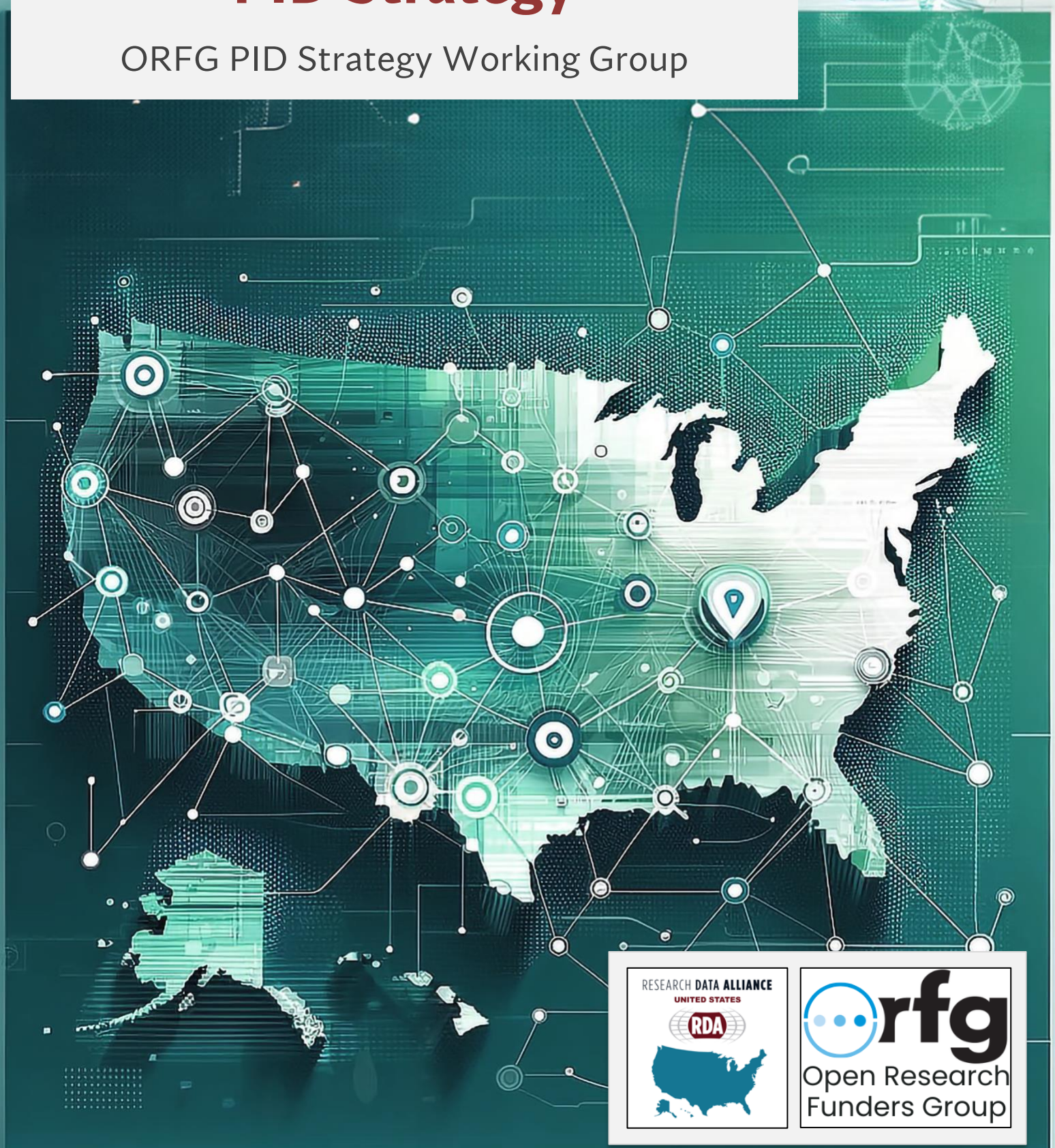
Developing a US National PID Strategy

ORFG PID Strategy Working Group



Developing a US National PID Strategy

by ORFG PID Strategy Working Group

Cite As: Developing a US National PID Strategy. ORFG PID Strategy Working Group. 2024. Zenodo. https://doi.org/10.5281/zenodo.10811008

1. Context Setting for Developing a US National PID Strategy

1.1 Document Purpose and Process

This document offers a set of recommendations regarding persistent identifiers (PIDs) that can be applied to numerous elements of the research landscape across the US, including governmental entities, infrastructure providers, research institutions, public and private funding bodies, publishers, content providers, and other stakeholders. Ideally, this guidance will be widely adopted by organizations throughout the research ecosystem in the US and potentially adapted globally in other national contexts around the world, as part of a growing movement to deploy national persistent identifier strategies.

Utilizing the framework created by the Research Data Alliance,[1] this report was created in collaboration with members of the Higher Education Leadership Initiative for Open Scholarship (HELIOS Open) and the Community Effort on Research Output Tracking workstreams organized by the Open Research Funders Group (ORFG).[2] Members of the working group are included in the appendix to this document.[3] The report outlines the benefits of PIDs, their associated metadata, and the systems that connect them in advancing open scholarship goals in the United States. It provides information on the research and policy landscape associated with PIDs, discusses the value of PID infrastructure, and offers recommendations for effective utilization of PIDs in connecting and tracking research outputs.

The ultimate goal of this document is to further develop its recommendations into a national PID strategy, by way of presenting it as the basis for formalization as a U.S National Standard via the ANSI-accredited standards development process of the National Information Standards Organization (NISO).[4] After this current document has been circulated publicly for input and feedback, the draft will be included as the foundation of a proposed text of a National Standard. If approved as a project, the text will be vetted by a NISO working group and undergo a public comment period for broadest possible consensus before it is published. The expectation is that this document will also be shared via the Research Data Alliance (RDA) National PID Strategies Interest Group[5] and RDA-US[6] so that other national and international

---

[1] Simons, N., Brown, C., Bangert, D., & Sadler, S. (2023). National PID Strategies Guide and Checklist (Version 1.0). Research Data Alliance. Zenodo. https://doi.org/10.15497/RDA/00091

[2] https://www.orfg.org/news/2023/6/5/orfg-advances-efforts-to-improve-research-output-tracking

[3] See https://docs.google.com/document/d/1KeKDCo_-BDllxNIklzHiK4ea89oX9eRaIgjs0hCKr9M/ for the original charge of this multi-stakeholder group.

[4] https://www.niso.org/creating-niso-standards

[5] https://www.rd-alliance.org/groups/national-pid-strategies-interest-group

[6] https://rdaus.org

organizations can learn from the application of PIDs in the US context. In this way, the efforts of this group can be seen as contributing to and benefiting from the ongoing conversation about improving the ecosystem of shared research outputs.

1.2 Policy Landscape in the US

Our effort to create recommendations for a PID national strategy for the US comes at a time of clarity and forward thinking regarding policy in the US. Policymakers have increasingly recognized the importance and potential value to be gained using PIDs. In the past years, there have been several policy and guidance documents issued that reference this. Specifically, the White House Office of Science and Technology Policy (OSTP) issued two significant memos described below, the 2013 OSTP "Holdren" Memo and the 2022 "Nelson" Memo, which have had far-reaching implications for the sharing of research outputs and which, by extension, touch on the need for wider utilization of PIDs. A range of agency guidance developed in response to these memos relates to the specific use of PIDs in aspects of the US research ecosystem, such as National Science Foundation (NSF) policies related to data management plans,[7] effective data publishing,[8] and data sharing[9] and the Department of Energy's Policy for Digital Research Data Management.[10] The NSF is also using PIDs as an optional tool for its Biographical Sketch Common Form[11] and its Current and Pending (Other) Support Form.[12] Below is a brief overview of these two OSTP memos.

2013 OSTP Memo on Public Access to Federally Funded Research

"*Increasing Access to the Results of Federally Funded Scientific Research*," issued on February 22, 2013, emphasized open access to scientific publications and data resulting from federally funded research.[13] While not explicitly focused on PIDs, it indirectly influenced their adoption through data management and sharing requirements:

- Data management plans (DMPs): The 2013 memo called for federal agencies to require DMPs for funded research projects. To do this, researchers need to describe the data created and plans for their long-term access. PIDs play a vital role in tracking, citing, and linking datasets, ensuring transparency and reproducibility, as advocated by the memo.
- Data sharing: Federal agencies were encouraged to make research data publicly accessible, which in turn will improve reuse, leading to a greater return on research investments.

