

Vers une stratégie nationale des identifiants persistants : coordonner, renforcer l'interopérabilité et accroître la visibilité de la recherche

Véronique STOLL
veronique.stoll@recherche.gouv.fr



Contexte national : la politique des données

2016 : Ouverture des données publiques - Loi pour une république numérique

Passage d'une logique de la demande d'un accès à une logique de l'offre de données publiques

2020 : AMDAC au MESR

2021 : Une politique publique de la donnée déclinée dans chaque ministère

2022 : Le partage d'informations entre administration devient obligatoire : principe *dites-le nous une fois* (loi 3DS)

2024- : Au sein du MESR, mise en place des instances de gouvernance



Contexte international : des bénéfices mesurés

Des stratégies d'adoption et d'usages actives dans certains pays (Australie, Canada, Royaume-Uni, République tchèque...)

Bénéfices en termes de gestion et de circulation des données sur la recherche et l'innovation

- Garantir l'identification, et la traçabilité des productions, des activités et des acteurs tout au long du cycle de recherche,
- Renforcer l'intégrité de la recherche, la confiance dans les résultats,
- Eviter de multiples saisies de mêmes informations,
- Développer des services à valeur ajoutée (CV complet),
- Cartographier la capacité de recherche et mesurer son impact,
- ...



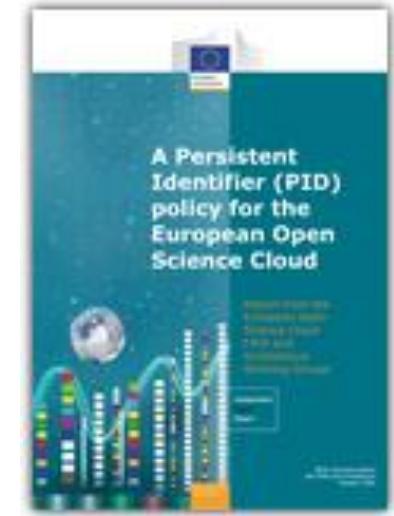
Australie : coût de la ressaisie des métadonnées relatives aux projets, publications, subventions

38 000 jours/an

24M\$

(pour 109 000 chercheurs)

<https://doi.org/10.5281/zenodo.7100578>



RESEARCH DATA ALLIANCE

Les PIDs, pivots de la politique française sur l'ouverture des données de la recherche

Innovation : ouvrir les données

- pour favoriser leur réutilisation par les chercheurs, les enseignants, les citoyens et par les entreprises
- pour qu'elles deviennent créatrices de valeur scientifique et économique

Confiance : pour plus de transparence de l'action publique

Simplification : faire de la circulation des données un outil pour simplifier et rendre plus efficents les actions et les processus administratifs (« Dites-le nous une fois ! »)

Les PIDs, pivots de la politique française sur l'ouverture des données de la recherche

Un **identifiant persistant** : référence durable à une entité telle qu'un document, un fichier ou tout autre élément. Il continue d'exister aussi longtemps que l'infrastructure qui le maintient existe

Unique : un identifiant ne doit identifier qu'une seule entité. Une entité ne doit avoir qu'un seul identifiant d'un type particulier

Machine-actionable : un identifiant doit pouvoir être utilisé pour récupérer des informations sur l'entité à laquelle il fait référence

Associé à un **schéma de métadonnées**

Les PIDs, pivots de la politique française sur l'ouverture des données de la recherche

Chaîne de caractères alphanumériques non signifiante reliée à des registres d'information

Métadonnées permettant de décrire l'objet associé au PID

Données interopérables, lisibles par machine

Méthode universelle pour identifier & connecter

- » de manière unique et sans ambiguïté
- » dans la durée
- » les données de et sur la recherche

**Composante essentielle de
l'écosystème mondial de la recherche**

Éléments pivots pour la circulation des données



Bénéfices de l'usage des PIDs dans la recherche

Améliorer la visibilité et la traçabilité de la recherche

Les PIDs facilitent la découverte des produits de la recherche

Bénéfices de l'usage des PIDs dans la recherche

Améliorer la visibilité et la traçabilité de la recherche

Les PIDs facilitent la découverte des produits de la recherche

Faciliter la réutilisation des produits de la recherche

Les PIDs facilitent la réutilisation des produits de la recherche

Bénéfices de l'usage des PIDs dans la recherche

Améliorer la visibilité et la traçabilité de la recherche

Les PIDs facilitent la découverte des produits de la recherche

Faciliter la réutilisation des produits de la recherche

Les PIDs facilitent la réutilisation des produits de la recherche

Favoriser l'intégrité scientifique

Les PIDs agissent comme des points de référence uniques pour les informations de recherche

Bénéfices de l'usage des PIDs dans la recherche

Améliorer la visibilité et la traçabilité de la recherche

Les PIDs facilitent la découverte des produits de la recherche

Faciliter la réutilisation des produits de la recherche

Les PIDs facilitent la réutilisation des produits de la recherche

Favoriser l'intégrité scientifique

Les PIDs agissent comme des points de référence uniques pour les informations de recherche

Soutien des principes FAIR

Les PIDs fournissent des points d'ancrage techniques et organisationnels

Bénéfices de l'usage des PIDs dans la recherche

Améliorer la visibilité et la traçabilité de la recherche

Les PIDs facilitent la découverte des produits de la recherche

Faciliter la réutilisation des produits de la recherche

Les PIDs facilitent la réutilisation des produits de la recherche

Favoriser l'intégrité scientifique

Les PIDs agissent comme des points de référence uniques pour les informations de recherche

Soutien des principes FAIR

Les PIDs fournissent des points d'ancrage techniques et organisationnels

Réduire les charges administratives

Les PIDs automatisent les échanges d'informations et standardisent les (méta)données

