rene, Matthiesen, Martin, Parland-von Essen, Jessica, Béquet, Gaëlle, & KE Task & Finish Group For PID Risk & Trust. (2021). *Risks and Trust in Pursuit of a Well-functioning Persistent Identifier Infrastructure for Research.* https://doi.org/10.5281/ZENODO.5018216

Bilder, G., Lin, J., & Neylon, C. (2015). *Principles for Open Scholarly Infrastructures-v1.* https://doi.org/10.6084/m9.figshare.1314859.v1

Brown, J. (2020a, April 8). *Developing a persistent identifier roadmap for open access to UK research* [Other]. https://repository.jisc.ac.uk/7840/

Brown, J. (2020b). *PID Federation scoping study: Final report.* Zenodo. https://doi.org/10.5281/ZENODO.4059557

Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022). *Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems.* Zenodo. https://doi.org/10.5281/ZENODO.7100578

Bütikofer, N. (2009). *Catalogue of criteria for assessing the trustworthiness of PI systems* (Nr. 13; nestor-studies). nestor c/o Deutsche Nationalbibliothek. https://d-nb.info/1047610442/34

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), Art. 0. https://doi.org/10.5334/dsj-2017-013

Chan, L. (2018). *Asymmetry and inequality as a challenge for open access – an interview with Leslie Chan, (interview by Joachim Schöpfel).* Litwin Books. https://tspace.library.utoronto.ca/handle/1807/87296

Cousijn, H., Braukmann, R., Fenner, M., Ferguson, C., van Horik, R., Lammey, R., Meadows, A., & Lambert, S. (2021). Connected Research: The Potential of the PID Graph. *Patterns*, *2*(1), 100180. https://doi.org/10.1016/j.patter.2020.100180

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

European Commission. Directorate General for Research and Innovation. (2017). *Evaluation of research careers fully acknowledging Open Science practices: Rewards, incentives and/or recognition for researchers practicing Open Science.* Publications Office. https://data.europa.eu/doi/10.2777/75255

Frank, R. D. (2020). The Social Construction of Risk in Digital Preservation. *Journal of the Association for Information Science and Technology*, *71*(4), 474–484. https://doi.org/10.1002/asi.24247

Franken, J., Birukou, A., Eckert, K., Fahl, W., Hauschke, C., & Lange, C. (2022). Persistent Identification for Conferences. *Data Science Journal*, *21*, 11. https://doi.org/10.5334/dsj-2022-011

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly*, *27*(1), 51–90. https://doi.org/10.2307/30036519

Hellström, M., Heughebaert, A., Kotarski, R., Manghi, P., Matthews, B., Ritz, R., Conrad, A. S., Weigel, T., Wittenburg, P., & Valle, M. (2020). *A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC).* https://doi.org/10.2777/926037

Hellström, M., Johnsson, M., & Vermeulen, A. (2020). Identification and Citation of Digital Research Resources. In Z. Zhao & M. Hellström (Hrsg.), *Towards Interoperable Research Infrastructures for Environmental and Earth Sciences* (Bd. 12003, S. 162–175). Springer International Publishing. https://doi.org/10.1007/978-3-030-52829-4_9

Jones, P., & Murphy. (2021). *Openness Profile: Modelling research evaluation for open scholarship*. Zenodo. https://doi.org/10.5281/ZENODO.4581490

Kiley, R., Fentrop, N., & Hendricks, G. (2018). *Wellcome explains the benefits of developing an open and global grant identifier* [CrossRef blog]. https://www.crossref.org/blog/wellcome-explains-the-benefits-of-developing-an-open-and-global-grant-identifier/

Leveson, N., Dulac, N., Marais, K., & Carroll, J. (2009). Moving Beyond Normal Accidents and High Reliability Organizations: A Systems Approach to Safety in Complex Systems. *Organization Studies*, *30*(2–3), 227–249. https://doi.org/10.1177/0170840608101478

Lippert, S. K., & Michael Swiercz, P. (2005). Human resource information systems (HRIS) and technology trust. *Journal of Information Science*, *31*(5), 340–353. https://doi.org/10.1177/0165551505055399

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 12:1-12:25. https://doi.org/10.1145/1985347.1985353

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1), 1–28. https://doi.org/10.1037/h0092976

Science Europe. (2016). *Science Europe Position Statement on Research Information Systems*. Science Europe. https://www.scienceeurope.org/our-resources/position-statement-on-research-information-systems