---

[7] https://www.nsf.gov/bfa/dias/policy/dmpdocs/dmref.pdf
[8] https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp
[9] https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf
[10] https://www.energy.gov/datamanagement/doe-policy-digital-research-data-management
[11] https://www.nsf.gov/bfa/dias/policy/researchprotection/commonform_biographicalsketch.pdf
[12] https://www.nsf.gov/bfa/dias/policy/researchprotection/commonform_cps.pdf
[13] https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

- Long-term accessibility: PIDs are instrumental in ensuring the long-term accessibility of data, aligning with the memo's emphasis on preserving and sharing research results.

2022 OSTP Memo on Public Access to Federally Funded Research

"Ensuring Free, Immediate, and Equitable Access to Federally Funded Research," issued on August 25, 2022, provides policy guidance to federal agencies with research and development expenditures on updating their public access policies.[14] Key implications for PIDs are as follows:

- Public access: The memo urges federal agencies to update their public access policies by no later than December 31, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without embargoes.
- Scientific data accessibility: Scientific data underlying scholarly publications should also be made freely available and publicly accessible at the time of publication.
- Metadata and attribution: Federal agencies are instructed to make appropriate metadata associated with scholarly publications and data publicly available. This metadata includes author information, funding sources, and persistent identifiers, reinforcing the role of PIDs in uniquely identifying and tracking research outputs. The availability of these elements are necessary components of any research output to achieve the goals of free, immediate, and equitable access.
- Research integrity: Ensuring the integrity of publicly accessible research outputs aligns with the principles of openness and accountability.

In addition to these two memos, in 2022, the US National Science and Technology Council issued a report titled "Guidance for Implementing National Security Presidential Memorandum 33 (NSPM-33) on National Security Strategy for United States Government-Supported Research and Development," specifically related to PIDs and in part focusing on "how research agencies will incorporate persistent identifiers into disclosure processes to bolster research security and integrity while reducing administrative burden."[15] Combined, the documents have had a substantial impact on discourse regarding the promotion and utilization of PIDs in the United States and beyond. They, and other policy and guidance documents that have emerged, prioritize immediate public access to federally funded research, data transparency, and research

[14] https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf

[15] https://www.whitehouse.gov/wp-content/uploads/2022/01/010422-NSPM-33-Implementation-Guidance.pdf. Note: The NSPM-33 memorandum uses the term "digital persistent identifier (DPI)." This term and "PID" are synonymous. Ideally, the community will converge on the use of the acronym "PID".

integrity, all of which are closely tied to and in many cases depend upon the effective use of PIDs in advancing open scholarship goals in the US.

2. PIDs in the Research Ecosystem

PIDs have long provided the basis for citation linking, cataloging systems, entity identification, and description, as well as supporting research discovery via their associated metadata. PIDs uniquely identify a given resource (or research entity) and enable robust digital access to digital objects, their metadata, and/or related services. All the benefits associated with PIDs combine to enable the navigation and analysis of research and innovation.

The US-based research community has played a pivotal role in the development and adoption of PID infrastructure, with collaborative efforts driven by investments from federal agencies, research institutions, industry, and nonprofit community organizations. These endeavors are well adopted and create a robust information infrastructure that elevates the resilience, efficiency, and excellence of the global research sector.

2.1 PIDs in the US Research Ecosystem

The research ecosystem is vast and multifaceted, encompassing government agencies, universities, corporate labs, philanthropic funders, publishers, and various support services, all dedicated to conducting and sharing research findings. Within this global system, PIDs already play a crucial role in tracking research outputs and their interconnections. These identifiers are used billions of times each month to link research elements, enhancing the ecosystem's efficiency.[16] However, there is potential to further improve and expand their impact with increased investment in their integration.

Nationally adopting recommendations for PID usage benefits all stakeholders by streamlining interactions and reducing redundant efforts. For instance, philanthropic organizations that fund research can more easily ensure adherence to their guidelines and assess the impact of their investments. Likewise, PIDs can help publishers and data repositories enhance user engagement with their content through easier

[16] The research ecosystem currently uses PIDs extensively in identifying textual outputs (>150 million Crossref DOIs assigned; see https://www.crossref.org/06members/53status.html), data and software objects (>50 million DataCite DOIs assigned; see https://datacite.org/blog/datacites-thriving-community-3000-repositories-and-counting/), people (15 million ORCIDs assigned; see https://info.orcid.org/orcidat10-celebrating-10-years-of-the-orcid-galaxy/), and many other elements of the research ecosystem. Beyond assignment, these systems support billions of data connections to these research objects each month. While broadly adopted, the application and use of these systems could be improved, and ongoing investments in integration could spur advancements in expanding their impact.

discovery, citation, and reuse. They also can simplify reporting on content usage to various stakeholders.

Moreover, integrating PIDs across the research lifecycle eases publication processes, such as filling in author details automatically and verifying researchers' identities. This is particularly important in today's global research environment, which includes both US and international participants. The structured data format of PIDs enables cost-effective, multilingual searches across databases, making research more accessible to English and non-English speakers and facilitating broader participation.

2.2 PID Strategies in the International Landscape

US engagement in international research activities is extensive, robust, and impactful. Involvement includes international researchers working on US campuses, direct international cross-institutional partnerships, and other global scientific collaborations. Many research infrastructures (e.g., publication and software systems) are set up to serve global audiences and utilize international standards. Therefore, guidance provided by this proposed standard should align and be consistent with international standards and efforts[17] as much as possible, given the specific contexts within the US research environment noted above. As research and scholarly communications are worldwide endeavors, consistent use of PIDs and—at a minimum—interoperable systems for data exchange in this global research ecosystem are critical to a healthy and effective research ecosystem.

Several international efforts to improve scholarly communications have identified the need for robust PID infrastructure for connecting this ecosystem. For example, the 2022 UNESCO Open Science Toolkit provides useful guidance.[18] It specifically calls out the need for persistent identifiers, stating that "Due attention should be given to unique persistent identifiers of digital objects. Examples include the definition and attribution of persistent identifiers as appropriate for each type of digital object, the necessary metadata for their efficient assessment, access, use and reuse, and proper stewardship of data by a trusted regional or global networks of data repositories."

[17] 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.* Research Data Alliance. https://doi.org/10.5281/zenodo.1116189

[18] https://unesdoc.unesco.org/ark:/48223/pf0000383711

Groups across other nations have crafted national PID strategies, including cross-stakeholder teams in Australia,[19] Canada,[20] the United Kingdom,[21] and elsewhere. Additionally, several countries have undertaken economic analyses to quantify the potential savings that could be realized from the application of PIDs in their respective research ecosystems. The potential savings have been estimated to be in the millions of dollars, even in countries with much smaller research ecosystems than the US.[22]

3. Proposed Components of a National PID Strategy

The sections that follow present the reasoning and rationale for embracing PIDs as well as a set of recommendations designed to significantly improve the implementation and efficacy of PIDs within the diverse spheres of the US research landscape[23]. These recommendations are the product of extensive collaboration among key stakeholders. They are carefully formulated to leverage the benefits of PIDs in enhancing the visibility, traceability, and interoperability of research outputs, thereby facilitating a more interconnected and transparent research ecosystem. These recommendations create a pathway not only towards strengthening the research infrastructure in the United States but also towards contributing valuable insights to the international community, aligning with global efforts to harmonize research output management.

