



HAL
open science

Recommandations pour l'adoption des identifiants persistants dans l'enseignement supérieur et la recherche

Véronique Stoll, Frédéric de Lamotte

► To cite this version:

Véronique Stoll, Frédéric de Lamotte. Recommandations pour l'adoption des identifiants persistants dans l'enseignement supérieur et la recherche. Comité pour la Science Ouverte. 2025. <hal-05266777v2>

HAL Id: hal-05266777

<https://hal-lara.archives-ouvertes.fr/hal-05266777v2>

Submitted on 18 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



**Recommandations
pour l'adoption
des identifiants
persistants dans
l'enseignement
supérieur et la
recherche**

Comité pour la science ouverte

Collège des Données de la recherche

Octobre 2025



Véronique STOLL

Observatoire de Paris-PSL, Bibliothèque, F-75014 Paris, France
ORCID : 0000-0002-4118-3574

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

La première version de ce document a été publiée en septembre 2025.

La présente version tient compte des évolutions des travaux de la Research Data Alliance autour des identifiants des instruments, équipements et infrastructures, présentés en octobre 2025 lors de la 25^e plénière.

DOI : 10.52949/84

Octobre 2025

Conception graphique : opixido



Except where otherwise noted, this work is licensed under
<https://creativecommons.org/licenses/by/4.0/deed.fr>

Sommaire

Introduction	4
Contexte et enjeux	5
Diversité des identifiants persistants	5
Principes directeurs et objectifs stratégiques	7
Axes opérationnels et préconisations	8
Conclusion	17
Synthèse des recommandations	18
Table des abréviations	19
Bibliographie indicative	20

Introduction

Le paysage actuel de la recherche produit un volume croissant de publications, de données et, plus largement, de résultats et objets scientifiques (numériques et/ou physiques) diversifiés. Dans ce contexte, la capacité à identifier de manière unique et fiable les différents éléments de l'écosystème scientifique – équipes de recherche, publications, ensembles de données, logiciels, etc. - à travers une multiplicité de systèmes d'information, est devenue un enjeu central pour la structuration et la valorisation de l'activité scientifique notamment pour permettre une traçabilité des objets scientifiques.

Les identifiants persistants (PID – *Persistent Identifiers*) répondent à cet enjeu. Il s'agit de codes numériques ou alphanumériques, uniques et permanents, lisibles à la fois par l'homme et par la machine. Contrairement aux adresses URL, susceptibles de changer ou de devenir obsolètes, les PIDs sont conçus pour constituer des références durables, assurant la stabilité de l'accès à une entité (numérique ou non). Ils permettent ainsi son identification, sa découverte, sa traçabilité et sa citation normalisée, tout au long du cycle de la recherche.

L'usage des PIDs participe directement à la mise en œuvre des principes FAIR (*Findable, Accessible, Interoperable, Reusable*), en rendant les objets numériques plus facilement repérables, accessibles, interopérables et réutilisables. Par ailleurs, ils favorisent l'automatisation des échanges de données entre systèmes d'information, contribuant à la simplification administrative, dans une logique de réutilisation et de non-duplication des informations ("Dites-le nous une fois"). À ce titre, les PIDs sont indispensables pour garantir la durabilité, la cohérence et l'interopérabilité des données dans les environnements numériques, notamment ceux de l'enseignement supérieur et de la recherche.

La reconnaissance des PIDs comme instruments structurants de la science ouverte s'inscrit dans une dynamique internationale. Plusieurs initiatives convergent en ce sens, notamment la feuille de route pour la science ouverte du gouvernement fédéral canadien¹, les orientations de l'*Office of Science and Technology Policy* (OSTP) aux États-Unis², ou encore la politique PID développée dans le cadre de l'*European Open Science Cloud* (EOSC)³. Le Royaume Uni et l'Australie ont mesuré les bénéfices de l'adoption des PIDs en termes de nombre de jours de travail administratif évités par les chercheurs⁴. Ces pays, ainsi que la Finlande, le Canada, les Pays-Bas, l'Allemagne, la République tchèque, la Corée du Sud et la Nouvelle-Zélande se sont dotés de politiques ou de feuilles de route en la matière pour améliorer la qualité et l'efficacité de la recherche⁵. Le Pacte pour la recherche du G7 (2021) engage également les pays membres à renforcer la disponibilité, la durabilité, l'interopérabilité et l'accessibilité des données, technologies et infrastructures scientifiques⁶. Enfin, les PIDs sont explicitement mentionnés dans les recommandations de l'UNESCO sur la science ouverte, en tant qu'éléments fondamentaux pour une gouvernance ouverte, fiable et durable de la recherche⁷.

L'ensemble de ce document s'inscrit dans le cadre des travaux initiés en 2024 par le MESR autour de la feuille de route des « Données pour la simplification et le pilotage de la recherche », dont les principes directeurs visent à garantir la circulation et l'interopérabilité des données, tout en respectant le principe d'autonomie des établissements. Cette feuille de route repose sur un plan d'action élaboré collectivement par les parties prenantes et s'appuie sur le principe « dites-le-nous une fois », qui illustre la volonté d'alléger le fardeau administratif des équipes de recherche, en s'appuyant sur un ensemble de données qualifiées à partager entre les systèmes d'information, selon des standards et principes de qualité communs et une gouvernance collective. Les objectifs visés sont l'interopérabilité accrue entre systèmes, la consolidation et la fiabilité des données partagées, une meilleure coordination entre tutelles et la réduction des collectes et enquêtes répétitives. Dans ce contexte, les identifiants persistants jouent un rôle clé pour

¹ Feuille de route pour la science ouverte, 2020. Disponible sur : <https://science.gc.ca/site/science/sites/default/files/attachments/2022/La-Feuille-de-route-pour-la-science-ouverte.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 Voir par exemple <https://zenodo.org/records/10656276> pour l'Australie,

https://repository.jisc.ac.uk/7840/2/PID_roadmap_for_open_access_to_UK_research.pdf pour le Royaume-Uni

⁶ Pacte du G7 pour la recherche, 2021. Disponible sur : <https://www.consilium.europa.eu/media/50365/g7-2021-research-compact-pdf-356kb-2-pages-1.pdf>

⁷ UNESCO Recommendation on Open Science, 2021.

assurer l'interopérabilité des données des systèmes hétérogènes de l'ESR, en garantissant traçabilité, fiabilité et réutilisabilité des informations.

