Recommendations for the Adoption of Persistent Identifiers in Higher Education and Research in France

Véronique Stoll, Frédéric De Lamotte

Ouvrir la science !

Recommendations for the Adoption of Persistent Identifiers in Higher Education and Research in France

French Committee for Open Science
Research Data College

September 2025

Dr. Véronique STOLL
Observatoire de Paris-PSL, Library, F-75014 Paris, France
ORCID : 0000-0002-4118-3574

Dr. Frédéric de LAMOTTE
UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France & Institut Français de Bioinformatique (IFB), CNRS, INSERM, INRAE, CEA, 91057 Evry, France
ORCID : 0000-0003-4234-1172

DOI : 10.52949/85

September 2025

Conception graphique : opixido

Contents

Introduction

The current research landscape produces an increasing volume of publications, data and, more broadly, diverse scientific results and objects (digital and/or physical). In this context, the ability to uniquely and reliably identify different elements of the scientific ecosystem - researchers, publications, datasets, software, etc. - across multiple information systems has become a central challenge for structuring and enhancing scientific activity, particularly to enable traceability of scientific objects.

Persistent Identifiers (PIDs) address this challenge. These are unique and permanent digital or alphanumeric codes that are readable by both humans and machines. Unlike URLs, which can change or become obsolete, PIDs are designed to provide lasting references, ensuring stable access to an entity (digital or otherwise). They enable identification, discovery, traceability, and standardized citation throughout the research lifecycle.

The use of PIDs directly supports the implementation of FAIR principles (Findable, Accessible, Interoperable, Reusable) by making digital objects more easily discoverable, accessible, interoperable, and reusable. Furthermore, they promote automation of data exchange between information systems, contributing to administrative simplification through a logic of reuse and non-duplication of information ("tell us once" principle). As such, PIDs are essential for ensuring sustainability, consistency, and interoperability of data in digital environments for higher education and research.

Recognition of PIDs as structural instruments of open science is part of an international movement. Several initiatives converge in this direction, notably the Canadian federal government's roadmap for open science[1], guidelines from the Office of Science and Technology Policy (OSTP) in the United States[2], and the PID policy developed within the European Open Science Cloud (EOSC) framework[3]. The United Kingdom and Australia have measured the benefits of adopting PIDs in terms of the number of days of administrative work saved for researchers[4]. These countries, as well as Finland, Canada, the Netherlands, Germany, the Czech Republic, South Korea and New Zealand, have implemented policies or roadmaps in this area to improve the quality and efficiency of research[5]. The G7 Research Compact (2021) also commits member countries to strengthening the availability, sustainability, interoperability, and accessibility of scientific data, technologies, and infrastructures[6]. Finally, PIDs are explicitly mentioned in UNESCO recommendations on open science as fundamental elements for open, reliable, and sustainable research governance[7].

This document is part of the work launched in 2024 by the MESR on the roadmap “Data for simplification and research management”, whose guiding principles aim to ensure the circulation and interoperability of data while respecting the autonomy of institutions.

The roadmap is based on an action plan developed collectively by stakeholders and follows the principle of “tell us once”, reflecting the commitment to reduce the administrative burden on research teams. It relies on a set of qualified data to be shared across information systems, according to common quality standards and principles, under a framework of collective governance.

The objectives are to strengthen interoperability between systems, consolidate and improve the reliability of shared data, enhance coordination between supervisory bodies, and reduce repeated data collections and surveys.

In this context, persistent identifiers play a key role in ensuring the interoperability of data across heterogeneous higher education and research systems, by guaranteeing the traceability, reliability, and reusability of information.

[1] Roadmap for Open Science, 2020. https://science.gc.ca/site/science/sites/default/files/attachments/2022/Roadmap-for-Open-Science.pdf
[2] National Science and Technology Council, Subcommittee on Open Science, Desirable Characteristics of Data Repositories for Federally Funded Research 2022. DOI : https://doi.org/10.5479/10088/113528
[3] Kotarski, R., et al., *A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC)*, Publications Office, 2020.
[4] https://doi.org/10.5281/zenodo.7356219
https://doi.org/10.5281/zenodo.4772627
[5] https://zenodo.org/records/10656276 https://repository.jisc.ac.uk/7840/2/PID_roadmap_for_open_access_to_UK_research.pdf
[6] G7 Research Compact, 2021. https://www.consilium.europa.eu/media/50365/g7-2021-research-compact-pdf-356kb-2-pages-1.pdf
[7] UNESCO Recommendation on Open Science, 2021.