3.1 Benefits of Embracing PIDs

Ever greater, more consistent adoption of PIDs is important for a variety of entities, including funding organizations, individuals, and organizations, who manage grants, journal articles, datasets, software, and other research outputs. This growing significance emphasizes their usefulness throughout the entire research lifecycle, from securing initial funding to disseminating findings, and their broader implications for the global research landscape. At the heart of this movement is the recognition that research is a dynamic and interconnected process, involving various stakeholders and stages.[24] While many in these stakeholder groups have already adopted PIDs, broader uptake could realize even more benefits. PIDs provide a critical framework that bridges these stages seamlessly, offering a way to track, link, and manage research outputs

---

[19] https://ardc.edu.au/project/australian-national-persistent-identifier-pid-strategy-and-roadmap/

[20] Brown, J., Jones, P., Meadows, A., & Murphy, F. (2022). *Towards a national PID strategy for Canada—Vers une stratégie nationale sur les PID pour le Canada.* https://doi.org/10.5281/zenodo.7217469

[21] Brown, J. (2020). *Developing a persistent identifier roadmap for open access to UK research.* https://repository.jisc.ac.uk/7840/2/PID_roadmap_for_open_access_to_UK_research.pdf

[22] Brown, J., Jones, P., Meadows, A., & Murphy, F. (2022). *Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems.* The MoreBrains Cooperative. https://doi.org/10.5281/zenodo.7100577

[23] Meadows, A. (2023, January 25). Why PID strategies are having a moment — and why you should care. *The Scholarly Kitchen.* https://scholarlykitchen.sspnet.org/2023/01/25/why-pid-strategies-are-having-a-moment-and-why-you-should-care/

[24] de Castro, P., Herb, U., Rothfritz, L., & Schöpfel, J. (2023). *Building the plane as we fly it: The promise of persistent identifiers.* Scidecode science consulting publications & data. https://doi.org/10.5281/zenodo.7258286

and entities as well as their metadata, the key information about each output, at every step and over time.[25]

3.1.1 PID Stakeholders

Such a PID framework serves the needs of several main stakeholder groups[26]:

- Funders: Funding agencies and organizations allocate substantial resources to research projects around the globe, and they need effective means to monitor progress and outcomes. PIDs allow funders to precisely track the impact of their investments, facilitating accountability and informed decision-making. Researchers can use PIDs to link their grant proposals, progress reports, publications, software, and datasets, creating a transparent record of their work. This, in turn, enhances trust between funders and researchers and fosters a culture of responsible research conduct.
- Researchers: PIDs are indispensable tools for establishing and maintaining researchers' scholarly identities. PIDs for researchers, like ORCIDs (Open Researcher and Contributor IDs),[27] have gained widespread acceptance, enabling them to create unique identifiers that connect them to their body of work. These PIDs associate scholars with their publications, software, datasets, and other research outputs, ensuring that their contributions are consistently attributed to them. This not only aids in securing academic positions and promotions, grants, and collaborations, but also helps them build a comprehensive and easily updatable view of their impact and engagement in their respective fields.
- Publishers: In the world of academia, PIDs have revolutionized the way research articles are disseminated and tracked. Publishers have almost universally adopted Digital Object Identifier (DOI) systems to assign PIDs to individual works, ensuring that they remain discoverable and accessible long after publication.[28] This permanence greatly benefits both authors and readers, as it helps to ensure that scholarly outputs can be reliably cited and referenced in perpetuity. Furthermore, PIDs facilitate the tracking of research impact such as article-level metrics, providing valuable insights into the impact and reach of research publications.
- Research institutions: PIDs extend their reach beyond individual researchers and are also applicable and beneficial to the work of research institutions. They are instrumental in managing and preserving research data, software, projects, and other research outputs. These are all important aspects of the work conducted by research institutions. Moreover, PIDs better enable these

[25] Klein, M., & Balakireva, L. (2020). *On the persistence of persistent identifiers of the scholarly web (Version 1)*. https://doi.org/10.48550/ARXIV.2004.03011
[26] Praetzellis, M., & Gould, M. (2021). Open persistent identifiers: The building blocks of sustainable scholarly infrastructure. *Research Library Issues: Sustaining Open Content and Infrastructure*, 302, 8. https://doi.org/10.29242/rli.302.2
[27] https://orcid.org/
[28] DOIs are administered by the DOI Foundation (http://doi.org), which handles the overarching governance and resolution of DOIs across multiple registration agencies.

institutions to manage their research administration and track research outputs, for example, through research information systems.[29]

- Research support organizations: Aggregators, distributors, or service providers who support the enterprise can improve service to their customers, be they researchers themselves, administrators, or funding or governmental agencies. Many of these organizations already support some elements of this infrastructure, but they would benefit from greater consistency and compliance with existing practice.
- PID infrastructure providers: It is also important to consider the providers of PID infrastructure—encompassing platforms, service providers, and knowledge bases—as a key stakeholder group that occupies a significant role. To be most effective, infrastructure requires investment, consistency in adoption, and ongoing community consensus as it develops and evolves. The providers benefit by efficiencies of scale and interoperability across the ecosystem, just as the community benefits from their stability and sustainability. By making the PIDs and the associated metadata they provide open, they streamline the accessibility and usability of their services.[30] This not only enhances PID infrastructure providers' offerings but also contributes to the overall efficiency and effectiveness of the research ecosystem.

3.1.2 Key Benefits of PID Infrastructure

PIDs are a transformative force in scholarship, offering a myriad of benefits that extend across the entire research lifecycle. Embracing PIDs can offer significant cost savings, streamlined metadata management, automated processes, and enhanced data analysis capabilities.[31] By providing a standardized and interoperable framework, PIDs also open avenues for scalability and improved research assessment while reducing administrative costs, reducing error, and improving data quality, and offering scalable approaches to policy compliance. Below are examples of real-world benefits of adopting PIDs.

3.1.2.1 Reduction of Administrative Burden

The process of research administration can be extremely time consuming. Persistent identification of people, objects, institutions, and projects can significantly reduce the burden on researchers and administrative staff alike, at every stage of the research process—grant application, preregistration, repository deposition of preprints or datasets, manuscript submission, grant reporting, promotion and tenure dossier preparation, and administrative review. Researchers' time is best utilized conducting research and pursuing novel discoveries, not completing forms and submitting

[29] Bryant, R., Fransen, J., de Castro, P., Helmstutler, B., & Scherer, D. (2021, November). Research information management in the United States. *OCLC Research Reports*. https://doi.org/10.25333/8hgy-s428

[30] Open PIDs are defined as standards that are "community governed" and "allow use and reuse of the metadata."

[31] Wittenburg, P. (2019, March 1). From persistent identifiers to digital objects to make data science more efficient. *Data Intelligence*, 1, 6–21. https://doi.org/10.1162/dint_a_00004

reports. Interoperable PIDs provide an opportunity to reduce this burden by allowing for reuse of existing data and pre-population of information associated with the PID (i.e., metadata) in submission forms, grant applications, manuscript submission systems, and final output reporting.[32]

3.1.2.2 Cost Savings and Scalability

Investing in scholarly infrastructure and PIDs yields significant time and financial advantages.[33] Efficiency gains resulting from PIDs become evident and lead to tangible cost savings for organizations when PIDs are open and interoperable across systems. This is achieved by streamlining manual data management efforts, minimizing the risk of data errors, and enhancing data analysis capabilities. As a result, businesses and research institutions can optimize their operations and allocate resources more effectively. Importantly, the scalability of PIDs ensures that, as data volumes increase, the benefits persist without a corresponding rise in overhead costs.

3.1.2.3 Improved Research Assessment

With research investments in the US counted in the billions of dollars[34], the necessity of tracking the productivity and impact of these investments is critical. The leadership of the research enterprise has significant interest in regular and efficient tracking of outputs and outcomes. Research funders—federal, state, and philanthropic, as well as institutional—collect a variety of information from various sources. Connecting outputs with investments is never an easy process, but it could be done more efficiently through the application of PIDs to grants, institutions, people, infrastructure, projects, and outputs. This would make the aggregation of this data and its use feasible in a way that is presently time consuming and challenging, if it is possible at all.[35]

3.1.2.4 Research Integrity

Trust and credibility are paramount in academia, and PIDs facilitate the traceability of research sources.[36] They provide access to a broader context for a research project,

---

[32] Goddard, L. (2021). Persistent identifiers as open research infrastructure to reduce administrative burden. *Pop! Public. Open. Participatory*, 3. https://doi.org/10.54590/pop.2021.006

[33] Brown, J., Jones, P., Meadows, A., & Murphy, F. (2022). *Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems.* The MoreBrains Cooperative. https://doi.org/10.5281/zenodo.7100578

[34] West, Darrell M. R&D for the public good: Ways to strengthen societal innovation in the United States (2022). Brookings Institute https://www.brookings.edu/articles/rd-for-the-public-good-ways-to-strengthen-societal-innovation-in-the-united-states/

[35] Brown, J., Demeranville, T., & Meadows, A. (2016). Open access in context: Connecting authors, publications and workflows using ORCID identifiers. *Publications* 4, 30. https://doi.org/10.3390/publications4040030