Un identifiant persistant unique (Persistent Identifier - PID) est une chaîne de caractères alphanumériques qui identifie de façon non ambiguë et persistante dans le temps un objet numérique ou physique (ex: DOI Crossref pour une publication, DOI Datacite pour une donnée, SWHID pour un code ou logiciel, ORCID pour un contributeur, ROR pour une structure...). Interopérables et lisibles par les machines, ces identifiants facilitent notamment l'échange entre différents systèmes d'information.

Contexte et enjeux

Dans le contexte de transformation numérique de l'enseignement supérieur et de la recherche, la gestion des données, des algorithmes et des codes sources fait l'objet d'une attention croissante, portée notamment par la *Politique des données, des algorithmes et des codes sources* de l'État, qui promeut leur ouverture, leur traçabilité et leur réutilisabilité⁸. Cette dynamique s'inscrit en cohérence avec les plans nationaux pour la science ouverte et les prescriptions européennes (ex. Horizon Europe, EOSC). Par ailleurs, le rapport Gillet sur la simplification administrative dans l'ESR souligne la nécessité de mieux structurer l'écosystème d'information de la recherche, notamment à travers une adoption coordonnée des identifiants persistants, pour améliorer la gestion, le suivi et la valorisation des productions scientifiques, tout en réduisant les charges administratives pesant sur les équipes de recherche et les institutions⁹.

L'adoption généralisée des identifiants persistants génère des bénéfices significatifs pour l'ensemble des acteurs de l'écosystème de la recherche. Elle contribue notamment :

- À renforcer la visibilité, la citabilité et la traçabilité des unités, contributeurs et productions de recherche : les PIDs facilitent la découverte, l'accès et la réutilisation des ressources scientifiques (articles, données, logiciels, ressources éducatives libres, etc.) dans le cadre de nouvelles recherches ;
- À garantir l'intégrité scientifique et à renforcer la confiance dans les résultats produits : en agissant comme points de référence univoques pour l'échange et la consolidation des informations, les PIDs permettent de documenter et de suivre précisément les structures, contributions, les processus et les objets de recherche, consolidant ainsi la fiabilité des résultats scientifiques ;
- À faciliter la réutilisation des objets scientifiques, qu'il s'agisse de données, de publications, de logiciels ou d'autres ressources ;
- À alléger les charges administratives pesant sur les équipes de recherche, les unités et leurs tutelles, grâce à l'automatisation des échanges d'information, la standardisation des données et la limitation des ressaisies multiples ;
- À appuyer l'application des principes FAIR (Findable, Accessible, Interoperable, Reusable), en fournissant des points d'ancrage techniques et organisationnels pour une gestion durable et interopérable des ressources de recherche.

Diversité des identifiants persistants

Malgré leur importance stratégique, la prolifération et l'hétérogénéité des identifiants persistants peuvent en limiter la lisibilité, la pertinence et l'efficacité. À l'échelle internationale, leur multiplication a engendré un écosystème complexe, marqué par la coexistence d'identifiants complémentaires, voire redondants, sans qu'une vision d'ensemble unifiée soit toujours disponible.

Certains identifiants persistants se sont imposés comme des standards largement reconnus au sein de certaines communautés scientifiques et académiques. Toutefois, la diversité des PIDs disponibles conduit

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>

à la coexistence de plusieurs identifiants pour une même entité, en fonction du contexte d'usage et des besoins spécifiques des acteurs. L'écosystème des PIDs se caractérise ainsi par une forte densité, avec des niveaux hétérogènes de reconnaissance, de maturité et d'intégration dans les pratiques.

Dans ce contexte, il apparaît essentiel de disposer d'une compréhension claire des typologies existantes et des caractéristiques propres à chaque type de PID, afin de permettre une utilisation cohérente, rationnelle et efficiente au service de la recherche.

On distingue généralement deux grands types de PID :

- **Les identifiants objet** sont utilisés pour identifier les productions scientifiques. Cela inclut les publications (articles, chapitres de livre, actes de conférence, thèses, posters...), les jeux de données, les logiciels, les ressources éducatives libres (REL), les instruments de recherche, les échantillons, ou les projets ;
- **Les identifiants contributeur** identifient les acteurs de la recherche. Il s'agit principalement des auteurs des publications scientifiques et des organisations ou institutions de recherche (structures de recherche, universités, organismes de recherche) et des financeurs.

Le choix d'un PID dépend avant tout des besoins de l'utilisateur et de la nature de la ressource à identifier. Plusieurs critères peuvent être pris en compte pour évaluer un PID :

- Sa catégorie (objet, contributeur, ou mixte) ;
- Les dispositifs d'allocation et système qui le sous-tendent et assurent sa gestion et sa pérennité (ex : DataCite, Crossref, Handle...) ;
- Son domaine principal d'application (scientifique, culturel, etc.) ;
- L'existence de standards ou de conventions associés (ISO, Convention Climate and Forecast ...) ;
- Sa syntaxe (signifiante ou non), qui peut affecter la compréhension et la pérennité en cas de changement de nom de l'entité ;
- Sa capacité de granularité, c'est-à-dire à identifier différents niveaux d'une ressource ;
- Les métadonnées qui lui sont associées, décrivant l'entité identifiée ;
- Sa résolubilité (i.e. la capacité à retrouver la ressource via l'identifiant) ;
- Son mode d'hébergement (centralisé ou décentralisé) ;
- Son coût (gratuit, payant, variable).