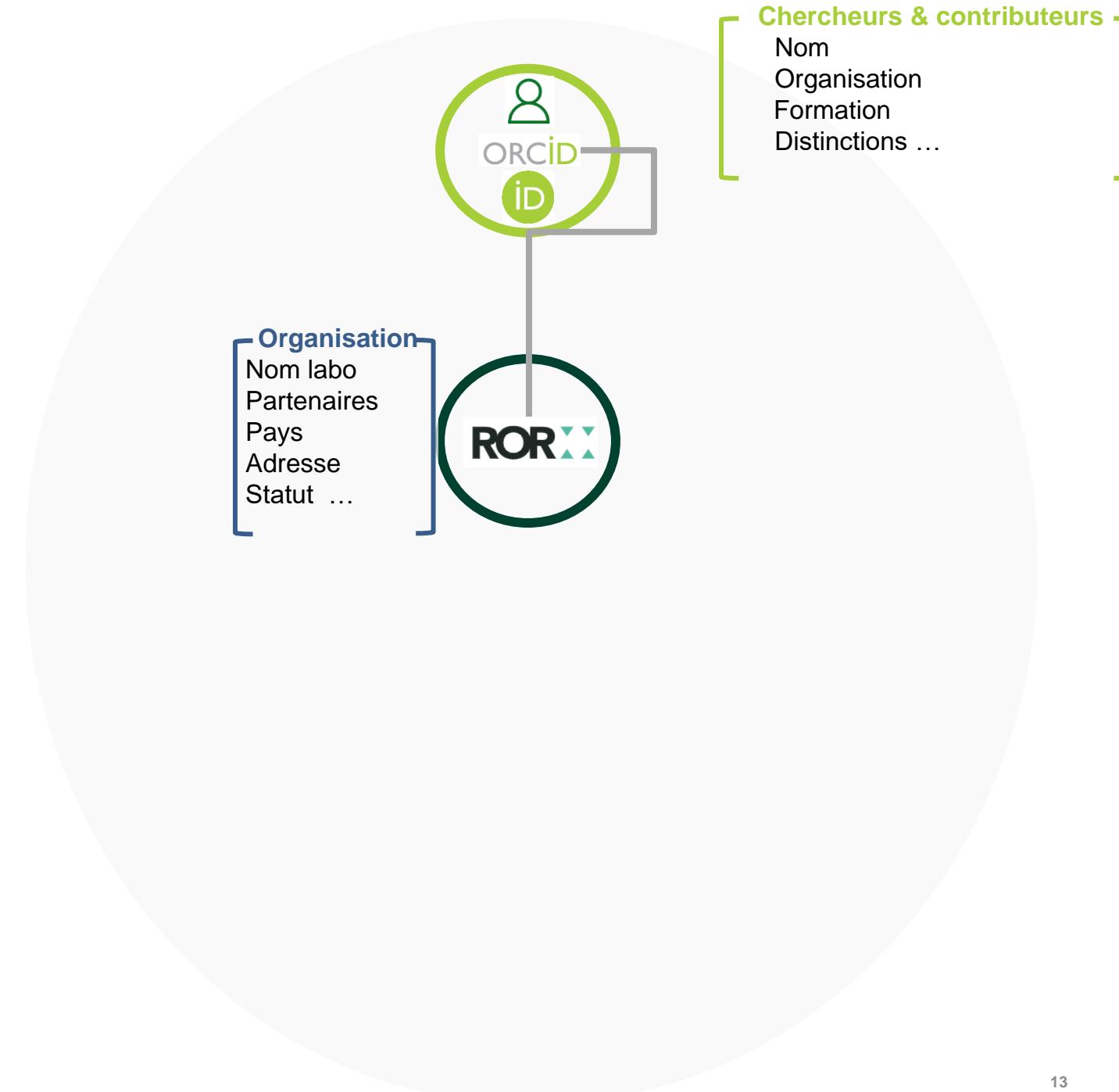
Ecosystème des PIDs



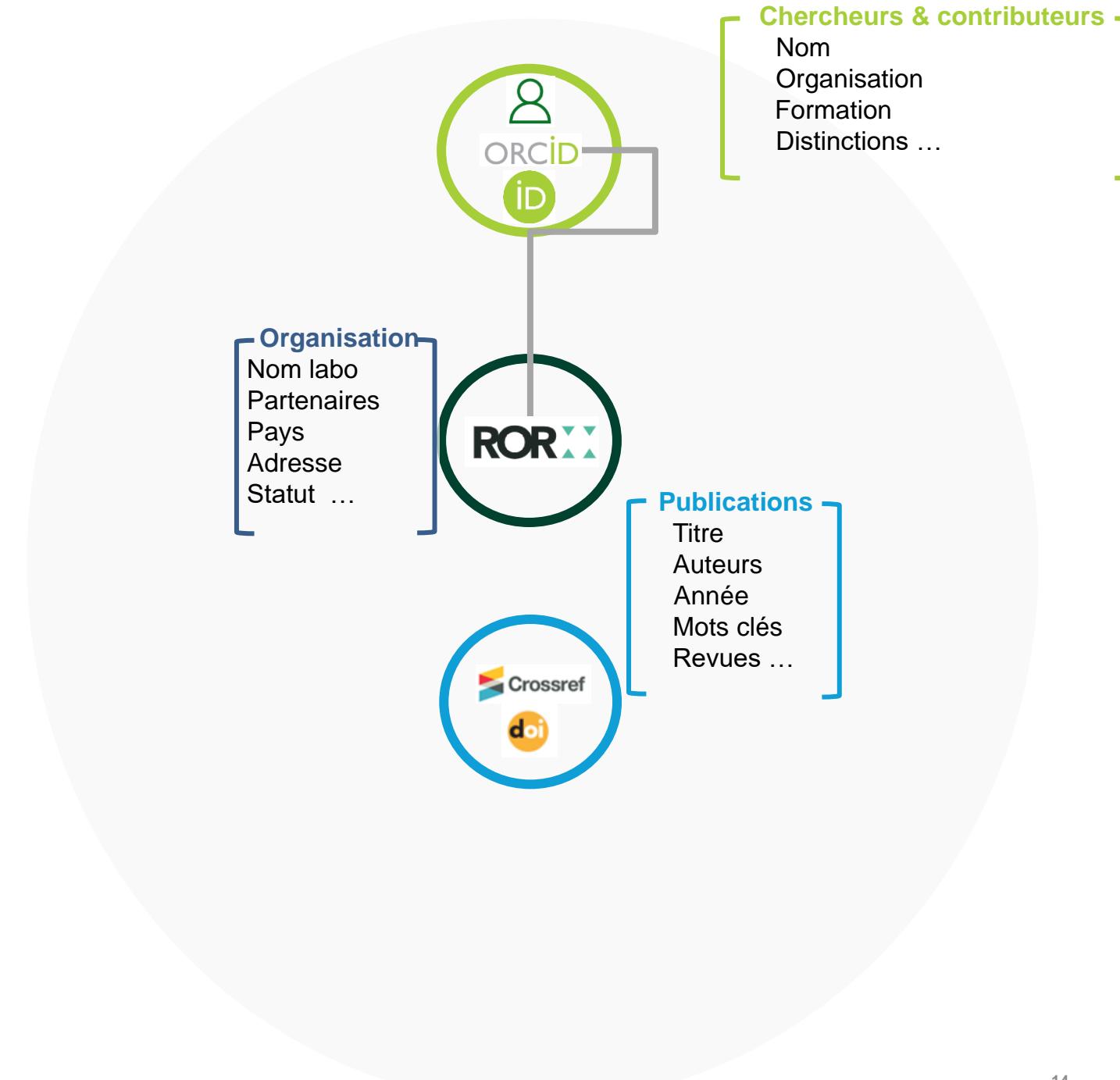
Chercheurs & contributeurs

Nom
Organisation
Formation
Distinctions ...