[36] Belsø, R., Matthiesen, M., Parland-von Essen, J., Béquet, G., & 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

such as links to preregistration of protocols, related datasets, software code, and publications, and to the originating award that supported the project. The visibility of these corroborating objects provides a more complete record of a research object's provenance or supporting evidence for conclusions. They also can assist in improving research integrity by mitigating the risks associated with paper mills and conflicts of interest during peer review, through connecting research outputs to source materials, pre-registration of research, data management plans, or research funding. PID infrastructure allows for third-party analysis and monitoring of potentially questionable practices, at scale.[37] While falsifying a single research output is comparatively easy, it is more difficult to falsify an entire interlinked chain of related research outputs and documentation. PIDs also provide additional context to the research process that could be an indicator of malign behavior based on connecting related PID activity data, i.e., when records were created/updated and by whom. This transparency not only strengthens the integrity of scholarly work but also encourages a culture of responsible research practices.[38]

3.1.2.5 Simplifying and Automating Processes

Efficient metadata management is a cornerstone of effective data organization. Without high-quality and well-organized metadata, data can become a chaotic jumble of information, making it challenging to locate, understand, and utilize effectively. PIDs provide a structured approach to metadata management by attaching standardized information to each digital object or entity that is retrievable as a resolvable link. This metadata includes crucial details like authorship, affiliation, licensing information, creation date, data source, and relevant contextual information. With PIDs in place, organizations can maintain consistent and up-to-date metadata, simplifying data categorization and retrieval.[39] This ensures that valuable information remains accessible and comprehensible, mitigating the risk that data about research activities becomes an unwieldy resource that hinders rather than facilitates decision-making.

3.1.2.6 Enabling FAIR Data Exchange

PIDs also offer a foundation for FAIR (Findable, Accessible, Interoperable, and Reusable) data.[40] They ensure that data can be uniquely and permanently linked to,

---

[37] French, A. (2023, November 27). *Case study: Clear skies, research integrity, data science, and ROR.* ROR. https://ror.org/blog/2023-11-27-clear-skies-case-study/

[38] PIDs facilitate the linkage of research and its outputs to ethical standards and necessary approvals, such as those from Institutional Review Boards, enhancing the scientific process's integrity. They also enable the association of research with ethical practices, like the CARE Principles for Indigenous Cultural Governance (https://www.gida-global.org/care) supporting both automated compliance checks and studies on ethical practice adoption.

[39] Klein, M, Van de Sompel, H, Sanderson, R, Shankar, H, Balakireva, L, Zhou K, & Tobin, R. (2014, December 26). Scholarly context not found: One in five articles suffers from reference rot. *PLoS ONE*, 9, e115253. https://doi.org/10.1371/journal.pone.0115253

[40] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L. Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., …

enhancing its discoverability (Findable) and ensuring stable, long-term accessibility (Accessible). Furthermore, by adhering to standardized formats and metadata schemas, PIDs support the integration and compatibility of data from diverse sources, making it interoperable. Lastly, PIDs, by providing clear, persistent links to datasets and their associated metadata, facilitate the understanding and reuse of data (Reusable), thereby fostering collaboration and innovation in the research community. Through these mechanisms, persistent identifiers are foundational to realizing the FAIR data principles, propelling the efficiency and effectiveness of data exchange and knowledge creation.

3.1.2.7 Resilience and Discoverability

The nature of research demands efficient ways to manage and preserve scholarly knowledge over time. This underscores the importance of PIDs, which provide a consistent and unchanging means of identifying and discovering research objects. PID systems maintain their constancy as a source of reference, flexibly incorporating new metadata and information about the referenced entity, without succumbing to the chaotic and inconsistent nature of traditional citation practices. This stability allows researchers, institutions, and libraries to organize, access, and share information with confidence. The consolidation of data through PIDs promotes the efficient dissemination of knowledge, reducing redundancy and enhancing collaboration across academic disciplines.

3.2 Desirable Characteristics of PIDs

Several organizations have described ideal principles for PID infrastructure, such as the European Open Science Cloud (EOSC)[41] and its PID Policy[42] and the ISO Subcommittee on Identification and Description (ISO TC 46/SC 9)[43] and its Principles of Identification.[44]

Across these many efforts, PIDs are similarly defined as stable references to a digital object designed to address the two core functions of *identification* (by assigning a unique and unambiguous designation to a digital object or entity) and *persistence* (providing long-term and stable access to that thing regardless of where it might be located or how it might change over time).

Furthermore, as is defined on the Department of Energy (DOE)'s Office of Scientific and Technical Information (OSTI)'s website, "A persistent identifier (PID) is a digital identifier that is globally unique, persistent, machine resolvable, has an associated metadata schema, identifies an entity (e.g. person, researcher, publication, award,

---

Mons, B. (2016, March 15). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18

[41] https://eosc.eu

[42] See https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1

[43] See https://www.iso.org/committee/48836.html

[44] See https://www.iso.org/standard/83121.html

organization, or research output), and is frequently used to disambiguate between entities. PIDs are long-lasting, managed, and registered unique digital references (often in the form of a URL) to an object that can be represented or described online."[45]

This means that to be considered a "persistent identifier" or a "PID", an identifier and its underlying infrastructure must capture robust metadata and make that metadata available in consistent and reliable ways. In addition, this means that to be considered a PID, it also must exhibit persistence over time, which can be facilitated and perhaps best achieved when there is community ownership and public governance to ensure sustainable infrastructure support and widespread adoption. When working correctly, PIDs interoperate and reference each other within research discovery and management systems, linking descriptive information to other objects, without restrictions.[46]

Within the umbrella category of PIDs, there is a diverse array of identifiers, which differ in terms of what they identify, why they are used, how they are implemented and managed, and other characteristics. Each PID system has a defined metadata standard describing the objects they reference, and that information can then be read and used by machines and people.

While these distinctions are important to understand, it is also helpful to take a broader view and outline shared characteristics and principles to guide strategies. To help with this, the foundational article "Identifiers for the 21st Century"[47] emphasized several key criteria for PID infrastructure:

- Built for connection and expansion
- Promoting interoperability
- Filling gaps specific to research communication
- Ensuring open availability of metadata
- Establishing persistent identifier trustworthiness
- Emphasizing community ownership and governance
- Fostering organizational sustainability

Using these definitions and criteria as background, this document recommends the following desirable characteristics as part of shared implementations of PIDs in the US:

---

[45] https://www.osti.gov/pids/using-pids

[46] Van de Sompel, H., Sanderson, R., Shankhar, H., & Klein, M. (2014). Persistent identifiers for scholarly assets and the web: The need for an unambiguous mapping. *International Journal of Digital Curation*, 9. https://doi.org/10.2218/ijdc.v9i1.320

[47] McMurry, J.A., Juty, N., Blomberg, N., Burdett, T., Conlin, T., Conte, N., Courtot, M., Deck, J., Dumontier, M., Fellows, D.K., Gonzalez-Beltran, A., Gormanns, P., Grethe, J., Hastings, J., Hériché, J.-K., Hermjakob, H., Ison, J.C., Jimenez, R.C., Jupp, S., … Parkinson, H. (2017, June 29). Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biology*, 15, e2001414. https://doi.org/10.1371/journal.pbio.2001414

- Open availability of core metadata: All PIDs should include freely available and open exchange of basic data (through services such as data dumps, feeds, APIs, or other forms of machine access). More robust access, higher-throughput services, and real-time data access may be provided for a fee to support the infrastructure system, but basic services and openly licensed kernel metadata should be available for integration and reuse in other systems.
- Use of well-established resolver services: A core value derived from using PIDs is the ability to use the identifier to link various elements of the ecosystem together, by resolving the PID to more information about the object to which it refers. Connecting the ecosystem in this way reduces duplication of data entry that could lead to errors and allows it to be easily maintained across disconnected systems. Examples of such services include doi.org, hdl.net, identifiers.org, and n2t.net. These resolver services help ensure that URIs remain functional even if underlying resources change.
- Documentation of identifier policies: PID providers should document and publish their identifier policies alongside schema descriptions, so that users and other actors can access and understand how identifiers are handled within their systems. This helps provide clarity and transparency about how identifiers are assigned, managed, and referenced. Policies should include information related to identifier management, versioning, handling changes, avoiding reassignment, and ensuring persistence. These policies are essential for ensuring consistency, transparency, and reliability in the use of PIDs in the scholarly ecosystem.
- Monitoring and reporting services: PID providers should actively monitor assigned PIDs to ensure that they remain functional. If any of the referenced URIs become “dead” or nonfunctional, PID providers should have ways to report the issue to the original data provider and/or the community.
- Ease of assignment/metadata creation and curation: The assignment of PIDs is a critical stage in the deployment of identification systems. While each PID system handles assignment and the record metadata creation process differently, this process should be as simple and as user friendly as possible to facilitate use of the system. PID assignment should be as closely associated with the creator and the creation event as practical. Responsibility for creating PIDs and maintaining metadata records after the PID is assigned should be managed using best practice for user-centered design. Users of PID systems should be engaged in the management process.
- Standardized structures, metadata, and services that allow for community input: Consistency in how the community accesses the data, how data is structured, and what services are made available is driven by standardization of the PID system. These structures need to be driven, in part, by community consensus processes to ensure the robustness of the service is suitable for a diverse user base.
- Extensibility: No system can be developed to serve every use case, nor can every implementation be projected. Therefore, PID systems should allow for

extensibility and have a process in place to extend the system to adapt to new use cases and demands on the system.
- Community governance: PID infrastructure systems should be accountable to the user community that adopts them. As such, a wider community should be involved in governance structures that manage the PID systems.