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). *Understanding the Formation of Trust in IT Artifacts* (F. G. Joey, Hrsg.). Association for Information Systems. http://aisel.aisnet.org/icis2012/proceedings/HumanBehavior/11/

Weigel, T., Plale, B., Parsons, M., Zhou, G., Luo, Y., Schwardmann, U., Quick, R., Hellström, M., & Kurakawa, K. (2018). *RDA Recommendation on PID Kernel Information*. https://doi.org/10.15497/RDA00031

Wittenburg, P., Hellström, M., Zwölf, C.-M., Abroshan, H., Asmi, A., Di Bernardo, G., Couvreur, D., Gaizer, T., Holub, P., Hooft, R., Häggström, I., Kohler, M., Koureas, D., Kuchinke, W., Milanesi, L., Padfield, J., Rosato, A., Staiger, C., van Uytvanck, D., & Weigel, T. (2017). *Persistent identifiers: Consolidated assertions. Status of November, 2017.* Zenodo. https://doi.org/10.5281/zenodo.1116189

Yoon, A. (2014). End users' trust in data repositories: Definition and influences on trust development. *Archival Science*, *14*(1), 17–34. https://doi.org/10.1007/s10502-013-9207-8

10. Appendices

A. Members of the Task & Finish Group

The activity was led by KE representatives Frank Manista (Jisc) and Josefine Nordling (CSC). The Task & Finish Group for this activity consists of experts from across each of the six KE partner countries:

▸ Gül Akcaova, SURF, Netherlands
▸ Pascal Aventurier, IRD, France
▸ Rene Belsø (Expert Lead), DeiC, Denmark
▸ Gaëlle Béquet, ISSN, France
▸ Britta Dreyer, Technische Informationsbibliothek, Germany
▸ Nathalie Fargier, CNRS, France
▸ Jürgen Kett, Deutsche Nationalbibliothek, Germany
▸ Martin Matthiesen (Expert Co-lead), CSC, Finland
▸ Hilda Muchando, Human Made / ALTIS, UK
▸ Stephanie Palek, Deutsche Nationalbibliothek, Germany
▸ Jessica Parland-von Essen, CSC, Finland
▸ Laurents Sesink, Leiden University, Netherlands
▸ Clifford Tatum, CWTS, Netherlands
▸ Adam Vials Moore, Jisc, UK
▸ Kirsty Wallis, University College London, UK

B. Acknowledgments

We would like to thank all interviewees for their time and openness spent on discussing risk and trust in the PID infrastructure.

C. PID roles

The following list contains PID Roles including definitions according to KE and EOSC as well as examples of persons/ entities holding these roles. It was compiled together with the Task & Finish Group. In the discussion, it turned out to be often difficult to establish a one-to-one correspondence between PID roles and actual organisations holding these roles.

Role as named by KE (Belsø, Rene et al., 2021) and EOSC (Hellström, Maggie et al., 2020)	Definition (KE)	Definition (EOSC)	Examples
PID Authority	▶ Under the control or part of international standardization body ▶ establishes and enforces processes for creating, approving, maintaining and terminating PID standards	A controller responsible for maintaining the rules for defining the integrity of PIDs within a PID Scheme. These rules may include setting standards for lexical formats, algorithms and protocols to ensure global uniqueness, together with setting quality of service conditions to enforce compliance to the rules. PID Authorities may be organisations (e.g. DOI.org), which enforce control over a PID infrastructure. But there may also be Authorities which do not have a central control, but provide a community standardisation mechanism that specifies the conformance of PIDs to a PID Scheme	▶ ARK Alliance (ARK) International DOI Foundation (DOI) ▶ IETF (URN) ▶ ISSN International Centre
PID Service Provider	▶ organisation with clear business model in agreement with PID authority ▶ implements and invests in financial, technological and human resources to sustain PID information system ▶ PID information system binds PIDs to objects and stores metadata, provider makes PIDs available to PID managers	An organisation which provides PID services in conformance to a PID Scheme, subject to its PID Authority. PID Service Providers have responsibility for the provision, integrity, reliability and scalability of PID Services, in particular the issuing and resolution of PIDs, but also lookup and search services.	▶ DataCite, Crossref, ▶ consortia like the German DOI consortium ▶ ISSN National Centres (93 countries)