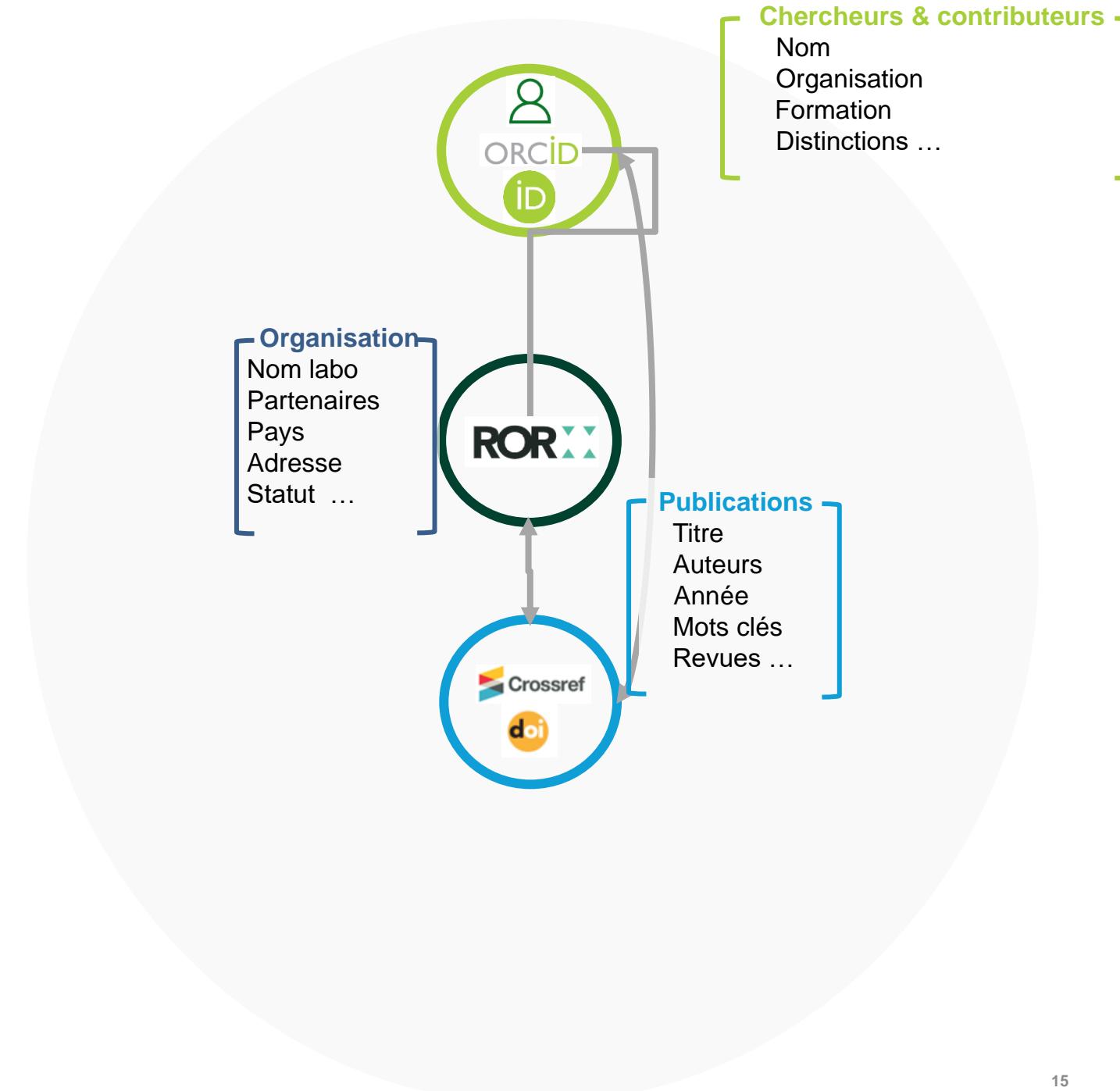
Ecosystème des PIDs



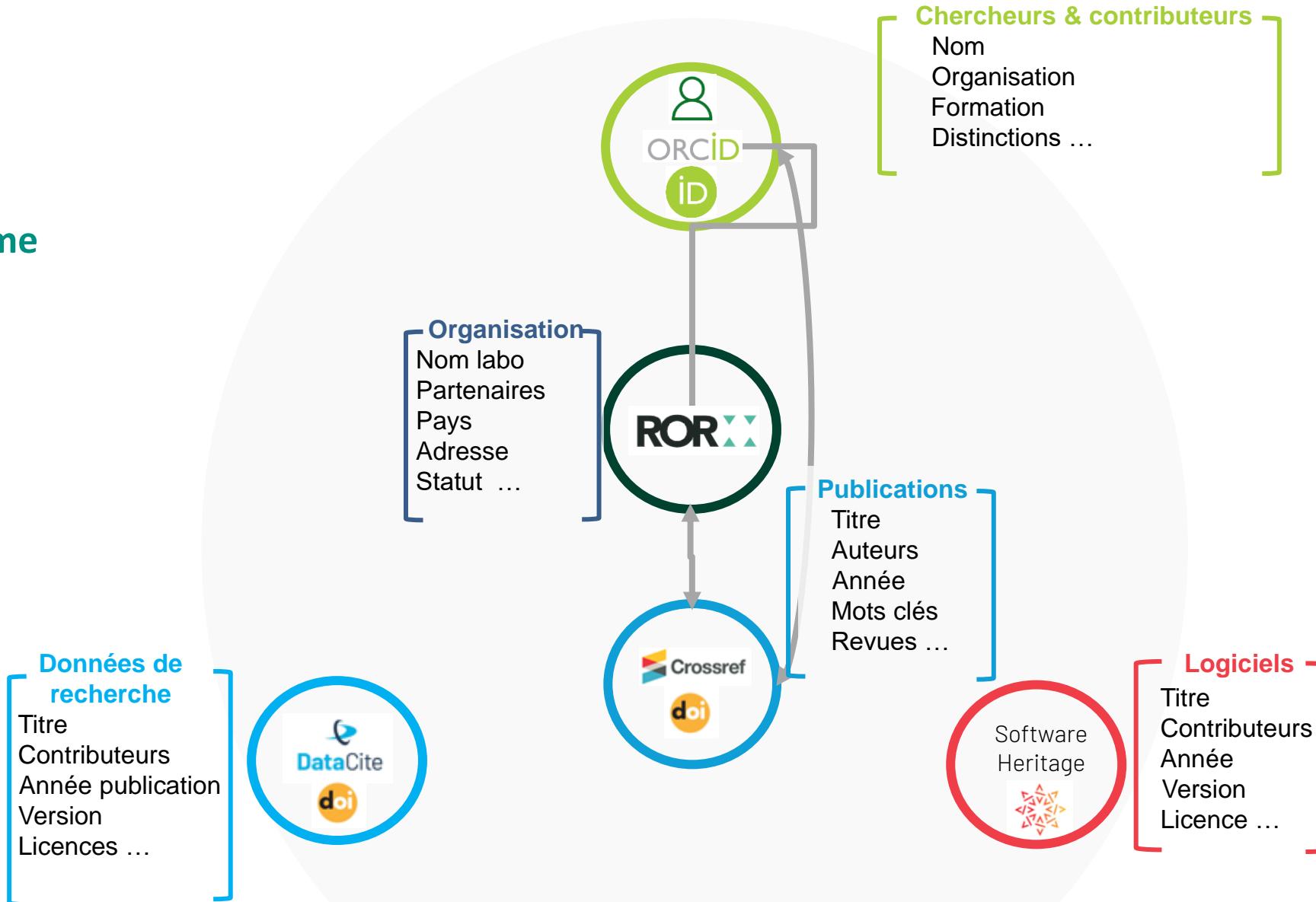
Ecosystème des PIDs



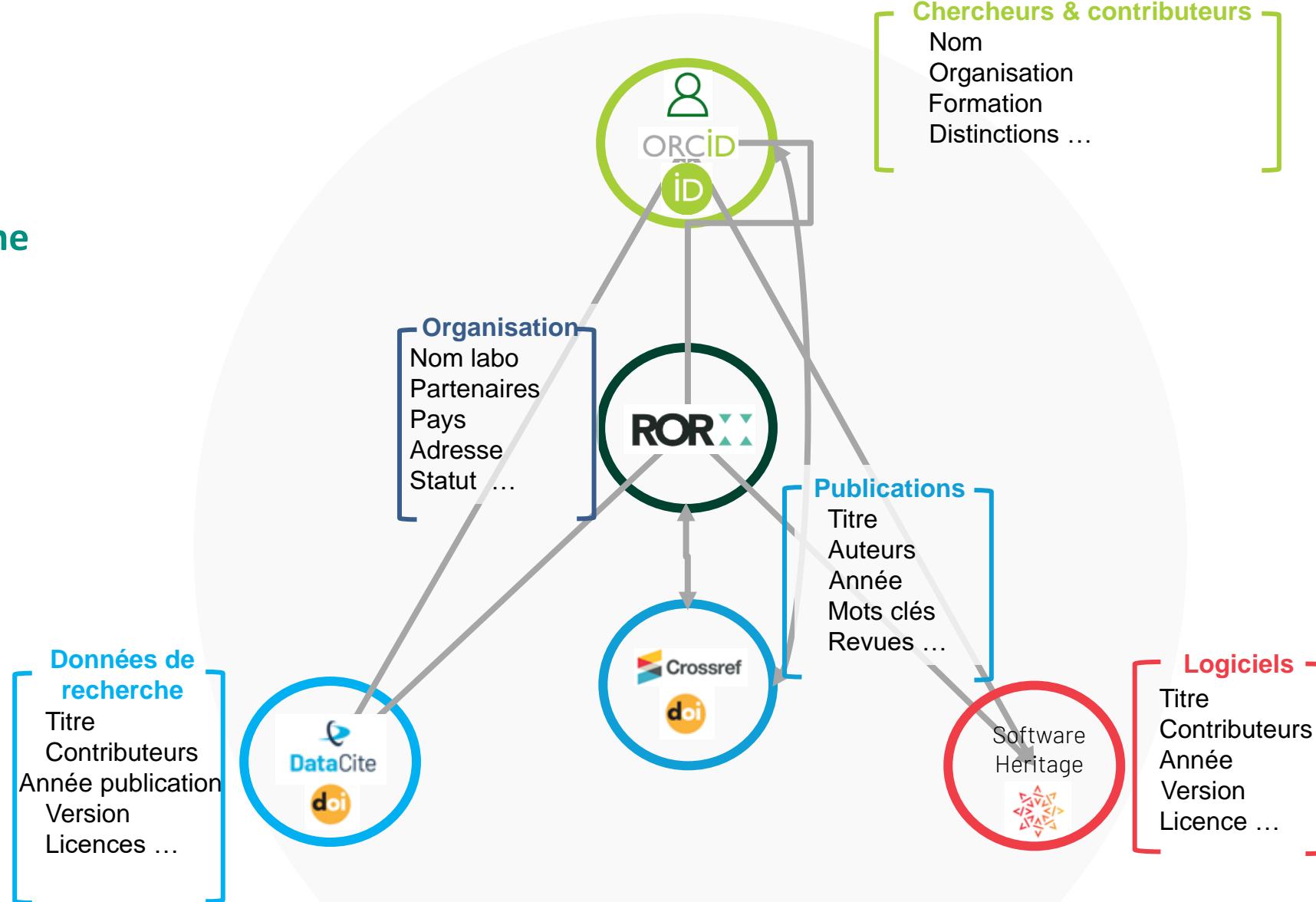
Ecosystème des PIDs



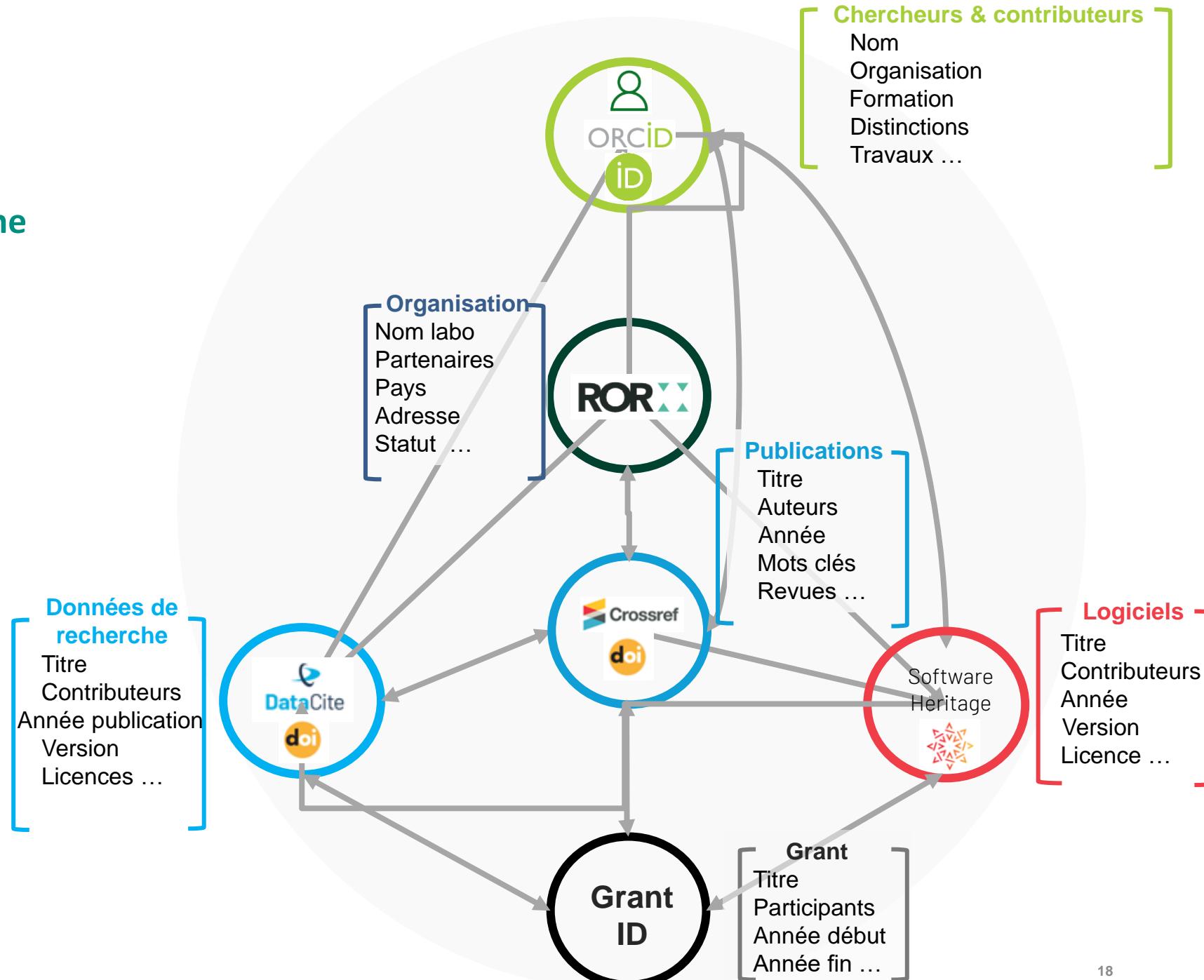
Ecosystème des PIDs



Ecosystème des PIDs



Ecosystème des PIDs

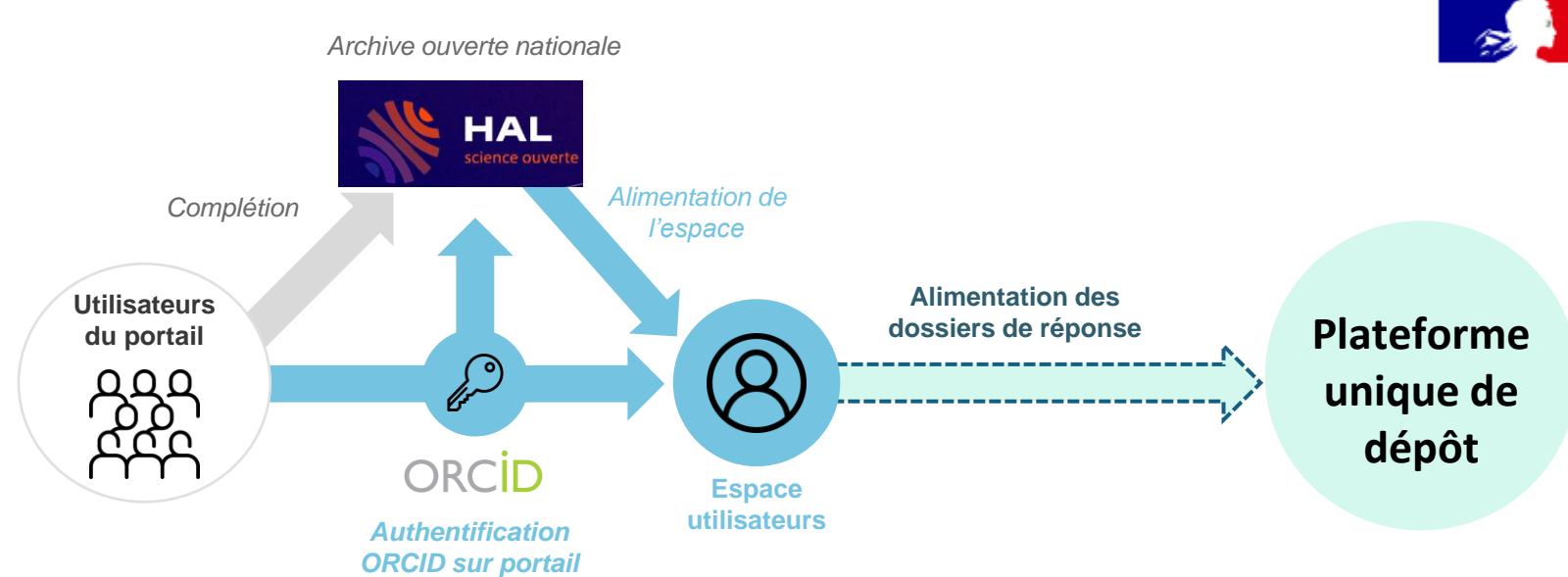


Exemple concret de simplification autour des PIDs

Portail des appels à projets :