Looking beyond the identifiers themselves, and in agreement with the desirable characteristics outlined above, PID systems should also draw fundamentally from the Principles of Open Scholarly Infrastructure (POSI),[48] UNESCO, and other PID guidance in how they are developed and managed. It is important that the organizations and services responsible for providing PIDs follow a similar set of best practices when it comes to their operations and governance models.

3.3 Recommendations for Evaluating and Adopting PID Infrastructure

It is imperative to take a coherent and consistent approach to PIDs across the diverse spectrum of US stakeholders.[49] A unified strategy ensures that the benefits of PIDs are harnessed uniformly, fostering a seamlessly interconnected scholarly ecosystem. This consistency is particularly crucial when considering compliance with public access policies at both the US federal level and within institutions and research groups nationwide.

At the national level, maintaining a standardized approach to PIDs is instrumental in tracking and ensuring compliance with public access policies, and in research security processes. PIDs serve as unique markers with consistent metadata standards, enabling the systematic tracking and reporting of research outputs, publications, data, and software associated with federally funded projects. This approach facilitates efficient monitoring and reporting, allowing agencies to assess the impact of research and ensure adherence to public access mandates. PID frameworks implemented across federal agencies would not only reduce costs and enhance accountability, but also promote transparency in the dissemination of publicly funded research.

The same is true within institutions and research groups across the US. A consistent PID approach is equally vital for adhering to public access policies and fostering collaborative research environments. PIDs provide a standardized means of identifying and linking various research outputs, such as publications, datasets, software, and research awards. This consistency aids in creating a transparent and traceable research landscape, enabling institutions to easily demonstrate compliance with federal mandates. It also facilitates efficient communication and collaboration among

[48] https://openscholarlyinfrastructure.org/
[49] de Castro, P., Herb, U., Rothfritz, L., & Schöpfel, J. (2023). *Building the plane as we fly it: The promise of persistent identifiers.* Scidecode science consulting publications & data. https://doi.org/10.5281/zenodo.7258286

researchers and institutions, ensuring that the broader research community benefits from the outcomes of federally funded projects.

In summary, the US should embrace a uniform approach to PID adoption and implementation across our diverse landscape of stakeholders. This will not only strengthen our work within open scholarship, but also play a pivotal role in tracking and ensuring compliance with public access policies, which is important both at the US federal level and within institutions and research organizations throughout the country. This consistency will enhance accountability, transparency, and the overall impact of publicly funded research.[50]

3.3.1 Recommended Strategy for Evaluating PID Infrastructures

Building on the above characteristics, an important component moving forward should be a consistent method for evaluating PID infrastructure. The following characteristics of this evaluation process are informed by the definitions, criteria, and characteristics provided above as well as by POSI:

- Promoting collaboration and inclusivity: Involving a diverse set of stakeholders is essential for ensuring that the resulting infrastructure serves the needs of all sectors of the scholarly community.
- Transparency and trust: Transparent governance practices will foster trust among users and stakeholders, ensuring that the infrastructure's decision-making processes are open and accountable.
- Advocacy for policy changes: Consistent and visible support for policies that support open scholarly practices, data sharing, and the free flow of research information.
- Sustainable funding and adaptation: Prioritizing sustainability ensures that the infrastructure can endure and adapt over time.
- Openness and accessibility: Promoting the values of open science can help ensure open and accessible scholarly infrastructure.
- Commitment to community interests: Prioritizing community needs and interests will ensure that the infrastructure remains supportive and not inhibitory.

3.3.1.1 Proposed Strategies to Consistently Evaluate PID Infrastructures in the US

Follow clear policies and guidelines: Utilize guidelines for implementing a national PID strategy, incorporating principles from this document as well from POSI.

[50] Many of the PIDs and activities recommended are generalized and can be used in conjunction with other PIDs and activities in specific contexts. The focus of this document is to outline a framework that can both cut across localized approaches and be in harmony with the international research ecosystem.

Collaborate with standardization initiatives: Engage with standardization initiatives that serve as the foundation for most PID providers, such as Crossref,[51] DataCite,[52] and ORCID to ensure alignment with best practices, foster interoperability, and contribute to the development of a standardized framework for PIDs.

3.3.2 Recommended PIDs for Common Use Cases

The research system, as noted above, is complex and heterogeneous. Each research domain may have its own specialty objects, identifiers, and use cases. Despite this, some PIDs and their associated metadata schemas are commonly used across fields, as outlined in Table 1. Rather than setting forth an exhaustive catalog of all identifiers in all domains and recognizing that any such list is bound to need constant updating, the following should be broadly adopted for each category.[53]

Table 1. Standardized PIDs for Common Research Entities

Entity	**PID(s)**	**Provider(s)**	**Notes**
Textual Outputs	DOI	Crossref DataCite	Crossref DOIs are widely used by the journal publishing community for articles and for an increasing number of books. However, both Crossref and DataCite support metadata schemas for a variety of written publications. Also, note that ISBNs and ISSNs are assigned for various types of publications in the supply chain.
Research data	DOI	DataCite	DataCite DOIs are widely used PIDs that support common metadata for a variety of data types. DataCite also supports metadata extensions for more robust metadata f specific use cases and/or data types.
Research software	DOI	DataCite	DataCite DOIs are available for use for software and code. This is primarily done when/if code is deposited into a digital repository.[54]
Organizations	ROR	ROR	ROR[55] is an institutional identifier for a researcher's organization (affiliation).
Funders	ROR	ROR	Institutional identifier for the funding body, distinct from the award ID. Crossref's Open

[51] https://www.crossref.org/
[52] https://datacite.org/
[53] Klump, J., & Huber, R. (2017). 20 years of persistent identifiers—which systems are here to stay? *Data Science Journal*, 16, 9. https://doi.org/10.5334/dsj-2017-009
[54] https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content
[55] ROR, or Research Organization Registry, is a global, community-led registry of open persistent identifiers for research organizations, a collaborative initiative from California Digital Library, Crossref, and DataCite. See https://ror.org/.

			Funder Registry IDs were previously promoted as the primary PID for this category but were recently deprecated in favor of ROR.[56]
Researcher	ORCID	ORCID	ORCID is in wide use and should be used as a primary identifier for researchers. In some specific use cases, ISNI[57] may be more appropriate (e.g., deceased researchers or non-academic researchers). Note that ISNIs are assigned after, not prior to, publication.
Data management and sharing plan	DMP-ID	DataCite	Identifier for a researcher's data management plan (DMP) that includes custom metadata information about the associated plan.
Archival records	ARK	ARK Alliance	Decentralized identifier used by libraries, archives, and museums for scholarly and cultural objects.
Research award/grant	Grant ID	Crossref DataCite	Identifier for a specific grant or research award that includes custom metadata information about the grant/award.
Research project	Project ID RAiD	DataCite	DataCite currently registers DOIs for projects. ARDC (Australian Research Data Commons) plans to register DataCite DOIs as part of its RAiD project (currently in development).[58]

3.3.2.1 Proposed Strategy for Consistently Utilizing PID Infrastructures for Common Use Cases in the US

Adopt standardized PIDs for common entities: Implement standardized PIDs, as listed in Table 1, for common research entities such as written publications, research data, software, organizations, funders, researchers, data management plans, research awards/grants, and research projects.