A persistent unique identifier (PID) is an alphanumeric string that provides an unambiguous and lasting identification of a digital or physical object (e.g., Crossref DOI for a publication, DataCite DOI for data, SWHID for code or software, ORCID for a contributor, ROR for an institution). Interoperable and machine-readable, these identifiers make it easier to exchange information between different information systems.

Context and Challenges

In the context of digital transformation of higher education and research, the management of data, algorithms, and source codes receives increasing attention, particularly driven by the State's Data, Algorithms, and Source Codes Policy, which promotes their openness, traceability, and reusability[8]. This dynamic aligns with national plans for open science and European requirements (e.g., Horizon Europe, EOSC). Furthermore, the Gillet report on administrative simplification in higher education and research emphasizes the need to better structure the research information ecosystem, particularly through coordinated adoption of persistent identifiers, to improve management, monitoring, and enhancement of scientific productions while reducing administrative burdens on researchers and institutions[9].

The widespread adoption of persistent identifiers generates significant benefits for all actors in the research ecosystem. It contributes notably to:

- **Strengthening visibility, citability, and traceability of research units, contributors and their outputs**: PIDs increase the visibility of national scientific production by facilitating discovery, access, and reuse of resources (articles, data, software, open educational resources, etc.) in new research contexts;
- **Ensuring scientific integrity and reinforcing trust in research results**: by acting as unambiguous reference points for information exchange and consolidation, PIDs enable precise documentation and tracking of research structures, contributions, processes, and research objects, thus consolidating the reliability of scientific results;
- **Facilitating reuse of scientific objects**, whether data, publications, software, or other resources;
- **Reducing administrative burdens on research teams, units, and their supervisory bodies**, through automation of information exchange, data standardization, and limitation of multiple data entries;
- **Supporting the application of FAIR principles** (Findable, Accessible, Interoperable, Reusable) by providing technical and organizational anchoring points for sustainable and interoperable management of research resources.

Diversity of Persistent Identifiers

Despite their strategic importance, the proliferation and heterogeneity of persistent identifiers can limit their readability, relevance, and effectiveness. At the international scale, their multiplication has created a complex ecosystem, marked by the coexistence of complementary or even redundant identifiers, without a unified overall vision always being available.

Some persistent identifiers have established themselves as widely recognized standards within certain scientific and academic communities. However, the diversity of available PIDs leads to the coexistence of several identifiers for the same entity, depending on the usage context and specific needs of actors. The PID ecosystem is thus characterized by high density, with heterogeneous levels of recognition, maturity, and integration into practices.

In this context, it appears essential to have a clear understanding of existing typologies and characteristics specific to each type of PID, to enable coherent, rational, and efficient use in service of research.

[8] https://www.ouvrirlascience.fr/politique-des-donnees-des-algorithmes-et-des-codes-sources-feuille-de-route-2021-2024/
[9] https://www.enseignementsup-recherche.gouv.fr/fr/remise-du-rapport-de-la-mission-gillet-sur-l-ecosysteme-de-la-recherche-et-de-l-innovation-91274

Two main types of PIDs are generally distinguished:

- **Object identifiers** are used to identify scientific productions. This includes publications (articles, book chapters, conference proceedings, theses, posters...), datasets, software, open educational resources (OER), research instruments, samples, or projects;
- **Contributor identifiers** identify the actors involved in research. They mainly refer to the authors of scientific publications, research organizations or institutions (research units, universities, research bodies) as well as funding agencies.

The choice of a PID depends primarily on user needs and the nature of the resource to be identified. Several criteria can be considered for evaluating a PID:

- Its category (object, contributor, or mixed);
- The allocation process and the system behind it, which ensures proper management and long-term sustainability (e.g., DataCite, Crossref, Handle...);
- Its main application domain (scientific, cultural, etc.);
- The existence of associated standards or conventions (ISO, Climate and Forecast Convention...);
- Its syntax (meaningful or not), which can affect understanding and persistence in case of entity name changes;
- Its granularity capacity, i.e., to identify different levels of a resource;
- The metadata associated with it, which describe the identified entity;
- Its resolvability (i.e., the ability to retrieve the resource via the identifier);
- Its hosting mode (centralized or decentralized);
- Its cost (free, paid, variable).

Beyond these criteria, the viability of a PID fundamentally depends on the robustness of its governance. Allocating and managing organizations must have a formalized policy covering the entire lifecycle of identifiers. This policy should explicitly define creation and management procedures, as well as responsibilities regarding reliability, persistence, and availability of identifiers. It should also set out the conditions of access (including any possible business model), the guarantees for long-term preservation, and the mechanisms ensuring continuity and resilience of the associated services.