Role as named by KE (Belsø, Rene et al., 2021) and EOSC (Hellström, Maggie et al., 2020)	Definition (KE)	Definition (EOSC)	Examples
PID Manager	▸ manages PID requests to service provider ▸ uploads and updaters metadata to PID information system ▸ populates local system with PIDs	PID Managers have responsibilities to maintain the integrity of the relationship between entities and their PIDs, in conformance to a PID Scheme defined by a PID Authority. A PID Manager will typically subscribe to PID services to offer functionality to PID Owners within the PID Manager's services. One example is a Service Provider which uses PID Services as part of its own service delivery. For example, PID Managers may include a provider of a data repository, a data catalogue, or a research workflow system.	▸ Repositories (Zenodo) AND/OR stakeholders operating repositories, also Publishers/ Database providers ▸ Library Catalogues and Repositories
PID Owner	▸ creates and updates PID's referent metadata within local system	An actor (an organisation or individual) who has the authority to create a PID, assign PID to an Entity, provide and maintain accurate Kernel Information for PID.	▸ Repository Managers
PID End User	▸ In the KE definition, characteristics of PIDs are described here	The end user of PID services and PID User Services. These can be for example researchers, or software, or services produced to support researchers. End users will use PIDs to cite and access resources or Kernel Information.	▸ Researchers, Research-performing organisations, Research funders, Software, Systems (e.g. Citation Counters, Reference Management Systems) ▸ Libraries

D. Participants

This is the list of all interviews which were conducted during this study. Rows marked in white indicate joint interviews.

First Name	Last Name	Organisation/ Service	Country	Role
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/ConflDent	GER	PID Authority
Juha	Hakala	URN representative	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

E. Interview protocol Risk & Trust in PID systems

As mentioned, some common perceptions of PIDs in general and of perceived risks for well-functioning PID systems were expected across all roles (PID Authority, PID Service Provider, PID Manager, PID Owner, PID End User), while some perceptions on trust were expected to be different depending on whether a person was speaking as a PID Manager, PID Owner, PID End User or for a PID Authority/ PID Service Provider. Therefore, questions 1-15 and 29-31 were addressed to all experts, whereas questions 16-22 were addressed to PID Managers, PID Owners and PID End Users only and questions 23-28 were addressed to PID Authorities and PID Service Providers only.

Introductory questions

1. Please tell me a bit about yourself and your background. How did you come to be involved in specific PIDs and PID systems?
    a. e.g. education, previous work etc.

2. How would you describe your role at (name of organisation)?
    a. How would you describe your expertise regarding PIDs?

PIDs

3. What kind of PIDs do you work with?
    a. How mature do you consider these PIDs (systems) to be at the moment, particularly in your country?

4. Are there any PIDs that you prefer using over other options?
    a. Are there any PIDs you refrain from using in general? If so, why?

5. For what purpose do you consider PIDs useful and how does your organisation use them?
    a. e.g. infrastructure, software, specific use
    b. if multiple: How are the identifiers interoperable?
    c. if just one: Which other identifiers do you plan on using and why?

6. Which systems are the identifiers incorporated in?
    a. Do you have any plans incorporating more PIDs? If so, which ones and why?

7. How would you describe your organisation's role within PID systems you use or offer?
    a. If clarification is needed: What kind of authority does your organisation have regarding the management of the PIDs you offer or use?

8. What do you think are the strengths of the identifiers?
    a. In terms of their government structure (e.g. centralised vs decentralised)?
    b. In terms of their interoperability with other identifiers on a national and international level?
    c. Other?

9. In general, what do you think are the greatest strengths of PID systems/infrastructures? What are they most useful for?

Risks

10. What do you think are the weaknesses of the identifier(s)?
    a. How did you identify these weaknesses?
    b. Can you remember or identify a time when the PID didn't work? What happened?

11. How vulnerable is your organisation to these weaknesses?
    a. What does your organisation do to counteract these weaknesses?

12. What do you think are the greatest risks or threats in PID structures in general and in your specific area of PID expertise in particular?
    a. Can you think of any specific
        i. technological risks? e.g. failure of resolving service, faulty metadata/kernel information
        ii. economical risks? e.g. loss of funding for PID system
        iii. social risks? e.g. lack of uptake in the community
        iv. political risks? e.g. organisational failure of PID providers, governance/organisational model
        v. other risks?

13. Why are these risks significant and what negative consequences do they have?

14. Does your organisation have a fallback plan or another way for risk management regarding PIDs?

15. Are there any other uncertainties concerning PIDs that you can think of?
    a. In terms of their government structure (centralised vs. decentralised)?
    b. In terms of their interoperability in the larger PID infrastructure?
    c. Other?