3.4 Recommended Strategies for Supporting Core PID Infrastructure

Navigating the ever-evolving landscape of open scholarship demands a concerted effort toward the effective implementation of PIDs, as pivotal tools for fostering collaboration and transparency in scholarly communications. Therefore, it is important that

[56] https://www.crossref.org/blog/open-funder-registry-to-transition-into-research-organization-registry-ror/
[57] ISNI, the International Standard Name Identifier (ISO 27729), is managed by the ISNI International registration authority for ISO. See https://isni.org.
[58] Research Activity Identifier (RAiD) is the persistent identifier dedicated to research projects and activities. See https://raid.org/.

individuals and organizations actively contribute to and support the core PID infrastructures that underpin this ecosystem.[59] By understanding and assuming key roles, stakeholders can play valuable roles in advancing open scholarship through the seamless integration of PIDs,[60] including through financial contributions, policy advocacy, in-house integration, and widespread adoption. These different strategies underscore the indispensable role of PIDs in cultivating a resilient and interconnected scholarly environment.

Table 2, below, outlines specific actions and responsibilities associated with various roles (users, champions, supporters, adopters), offering stakeholders a comprehensive guide to actively participating in shaping the future of open scholarship practices. People and organizations may play multiple roles in this ecosystem in different PID systems or different contexts.

Table 2. Stakeholder Roles and Related Actions

Role	**Stakeholder Definition**	**Proposed Strategies for Each Role**
Users	Users are members of the community who directly leverage the benefits of PID infrastructure. In some manner, almost every researcher, information provider, library staff member, author, student, or reader is a user.	Users of PID infrastructure are encouraged to share their experience, communicate their needs, and critique the services that are provided by PID infrastructure providers. Whenever possible, users should be engaged in needs assessment, systems development, and user testing and be actively sought out for general feedback on changes to the infrastructure to ensure fitness for purpose.
Champions	Champions are vocal advocates within the PID community who engage directly with developing and/or promoting the use of PIDs.	Champions can play a pivotal role in advocating for and extending the adoption of PID infrastructure, by referencing PIDs and PID-based practices in research policies and procedures and emphasizing their importance. Champions also advocate for PID best practices, offering guidance to colleagues and staff engaged in PID-related activities. Their support extends to the development and maintenance of PIDS\s systems, as well as facilitating the seamless integration of PIDs into research workflows within their work context.

[59] Bilder, G, Lin, J, & Neylon, C. (2016, January 28). Where are the pipes? Building foundational infrastructures for future services. *Science in the Open.* http://cameronneylon.net/blog/where-are-the-pipes-building-foundational-infrastructures-for-future-services/

[60] Chodacki, J., Cruse, P., Lin, J., Neylon, C., Pattinson, D., & Strasser, C. (2018, April 5). *Supporting research communications: A guide.* https://doi.org/10.5281/zenodo.3524663

Supporters	Supporters are involved in strategies to contribute to the robust implementation of PID infrastructure.	Supporters may join existing infrastructure providers as members and actively participate in the PID community, or provide financial support to PID infrastructure or initiatives, ensuring their continued development. Additionally, supporters are encouraged to consider PID options and practices when making integration and/or purchasing decisions, with the goal of including PIDs seamlessly into their operational framework.
Adopters	Adopters actively engage with the implementation of PIDs by employing a multifaceted approach to integrating PID infrastructure into their operations.	Adopters include, develop, or support in-house integration of PIDs, in recognition of the value of these identifiers in enhancing scholarly communications. Adopters may also adopt PID services to streamline their processes. Assigning PIDs to various research outputs, including publications, data, and software, is a key aspect of their commitment. Additionally, adopters can extend the use of PIDs to research awards, incorporating them into metadata associated with both outputs and awards. They can also require or encourage researchers to obtain personal identifiers like ORCID, fostering a culture of unique identification. Integration of PIDs in various research systems, including publishing, grants, and DMPs, is a comprehensive strategy adopted by these organizations to enhance the overall scholarly ecosystem. In addition, citing PIDs in publications serves as a foundational measure to acknowledge their importance and promote their widespread adoption.

3.5 Recommended Strategies for Moving Beyond Legacy Systems

As described in the sections above, PIDs play a crucial role in ensuring the long-term accessibility, discoverability, and usability of research outputs. These benefits are realized through leveraging persistent identifiers that meet the desirable characteristics outlined above. Unfortunately, many legacy identifier systems, classified here as systems that are siloed and/or do not meet the desirable

characteristics outlined above, often fall short.[61] When our communities continue to rely on these systems, the entire research ecosystem is restricted by their limitations:

- Limited granularity: Assigned at a single, coarse level and lacking the granularity required for precision in data retrieval. PIDs, on the other hand, offer a more detailed and flexible structure, allowing for fine-grained identification of various entities, such as datasets, individual files, or even specific versions of data.
- Lack of interoperability: Not designed with interoperability in mind, hindering the seamless integration of data across different platforms and repositories. PIDs, built on standardized frameworks, facilitate interoperability, enabling researchers to link and share data more effectively.
- Vulnerability to changes: Susceptible to changes, especially in scenarios where items are updated or modified. This can lead to confusion and data integrity issues. PIDs are designed to remain stable over time, ensuring consistency and reliability in referencing research resources.
- Inadequate metadata support: Lacking the capacity to store comprehensive metadata, impeding efforts to provide context and information about the associated data. PIDs come equipped to integrate with metadata standards from multiple institutions, allowing for the inclusion of crucial details that enhance the understanding and utility of the research data.
- Barriers to access: Lacking resolver services (i.e., a persistent URL), making it more difficult for the end user to access the item. This deficiency not only hampers the user experience but also creates obstacles in cross-platform navigation, hindering the smooth retrieval of research resources. In contrast, PIDs provide a robust infrastructure, including resolvers, that ensures a seamless and user-friendly access experience, promoting efficient engagement with research outputs across diverse platforms and repositories.

The limitations of legacy identifier systems in meeting these characteristics compared with PIDs underscore the critical need for increased investment in PIDs within the US research community. As the volume of research outputs continues to grow, there is a pressing need for a more robust and standardized system of persistent identification. By embracing and implementing robust PID systems, researchers can enhance data management practices, promote collaboration, and ensure the long-term impact and accessibility of their contributions to scientific knowledge.[62]

---

[61] One example of legacy identifier systems is accession numbers. "An accession number is a sequential number assigned to each record or item as it is added to a…collection or database and which indicates the chronological order of its acquisition." University of Liverpool Library. (2022). What is an accession number? Do I need it for referencing my work? *Library Help.* https://libanswers.liverpool.ac.uk/faq/181287

[62] Agosti, D., Benichou, L., Addink, W., Arvanitidis, C., Catapano, T., Cochrane, G., Dillen, M., Döring, M., Georgiev, T., Gérard, I., Groom, Q., Kishor, P., Kroh, A., Kvaček, J., Mergen, P., Mietchen, D., Pauperio, J., Sautter, G., & Penev, L. (2022). Recommendations for use of annotations and persistent identifiers in taxonomy and biodiversity publishing. *Research Ideas and Outcomes*, 8, e97374. https://doi.org/10.3897/rio.8.e97374

3.5.1 Proposed Strategies for Moving Beyond Legacy Identifier Systems in the US

Embrace robust PID systems:

- Increase investment in PID adoption within the US research community.
- Enhance data management practices, collaboration, and long-term accessibility by implementing robust PID systems.

Move away from legacy systems:

- Ensure our infrastructure meets the requirements of modern technical and information exchange standards.
- Invest in transitioning PIDs that do not meet defined principles to future-proofed, community-adopted PID standards.

