

ROR : une base d'identifiants rugissante

Base de données collaborative internationale, ROR¹ (*Research Organizations Registry*) œuvre à identifier de manière univoque toutes les entités liées à la recherche. Depuis l'été 2023, ROR a affiné son niveau de description en intégrant une grande partie des laboratoires publics français.

Ces dernières années, le patient travail des chartes de signature, élaboré dans la majorité des établissements, a consisté à tenter de normaliser la manière dont les informations d'affiliation issues des publications vont figurer dans les bases bibliographiques et bibliométriques. Mais même orthographiées de la même manière et scrupuleusement listées dans le même ordre, si ces informations précieuses restent des chaînes de caractères, cela ne suffit pas pour identifier de manière certaine et pérenne les structures qu'elles désignent. Comment boucher les trous de ce tonneau des Danaïdes pour enfin apparier sans erreur les publications aux organisations ?

ROR est l'acronyme du *Research Organizations Registry*, qu'on peut traduire par Répertoire des organisations liées à la recherche. Cette base de données vise à identifier toutes les entités liées à la recherche : universités, entreprises, organismes, hôpitaux, agences publiques de financement, infrastructures matérielles, etc. Le cas d'usage central de ROR est d'explicitier l'affiliation, dans les publications, c'est à dire le lien qui existe entre une personne physique et une organisation pour une production (le si fameux et difficilement traduisible « *research output* »), le plus souvent un article, mais aussi un jeu de données, par exemple.

L'identifiant ROR (ROR ID) est un PID (*Permanent Identifier, identifiant pérenne*) destiné à isoler et reconnaître de manière univoque, sur la scène internationale, une entité collective à laquelle peuvent être liées des personnes qui font de la recherche. Il se présente comme une chaîne non signifiante de 9 caractères alphanumériques à la suite du préfixe du registre ; ainsi l'ID ROR de l'Abes est <https://ror.org/027xymc69>. Le rôle de cet identifiant est de se diffuser dans tous les systèmes de gestion qui ont trait à la recherche : plateforme de publication, outil d'évaluation, demande de financement... Les données de ROR sont accessibles de plusieurs manières : par une recherche directe dans le registre², par une requête via l'API³ ou en téléchargeant le jeu de données complet⁴. À noter que les métadonnées visibles dans le moteur de recherche ne reflètent pas la totalité de celles fournies par l'API. ROR fonctionne également de manière intéressante pour la curation des données : tout un chacun peut

suggérer via un formulaire⁵ un ajout ou une modification. Le traitement de ces sollicitations est présenté, en toute transparence, sur une instance GitHub⁶. On peut ainsi observer en direct le travail d'aiguillage des demandes et les modifications opérées par la petite équipe dédiée à la curation, composée d'une personne à plein temps, deux en renfort, sous la houlette d'un groupe dénommé « *Curation Advisory Board* » qui juge de la pertinence des requêtes, notamment pour les nouvelles entités⁷.

ROR DANS L'ÉCOSYSTÈME DES IDENTIFIANTS PÉRENNES

La collectivité (dirions-nous en bons bibliothécaires) qui possède un ID ROR n'est jamais bien loin des personnes physiques, lesquelles peuvent posséder un identifiant ORCID. ROR, organisation dont la gestion est partagée par tous les acteurs du circuit de l'édition scientifique, des bibliothèques aux éditeurs privés, a nettement marqué dès le départ son intention d'être un identifiant ouvert, dont le registre soit bâti par et pour la communauté.

Les fées qui se sont penchées sur son berceau sont la *California Digital Library*, ainsi que CrossRef et DataCite, deux poids lourds parmi les agences d'attribution de DOI. Le premier état du référentiel a vu le jour en 2019. Il comprenait un bagage de près de 90 000 entrées provenant de GRID, un référentiel ouvert créé en 2015 par la société britannique *Digital Science* pour combler un vide dommageable à ses activités. ROR en a hérité une grande partie de son modèle de données (qu'il a néanmoins fait évoluer en 2022) et récupéré les alignements déjà effectués avec les référentiels généralistes que sont ISNI et Wikidata. Le référentiel s'est ensuite étoffé pour atteindre aujourd'hui près de 107 000 entrées. Les données sont placées sous licence *Creative Commons* CC0 1.0 (le plus proche qu'il soit possible du domaine public) et leur usage est libre et gratuit. Il en va de même pour l'API, gratuite et en *open source*.

COLLABORATION ET OUVERTURE COMME VERTUS CARDINALES

Concernant la gouvernance, ROR est conçu comme un service partagé, et non une entité autonome.

