

Des identifiants ouverts pour la science ouverte

Action concertée au bénéfice des chercheurs et des institutions

IdHal, OrcID, WosId, ArXivID, DOI, ISSN, ISBN, Handle, IdRef, VIAF, ISNI, ...

Pourquoi tant d'identifiants ? Pour qui ? Pour quoi faire ?

La quasi-totalité de la production scientifique mondiale est désormais signalée ou disponible sur le web. Des millions d'objets scientifiques (publications, données et autres objets numériques) produits par autant d'auteurs ou de contributeurs affiliés à des centaines de milliers d'organisations sont désormais trouvables grâce à des systèmes d'identifiants pérennes nombreux et variés.

Ces systèmes ont été développés au cours des années récentes, pour répondre à des besoins propres à chaque communauté ou à chaque usage, mais la plupart sont encore en phase de consolidation ou même de développement.

Les plus connus aujourd'hui sont souvent soutenus par des organisations à but non lucratif, mais financées par des fonds privés.

Pour répondre aux objectifs de la science ouverte, c'est-à-dire garantir la disponibilité libre et indépendante sur le long terme des informations scientifiques, il est nécessaire de s'assurer que les identifiants s'appuient sur une architecture ouverte, documentée, libre et qu'ils sont portés par et pour les communautés scientifiques.

La mise sur pied du Comité pour la Science ouverte a donné le coup d'envoi à une réflexion conduite par les opérateurs de l'Enseignement supérieur et de la recherche français destinée à améliorer la structuration des identifiants les plus utiles, à accélérer leur adoption par les communautés, et à les rendre plus libres et pérennes.

Un identifiant ?

C'est un numéro ou une étiquette alphanumérique, opaque ou explicite, lisible par des machines et par des humains, permettant de désigner et de retrouver de manière univoque et pérenne un objet, un document, une personne, un lieu, un organisme, ou toute entité, dans le monde réel et sur internet.

Les identifiants pérennes ou « **PIDs** » (Permanent IDentifiers) les plus connus sont aujourd'hui le DOI (Digital Object Identifier, <https://doi.org/>) pour les articles ou les chapitres de livres et les documents en général, ou encore les jeux de données, l'ORCID (Open Research or Contributor ID, <https://orcid.org/>) pour les auteurs de publications, ou l'ISSN (International Standard Serial Number, <https://issn.org>) pour les revues. Handle (<https://www.handle.net/>) est très répandu dans certaines communautés savantes.

Pour les chercheurs et les contributeurs aux productions scientifiques, plusieurs systèmes d'identifiants sont disponibles, dont l'usage varie selon les pays ou les communautés (ORCID, ScopusID, WosID, IdHal, ArXivID, ...). Grâce au travail des spécialistes de l'information ils sont déjà relativement bien reliés entre eux – on dit aussi alignés – en comparaison des autres types d'identifiants, au travers de grands référentiels, comme IdRef (<https://idref.fr>) en France, ou ISNI (<https://isni.org>) à l'échelle internationale.

En revanche, à ce jour il n'existe pas d'identifiant international standard pour les organisations ni pour les affiliations. Ces entités sont répertoriées dans des registres nationaux (en France, RNSR, AURÉHAL structures) et internationaux (ROR, GRID, Ringgold, ...) encore imparfaitement alimentés et connectés, même si les référentiels qui permettent l'alignement (IdRef, ISNI) permettent de progresser de jour en jour.

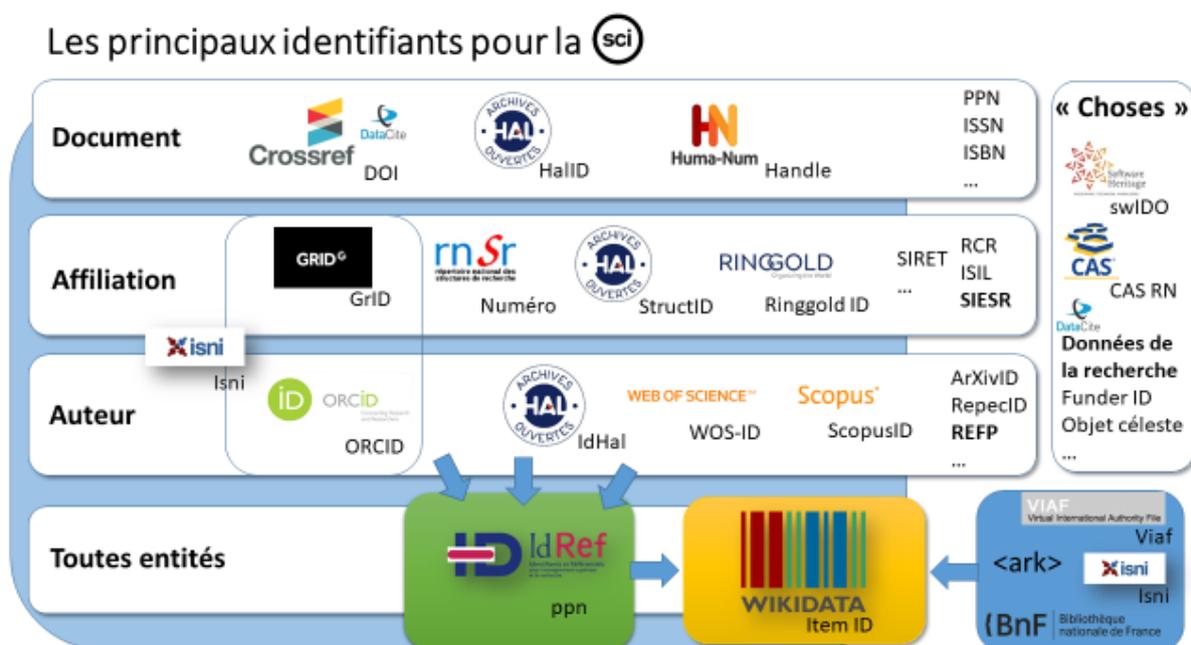
Malgré ces limitations, on observe chaque jour que de nouvelles communautés de recherche, de nouvelles institutions, et de plus en plus de chercheurs individuels adoptent les identifiants afin de faciliter le repérage de leurs publications et de leurs données, de leur propre chef ou à la demande des éditeurs ou des financeurs, et de plus en plus souvent pour favoriser l'accès ouvert aux résultats de la recherche.