Au-delà de ces critères, la viabilité d'un PID repose fondamentalement sur la robustesse de sa gouvernance. Les organisations allocataires et gestionnaires doivent impérativement disposer d'une politique formalisée qui encadre l'ensemble du cycle de vie des identifiants. Cette politique devrait définir de manière explicite les modalités de création, de gestion, ainsi que les responsabilités relatives à la fiabilité, à la persistance et à la disponibilité des identifiants. Elle doit également préciser les conditions d'accès (incluant un éventuel modèle économique), les garanties de conservation à long terme, ainsi que les mécanismes assurant la continuité et la résilience des services associés.

Principes directeurs et objectifs stratégiques

Principes directeurs d'une politique nationale des PIDs autour des produits de la recherche

Une stratégie nationale en matière d'adoption des identifiants persistants autour des produits de la recherche devrait s'articuler autour des principes directeurs suivants :

- **Interopérabilité et standardisation** : garantir la compatibilité entre les écosystèmes nationaux et internationaux par l'alignement des systèmes nationaux existants ou en création (tels que identifiants et référentiel des structures, Id HAL, Id Ref) avec les standards internationaux reconnus (tels que ROR, ORCID, DataCite). Cette compatibilité repose sur l'adoption de métadonnées normalisées, la mise en place d'API ouvertes, et la définition de référentiels fiables (« points de vérité ») pour chaque type d'entité concernée, mais également un alignement des vocabulaires contrôlés et des référentiels au sein des différentes communautés scientifiques ;
- **Gouvernance collective et structurée** : associer l'ensemble des parties prenantes de l'enseignement supérieur et de la recherche (établissements, organismes, infrastructures, réseaux professionnels) ainsi que les opérateurs concernés (tels que l'ABES, le CCSD...) dans une gouvernance partagée. A cet effet, le Comité orientation identifiants et circulation des données mis en place en 2024 évoluera pour exercer la coordination stratégique des PIDs au niveau national ;
- **Ouverture, indépendance et transparence** : privilégier l'usage de services PIDs opérés par des structures à but non lucratif, disposant d'une gouvernance ouverte reposant sur l'implication de la communauté scientifique, afin de garantir l'indépendance, la soutenabilité et la transparence des services proposés ;
- **Pérennité des infrastructures** : veiller à la durabilité à long terme des systèmes d'identification, en anticipant les évolutions technologiques et institutionnelles susceptibles d'en affecter le fonctionnement, afin de garantir la continuité des services associés aux PIDs ;
- **Approche systémique** : adopter une vision globale couvrant l'ensemble du cycle de vie de la recherche, en intégrant l'identification pérenne des contributeurs, des structures, des projets, ainsi que de toutes les productions scientifiques (publications, données, logiciels, etc.).

Objectifs stratégiques

Articulée avec les travaux en cours initiés en 2024 par le MESR autour de la simplification administrative, la stratégie nationale relative aux identifiants persistants poursuit les objectifs suivants :

- **Favoriser l'adoption de PIDs prioritaires**, en soutenant leur usage généralisé, tout en assurant une articulation cohérente avec les identifiants nationaux existants (IdRef, IdHal...) :
 - de l'identifiant ORCID pour les acteurs de la recherche scientifique (chercheurs, ingénieurs...),
 - des identifiants DOI, Handle, ARK ou SWIDH pour les productions scientifiques (publications, jeux de données, logiciels, etc.),
 - de l'identifiant ROR pour les structures de recherche ;
- **Mettre en place une gouvernance claire, coordonnée et partagée**, associant les acteurs ministériels concernés, les établissements d'enseignement supérieur et de recherche, les infrastructures nationales et les opérateurs. Cette gouvernance doit être étroitement connectée aux dynamiques internationales en matière de PIDs ;
- **Renforcer l'interopérabilité des systèmes d'information**, par l'alignement des identifiants et métadonnées utilisés à l'échelle nationale avec les standards internationaux (ORCID, ROR, DataCite) et par l'intégration fluide de ces identifiants via des interfaces de programmation ouvertes (API) dans les systèmes numériques ;
- **Soutenir la souveraineté et l'indépendance des infrastructures PIDs**, notamment par une participation active à leur gouvernance, afin d'en assurer la pérennité et l'alignement avec les intérêts de la communauté scientifique ;

- **Déployer le référentiel national des structures de recherche (RNeST)**, appelé à devenir le registre national des structures de recherche, en continuité et en évolution du RNSR. Ce registre constituera la source nationale de confiance, garantira une représentation cohérente dans ROR, en l'alimentant automatiquement, et évitera la double saisie de l'information. Le MESR a confié à l'ABES la gouvernance opérationnelle de RNeST ;
- **Accompagner les acteurs du paysage de la recherche**, à travers des actions de formation, de sensibilisation et de soutien, en s'appuyant notamment sur le réseau des centres de compétences de l'écosystème Recherche Data Gouv et les personnels chargés de la sensibilisation aux identifiants dans les établissements.

Axes opérationnels et préconisations

Dans l'écosystème des PIDs, certains sont désormais largement reconnus et utilisés à l'échelle internationale, comme le DOI (pour les publications et données) ou l'ORCID (pour les acteurs de la recherche). D'autres, plus récents, ont émergé pour répondre à de nouveaux besoins, tels que le SWHID pour les logiciels ou le RAiD pour les projets de recherche. Certains PIDs restent encore à consolider, que ce soit en termes d'adoption, de standardisation ou d'interopérabilité globale, comme les identifiants pour les instruments, ou les protocoles. Enfin, des besoins se font jour pour couvrir l'entièreté du cycle de vie d'un projet de recherche, notamment sur les flux financiers, le HPC ou le système de stockage. Cette diversité reflète un écosystème en construction, où la complémentarité et la coordination entre PIDs sont essentielles pour une science réellement ouverte et transparente.