faciliter les candidatures

remplissage automatique des informations déjà saisies dans les profils ORCID et HAL
(synchronisation automatique entre ORCID et HAL)



Principes directeurs et méthode

Alignment avec les standards internationaux

Alignment avec l'écosystème international des PIDs (ORCID, ROR, DOI...)

Stratégie collective à l'échelle nationale

Pilotage par le MESRE, avec instances de gouvernance dédiées

Respect du principe d'autonomie des établissements

Pas de SI unique

Démarche collaborative

Incluant établissements, ONR, opérateurs, agences de financement...
et réseaux métiers

Plan d'actions co-construit par les parties prenantes ;
feuille de route pragmatique



Ouvrir
la science !



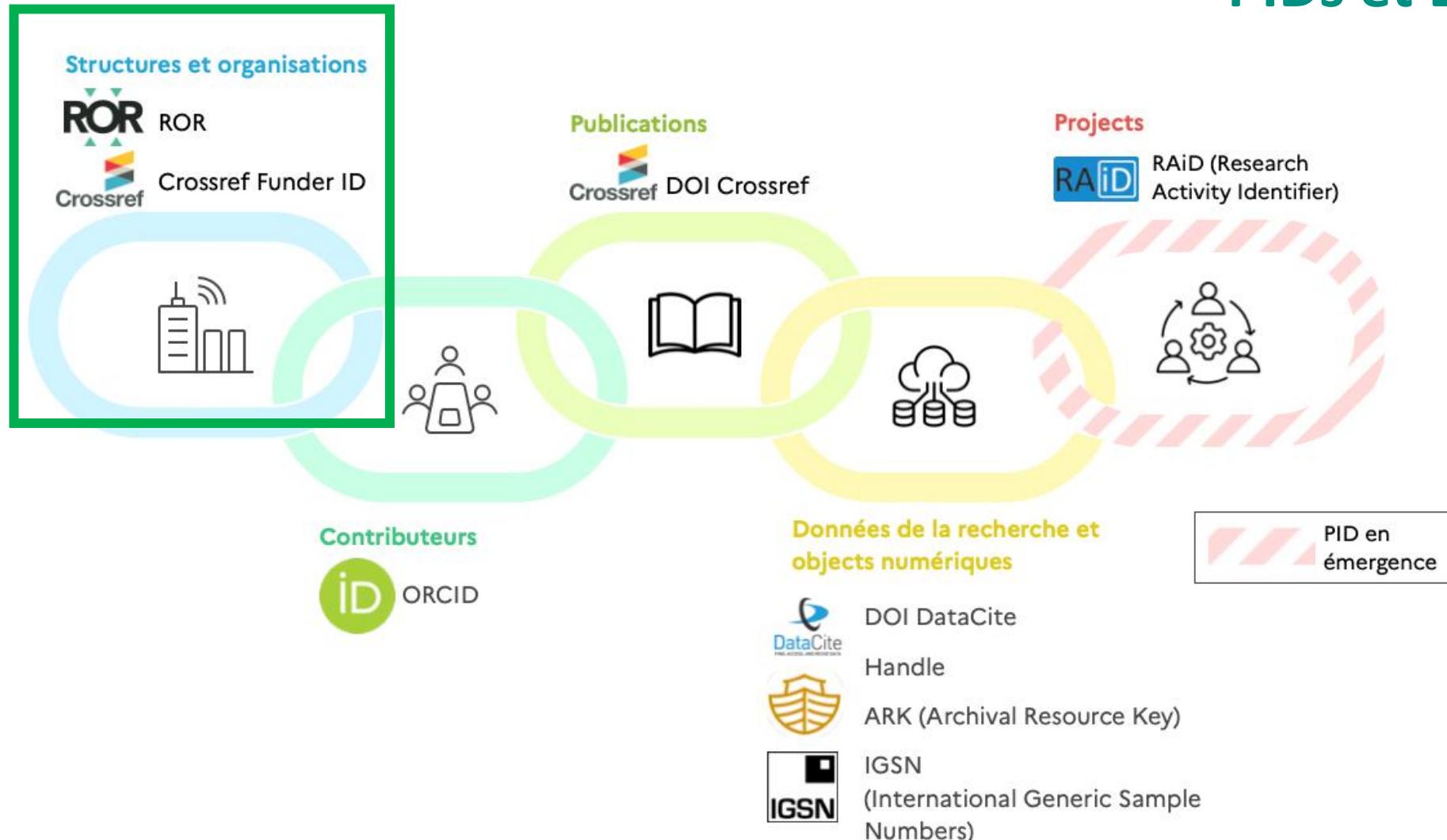
Recommandations
pour l'adoption
des identifiants
persistants dans
l'enseignement
supérieur et la
recherche