Guiding Principles and Strategic Objectives

Guiding Principles of a National PID Policy

A national strategy for persistent identifiers should be structured around the following guiding principles:

- **Interoperability and standardization**: ensure compatibility between national and international ecosystems by aligning existing or newly developed national systems (such as RNeST – national registry of French research organizations, or HAL – French national open archive) with recognized international standards (such as ORCID, ROR, DataCite). This compatibility relies on adoption of standardized metadata, implementation of open APIs, and definition of reliable references ("points of truth") for each type of entity concerned, as well as alignment of controlled vocabularies and repositories within different scientific communities;
- **Collective and structured governance**: involve all stakeholders in higher education and research (institutions, organizations, infrastructures, professional networks) as well as relevant operators (such as ABES, CCSD) within a shared governance framework. For this purpose, the Identifiers and Data Circulation Steering Committee, established in 2024, will evolve to take on the role of coordinating PID strategy at the national level;
- **Openness, independence, and transparency**: prioritize the use of PID services operated by non-profit structures with open governance based on scientific community involvement, to ensure independence, sustainability, and transparency of offered services;
- **Infrastructure sustainability**: ensure long-term durability of identification systems by anticipating technological and institutional developments likely to affect their functioning, to guarantee continuity of PID-associated services;

- **Systemic approach**: adopt a global vision covering the entire research lifecycle, integrating persistent identification of contributors, structures, projects, as well as all scientific productions (publications, data, software, etc.).

Strategic Objectives

Aligned with the work launched in 2024 by the MESR on administrative simplification, the national strategy on persistent identifiers pursues the following objectives:

- **Promote adoption of priority PIDs**, by supporting their widespread use while ensuring coherent articulation with existing national identifiers (IdRef, IdHal...):
  - ORCID identifier for research actors (researchers, engineers...);
  - DOI, Handle, ARK, or SWIDH identifiers for scientific productions (publications, datasets, software, etc.);
  - ROR identifier for research structures.
- **Implement clear, coordinated, and shared governance**, involving concerned ministerial actors, higher education and research institutions, national infrastructures, and operators. This governance must be closely connected to international dynamics regarding PIDs;
- **Strengthen interoperability of information systems**, through alignment of identifiers and metadata used at national scale with international standards (ORCID, ROR, DataCite) and through fluid integration of these identifiers via open programming interfaces (APIs) in digital systems;
- **Support sovereignty and independence of PID infrastructures**, particularly through active participation in their governance, to ensure their sustainability and alignment with scientific community interests;
- **Deploy the National Registry of Research Structures (RNeST)**, which is set to become the national registry of research structures. This registry will serve as the trusted national source, ensure consistent representation in ROR through automatic updates, and prevent duplicate data entry. The MESR has entrusted ABES with the operational governance of RNeST;
- **Support research landscape actors**, through training, awareness, and support actions, particularly relying on the network of competence centers of the Research Data Gouv ecosystems.

Operational Axes and Recommendations

In the PID ecosystem, some are now widely recognized and used internationally, such as DOI (for publications and data) or ORCID (for researchers). Others, more recent, have emerged to meet new needs, such as SWHID for software or RAiD for research projects. Some PIDs still need consolidation, whether in terms of adoption, standardization, or global interoperability, such as identifiers for instruments or protocols. Finally, needs arise to cover the entirety of a research project's lifecycle, particularly regarding financial flows, HPC, or storage systems. This diversity reflects an infrastructure under construction, where complementarity and coordination between PIDs are essential for truly open and transparent science.

Adopting International Persistent Identifiers in Higher Education and Research

In the context of Higher Education and Research, the Research Data College of the French Committee for Open Science primarily recommends using the following PIDs at the national level:

Structure and Organization Identifiers

ROR (Research Organization Registry)

Used to identify research structures and organizations, this identifier is widely adopted at the international level. A comprehensive and consistent registration of French structures in RNeST will ensure their automatic addition or update in ROR. As the national trusted source for ROR, RNeST will rely on a common reference framework shared by all institutions for the characterization and designation of research structures. This register will provide a more detailed representation of the organization of French research for national data circulation needs—something ROR does not offer—while preserving the use of ROR at the international level.

ROR and Crossref ID

Crossref Funder IDs document funders in the metadata of publications and projects, while ROR IDs identify research structures (institutions, units, etc.) as producers or project holders. These two systems are working towards a systematic alignment of their registries.

Person Identifiers

ORCID (Open Researcher and Contributor ID)

Among the identifier systems dedicated to research contributors, ORCID has established itself as a widely adopted international standard. Already well integrated into the HAL open archive, ORCID is managed by a global non-profit organization governed by a board drawn from the scientific community, in which France is represented. Its funding comes exclusively from institutional membership fees. In France, a national consortium, led by ABES, coordinates member institutions, provides them with first-level support, and facilitates the deployment of ORCID[10].