Adopter des identifiants persistants internationaux dans le cadre de l'ESR

Dans le cadre de l'ESR, le Collège des Données de la recherche préconise prioritairement l'usage des PIDs suivants au niveau national :

Identifiants de structures et d'organisations

ROR (Research Organization Registry)

Utilisé pour identifier les structures de recherche et les organisations de recherche, cet identifiant est largement adopté au niveau international. Un référencement exhaustif et homogène des structures françaises dans RNeST garantira leur ajout ou mise à jour automatique dans ROR. En tant que point de confiance national pour ROR, RNeST s'appuiera sur un cadre de référence commun à tous les établissements pour la caractérisation et la désignation des structures de recherche. Ce registre RNeST représentera plus finement l'organisation de la recherche française, pour des besoins nationaux de circulation des données, ce que ne permet pas ROR, tout en préservant l'usage du ROR à l'échelle internationale.

ROR et Crossref ID

Les Crossref Funder ID documentent les bailleurs de fonds dans les métadonnées de publications et projets, les ROR ID identifient les structures de recherche (établissements, unités...) en tant que producteurs ou porteurs de projets. Ces deux systèmes travaillent à un alignement systématique de leurs référentiels.

Identifiants de personnes

ORCID (Open Researcher and Contributor ID)

Parmi les systèmes d'identifiants dédiés aux contributeurs de la recherche, ORCID s'est imposé comme une référence largement adoptée à l'échelle internationale. Déjà bien intégré à l'archive ouverte HAL,

ORCID est géré par une organisation mondiale à but non lucratif, pilotée par un conseil d'administration (*board*) issu de la communauté scientifique au sein duquel la France est représentée, et dont les finances proviennent exclusivement de l'adhésion des membres institutionnels. En France, un consortium national, porté par l'ABES, coordonne les établissements adhérents, leur offre un support de premier niveau et soutient le déploiement d'ORCID¹⁰.

La création d'un identifiant et d'un profil ORCID peut relever de l'initiative individuelle d'un contributeur à la recherche (chercheur, ingénieur, doctorant, technicien, ou toute personne souhaitant disposer d'un identifiant ORCID), ou de l'initiative de l'établissement employeur¹¹. Elle requiert cependant l'autorisation de la personne concernée, conformément à la politique fondatrice d'ORCID qui garantit que chaque individu contrôle les informations figurant sur son profil. Les profils ORCID peuvent être alimentés par la personne elle-même ou par les membres du consortium d'ORCID (établissements, organismes, éditeurs, financeurs...), grâce à l'intégration d'ORCID dans leurs systèmes d'information. Dans tous les cas, les sources sont identifiables, et une coche verte signale les informations certifiées, reconnues comme « Trust Markers » par ORCID. L'utilisation d'un profil ORCID public favorise la visibilité et l'interopérabilité. Chaque profil peut être alimenté par son détenteur et, avec son accord, relié automatiquement au système d'information de son établissement employeur pour y inscrire des informations certifiées. ORCID opère à l'échelle mondiale en conformité avec le RGPD, et place la confiance des utilisateurs parmi ses priorités stratégiques. Ainsi, plus de 98 % des données sont publiques et archivées en France, assurant disponibilité et sécurité des informations.

¹⁰ <https://orcid-france.fr/>

¹¹ Seul l'établissement employeur est habilité à certifier les rattachements des profils ORCID.

Recommandations

Encourager systématiquement les acteurs de la recherche à créer et maintenir un profil public ORCID, en l'intégrant aux procédures et outils institutionnels (dépôt HAL, gestion des publications...). A cet égard, l'utilisation d'ORCID sur le portail appelsprojetsrecherche.fr, qui fédère six partenaires, constitue une illustration concrète et significative de la démarche de simplification.

Encourager systématiquement les établissements employeurs à intégrer ORCID dans leur système d'information, afin d'alimenter automatiquement les profils des contributeurs en informations certifiées¹².

Identifiants de publications

Les publications scientifiques sont principalement décrites par des DOI (Digital Object Identifier), attribués par des agences comme Crossref ou DataCite. Très répandu dans l'édition scientifique, ce standard soutient les citations ouvertes (OpenCitations).

Crossref est centré sur la valorisation et la traçabilité des publications académiques. Son schéma de métadonnées offre une description fine des références, auteurs (avec liens ORCID), affiliations, résumés, licences, financements et liens vers le texte intégral. Il propose par ailleurs des services avancés de suivi des citations, d'interconnexion avec ORCID et avec des plateformes bibliographiques, particulièrement adaptés aux workflows éditoriaux et à la citation classique des articles.

Le schéma de métadonnées de DataCite est plus générique et extensible, conçu pour décrire une grande diversité de ressources scientifiques, en particulier les jeux de données. Il insiste sur les relations entre objets (liens entre DOI), la granularité des types de ressources et la gestion des droits et licences. Il est ainsi bien adapté à la description de données numériques, de matériaux de recherche ou d'objets de support, même s'il reste moins détaillé pour les publications traditionnelles.

Les métadonnées Crossref sont donc orientées vers la traçabilité des publications, alors que celles de DataCite mettent l'accent sur la réutilisabilité et la citabilité des données de recherche

Recommandation

Utiliser de manière systématique des DOI **attribués par Crossref** pour les publications scientifiques.

Identifiants de données de recherche et objets numériques

Plusieurs systèmes d'identifiants persistants sont utilisés pour les données issues des travaux de recherche, en fonction des pratiques disciplinaires, les plus courants étant le DOI, le Handle et l'ARK.

Des données de recherche sont des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche

¹² Voir par exemple l'exemple de l'université de Lorraine : <https://factuel.univ-lorraine.fr/article/un-nouveau-service-orcid-pour-les-chercheurs-et-chercheuses-du-site-lorrain-2/>

Les DOI DataCite

DataCite est un consortium international à but non lucratif, permettant d'attribuer des DOI aux données de la recherche, mais également à d'autres types de productions scientifiques. En France, le consortium national est porté par l'INIST-CNRS, qui attribue les DOI pour le compte des établissements et infrastructures de recherche¹³.