La recherche a besoin d'identifiants

Le développement de la science ouverte s'accompagne de celui des principes « FAIR : findable, accessible, interoperable, reusable », en français : « facile à trouver, accessible, interopérable et réutilisable ». Chaque principe est décomposé en 4 exigences. La première exigence du premier principe porte précisément sur l'attribution à la donnée d'un identifiant unique (« F1. (meta)data are assigned a globally unique and eternally persistent identifier »). Le rôle de l'identification unique ne peut donc être sous-estimé.

L'Union européenne soutient des projets transnationaux visant à développer l'utilisation des identifiants de manière coordonnée par les chercheurs et les institutions des différents pays, parmi lesquels Freya - *Connected open identifiers for discovery, access, and use of research resources* (<https://www.project-freya.eu>), dans lequel la France est représentée au sein du programme des « ambassadeurs ». Freya semble être amené à jouer un rôle important dans la définition de l'usage des identifiants dans l'European Open Science Cloud (<https://www.eosc-portal.eu>), plateforme européenne qui fédère les outils permettant aux chercheurs de stocker, gérer, analyser et réutiliser de larges quantités de données issues de la recherche.

Depuis 2012 la base Wikidata, issue de la fondation Wikimedia qui porte Wikipedia, est devenue progressivement le point de convergence mondial des identifiants ouverts. Les principaux systèmes d'identifiants décrits ci-dessus sont alignés avec Wikidata ou y versent leurs propres données.



Que faire à l'échelle nationale ?

Le contexte national et international constitue un environnement propice à une action concertée en matière d'identifiants pour la science ouverte :

- La réflexion internationale sur les identifiants utiles pour la recherche (principalement ceux concernant personnes, organisations, publications, données) a conduit les communautés à produire des guides de bonne pratique et même des plans d'action.
- En France, le plan national pour la science ouverte, manifestation d'une volonté politique forte, a accéléré un mouvement concerté autour d'une adhésion nationale à ORCID, un soutien à Open Citations qui s'appuie sur les DOI, ainsi qu'à un soutien aux principes FAIR, qui comprennent l'adoption d'identifiants uniques. Par ailleurs, le plan a créé le Comité pour la science ouverte, qui peut aujourd'hui à la fois proposer des objectifs opérationnels et coordonner un plan d'action national impliquant les acteurs de terrain.
- L'expérience acquise dans le développement des identifiants de publications (DOI) et des données de la recherche, le chemin parcouru mais encore partiel dans l'adoption d'identifiants de personnes (ORCID et autres), et les tâtonnements dans l'identification des organisations, ont permis de comprendre l'importance d'une action politique en la matière. Ces acteurs de terrain ont acquis une connaissance assez fine des forces et des limitations des PIDs et des registres et veulent de nouveaux services pour mieux les exploiter.

L'année 2019 marque le lancement d'une action coordonnée au niveau national pour le développement en France des identifiants pérennes pour la science ouverte, qui conduira à la mise en oeuvre d'un réel plan d'action.

Considérant la nécessité et l'utilité d'adopter et de faire évoluer les systèmes internationaux d'identifiants, notamment les identifiants de personnes, de structures, de publications et de données de la recherche, le Comité pour la science ouverte définit plusieurs objectifs stratégiques :

1. disposer d'une stratégie nationale coordonnée, très fortement articulée avec le paysage international, notamment afin de rendre plus visible la recherche française,
2. accélérer l'adoption des identifiants par les chercheurs, les laboratoires et les institutions, d'une part, et par les services numériques utilisés par l'ESRI, d'autre part,
3. identifier des modèles économiques viables pour assurer un déploiement durable des identifiants,
4. améliorer l'interopérabilité et la normalisation des identifiants tout en s'assurant d'un contrôle par la communauté scientifique
5. contribuer au pilotage et à l'évolution des systèmes d'identifiants afin d'en garantir l'ouverture et l'indépendance sur le long terme.

Cette action nationale coordonnée vise de manière concrète

- le développement de services utiles à partir de cas réels permettant de démontrer les atouts des PIDs

- la communication adaptée vers les chercheurs pour les guider dans leur usage quotidien des identifiants

Compte tenu des besoins identifiés et du niveau de maturité des solutions, elle se décompose en 4 actions distinctes menées en parallèle, classées selon l'urgence à agir et qui concernent :

- Les identifiants de structures et d'organisations, autour du Répertoire national des structures de recherche RNSR
- Les identifiants de chercheurs, en s'appuyant sur une adhésion collective à ORCID
- Les identifiants de publications, en amplifiant les usages du DOI
- Les identifiants de données, logiciels et objets numériques, en s'appuyant sur le groupe RDA France et les collèges et groupes de travail du CoSo.

Pour en savoir plus : Lien vers la note complète