The creation of an ORCID identifier and profile may be initiated either by an individual research contributor (researcher, engineer, doctoral student, technician, or anyone wishing to obtain an ORCID identifier) or by their employing institution[11]. However, it always requires the consent of the person concerned, in line with ORCID's founding policy, which guarantees that each individual controls the

[10] https://orcid-france.fr/
[11] Only the employing institution is authorized to certify the affiliations on ORCID profiles.

information displayed on their profile. ORCID profiles may be updated directly by the individual or by members of the ORCID consortium (institutions, organizations, publishers, funders, etc.), thanks to ORCID's integration into their information systems. In all cases, sources are clearly identified, and a green check mark signals certified information, recognized by ORCID as "Trust Markers."

The use of a public ORCID profile enhances both visibility and interoperability. Each profile may be managed by its holder and, with their agreement, linked automatically to the employer's information system to record certified information. ORCID operates globally in compliance with the GDPR and places user trust among its strategic priorities. More than 98% of data are public and archived in France, ensuring both availability and security of information.

Recommandations

Systematically encourage research stakeholders to create and maintain a public ORCID profile by integrating it into institutional procedures and tools (HAL deposit, publication management, etc.). In this respect, the use of ORCID on the appelsprojetsrecherche.fr portal, which brings together six partners, provides a concrete and significant example of this simplification effort.

Systematically encourage employing institutions to integrate ORCID into their information systems, in order to automatically populate contributors' profiles with certified information.

Publication Identifiers

Scientific publications are primarily identified by DOIs (Digital Object Identifiers), assigned by agencies such as Crossref or DataCite. This widely used standard in scholarly publishing also supports open citations (OpenCitations).

Crossref focuses on the visibility and traceability of academic publications. Its metadata schema provides detailed descriptions of references, authors (with ORCID links), affiliations, abstracts, licenses, funding, and links to full texts. It also offers advanced services for citation tracking and interoperability with ORCID and bibliographic platforms, making it particularly suited to editorial workflows and traditional article citation.

DataCite's metadata schema is more generic and extensible, designed to describe a wide range of scientific resources, particularly datasets. It emphasizes relationships between objects (links between DOIs), the granularity of resource types, and rights and license management. This makes it well suited for describing digital data, research materials, or supporting objects, even if it is less detailed for traditional publications.

In summary, Crossref metadata are oriented toward publication traceability, while DataCite metadata emphasize the reusability and citability of research data.

Recommandation

Systematically use DOIs **assigned by Crossref** for scientific publications.

Research Data and Digital Object Identifiers

Several persistent identifier systems are used for research data, depending on disciplinary practices, with the most common being DOI, Handle, and ARK.

> Research data are factual records (numbers, texts, images or sounds) used as primary sources in scientific research and generally recognized by the scientific community as necessary to validate research results

DataCite DOIs

DataCite is an international non-profit consortium that allows assigning DOIs to research data, but also to other types of scientific productions. In France, the national consortium is carried by INIST-CNRS, which assigns DOIs on behalf of research institutions and infrastructures.

The DataCite metadata schema evolves regularly to meet the scientific community's needs (latest version 4.6 published in December 2024). This schema is based on a modular and hierarchical architecture, founded on XML, with a JSON version also available. It includes mandatory elements (title, creator, DOI, publisher, publication date) and optional ones (contributors, affiliations, funding, inter-object relations, etc.) allowing adaptable granularity. Use of controlled vocabularies for certain fields ensures semantic consistency and a relational model allows documenting links between research objects.

> **IGSN IDs**
>
> An IGSN ID is the identifier specifically designed for physical samples. It ensures traceability and interoperability between samples and different entities in the research ecosystem (projects, instruments, publications, data, people, institutions).
>
> In September 2021, a strategic partnership was concluded between IGSN (International Generic Sample Number) and DataCite, entrusting the latter with responsibility for providing IGSN ID registration services as well as associated technological infrastructure. IGSN allocation agencies now join DataCite, either as direct members or via consortia. Furthermore, IGSN identifiers previously assigned as Handles are now registered as DOIs. Technically, this convergence relies on a specific adaptation of the DataCite metadata schema, modified to introduce new metadata necessary for describing physical samples, such as geolocation, sampling methods, or material characteristics.