Trust (PID Manager, PID Owner, PID User)
16. How well do you think are risks about identifiers communicated to
    a. PID managers?
    b. PID owners?
    c. PID service providers?
    d. PID users?

17. How confident are you that you are aware of all possible risks concerning the usage of PIDs? (*note: structural assurance*)
    a. Do you feel that the organisations governing and providing the PIDs document risk management techniques and guarantees well enough and in a transparent way?
    b. Do you believe that enough checks and balances are in place in case of outages/failures?

18. Do you feel that the organisation responsible for the PIDs you use/want to use has the best interest of the community in mind? (*note: benevolence, positive intentions*)
    a. How do you know? Please explain.

19. Do you think they are fit to act as a PID supplier/authority? (*note: competence, ability*)
    a. Why? Please explain.

20. Do you believe that they share enough information about their product? (*note: reliability, honesty*)
    a. Metadata schema
    b. Government structures
    c. Internal risk management
    d. Technical documentation
    e. other

21. Do you believe that in general the technical infrastructure is persistent enough? (*note: trustworthiness of technology*)
    a. Are the PIDs reliable (enough) for your system?
    b. How useful are the PIDs for your community?

22. How confident are you that the PIDs you use will stay usable/actionable/persistent over a sufficient amount of time? (*note: functionality, reliability, predictability of technology*)
    a. Would you change anything about their functionality? If so, what would that be?

Trust (PID Authority, PID Service Provider)

23. How well do you think you are doing in communicating risks about identifiers to
    a. PID managers?
    b. PID owners?
    c. PID Service providers?
    d. PID users?

24. How confident are you that you are aware of all possible risks concerning the usage of PIDs? (*note: structural assurance*)
    a. Do you have a written policy/protocol in place in case of system outages/failures?
    b. Has this protocol been shared to PID managers/owners/users?
    c. Do you have a fallback and/or persistence solution in place, in case your organisation encounters general failure?

25. How much work do you put into openly communicating about possible risks with your user community? (*note: benevolence, positive intentions + reliability, honesty*)
    a. How do you do this? Please explain.
    b. Are there any other community orientated activities that you do or are planning to do?

26. What makes your organisation fit to be a PID authority/PID supplier? (*note: competence, ability*)
    a. What kind of organisation is needed for this role? What are the requirements?

27. Do you believe that in general the technical infrastructure is persistent enough? (*note: trustworthiness of technology*)
    a. Are the PIDs reliable (enough) for your system?
    b. How useful are the PIDs for your community?

28. How confident are you that the PIDs you use will stay usable/actionable/persistent over a sufficient amount of time? (*note: functionality, reliability, predictability of technology*)
    a. Would you change anything about their functionality? If so, what would that be?

Wrap-up

29. What are in your view the clearest gaps/issues in the development of the general PID infrastructure at the moment?

30. Can you name three measures which in your opinion could improve the functionality and usefulness of the general PID infrastructure?

31. Is there anything I haven't asked you that you would like to discuss?

F. Case studies

1. **Adoption of the DAI in the Netherlands and subsequent superseding by ORCID/ISNI | Zenodo**

2. **The gradual implementation of organisational identifiers (OrgIDs) | Zenodo**

3. **Persistent identifiers for research instruments and facilities: an emerging PID domain in need of coordination | Zenodo**

4. **The role of research funders in the consolidation of the PID landscape | Zenodo**

5. **IGSN - building and expanding a community-driven PID system | Zenodo**

6. **RePEc Author Service: An established community-driven PID | Zenodo**

7. **Failed PIDs and unreliable PID implementations | Zenodo**

Glossary

ARK (Archival Resource Key): PID to identify any information objects, mostly used in libraries, data centres, archives, museums, publishers, and government agencies to provide references to scholarly, scientific, and cultural objects.

APC (Article processing charges): Fee charged to authors of manuscripts accepted in specific (paid-for Gold Open Access) journals in order to have their work published Gold Open Access under a Creative Commons (CC) licence.

ConfIDs: PIDs issued by DataCite to identify scientific events such as conferences or conference series.

CRIS (Current Research Information System): A database or other information system that stores and provides metadata for the research activity funded by a RFO or conducted at an RPO (source: Wikipedia)

Crossref: A registration agency for the Digital Object Identifier (DOI). Crossref is run by the Publishers International Linking Association Inc (PILA).

DAI (Digital Author Identifier): A national-level PID to uniquely identify researchers used in the Dutch research system until 2016.