Collège des Données de la recherche

Septembre 2025

doi. 10.52949/84

PIDs et ESR



ROR et le défi des structures de recherche françaises

Créer un registre national aligné sur ROR **RNeST**



**Source nationale de données
institutionnelles de confiance**



Alimentation automatique de ROR



Données plus riches que ROR

granularité fine, historique des
transformations, liens hiérarchiques, contacts
des responsables d'unités, spécificités
françaises...

2025

**Cadre de référence finalisé
définit**

- ✓ Types d'objets (établissements, unités de recherche...)
- ✓ Relations
- ✓ Cycle de vie des structures
- ✓ Gouvernance des données (niveau local/national)

2026

Déploiement

PIDs et ESR

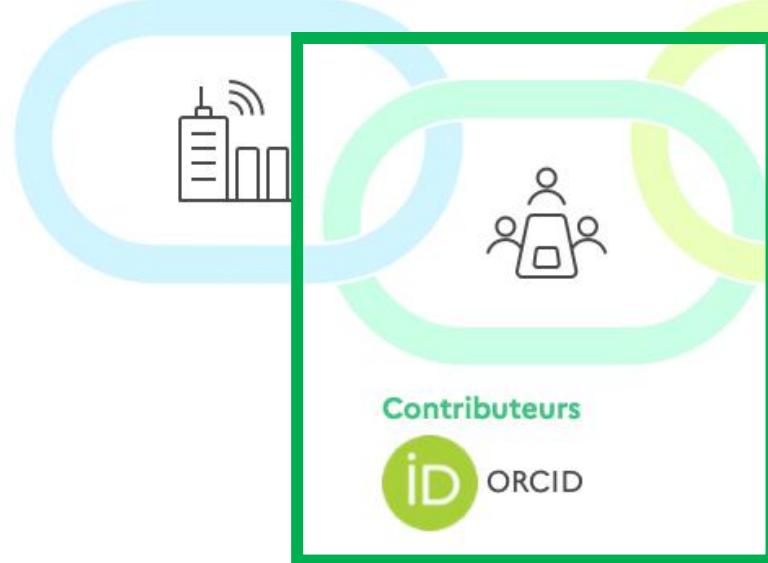
Structures et organisations



ROR



Crossref Funder ID



Publications



DOI Crossref

Projects



RAiD (Research Activity Identifier)



Données de la recherche et objects numériques



DOI DataCite



Handle



ARK (Archival Resource Key)



IGSN
(International Generic Sample Numbers)

PID en
émergence



Identifiants pour les contributeurs de la recherche

Le nom n'est pas un indicateur fiable :

- Homonymie
 - Confusion nom/prénom
 - Translittération
 - etc.

Identifiants pour les contributeurs de la recherche

De nombreux identifiants pour les contributeurs de la recherche



Choix du standard international **ORCID**, ouvert et **interopérable**

ORCID permet de lier plusieurs informations, notamment :

- les publications et les autres produits de la recherche
- les informations administratives (affiliation, agences de financement etc.)

Un consortium : ORCID France (ABES)