Handles

Designed in the 1990s, the Handle System constitutes one of the first technical devices developed for assigning and managing persistent identifiers. Since 2014, Handle System governance has been ensured by the DONA Foundation, an international non-profit organization. The Handle System is distinguished mainly from DataCite by a decentralized infrastructure (each institution can host its own Handle server), allowing local identifier management, open source management software, and absence of mandatory metadata.

Handles are particularly used in contexts where institutional independence and flexibility are sought, such as in archival repositories, digital libraries, or certain research infrastructures.

However, low standardization of metadata and absence of structured community governance constitute barriers to their adoption in certain international use cases.

DOIs and Handles: similarities and differences

DOIs rely technically on Handle System infrastructure for their operation.

DOIs assigned by DataCite are based on Handle System technology, but unlike simple Handles, DOIs are associated with standardized and mandatory metadata, promoting their interoperability, visibility, and reuse, and are managed within a community governance framework, which guarantees transparency, openness, and adaptation to research needs and FAIR principles.

ARKs (Archival Resource Key)

Created in 2001 by the California Digital Library and recognized since 2019 as a URI (Uniform Resource Identifier) scheme, the ARK system was designed to provide a stable identifier to any digital or physical resource with a digital representation, regardless of its nature (document, bibliographic record, corpus, semantic web entity, etc.).

Like the Handle System, the ARK system relies on an open and decentralized infrastructure, allowing any institution to freely assign its own identifiers. Widely used by cultural and heritage institutions, it offers implementation flexibility, without obligation for centralized deposit or standardized metadata. Long-term persistence and accessibility of ARKs rely exclusively on the commitment of issuing institutions.

Several French institutions are naming authorities (NAAN), notably the Bibliothèque nationale de France, which uses ARKs for its digital and bibliographic resources, and has hosted a replica of the international registry since 2008.

Recommandation

Due to their scientific governance framework and standardized metadata schema, it is recommended to use DOIs assigned by DataCite for identifying datasets produced in higher education and research.

Software Identifiers

Several persistent identifiers are used for software, mainly:

- DOIs (or HAL IDs), assigned to a fixed version of software, often linked to a publication;
- SWHID (Software Heritage ID): cryptographically verifiable identifier of a code artifact (commit, file, directory, etc.);
- RRID, frequently used in life sciences and multi-purpose (software, databases, materials...).

The Software Heritage ID (SWHID) is calculated by Software Heritage to provide an intrinsic, unique identifier specific to the code content itself, resistant to platform changes (e.g., GitHub, GitLab) and location. Based on a cryptographic fingerprint (SHA-1), derived from a Merkle DAG structure, it allows, thanks to fine granularity, identification of an entire repository, a commit, a directory, a file, or even a line of code. Recommended by several funding agencies and institutions, it thus enables software citation in scientific publications.

Since April 2025, SWHID is published as international standard ISO/IEC 18670.

Use Cases and Complementarity between DOI and SWHID for Software Identifiers

DOI is a citation of a published software version, described through the DataCite metadata schema. It primarily targets the scientific publication context.

SWHID is identification of a specific code artifact. Its persistence is ensured by hashing and Software Heritage. By default, its technical scope does not provide for entering metadata as rich as DataCite. SWHID therefore targets the technical content of software.

Recommandation

It is recommended to use Software Heritage ID (SWHID) identifiers for identifying software produced in higher education and research or, in the context of a publication, DOIs associated with SWHID. Offering fine granularity, assured traceability, and persistence guaranteed by long-term archiving, SWHIDs constitute a robust and interoperable solution for valorizing, referencing, and preserving source code, in coherence with open science principles.

Project Identifiers

RAiD (Research Activity Identifier) is a PID intended to reference in a stable and univocal manner a research activity as a whole (project, program, initiative...). Initially developed by the Australian Research Data Commons (ARDC), the system is now framed by standard ISO 23527:2022. It is progressively integrating into the global open science ecosystem as a structuring identifier of scientific activities.

RAiD aims to persistently identify a research project in its entirety and link all involved digital objects and actors (people, institutions, results, etc.). RAiD therefore designates an activity as a whole (project, sub-project), without limiting itself to results, linking third-party identifiers (ORCID, ROR, DOI, SWHID...). It contains information on title, scope, stakeholders, dates, associated productions.

Its open governance is under development, in collaboration with the international scientific community.

RAiD is not intended to replace other identifiers, but to link them around the same activity by acting as a transversal identifier. Through a partnership with DataCite, RAiD is progressively integrated into EOSC (EOSC RAiD), thus guaranteeing its interoperability with data repositories, publications, and infrastructures[12]. Complementary to existing PIDs, it identifies and valorizes funded projects, contributes to automating reporting processes and monitoring research activities, and promotes open science through systematic documentation of productions and contributions.