DataCite: A registration agency for Digital Object Identifier (DOI). DataCite is run by the not-for-profit DataCite Consortium (consisting of infrastructure organisations).

DOI (Digital Object Identifier): PID to uniquely identify any information or physical object. DOIs are widely used in Scholarly Communication and mostly assigned to research outputs such as publications or datasets, but also to grants and other entities.

Economic: Relating to the process or system by which goods and services are produced, sold, and bought (source: Merriam-Webster dictionary).

Ecosystem: In the context of this report a dynamic socio-technical environment with identifiable but open borders, consisting of interacting elements (such as organisations, technical services, persons) and rules, definitions, contractually fixed and unspoken agreements linking these elements, and including all interactions between the different elements.

Emerging PID: Areas in which the use of PIDs for the identification of objects/entities is in its early stage of development and not yet widespread, e.g. PIDs for instruments and facilities, conferences, grants.

EOSC (European Open Science Cloud): An environment for hosting and processing research data to support EU science. (Source https://eosc-portal.eu/about/eosc).

FAIR: A set of principles to guarantee Findability, Accessibility, Interoperability, and Reusability of data and/or scientific outputs.

Federation: An encompassing entity formed by a union of smaller or more localised entities with a defined purpose, shared interests and a common vision.

GDPR (General Data Protection Regulation): Data protection legislation issued by the European Commision. The GDPR law entered into force in 2016 and as of May 25, 2018, all organisations were required to be GDPR-compliant. (source: https://gdpr.eu/what-is-gdpr/).

Grant ID: PID to identify research grants issued by a research funder. Grant IDs specify the funder's research grant or contract number for a funded research project they have awarded. Projects themselves are persistently identified via RAiDs, see below.

Handle ID: PID to identify any information object. The Handle system underpinning its operation is run by the Corporation for National Research Initiatives (CNRI).

HEI (Higher education institution): European organisations providing higher, postsecondary, tertiary, and/or third-level education. Research is typically part of their remit too. (source: IGI Global).

IGSN (International Generic Sample Number): PID to identify samples, context objects of samples or collections of samples. The same acronym was originally used to denote the International Geo Sample Number.

Infrastructure: A) the resources (such as personnel, buildings, or equipment) required for an activity B) the underlying foundation or basic framework (as of a system or organisation) (taken from the Merriam-Webster dictionary). In this report infrastructures are defined as socio-technical ecosystems made up of people, technologies, and institutions (organisations), which generate, manage and preserve information and knowledge (based on Edwards, 2010; Bowker et al., 2010). In the best case scenario, infrastructure is invisible to its end-user.

ISRCTN (International Standard Randomised Controlled Trial Number): A registry and curated database containing the basic set of data items deemed essential to describe a study at inception, as per the requirements set out by the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) and the International Committee of Medical Journal Editors (ICMJE) guidelines. All study records in the database are freely accessible and searchable and have been assigned an ISRCTN ID.

KE (Knowledge Exchange): Collaboration network across six key national organisations within Europe tasked with developing infrastructure and services to enable the use of digital technologies to improve higher education and research (source: https://knowledge-exchange.info/about-us).

Macro-Meso-Micro: Levels at which an analysis is conducted. At a macro level, large aggregates or systems are examined (such as State, government, nation…). At a meso level, the focus is on the parts or sectors of these systems (institutions, organisations, networks…). At a micro level, the actions and decisions of the actors and/or the relationships between the

actors are of interest (scientists, researchers, librarians, system managers…). (source: Schubert, Klaus/Martina Klein: Das Politiklexikon. 7., aktual. u. erw. Aufl. Bonn: Dietz 2020).

MPA (Multi-Primary Administrator): Credentialed organisation involved in the management of the Global Handle Registry (GHR). GHR operation is collaboratively managed by the DONA Foundation and MPAs (source: DONA Foundation, https://www.dona.net/handle-system).

NREN (National Research and Education Network): A specialised internet service provider dedicated to supporting the needs of the research and education communities within a country.

Observatory: A modified proposal: For this report, a virtual location used for observing PID related events and, in particular, for listing new and established PIDs including use cases, specifications and possible development scenarios to enable PID experts and policy makers to make informed technical and strategic decisions regarding e.g. the adoption of specific PIDs.

Open Infrastructure: Scholarly communication resources and services, including software, that the research and scholarly community of users depends upon to collect, store, organise, access, share, and assess research (source: Defining open infrastructure – SCOSS – The Global Sustainability Coalition for Open Science Services).