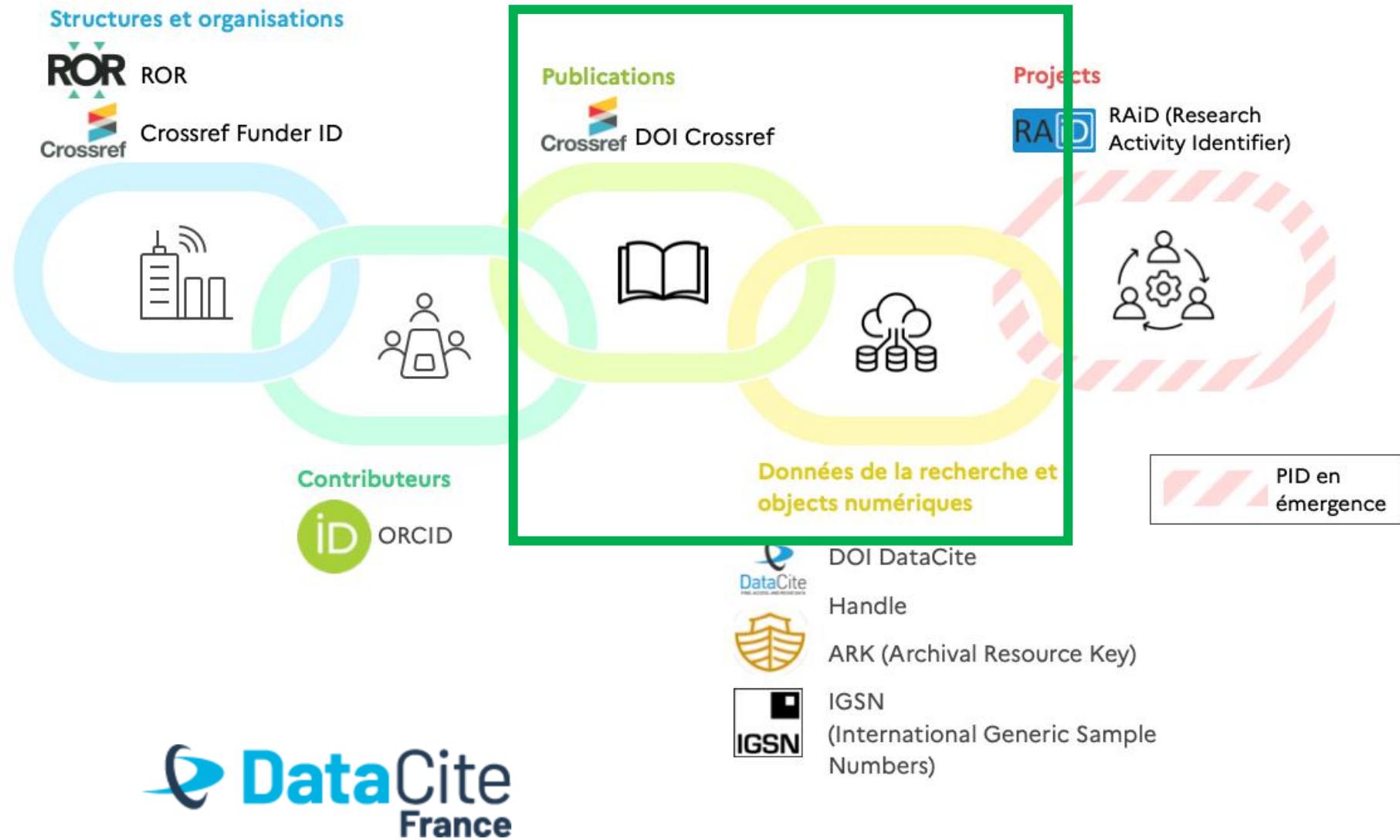


ORCID
•FRANCE•
**Recommandations pour
créer et mettre à jour son
profil ORCID :
que renseigner dans
la section "Emploi"**

ORCID
•FRANCE•
**Recommandations à
destination des
établissements pour une
meilleure saisie des emplois
dans les profils ORCID**

ORCID
•FRANCE•
**Recommandations pour la
création, la mise à jour et la mise
en qualité des données ROR**

PIDs et ESR



Identifiants pour les objets de la recherche

Plusieurs identifiants pour les produits issus des travaux de recherche (DOI, Handle, ARK...)

12 agences d'enregistrement des DOI au niveau international (dont Crossref et DataCite)

Choix des DOI attribués par **DataCite** pour les **données**

Schéma de **métadonnées** évolutif

Un consortium : DataCite France (CNRS-INIST)

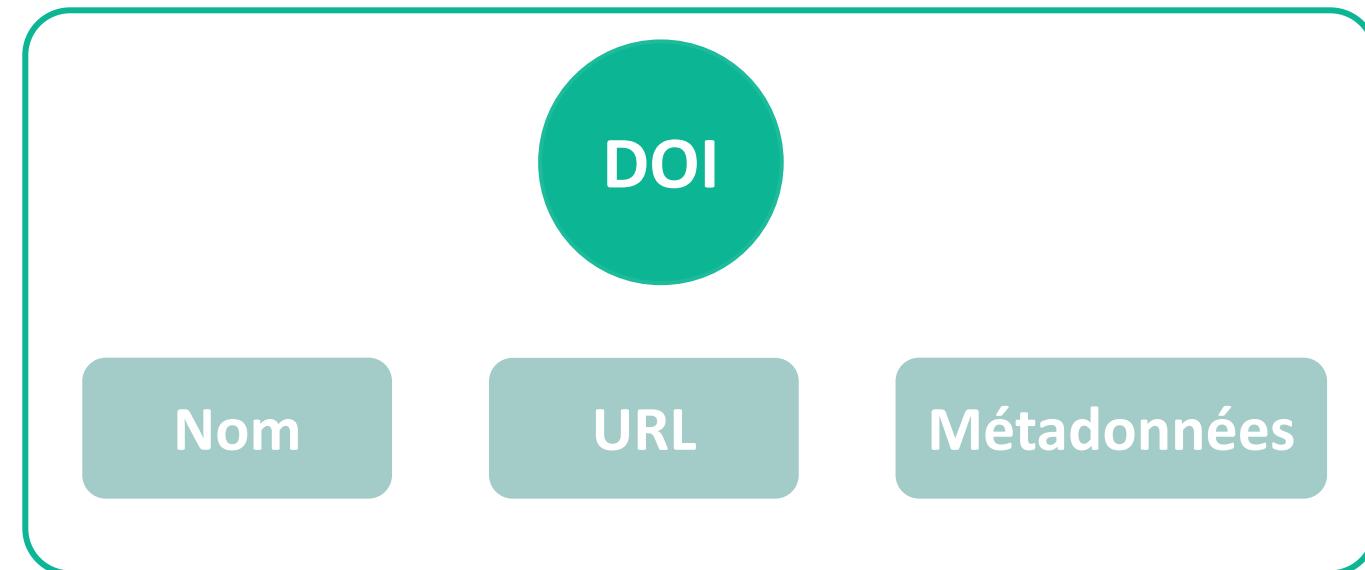


Qu'est-ce qu'un DOI ?

DOI = Digital Object Identifier

Chaîne alphanumérique

- affectée à **l'identification unique d'un objet** (articles scientifiques, jeux de données, logiciels, etc.)
- liée à une **description des métadonnées** de l'objet
- liée à un **emplacement numérique**, tel qu'une URL, où tous les détails de l'objet sont accessibles



Crossref VS DataCite



- Spécialisé dans les **publications scientifiques** articles, actes de conférences, chapitres d'ouvrage...
- Utilisé majoritairement par les éditeurs scientifiques

- Spécialisé dans les **données de recherche** et autres objets scientifiques
- Devenu la **référence pour les DOI de jeux de données**

Utilité du DOI : un outil clé pour mettre en œuvre les principes FAIR

Findable

- Le DOI est un identifiant unique, référencé dans des moteurs (Google Dataset Search, DataCite, etc.)
- Les DOI assurent un lien persistant vers le contenu, même si l'URL change, ce qui garantit un accès fiable aux objets
- Les DOI permettent des citations standardisées, ce qui améliore la visibilité, la découverbarilité et la valorisation des données

Accessible

- Le DOI pointe vers une landing page toujours accessible (avec les métadonnées)

Interoperable

- Les DOI facilitent l'intégration des ressources dans des écosystèmes numériques mondiaux, notamment pour le partage, la découverte et la réutilisation des données de recherche.
- Les métadonnées sont standardisées

Reusable

- Le DOI permet de donner des informations indispensables à la réutilisation : licences, contexte, versions, provenance etc...

Utilité du DOI : quels bénéfices pour le chercheur ?

Visibilité et valorisation

Le DOI facilite la citation et la découverte des données

Partage et accessibilité

En déposant ses données avec un DOI, le chercheur permet aux autres de retrouver, comprendre, et réutiliser son travail.