Recommandation

[12] Interoperability is defined as the ability of a set of computer systems, applications, or components of the same type to communicate, exchange, or use data or information without ambiguity.

Instruct the deployment of RAiDs at the national level in consultation with funding agencies and main stakeholders in the research ecosystem, to guarantee their coherent integration into funding, monitoring, steering, and evaluation systems for scientific projects.

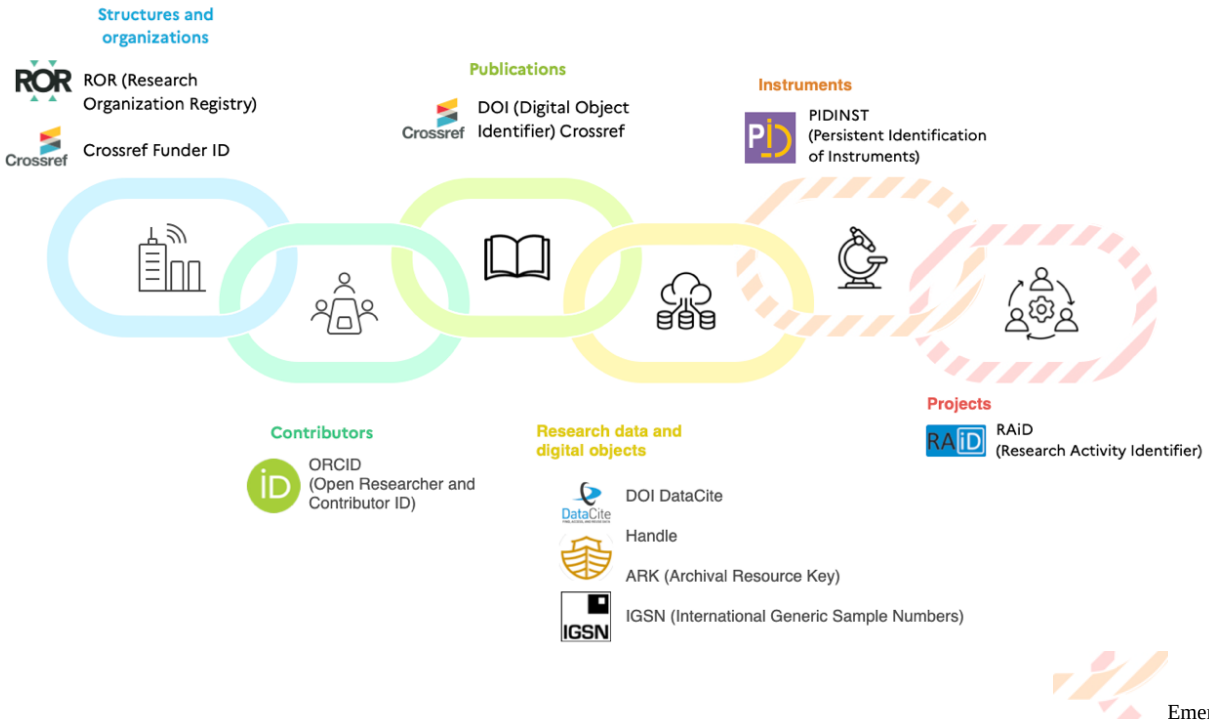
Specific Objects Identifiers

PIDINST (Persistent Identification of Instruments) is specifically designed to reference scientific instruments used in research activities, such as sensors, microscopes, spectrometers, telescopes, etc. The PIDINST schema is supported by the Research Data Alliance (RDA), with a first version of its metadata schema published in 2020. It remains an emerging identifier, still being adopted and consolidated, but it responds to a growing need in domains where instrumentation plays a central role in generating scientific data.

Its main objective is to facilitate traceability, reusability, and transparency of data produced by these instruments, regardless of their place of use or owner. It thus allows enriching dataset metadata with precise references to used instruments (type, manufacturer, model, owner, location, usage periods, links with produced data...). It is generally based on a DOI assigned via DataCite.

Recommandation

Conduct national reflection in connection with concerned disciplinary communities to evaluate the relevance and adoption procedures of PIDINST and adapt its deployment to specific needs regarding identification, traceability, and valorization of scientific instruments and infrastructures.



Main international PIDs used in High Education and Research

Identifier interoperability: a key factor for ensuring effective data exchange by aligning local, national and international identifiers

The heterogeneity of identifier systems constitutes a significant obstacle to their widespread adoption and effectiveness, particularly regarding fluid data circulation and simplification of administrative processes.

To guarantee their optimal integration into information systems, persistent identifiers and their associated metadata must be made available in structured, standardized, and reusable formats. This implies systematic use of open programming interfaces (APIs) and metadata conforming to international best practices.