ORCID (Open Researcher and Contributor ID): A PID to uniquely identify authors in the research and academic domains. The ORCID identifier is run by the international, interdisciplinary, open, non-proprietary, and not-for-profit ORCID organisation.

OS-CAM (Open Science Career Assessment Matrix): Systematic framework for the evaluation of research careers in which Open Science practices are fully acknowledged.

PID (Persistent Identifier): A globally unique and permanent reference to a document, file, web page, or other (digital or non-digital) entity. In addition to access to these objects, a PID usually provides metadata describing the content or technical attributes of the

object identified by the PID and, if necessary, sets it in relation to other objects.

PID Graph: A network of interconnected PID entitites, exploiting the metadata registered with them, e.g., to connect outputs associated with a particular researcher, grant, institution or funder, for discovery and impact assessment. (based on Introducing the PID Graph – DataCite Blog).

PID Roles: PID Authorities (e.g. ISSN International Centre, DOI Foundation), PID Service Provider (e.g. DataCite, Crossref), PID Manager (e.g. databases), PID Owner (e.g. repository managers), PID End User (e.g. researcher), for details see appendix 9 c.

PID System: The mutually referenced combination of definitions, policies, services and data sources which are used for the administration and use of PIDs. (based on Bütikofer, 2009).

PIDINSTs: PIDs to identify research instruments and facilities.

Political: Relating to, involving, or involved in politics and policies (for this report primarily science politics). (source: Merriam-Webster dictionary).

PURL (Persistent Uniform Resource Locator): a PID to identify web resources, run by the Internet Archive.

OrgIDs (Organisation IDs): PIDs to identify organisations, usually research-performing ones such as universities or research centres or institutes, but also companies and other entities.

RAiD (Research Activity Identifier): A PID primarily aimed for research projects. A RAiD is an envelope that includes other PIDs in it such as ORCIDs, RORs, grant IDs and DOIs for research outputs. The RAiD standard is currently under development by the ISO, https://www.iso.org/standard/75931.html .

RDA (Research Data Alliance): An organisation aimed to enable the open sharing and re-use of research data.

RePEc (Research Papers in Economics): Not-for-profit initiative that seeks to enhance the dissemination of research in Economics and related disciplines.

RFOs (Research Funding Organisations): Also known as research funders, these are the public or private-sector organisations devoted to funding research. This funding usually takes the form of funded projects but RFOs also fund research-performing organisations and research instruments and facilities.

Ringgold: PID to identify organisations. Ringgold ID is a PID for organisations in the publishing industry supply chain (publishers, funders, universities, corporations, government entities...), owned and administered by a corporate company (Ringgold) which is part of the US Copyright Clearance Center (CCC).

Risk: The likelihood of an event multiplied by the magnitude of its (damaging) consequences. For individuals a risk is present in a situation, where the possible damage would be greater than the advantage sought (Luhmann, 1988). Taking a risk is an outcome of trust.

ROR (Research Organisation Registry): A PID to identify research organisations. ROR is run as a not-for-profit service by the California Digital Library, Crossref and DataCite.

RPO (Research Performing Organisations): Public or private-sector organisations in which research is conducted. Universities and research institutes are the most frequent RPOs.

RRID (Research Resource Identifier): A PID for referencing research resources especially in the biomedical field, such as antibodies, organisms, tools, plasmids and cell lines.

Situational Normality: Belief that success is likely, because the situation is normal (McKnight et al., 1998).

Socio-technical: Relating to technology-using processes between organisations, individuals, groups, the consequences of which are not predetermined by the technology, but are subject to social interpretations, negotiations and sanctions. Also taking into consideration that technology is socially constructed and vice versa, technology has an influence on the construction of the social world.

Structural Assurance: Belief that success is likely because such contextual conditions as promises, contracts, regulations and guarantees are in place (McKnight et al., 1998).

Sustainability: A method of harvesting or using a resource so that the resource is not depleted or permanently damaged (source: Merriam-Webster dictionary). In the context of scholarly communications, a framework where the appropriate mechanisms are in place to guarantee the mid- and long-term operation of a service.

Trust: Willingness of a party to be vulnerable to the action of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other party (source: Mayer et al., 1995).

URL (Uniform Resource Locator): Reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it (source: Wikipedia).

URN (Uniform Resource Name): PID to identify any information or physical object. URN NBNs are a subset of URNs used by national libraries to assign National Bibliography Numbers.