Le schéma de métadonnées DataCite évolue régulièrement pour répondre aux besoins de la communauté scientifique (dernière version 4.6 publiée en décembre 2024). Ce schéma repose sur une architecture modulaire et hiérarchique, fondée sur XML, avec une version JSON également disponible. Il comprend des éléments obligatoires (titre, créateur, DOI, éditeur, date de publication) et facultatifs (contributeurs, affiliations, financement, relations inter-objets, etc.) permettant un niveau de granularité adaptable. L'utilisation de vocabulaires contrôlés pour certains champs garantit une cohérence sémantique et un modèle relationnel permet de documenter les liens entre objets de la recherche.

Les IGSN ID

Un IGSN ID est l'identifiant spécifiquement conçu pour les échantillons physiques. Il permet d'assurer la traçabilité et l'interopérabilité entre les échantillons et les différentes entités de l'écosystème de la recherche (projets, instruments, publications, données, personnes, institutions).

En septembre 2021, un partenariat stratégique a été conclu entre IGSN (International Generic Sample Number) et DataCite, confiant à ce dernier la responsabilité de fournir les services d'enregistrement des identifiants IGSN ID ainsi que l'infrastructure technologique associée. Les agences d'allocation IGSN rejoignent désormais DataCite, soit en tant que membres directs, soit via des consortia. Par ailleurs, les identifiants IGSN précédemment attribués sous forme de Handles sont désormais enregistrés sous forme de DOI. Sur le plan technique, ce rapprochement s'appuie sur une adaptation spécifique du schéma de métadonnées DataCite, modifié pour l'occasion en introduisant de nouvelles métadonnées nécessaires à la description des échantillons physiques, telles que la géolocalisation, les méthodes de prélèvement, ou encore les caractéristiques des matériaux.

Les Handle

Conçu dans les années 1990, le Handle System constitue l'un des premiers dispositifs techniques développés pour l'attribution et la gestion d'identifiants persistants. Depuis 2014, la gouvernance du Handle System est assurée par la DONA Foundation, organisation internationale à but non lucratif. Le Handle System se distingue principalement de DataCite par une infrastructure décentralisée (chaque institution peut héberger son propre serveur Handle), permettant la gestion locale des identifiants, un logiciel de gestion en open source et l'absence de métadonnées obligatoires.

Les Handle sont particulièrement utilisés dans des contextes où l'indépendance institutionnelle et la flexibilité sont recherchées, comme dans les dépôts d'archives, les bibliothèques numériques ou certaines infrastructures de recherche.

Toutefois, la faible standardisation des métadonnées et l'absence de gouvernance communautaire structurée constituent des freins à leur adoption dans certains cas d'usage internationaux.

¹³ <https://datacite-france.fr/>

DOI et Handle : similitudes et différences

Les DOI s'appuient techniquement sur l'infrastructure du Handle System pour leur fonctionnement.

Toutefois, les DOI attribués par DataCite reposent sur la technologie du Handle System, mais à la différence des Handle simples, les DOI sont associés à des métadonnées normalisées et obligatoires, favorisant leur interopérabilité, leur visibilité et leur réutilisation, et sont gérés dans un cadre de gouvernance communautaire, ce qui garantit transparence, ouverture et adaptation aux besoins de la recherche et aux principes FAIR.

Les ARK (Archival Resource Key)

Créé en 2001 par la California Digital Library et reconnu depuis 2019 comme schéma d'URI (Uniform Resource Identifier), le système ARK a été conçu pour offrir un identifiant stable à toute ressource numérique ou physique disposant d'une représentation numérique, quelle que soit sa nature (document, notice bibliographique, corpus, entité du web sémantique, etc.).

Tout comme le Handle System, le système ARK repose sur une infrastructure ouverte et décentralisée, permettant à toute institution d'attribuer librement ses propres identifiants. Largement utilisé par les institutions culturelles et patrimoniales, il offre une souplesse d'implémentation, sans obligation de dépôt centralisé ni de métadonnées standardisées. La persistance et l'accessibilité à long terme des ARK reposent exclusivement sur l'engagement des institutions émettrices.

Plusieurs institutions françaises sont autorités nommantes (NAAN), notamment la Bibliothèque nationale de France, qui utilise les ARK pour ses ressources numériques et bibliographiques, et héberge depuis 2008 une réplique du registre international.

Recommandation

En raison de leur cadre de gouvernance scientifique, et de leur schéma de métadonnées standardisé, il est recommandé d'utiliser des DOI attribués par DataCite pour l'identification des jeux de données produits dans l'ESR. La plupart des entrepôts de confiance (Recherche Data Gov, Data Terra...) permettent l'attribution de DOI.

Identifiants de logiciels

Plusieurs identifiants persistants sont utilisés pour les logiciels, principalement :

- Les DOI (ou les HAL ID), attribué à une version figée d'un logiciel, souvent lié à une publication ;
- Les SWHID (Software Heritage ID) : identifiant cryptographiquement vérifiable d'un artefact de code (commit, fichier, répertoire, etc.) ;
- Le RRID, fréquemment utilisé dans les sciences de la vie et multi-usage (logiciels, bases de données, matériels...).

Le Software Heritage ID (SWHID) est calculé par Software Heritage, afin de fournir un identifiant intrinsèque, unique et propre au contenu du code lui-même, résistant aux changements de plateforme (forges logicielles, ex. GitHub, GitLab) et d'emplacement. Basé sur une empreinte cryptographique (SHA-1), dérivée d'une structure Merkle DAG, il permet, grâce à une granularité fine, d'identifier un dépôt entier, un commit, un répertoire, un fichier, voire une ligne de code. Recommandé par plusieurs agences de financement et institutions, il rend donc possible la citation logicielle dans les publications scientifiques.

Depuis avril 2025, le SWHID est publié en tant que norme internationale ISO/IEC 18670.

Cas d'usage et complémentarité entre DOI et SWHID pour les identifiants logiciels

Le DOI est une citation d'une version logicielle publiée, décrite grâce au schéma de métadonnées DataCite. Il cible principalement le contexte de la publication scientifique.

Le SWHID est une identification d'un artefact de code spécifique. Sa pérennité est assurée par le hachage et Software Heritage. Par défaut, sa portée technique ne prévoit pas de renseigner des métadonnées aussi riches que DataCite. Le SWHID cible donc le contenu technique du logiciel.