Adoption of shared metadata standards, as well as development of a common vocabulary, are indispensable conditions for interoperability of national systems with international infrastructures and repositories. This convergence is essential to ensure coherence, reliability, and sustainability of information within the digital research ecosystem.

Recommandations

- Deploy open APIs and provide structured metadata to promote interoperability, reuse, and automation of information exchanges.
- Align national systems with internationally recognized metadata schemas to guarantee coherent integration into the global ecosystem of persistent identifiers.

Governance et management

From a perspective of sustainability and adequacy with research and open science needs, governance of persistent identifier providers based on scientific community involvement should be prioritized. Recourse to non-profit organizations constitutes a guarantee of openness, independence, and neutrality. It is also essential to ensure complete transparency, particularly on economic models of PID services, documentation of underlying data, and operational processes.

Risk analysis related to persistent identifiers, whether social, political, economic, or technological, should constitute an essential component of the national strategy. The risk of service interruption, particularly in case of organizational restructuring or commercial acquisition of PID providers, represents a major political issue. The question of digital sovereignty is particularly sensitive, especially regarding transfer of personal data to foreign jurisdictions, as is the case for ORCID data hosted in the United States. It is imperative to anticipate these risks, document them clearly, and support concerned actors with appropriate legal recommendations.

Furthermore, the national strategy for PIDs must be based on concerted governance, carried by stakeholders. Identification and mobilization of key actors at the national scale (funders, research organizations, infrastructures) are indispensable for collectively defining implementation procedures. Creation of a dedicated governance body, such as a PID Advisory Council, is among best practices identified internationally. This approach must build on lessons learned from foreign experiences, particularly through participation in coordination initiatives such as the RDA working group on national PID strategies.

Finally, success of this strategy relies on clear and targeted communication of concrete PID benefits to researchers. Institutions have a central role to play in supporting and raising awareness of regular use of persistent identifiers. Complementary communication systems should be implemented - both institutional (top-down) and participatory (bottom-up). The strategy should ensure bridges between communities and disciplines, taking into account disciplinary specificities.

Recommandations

Conduct risk analysis related to deployment and dependence on persistent identifier infrastructures.

Implement dedicated, concerted and representative national governance of stakeholders.

Support the national PID strategy with sustainable and structuring funding.

Implement communication, training, and awareness actions adapted to specificities of disciplinary communities, to ensure broad and sustainable appropriation of identifiers throughout higher education and research.

Conclusion

Widespread adoption and coordinated deployment of persistent unique identifiers constitute an essential condition for realizing open science ambitions in France. PIDs form the backbone of a research digital infrastructure that is simultaneously efficient, transparent, and interoperable. They not only increase visibility and citability of French scientific productions, but also strengthen traceability, integrity, and reusability of data, software, publications, projects, instruments, and contributors. By facilitating automation of management and reporting processes, they also contribute to reducing administrative burdens for researchers and institutions.

To build a scientific information landscape structured by PIDs, it is necessary to:

- Adopt internationally recognized persistent identifiers within research communities;
- Coordinate national efforts around common repositories;
- Ensure alignment with international standards (ORCID, DOI, ROR, RAiD, SWHID, etc.);
- Guarantee independence, sustainability, and governance of used registries;
- Train and support scientific communities and institutions;
- Valorize uses and demonstrate concrete benefits (steering, evaluation, valorization, reproducible science).

Several persistent unique identifiers have now reached maturity and are widely adopted internationally, such as DOI for publications and datasets, ORCID for researchers, and ROR for institutions. Others are emerging or being structured, such as RAiD for research activities or PIDINST for scientific instruments. Some still need consolidation, particularly regarding their governance, international recognition, or interoperability with existing systems. Beyond already well-established identifiers, specific needs persist in certain sectors or disciplinary communities, where common repositories must still be identified or developed. It is essential that building the PID ecosystem relies not only on a common foundation shared at national and international scales, but also on fine consideration of practices, uses, and realities specific to each scientific domain. This concerted and differentiated approach is indispensable for ensuring relevance, appropriation, and effectiveness of PIDs throughout the research ecosystem.