Réponse aux exigences des financeurs et éditeurs

Le DOI est souvent requis par les agences de financement ou les revues dans les politiques de science ouverte

Suivi des usages et mesures d'impact

Le DOI permet de mesurer les citations ou les réutilisations des données

Les DOI : à l'interface de l'ensemble des acteurs de la recherche

Entrepôts

Attribution des DOI ; vérification éventuelle des métadonnées



DOI

Personnels d'appui

Contribuent à la qualité et à la complétude des métadonnées

Chercheurs

- Fournissent et/ou valident les métadonnées lors du dépôt des jeux de données
- Signalent les mises à jour nécessaires

Pourquoi un DOI ne suffit pas : l'importance des métadonnées dans les entrepôts de données

Attribution d'un DOI : nécessaire, mais non suffisant

un DOI n'est réellement efficace que si les métadonnées sont complètes, précises et bien structurées



Données peu visibles, mal contextualisées, difficilement citables ou exploitables
Impact scientifique moindre

Entrepôts, DOI et métadonnées

Perte de découvrabilité : une donnée existe mais reste invisible

Indexation des métadonnées dans les moteurs de recherche, les portails thématiques

 La visibilité chute drastiquement, même si les données sont de haute qualité

 Données trouvables par multiples entrées thématiques, géographiques

Entrepôts, DOI et métadonnées

Perte de découvrabilité : une donnée existe mais reste invisible

Indexation des métadonnées dans les moteurs de recherche, les portails thématiques

 La visibilité chute drastiquement, même si les données sont de haute qualité

 Données trouvables par multiples entrées thématiques, géographiques

Contexte scientifique insuffisant : la donnée ne peut pas être comprise ni évaluée

Les métadonnées donnent le contexte de la donnée :

- qui l'a produite
- comment elle a été acquise
- où et quand
- dans quel cadre

Sans ces informations, l'utilisateur :

- ne peut pas juger la qualité des mesures
- ne peut comprendre le contexte d'acquisition
- ne peut pas évaluer la pertinence pour ses propres analyses
- ne peut combiner avec d'autres jeux de données de manière fiable
- risque d'abandonner la réutilisation faute de confiance

Entrepôts, DOI et métadonnées

Perte de découvrabilité : une donnée existe mais reste invisible

Indexation des métadonnées dans les moteurs de recherche, les portails thématiques

 La visibilité chute drastiquement, même si les données sont de haute qualité

 Données trouvables par multiples entrées thématiques, géographiques

Contexte scientifique insuffisant : la donnée ne peut pas être comprise ni évaluée

Les métadonnées donnent le contexte de la donnée :

- qui l'a produite
- comment elle a été acquise
- où et quand
- dans quel cadre

Sans ces informations, l'utilisateur :

- ne peut pas juger la qualité des mesures
- ne peut comprendre le contexte d'acquisition
- ne peut pas évaluer la pertinence pour ses propres analyses
- ne peut combiner avec d'autres jeux de données de manière fiable
- risque d'abandonner la réutilisation faute de confiance

Conséquence : données scientifiquement inutilisables malgré leur valeur potentielle

Extension du périmètre des DOI DataCite



Initialement :
jeux de données



Aujourd’hui :
infrastructure technique d’identifiants persistants pour tous types d’objets de recherche

Types d’objets désormais pris en charge par DataCite :

- » Échantillons / objets physiques → IGSN
- » Projets et activités de recherche → Project DOI, RAID, GRANT ID
- » Instruments et équipements scientifiques
- » Images, vidéos, sons et autres médias
- » Collections multi-objets
- » Documents de recherche (rapports, protocoles...)

Partenariat IGSN/DataCite

Depuis janvier 2023, les identifiants IGSN ne peuvent être enregistrés **que sous forme de DOI DataCite (IGSN-ID)** (type de ressource PhysicalObject).

Chaque membre DataCite ou membre du consortium DataCite peut attribuer des IGSN-ID DataCite
Le schéma de métadonnées IGSN est **mappé** au schéma DataCite

Exemple :

- Format ancien : ABC1234
- Format nouveau : 10.58052/ABC1234 (avec préfixe DOI)
- Citation complète : igsn:<http://doi.org/10.58052/ABC1234>



Partnership between IGSN and DataCite

October 28, 2021 Matt Buys and Kerstin Lehnert

<https://doi.org/10.5438/7270-1155>

Earlier this year, DataCite and IGSN announced their roadmap towards a partnership to support the global adoption, implementation, and use of physical sample identifiers. Today, we are pleased to share the announcement of the partnership agreement.

DataCite is a community-led organisation with a vision to connect research and identify knowledge. We have been providing the means to create, find, cite, connect, and use research across 48 countries globally since 2009. In addition to DataCite core services, we support the scaling efforts of several identifier communities through governance, sustainability, insurance, and technical implementation.

IGSN e.V. is an international organization that provides a central registration system for physical sample identifiers. It connects physical samples to grants, data, publications, and other resources. IGSN and DataCite benefit to our users from years of experience and engagement, and we are excited to work together to develop best practices for IGSN IDs over time.

IGSN ID Metadata Recommendations

IGSN IDs are registered with metadata encoded in the DataCite Metadata Schema. The following guidance for IGSN ID metadata in the DataCite Metadata Schema was developed by the IGSN–DataCite Crosswalk and Metadata Management Working Group and sourced from the [IGSN–DataCite Crosswalk Recommendation](#). DataCite, the IGSN e.V., and the Working Group will continue to expand and refine metadata best practices for IGSN IDs over time.

IGSN ID recommendations for mandatory properties in DataCite Metadata Schema

IGSN IDs must be registered with at least the six mandatory properties in the DataCite Metadata Schema.

Stratégie nationale PIDs prochaines étapes (2026)

- Déploiement de **RNeST**, synchronisé automatiquement avec ROR
- **ORCID** : encourager l'intégration dans les SI
- Définir les modalités d'adoption d'un identifiant adapté aux instruments, équipements et infrastructures de recherche (**facilities**)
- Définir les modalités d'adoption de recommandations nationales autour des **échantillons** (IGSN/DOI)

National PID Strategies Interest Group

Persistent Identification of Instruments WG

Physical Samples and Collections in the Research Data Ecosystem IG



Consortium DataCite France

Périmètre et modèle économique



Consortium DataCite France

Consortium DataCite France

Périmètre et modèle économique

Situation actuelle

Une stratégie nationale

Un consortium en évolution

Un nombre croissant d'adhérents (y compris hors ESR ou hors France)

Un changement de modèle économique de DataCite international

Consortium DataCite France

Périmètre et modèle économique

Un soutien financier du MESRE

Dans le cadre de la politique nationale de simplification et la stratégie PID
Pour les adhérents français de l'ESR dès 2026

Une nécessaire révision des statuts et de l'organisation du consortium

Définition du périmètre des adhésions du consortium

Définition d'un modèle économique pour les adhérents hors ESR

Définition d'une charte de bonnes pratiques pour l'usage des DOI