Recommandation

Il est recommandé d'utiliser les identifiants Software Heritage ID (SWHID) pour l'identification des logiciels produits dans l'ESR ou, dans le cadre d'une publication, des DOI (ou des HAL-ID) associés au SWHID. Offrant une granularité fine, une traçabilité assurée et une pérennité garantie par un archivage à long terme, les SWHID constituent une solution robuste et interopérable pour valoriser, référencer et préserver le code source, en cohérence avec les principes de la science ouverte.

Identifiants de projets

Le RAiD (Research Activity Identifier) est un PID destiné à référencer de manière stable et univoque une activité de recherche dans son ensemble (projet, programme, initiative...). Développé initialement par l'Australian Research Data Commons (ARDC), le système est aujourd'hui encadré par le standard ISO 23527:2022. Il s'intègre progressivement dans l'écosystème mondial de la science ouverte, en tant qu'identifiant structurant des activités scientifiques.

Le RAiD vise à identifier de manière pérenne un projet de recherche dans sa globalité et relier tous les objets numériques et acteurs impliqués (personnes, institutions, résultats, etc.). Le RAiD désigne donc une activité dans son ensemble (projet, sous-projet), sans se limiter aux résultats, reliant des identifiants tiers (ORCID, ROR, DOI, SWHID...). Il contient des informations sur le titre, le périmètre, les parties prenantes, les dates, les productions associées.

Sa gouvernance ouverte est en cours de développement, en collaboration avec la communauté scientifique internationale.

Le RAiD n'a pas vocation à remplacer les autres identifiants, mais à les relier autour d'une même activité en agissant comme identifiant transversal. Grâce à un partenariat avec DataCite, RAiD est progressivement intégré dans l'EOSC (EOSC RAiD), ce qui garantit son interopérabilité avec les dépôts de données, publications et infrastructures.

Complémentaire des PIDs existants, il identifie et valorise les projets financés, et s'appuie sur l'interopérabilité¹⁴ des identifiants pour permettre des remontées automatiques d'information entre systèmes (par exemple, l'actualisation des profils ORCID à partir des publications). Il contribue ainsi à automatiser le *reporting* et le suivi des activités de recherche.

¹⁴ L'interopérabilité se définit comme la capacité d'un ensemble de systèmes informatiques, applications ou composants d'un même type à communiquer, échanger ou utiliser des données ou informations sans ambiguïté.

Recommandation

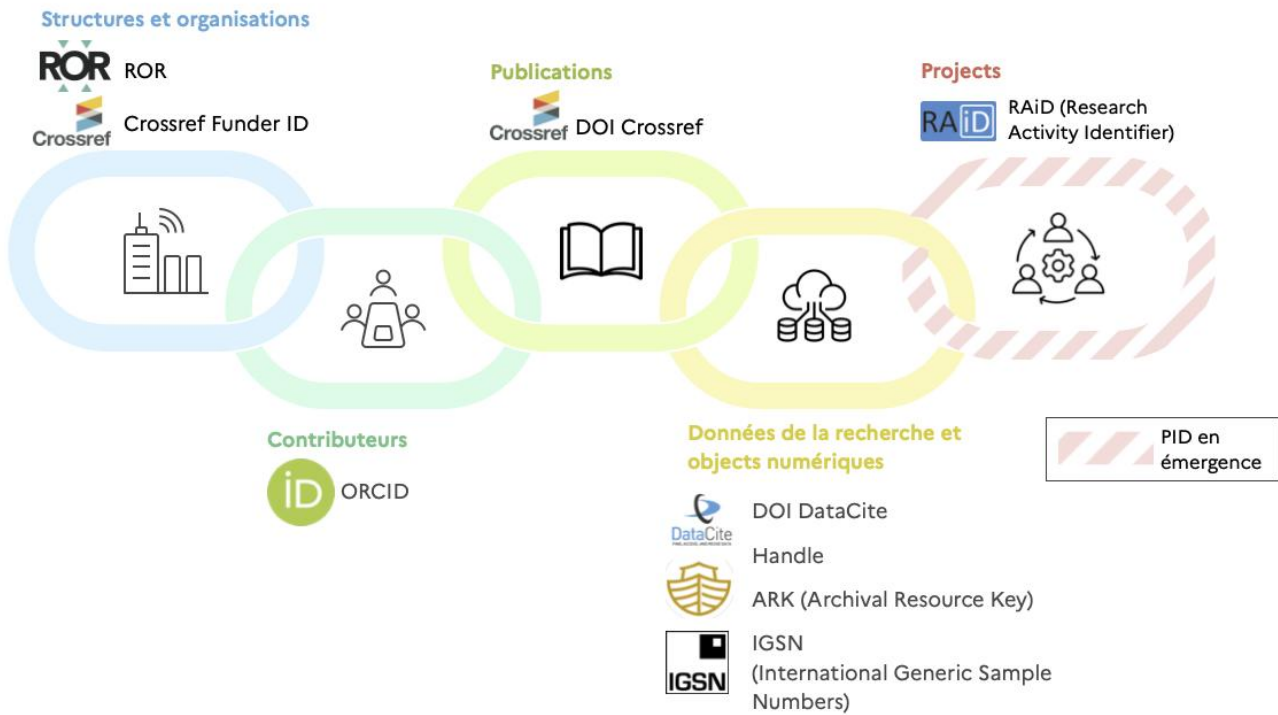
Instruire le déploiement des RAiD au niveau national en concertation avec les agences de financement et les principales parties prenantes de l'écosystème de la recherche, afin de garantir leur intégration cohérente dans les dispositifs de financement, de suivi, de pilotage et d'évaluation des projets scientifiques.