3.6 Recommended Strategy for Supporting Centralized PID Infrastructure

Centralized approaches to PIDs, in which central registries and management oversight can take place, are particularly valuable in the United States. This approach to PID infrastructure connects research outputs also offering a clear path to address the needs of the diverse ecosystem, including federal, state, local, institutional, public, and private entities. Below are several US-specific drivers for why it is important to adopt and support centralized PIDs:

- Heterogeneity in the US: The US has a highly diverse and decentralized marketplace, with a wide range of stakeholders, including federal agencies, state and local governments, academic institutions, public libraries, private enterprises, and nonprofit organizations. Centralized PIDs provide a standardized and unifying mechanism for these diverse entities to manage, reference, and share research outputs. This consistency is essential for bridging gaps and fostering collaboration across various sectors.
- Cross-agency collaboration: Multiple US federal agencies fund and conduct research, each with its own information systems and repositories. A centralized PID system helps these agencies work together more effectively by providing a common identifier for research outputs. This is particularly important for interdisciplinary research that often involves multiple agencies.
- Public-private partnerships: The US has a robust environment of public–private partnerships in research and innovation. A centralized PID system ensures that both public and private organizations can use a common identifier to reference researchers and research outputs. This promotes transparency, trust, and cooperation between government, industry, and academic sectors.
- Interoperability in the research ecosystem: The US research ecosystem encompasses a vast number of academic institutions, libraries, publishers, and research centers, each with its own digital repositories and systems. Centralized PIDs like DOIs enable interoperability among these disparate systems,

facilitating the discovery, sharing, and tracking of research outputs across the ecosystem.[63]

- National and international collaboration: Centralized PIDs are essential for promoting collaboration at the national and international levels. US researchers often collaborate with colleagues from around the world. A consistent approach to identifying and referencing research outputs through centralized PIDs simplifies the integration of US research into the global scholarly network.
- Regulatory and compliance requirements: Federal and state agencies often have specific regulatory and compliance requirements related to research reporting and data sharing. Centralized PIDs help streamline compliance efforts by providing a standard means of tracking and referencing research outputs, making it easier for organizations to meet these requirements.
- Preservation of research integrity: The US places a high value on research integrity and the responsible conduct of research. Centralized PIDs enhance the transparency and trustworthiness of research outputs, supporting efforts to maintain the highest ethical standards in research.
- Openness to innovation: PIDs position indexed resources for remixing in new contexts and unexpected novel use cases. Early adopters are likely to be in business analytics, where, for example, many non-research stakeholders have interests in following research trends. Longer-term use cases include research in economics, political science, sociology, and other social sciences, where PIDs enable inexpensive storytelling at scale about research by region, time, and field.

3.6.1 Proposed Strategy for Promoting Centralized Approaches to PIDs in the US

Embrace centralized PIDs:

- Recognize centralized PID systems, such as DOI, ORCID, and ROR, and their associated metadata as community standards that should be leveraged in both US and global contexts.
- Highlight the benefits of centralized PIDs for diverse stakeholders, including federal agencies, state and local governments, academic institutions, public libraries, private enterprises, and nonprofit organizations.

3.7 Recommended Areas for Additional Investment

In this interconnected era of research, where knowledge knows no boundaries, supporting initiatives like DataCite, Crossref, ROR, and ORCID becomes not only a strategic investment but a collaborative endeavor that benefits the United States and all global partners. Failure to do so poses a risk. Without a concerted effort to expand the adoption of PID infrastructure globally, there is a threat of fragmented and duplicative systems emerging in isolation. This fragmentation could lead to the development of siloed databases and hinder the seamless collaboration and sharing of

[63] Page, R. (2023). Ten years and a million links: Building a global taxonomic library connecting persistent identifiers for names, publications, and people. *Biodiversity Data Journal*, 11, e107914. https://doi.org/10.3897/BDJ.11.e107914

research outcomes. It is imperative to recognize that investing in these initiatives is not just a strategic choice for the US research community but also a vital collaborative endeavor. By doing so, our communities safeguard against the potential loss of a unified and interconnected scholarship, where the risk of duplication and isolation is mitigated, and the collective progress of the international scientific community is fostered.

3.7.1 Investment in Grant DOIs

In the landscape of research funding, the adoption of PIDs for grants has emerged as a pivotal step toward enhancing research management, reporting, and evaluation.[64] Most grant organizations do not use a consistent grant identification structure, nor is that information shared publicly for others to use. As underscored by the ORFG-organized Working Group on DOIs for Grants,[65] this shift not only brings advantages to funders but also aligns with broader initiatives, particularly within the US.

Grant DOIs offer a robust solution to the limitations posed by existing grant IDs. While traditional grant identifiers may lack standardization across funders and may be non-unique, grant DOIs provide a standardized, unique, and machine-actionable identifier. This ensures clarity and efficiency in tracking activities related to the grant, such as publications, software, and datasets.

Moreover, the rich metadata associated with grant DOIs enables funders to provide detailed information about their funded projects. This metadata encompasses award types, durations, currency and amounts, researcher affiliations, and project descriptions. The integration of this metadata with third-party systems, as seen in the PID Graph[66] services or services like the DataCite GraphQL API, contributes to streamlined compliance checking, improved discoverability, and reduced administrative burden.

The financial costs of adopting grant DOIs, as outlined by organizations such as Crossref and DataCite, are cost-effective relative to the accrued benefits. Annual membership fees and per-record fees are well-justified when considering the enhanced tracking capabilities, compliance facilitation, and potential for cost and time savings in the long run. Funders can leverage reports, like the one commissioned by the

---

[64] Kiley, R., Fentrop, N., & Hendricks, G. (2018, February 16). 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/

[65] Mader, C., Hendricks, G., Chandler, Z., Chen, X., Whylly, K. E., Jones, A., Chodacki, J., Cousijn, H., Tananbaum, G., McKiernan, E., Farley, I., & Carpenter, T. (2023). Concept note: Working group on DOIs for grants. *ORFG research output tracking community project.* OSF. https://osf.io/p87us

[66] https://www.project-freya.eu/en/pid-graph/the-pid-graph

Australian Research Data Commons,[67] to evaluate the incentives and cost-effectiveness of integrating grant DOIs into their overall PID strategy.

3.7.1.1 Proposed Strategy for Adopting Grant DOIs in the US

Adopt grant DOIs:

- Funders should evaluate and prioritize adoption of grant DOIs within their overall PID strategy.
- All stakeholders should follow key steps for leveraging grant DOIs in their systems.

3.7.2 Other Emergent Needs

In the ever-evolving landscape of scholarly research and academic collaboration, it is crucial to stay attuned to emerging needs within the community. One significant aspect of this is ensuring the effective leveraging of PIDs in a manner that aligns with community-focused goals. While PIDs have proven invaluable in enhancing the discoverability, accessibility, and overall integrity of scholarly outputs, there are identified gaps in the PID ecosystem that require attention and development.

- PIDs for projects: One of these gaps is in the tracking of research projects and their related research activities.[68] While many people and systems have leveraged DataCite DOIs or grant identifiers for some related use cases, existing communities of practice have fallen short in capturing the diverse and dynamic nature of scholarly projects. The Research Activity Identifier (RAiD) has been proposed as a potential new approach to this use case.[69] RAiD aims to provide a structured and standardized way of identifying and referencing research activities. The implementation of RAiD is still in its nascent stages, and the scholarly community is exploring ways to best incorporate this identifier into existing PID infrastructure effectively.
- PIDs for licenses: At present, there is a noticeable gap in establishing persistent identifiers for licenses assigned to scholarly works. While licensing practices have evolved, our communities lack a consistent PID convention that translates directly across all resources. Future development efforts should focus on devising a standardized approach for assigning PIDs to licenses. This will enhance transparency, traceability, and compliance, ultimately contributing to a more robust scholarly communications ecosystem.

Other areas where PIDS are missing in the research ecosystem:

---

[67] Brown, J., Jones, P., Meadows, A., & Murphy, F. (2022). *Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems.* The MoreBrains Cooperative. https://doi.org/10.5281/zenodo.7100578

[68] Robinson, E., & Habermann, T. (2023, October 17). Building a community of practice: Observations of the current use of DataCite DOIs as project IDs. *Upstream*. https://doi.org/10.54900/g4928-wva21

[69] https://raid.org/

- Facilities and infrastructure: Research Resource Identifiers (RRIDs)[70] are available for use for resources used during research; however, they lack broad adoption. The lack of a standardized approach results in challenges in accurately citing and acknowledging the use of specialized equipment and facilities in scientific publications. This gap undermines the traceability of research outputs to their source environments, complicating the replication of experiments and the validation of findings.
- Methods and protocols: Without PIDs for research protocols, there is a significant barrier to the seamless sharing and validation of experimental methods, leading to inefficiencies in research reproducibility and a hindrance to the cumulative advancement of scientific knowledge. This absence hampers the ability to link protocols directly to research outputs, diminishing the transparency and accountability in research practices.
- Software and code: DataCite DOIs are available for use for software and code; however, they lack broad adoption. The absence of universally adopted PIDs for software and code impedes the proper attribution and citation of computational tools and algorithms, critical components in modern research, leading to a lack of recognition for software developers and a gap in the research record. This gap also affects the traceability and reuse of software, essential for verifying research findings and fostering innovation.
- Images: In the research ecosystem, the lack of PIDs for images restricts the ability to accurately reference and retrieve scientific images, leading to potential misinterpretation of data and findings. This gap undermines the integrity of the scholarly record and complicates the process of building upon previous work, as images are pivotal in disciplines ranging from biomedical sciences to environmental science.