Summary of recommandations

- Systematically encourage research stakeholders to create and maintain a public ORCID profile, integrating it into institutional procedures and tools (HAL deposit, publication management, etc.).
- Systematically encourage employing institutions to integrate ORCID into their information systems, in order to automatically populate contributors' profiles with certified information.
- Systematically use DOIs assigned by Crossref for scientific publications.
- Due to their scientific governance framework and standardized metadata schema, it is recommended to use DOIs assigned by DataCite to identify datasets produced within higher education and research. Most trusted repositories (Recherche Data Gouv, Data Terra, etc.) allow DOI assignment.
- It is recommended to use Software Heritage IDs (SWHID) for software produced within higher education and research or, in the context of a publication, DOIs (or HAL IDs) associated with the SWHID.
- Promote the deployment of RAiD identifiers at the national level in coordination with funding agencies and key research ecosystem stakeholders, to ensure their coherent integration into funding, monitoring, management, and evaluation systems for scientific projects.
- Conduct a national review in consultation with relevant disciplinary communities to evaluate the relevance and adoption modalities of PIDINST, and adapt its deployment to specific needs for identification, traceability, and valorization of scientific instruments and infrastructures.
- Deploy open APIs and provide structured metadata to promote interoperability, reusability, and automation of information exchanges.
- Align national systems with internationally recognized metadata schemas to ensure coherent integration into the global persistent identifier ecosystem.
- Conduct a risk analysis related to the deployment of and dependence on persistent identifier infrastructures.
- Establish a dedicated national governance structure that is consultative and representative of stakeholders.
- Support the national PID strategy with sustainable and structured funding.
- Implement communication, training, and awareness-raising actions tailored to the specificities of disciplinary communities, to ensure broad and lasting adoption of identifiers across higher education and research.

List of abbreviations

ABES : Agence bibliographique de l'enseignement supérieur (Higher Education Bibliographic Agency)

API : Application Programming Interface

ARK : Archival resource Key

CCSD : Centre pour la communication scientifique directe (Center for Direct Scientific Communication)

DOI : Digital Object Identifier

EOSC : European Open Science Cloud

FAIR (principles) : Findable, Accessible, Interoperable, Reusable

HAL : Hyper Articles en Ligne (National open archive repository)

HPC : High Performance Computing

IGSN : International Generic Sample Numbers

ISO : International Organization for Standardization

MESR : Ministère de l'Enseignement supérieur et de la Recherche (French Ministry of Higher Education and Research)

OER : Open Educational Resources

ORCID : Open Researcher and Contributor ID

PID : Persistent Identifier

PIDINST : Persistent Identifier Instrument

RaiD : Research Activity Identifier

RDA : Research Data Alliance

RNeST : Registre national des structures de la recherche (national Registry of French research organizations)

ROR : Research Organization Registry

SWHID **:** Software Heritage ID

Suggested reading

Research Data Alliance (RDA), Persistent Identification of Instruments (PIDINST) working group. (2025) White paper. https://docs.pidinst.org/en/latest/white-paper/index.html#white-paper

Gruenpeter, M. (2025). SoftWare Hash IDentifier (SWHID) and use cases. Online seminar PIDs for Software, Online. Zenodo. https://doi.org/10.5281/zenodo.15750889

Australian Research Data Commons. (2024). Australian national Persistent Identifier (PID) strategy 2024. Zenodo. https://doi.org/10.5281/zenodo.10656276

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

Akcaova, G., Slot, P., de Smaele, M., Hoogerwerf, M., Sesink, L., Baars, C., de Jong, M., Coombs, S., van de Sanden, M., & Voorburg, R. (2022). Towards a national PID roadmap. Zenodo. https://doi.org/10.5281/zenodo.584931 [Pays-Bas]

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. Zenodo. https://doi.org/10.5281/zenodo.7217469

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. Zenodo. https://doi.org/10.5281/zenodo.7100578

Leggott, Aspler, Jenkyns, Corrie, Nicholson, Liu, & Huitema. (2022). Persistent Identifiers: Current Landscape and Future Trends (2.0). Zenodo. https://doi.org/10.5281/zenodo.7065515

Tatum, C., Doove, J., Bijsterbosch, M., Akcaova, G., Slot, P., Bouwhuis, M., Veenstra, N., Vanderfeesten, M., & van den Hoogen, H. (2022). From ORCID Pilot to a PID-centric framework for Research Information. Zenodo. https://doi.org/10.5281/zenodo.5836056 [Pays-Bas]

Permaliens et identifiants pérennes, France Archives, https://francearchives.gouv.fr/fr/article/339695117

Brown, J., Jones, P., Meadows, A., Murphy, F., & Clayton, P. (2021). UK PID Consortium: Cost-Benefit Analysis (1.0). Zenodo. https://doi.org/10.5281/zenodo.4772627

Commission européenne, Direction générale de la recherche et de l'innovation, EOSC Executive Board, Hellström, M., Heughebaert, A., Kotarski, R., et al., *A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC)*, Publications Office, 2020, https://data.europa.eu/doi/10.2777/926037