Identification d'objets spécifiques

Le schéma PIDINST, développé par la Research Data Alliance depuis 2020, fournit un modèle de métadonnées standardisé permettant de décrire les instruments scientifiques et de leur associer des identifiants persistants existants, tels que les DOI ou Handles. Son objectif principal est de faciliter la traçabilité, la réutilisabilité et la transparence des données produites par ces instruments, indépendamment de leur lieu d'usage ou de leur propriétaire. Il permet ainsi d'enrichir les métadonnées des jeux de données avec des références précises aux instruments utilisés (type, fabricant, modèle, propriétaire, localisation, périodes d'usage, liens avec les données produites...). Initialement conçu comme un identifiant, le terme PIDINST désigne désormais *Persistent Identification of Instruments*, marquant une évolution sémantique vers une approche plus descriptive de la relation entre les instruments et leur identification. Il se propose de fournir un cadre permettant d'utiliser les PIDs existants de manière cohérente et interopérable pour décrire les instruments scientifiques (comme les capteurs, microscopes, spectromètres, télescopes, etc.). Plusieurs infrastructures européennes et internationales ont commencé à l'adopter, et son application tend à s'étendre progressivement à l'identification des infrastructures de recherche.

Recommandation

Conduire une réflexion nationale en lien avec les communautés disciplinaires concernées afin d'évaluer la pertinence et les modalités d'adoption d'un identifiant adapté aux instruments, équipements et infrastructures de recherche



Principaux PIDs internationaux utilisés dans l'ESR

Promouvoir l'alignement et l'interopérabilité pour permettre la circulation des données

Interopérabilité des identifiants s'avère un élément clé pour assurer une circulation efficace des données, en alignant les identifiants locaux, nationaux et internationaux. L'hétérogénéité des systèmes d'identifiants forme un obstacle significatif à leur adoption généralisée et à leur efficacité, en particulier en ce qui concerne la circulation fluide des données et la simplification des processus administratifs.

Pour garantir leur intégration optimale dans les systèmes d'information, les identifiants persistants et les métadonnées qui leur sont associées doivent être rendus disponibles sous des formats structurés, normalisés et réutilisables. Cela implique le recours systématique à des interfaces de programmation ouvertes (API) et à des métadonnées conformes aux bonnes pratiques internationales.

L'adoption de standards de métadonnées partagés, ainsi que l'élaboration d'un vocabulaire commun, sont des conditions indispensables à l'interopérabilité des systèmes nationaux avec les infrastructures et référentiels internationaux. Cette convergence est essentielle pour assurer la cohérence, la fiabilité et la pérennité des informations au sein de l'écosystème numérique de la recherche.

Recommandations

Déployer des API ouvertes et fournir des métadonnées structurées afin de favoriser l'interopérabilité, la réutilisation et l'automatisation des échanges d'informations.

Aligner les systèmes nationaux sur les schémas de métadonnées reconnus à l'échelle internationale, afin de garantir une intégration cohérente dans l'écosystème mondial des identifiants persistants.

Gouvernance et pilotage

Dans une perspective de pérennité et d'adéquation avec les besoins de la recherche et de la science ouverte, une gouvernance des fournisseurs d'identifiants persistants reposant sur l'implication de la communauté scientifique est à privilégier. Le recours à des organisations à but non lucratif constitue un gage d'ouverture, d'indépendance et de neutralité. Il est également essentiel d'assurer une transparence complète, notamment sur les modèles économiques des services PID, la documentation des données sous-jacentes et les processus opérationnels.

L'analyse des risques liés aux identifiants persistants, qu'ils soient sociaux, politiques, économiques ou technologiques, devrait constituer un volet essentiel de la stratégie nationale. Le risque d'interruption de service, notamment en cas de restructuration organisationnelle ou d'acquisition commerciale des fournisseurs de PID, représente un enjeu politique majeur. La question de la souveraineté numérique est particulièrement sensible, en particulier lorsqu'il s'agit du transfert de données personnelles vers des juridictions étrangères. Il est impératif d'anticiper ces risques, de les documenter clairement et d'accompagner les acteurs concernés par des recommandations juridiques appropriées.

Par ailleurs, la stratégie nationale relative aux PIDs doit être fondée sur une gouvernance concertée, portée par les parties prenantes. L'identification et la mobilisation des acteurs clés à l'échelle nationale (financeurs, organismes de recherche, infrastructures) sont indispensables pour définir collectivement les modalités de mise en œuvre. La création d'un organe de gouvernance dédié, tel qu'un Conseil consultatif PID, s'inscrit parmi les bonnes pratiques identifiées à l'international. Cette démarche doit s'appuyer sur les enseignements tirés d'expériences étrangères, en particulier via la participation à des initiatives de coordination telles que le groupe de travail de la RDA sur les stratégies PID nationales.

Enfin, la réussite de cette stratégie repose sur une communication claire et ciblée des bénéfices concrets des PIDs à destination des équipes de recherche. Les établissements ont un rôle central à jouer dans l'accompagnement et la sensibilisation à l'usage régulier des identifiants persistants. Il convient de mettre en place des dispositifs de communication complémentaires — à la fois institutionnels (*top-down*) et participatifs (*bottom-up*). La stratégie devra veiller à favoriser les passerelles entre les communautés, les disciplines, en tenant compte des spécificités disciplinaires.

Recommandations

Mener une analyse des risques liés au déploiement et à la dépendance vis-à-vis des infrastructures d'identifiants persistants.

Mettre en place une gouvernance nationale dédiée, concertée et représentative des parties prenantes.

Soutenir la stratégie nationale des PIDs par un financement pérenne et structurant.

Mettre en place des actions de communication, de formation et de sensibilisation adaptés aux spécificités des communautés disciplinaires, afin d'assurer une appropriation large et durable des identifiants dans l'ensemble de l'ESR.

Conclusion

L'adoption généralisée et le déploiement coordonné des identifiants uniques persistants constituent une condition essentielle pour concrétiser les ambitions de la science ouverte en France. Les PIDs forment l'ossature d'une infrastructure numérique de la recherche à la fois efficiente, transparente et interopérable. Ils permettent non seulement d'accroître la visibilité et la citabilité des productions scientifiques françaises, mais aussi de renforcer la traçabilité, l'intégrité et la réutilisabilité des données, logiciels, publications, projets, instruments et contributeurs. Au-delà des équipes de recherche, les PIDs se révèlent également précieux pour l'administration et le pilotage de la recherche : en facilitant l'automatisation des processus de gestion et de *reporting*, ils permettent de réduire les charges administratives pour les équipes et les institutions.