Tracking these and other emerging PID discussions, as well as addressing identified gaps in the PID landscape, requires collaborative efforts and additional financial investments from the scholarly community, technology developers, and infrastructure providers. The path forward is not necessarily for additional organizations to emerge or financial obligations to burden to our communities. Instead, we should leverage existing PID infrastructure to scale and adapt to these emerging needs. The PID landscape is fluid, and embracing new uses for the existing, collective infrastructure will strengthen the foundation of the scholarly community, foster collaboration, and ensure the continued evolution of a dynamic and responsive PID ecosystem.

3.8 Recommended Approach to Measuring Success

As noted above, the goal of this process is to set the stage for the further consultation with the community and formal adoption of this strategy as a US national standard via the consensus development process at NISO. This work must ensure inclusivity, adaptability, and widespread benefit across the spectrum of participating organizations, as well as harmony with the international research landscape:

[70] https://scicrunch.org/resources

- Inclusivity serves as a measurable cornerstone within this approach. This document actively solicited input from a diverse array of stakeholders, including federal agencies, research institutions, publishers, data repositories, PID providers, and others. The effectiveness of future efforts will be measured by the ability to incorporate the voices of community practitioners and experts from diverse communities and in addressing their varied needs and challenges.
- Adaptability stands out as another measurable facet inherent in this approach. In recognition of the ever-evolving nature of technology, organizational structures, and research practices, future efforts must also be formulated to be quantifiably adaptable over time. The effectiveness will be measured by the ability to remain relevant and useful amid the shifting landscape of PID implementation.

Moreover, the impact will be measured not only for those from various stakeholder groups that are deeply entrenched in PID implementation, but also for those at the nascent stages of PID adoption. The execution and effectiveness of this approach must be gauged in terms of its role as a quantifiable foundational resource that provides clear and measurable guidance for stakeholders who are just embarking on their PID journey. Simultaneously, for stakeholders already proficient in PID practices, the standard's impact must be measured by its ability to offer advanced insights and strategies that quantifiably contribute to the refinement and optimization of their work.

Measuring the impact of PID adoption strategies in the US requires a holistic approach that considers diverse metrics, case studies, feedback mechanisms, and longitudinal analysis. By aligning with global standards and leveraging the collaborative efforts of stakeholders, the US research community can continue to pave the way for a robust and effective PID infrastructure that advances open scholarship goals.

3.8.1 Measuring the Impact of Increased PID Adoption Over Time

Measuring the impact of PID strategies involves assessing key metrics such as:

- Adoption rate: Tracking research outputs and entities assigned PIDs that meet the desirable characteristics listed above.
- Membership growth: Tracking the number of US-based groups that are members of key PID infrastructure organizations over time.
- Cost savings: Quantifying the burden reduction in manual efforts and associated costs.
- Interoperability: Evaluating the seamless exchange of information between different PID systems.
- Research integrity: Assessing the impact on transparent communication and verification of research sources.

Several methods to gather information on these metrics include:

- Case studies: Conducting case studies across diverse research domains will provide insights into the effectiveness of PID adoption. These studies should cover academic institutions, government agencies, and private entities to capture a comprehensive view of the impact.
- Feedback mechanisms: Establishing feedback mechanisms, such as surveys and interviews, will enable stakeholders to provide input on the practicality and effectiveness of PID adoption strategies. Feedback will contribute to continuous improvement and refinement of the PID framework.
- Longitudinal analysis: A longitudinal analysis of PID adoption over time will offer a comprehensive understanding of the evolving impact. This analysis should consider changes in research workflows, collaboration patterns, and information accessibility.

These and other tactics can be used to monitor effectiveness over time.

4. Conclusion

This community-generated report, created in partnership with ORFG, HELIOS Open, and RDA, offers guidance on using PIDs across various aspects of the research environment. It underscores the wide-ranging uses of PIDs in research and the international benefits of a cohesive approach to these identifiers. The strategies recommended here aim to standardize approaches in the US and align them with global norms to encourage worldwide cooperation and knowledge exchange. The advantages highlighted, such as reduced workload, easier compliance monitoring, more efficient metadata handling, automation of processes, improved data analysis, resilience, cost effectiveness and compliance checking, underline the significant impact of integrating PIDs throughout the research process.

This report represents a vision from stakeholders committed to promoting open scholarship. It advocates for the adoption of PIDs and common principles to strengthen the research infrastructure, ensuring a smooth, interconnected system where scholarly work is accessible and preserved. As the US research community considers making this strategy a national standard, ongoing discussion, inclusion of various viewpoints, and further refinement of these PID recommendations are crucial. The initiative's future success will be gauged by how widely PIDs are adopted and incorporated, as well as through the observable enhancements in research management, academic communication efficiency, and the overall contribution to scientific progress.

Appendix A: Examples of PID Infrastructure

PIDs serve as a core element of many modern information systems, ensuring the longevity, accessibility, and traceability of digital resources.[71] In the realm of scholarly research and open data, three key examples have emerged that illustrate the significance of global, nonprofit, community-driven governance for PIDs: Crossref, ORCID, and DataCite.

While several other PID organizations and initiatives exist, these three play a pivotal role in strengthening the United States' commitment to fostering open and sustainable information ecosystems. They do so by providing core PID services addressing people, publications, data outputs, funding, and institutional identification in the research community.

- Crossref, established in 2000, primarily focuses on supporting identification of published research outputs. It registers DOIs for scholarly content, including articles and books. Crossref also manages the Research Organization Registry (ROR) as well as assignment for grants and funding organizations.
- DataCite, established in 2009, specializes in supporting PIDs for the vast use cases of research institutions and universities. It registers DOIs for datasets, software, instruments, projects, samples, etc., ensuring their long-term accessibility and proper attribution.
- ORCID, established in 2010, assigns unique identifiers to individual researchers, enabling them to distinguish themselves from others, and to connect their works and other professional information across various platforms, institutions, and publishers.

DataCite, Crossref, and ORCID are collaborative entities that serve as pillars in fortifying open and sustainable information systems. Each of these three entities are nonprofit organizations that embrace community-driven governance for persistent identifiers, which is paramount. These community-led organizations actively engage researchers, repositories, publishers, and institutions, empowering them to influence and shape best practices. This approach fosters transparency, encourages the reuse of information, and, most importantly, propels the frontier of scientific discovery.

By partnering with established organizations such as DataCite, Crossref, and ORCID, funders, institutions, government agencies, and others can ensure a unified and efficient approach to persistent identifiers, avoiding fragmentation and promoting a collaborative environment that benefits the entire research community.

[71] Buys, M., & Hendricks, G. (2023, October 11). Working for global equity through digital object identifiers. *Upstream.* https://doi.org/10.54900/6sz4q-47185

Appendix B: ORFG PID Strategy Working Group Members

Working group co-chairs:

Todd Carpenter, National Information Standards Organization (NISO)
John Chodacki, California Digital Library, University of California Office of the President, and RDA-US

Core members of the working group (listed alphabetically):

Yvonne Campfens, OA Switchboard
Caitlin Carter, Open Research Funders Group and HELIOS Open
Zach Chandler, Stanford University and RDA-US
Helena Cousijn, DataCite
Harry Dimitropoulos, OpenAIRE and Athena Research Center Team
Chris Erdmann, Independent
Maria Gould, California Digital Library
Ioanna Grypari, OpenAIRE and Athena Research Center Team
Gretchen Gueguen, Center for Open Science and Open Science Framework
Ted Habermann, Metadata Game Changers
Alex May, Oak Ridge National Laboratory
Erin McKiernan, Universidad Nacional Autónoma de México
Katherine McNeill, Independent
Alice Meadows, MoreBrains Cooperative
Christina Miskey, University of Nevada, Las Vegas
Lucy Ofiesh, Crossref
Elli Papadopoulou, OpenAIRE and Athena Research Center Team
Leonidas Pispiringas, OpenAIRE and Athena Research Center Team
Sheila Rabun, Lyrasis
Lane Rasberry, Wikidata
Kristen Ratan, Stratos and ICOR
Howard Ratner, CHORUS
Erin Robinson, Metadata Game Changers
Shawna Sadler, ORCID
Chris Shillum, ORCID
Carly Strasser, Chan Zuckerberg Initiative
Greg Tananbaum, Independent
Ana Van Gulick, Figshare