Ces enjeux soulignent la nécessité d'établir des recommandations claires, destinées à structurer le paysage de l'information scientifique autour des PIDs. Sur la base de ces orientations, il est proposé de :

- Adopter des identifiants persistants internationaux, reconnus au sein des communautés de recherche ;
- Coordonner les efforts nationaux autour de référentiels communs ;
- Assurer l'alignement des identifiants locaux avec les standards internationaux (ORCID, DOI, ROR, RAiD, SWHID, etc.) ;
- Garantir l'indépendance, la pérennité et la gouvernance des identifiants et registres utilisés ;
- Former et accompagner les communautés scientifiques et les établissements ;
- Valoriser les usages et démontrer les bénéfices concrets (pilotage, évaluation, valorisation).

Plusieurs identifiants uniques persistants sont aujourd'hui arrivés à maturité et font l'objet d'une adoption large à l'échelle internationale, à l'image du DOI pour les publications et jeux de données, de l'ORCID pour les acteurs de la recherche et du ROR pour les institutions. D'autres identifiants sont en cours de structuration, tels que le RAiD pour les activités de recherche. Certains restent toutefois à consolider, en particulier en ce qui concerne leur gouvernance, leur reconnaissance internationale ou leur interopérabilité avec les systèmes existants. Au-delà des identifiants déjà bien établis, des besoins spécifiques persistent dans certains secteurs ou communautés disciplinaires, où des référentiels communs doivent encore être identifiés ou développés. Il est essentiel que l'édification de l'écosystème des PIDs s'appuie non seulement sur un socle commun partagé à l'échelle nationale et internationale, mais aussi sur une prise en compte fine des pratiques, des usages et des réalités propres à chaque domaine scientifique. Cette approche concertée et différenciée est indispensable pour assurer la pertinence, l'appropriation et l'efficacité des PIDs dans l'ensemble de l'écosystème de la recherche.

Synthèse des recommandations

- Encourager systématiquement les acteurs de la recherche à créer et maintenir un profil public ORCID, en l'intégrant aux procédures et outils institutionnels (dépôt HAL, gestion des publications...) ;
- Encourager systématiquement les établissements employeurs à intégrer ORCID dans leur système d'information, afin d'alimenter automatiquement les profils des contributeurs en informations certifiées ;
- Utiliser de manière systématique des DOI attribués par Crossref pour les publications scientifiques ;
- En raison de leur cadre de gouvernance scientifique, et de leur schéma de métadonnées standardisé, il est recommandé d'utiliser des DOI attribués par DataCite pour l'identification des jeux de données produits dans l'ESR. La plupart des entrepôts de confiance (Recherche Data Gouv, Data Terra...) permettent l'attribution de DOI;
- Il est recommandé d'utiliser les identifiants Software Heritage ID (SWHID) pour l'identification des logiciels produits dans l'ESR ou, dans le cadre d'une publication, des DOI (ou des HAL-ID) associés au SWHID ;
- Instruire le déploiement des RAiD au niveau national en concertation avec les agences de financement et les principales parties prenantes de l'écosystème de la recherche, afin de garantir leur intégration cohérente dans les dispositifs de financement, de suivi, de pilotage et d'évaluation des projets scientifiques ;
- Conduire une réflexion nationale en lien avec les communautés disciplinaires concernées afin d'évaluer la pertinence et les modalités d'adoption d'un identifiant adapté aux instruments, équipements et infrastructures de recherche ;
- Déployer des API ouvertes et fournir des métadonnées structurées afin de favoriser l'interopérabilité, la réutilisation et l'automatisation des échanges d'informations ;
- Aligner les systèmes nationaux sur les schémas de métadonnées reconnus à l'échelle internationale, afin de garantir une intégration cohérente dans l'écosystème mondial des identifiants persistants ;
- Mener une analyse des risques liés au déploiement et à la dépendance vis-à-vis des infrastructures d'identifiants persistants ;
- Mettre en place une gouvernance nationale dédiée, concertée et représentative des parties prenantes ;
- Soutenir la stratégie nationale des PIDs par un financement pérenne et structurant ;
- Mettre en place des actions de communication, de formation et de sensibilisation adaptés aux spécificités des communautés disciplinaires, afin d'assurer une appropriation large et durable des identifiants dans l'ensemble de l'ESR.

Table des abréviations

ABES	: Agence bibliographique de l'enseignement supérieur
API	: Application Programming Interface
ARK	: Archival resource Key
CCSD	: Centre pour la communication scientifique directe
DOI	: Digital Object Identifier
EOSC	: European Open Science Cloud
ESR	: Enseignement supérieur et recherche
FAIR (principes)	: Findable, Accessible, Interoperable, Reusable
HAL	: Hyper Articles en Ligne
HPC	: High Performance Computing
IGSN	: International Generic Sample Numbers
ISO	: International Organization for Standardization
MESR	: Ministère chargé de l'enseignement supérieur et de la recherche
ORCID	: Open Researcher and Contributor ID
PID	: Persistent Identifier
PIDINST	: Persistent Identification Instrument
RaiD	: Research Activity Identifier
RDA	: Research Data Alliance
REL	: Ressources éducatives libres
RNeST	: Registre national des structures de la recherche
RNSR	: Répertoire national des structures de la recherche
ROR	: Research Organization Registry
SWHID	: Software Heritage ID

Bibliographie indicative

-
- 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). Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems. Zenodo. <https://doi.org/10.5281/zenodo.7100578>
- Brown, J., Jones, P., Meadows, A., & Murphy, F.** (2022). Revised cost-benefit analysis for the UK PID Support Network (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7356219>
- 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>
- 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]
- Permalien 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>