[1] ROR joue sur l'homophonie en anglais entre son nom et le verbe *to roar*, rugir (voir le blog <https://ror.org/blog>)

[2] <https://ror.org/search>

[3] <https://api.ror.org>

[4] <https://zenodo.org/records/8436953>

[5] https://docs.google.com/forms/d/e/1FAIpQLSdJYaMTCwS7muuTa-B_CnAtCSkKzt19lkirAKG4u7umH9Nosg/viewform

[6] <https://github.com/ror-community/ror-updates/projects/1>

[7] <https://ror.org/registry/#curation-advisory-board>

[8] <https://info.orcid.org/orcid-board>

Ses trois fondateurs se sont engagés⁹ à conserver un fonctionnement collaboratif, en s'interdisant de monnayer les données ou le service rendu, et de transférer tout ou partie du registre à une entité commerciale. ROR s'engage à souscrire aux POSI – acronyme des *Principles of Open Scholarly Infrastructures*¹⁰ qui articulent gouvernance, pérennité et ouverture. Ceci étant posé, le modèle économique de ROR est toujours fragile car il reste conditionné au soutien volontaire de la communauté et à des subventions, souvent non reconductibles. On peut néanmoins noter par exemple que le Fonds national pour la science ouverte (FNSO) l'a choisi en 2023 comme l'un des quatre bénéficiaires dans le cadre de la campagne d'appel à financement¹¹, donnant le signe plutôt engageant d'une infrastructure qui commence à devenir incontournable. Lors de la publication de la note sur les identifiants pérennes du COSO Europe en juin 2019¹², le terrain des organisations n'était pas considéré comme

nom, il va falloir être un peu plus précis. L'échelon auquel la recherche française se fait, se pense, se finance et s'écrit, donc se signe, c'est avant tout le laboratoire. Or, qui dit laboratoire dit, la plupart du temps, multiplicité des tutelles : une ou plusieurs universités, et/ou un ou plusieurs organismes de recherche (ce fameux problème de la *mixité*). La relation d'un laboratoire avec ses tutelles fait qu'il n'est pas contenu ou subordonné ; et la nature de l'établissement peut changer (par exemple, se muer en un EPE) sans que le laboratoire en soit fondamentalement affecté. Nous touchons là à une des limites actuelles de ROR, eu égard au contexte français : le modèle de données ne permet pas de décrire de manière complexe les liens entre structures.

Sans signifier officiellement un changement de braquet, ROR a tout bonnement fait le choix pragmatique d'ajouter dans son registre une grande partie des laboratoires publics français à l'été 2023¹⁵. Les données issues du NSRS (Répertoire national

des structures de recherche, coordonné par le ministère de l'Enseignement supérieur et de la Recherche) étant publiées de manière ouverte¹⁶, après concertation avec le ministère, ROR en a intégré une partie pour enrichir sa base – et méca-

Les trois fondateurs de ROR se sont engagés à conserver un fonctionnement collaboratif, en s'interdisant de monnayer les données ou le service rendu.

stabilisé. En ce jadis pourtant pas si lointain, GRID existait encore, RingGold était une base crédible (car utilisée par de nombreux éditeurs, bien que payante et fermée) et ISNI pouvait être présenté comme un choix officiel (par exemple par le Jisc, au Royaume-Uni¹³). Cinq ans plus tard, ROR a suffisamment consolidé ses positions pour avaler l'*Open Funder Registry* de CrossRef (référentiel des organismes de financement), devenir l'unique référentiel des organisations pour ORCID au détriment de RingGold¹⁴, et s'imposer. Les implémentations se comptent par dizaines, qu'il s'agisse d'éditeurs, comme pour la prestigieuse revue Science de AAAS, d'archives ouvertes, comme celle de la NASA, voire de CRIS (*Current Research Information System*) comme celui du CERN, ou de bases de données bibliographiques comme OpenAlex.

QUID DE LA FRANCE ?

ROR a toujours défendu l'idée de décrire uniquement les « *top-level institutions* ». Il faut entendre par là le refus d'entrer dans la description fine de la composition interne d'un établissement. Néanmoins, la France représente – même si elle n'est pas l'unique – le principal cas de figure où on ne peut pas se contenter de rattacher les productions de la recherche à une seule institution. Si tout le monde sait ce qu'est le CNRS, il est évident que dans un référentiel des structures de recherche digne de ce

nomiquement renforcer son attractivité pour le public concerné.

Au printemps 2023, l'Abes a mené un chantier d'alignement vers ROR pour tous les établissements habilités à délivrer le doctorat. Ce discret feulement appellera sans nul doute un mouvement plus ample dans les mois à venir. Le rugissement serait-il aussi communicatif que le bâillement ?

CAROLE MELZAC

Service Autorités et référentiels de l'Abes
melzac@abes.fr

[9] <https://ror.org/documents/ROR-Memorandum-of-Agreement-2022.pdf>

[10] <https://openscholarlyinfrastructure.org>

[11] <https://www.ouvrirlascience.fr/le-fonds-national-pour-la-science-ouverte-inscrit-son-soutien-aux-infrastructures-internationales-dans-la-duree>

[12] <https://hal-lara.archives-ouvertes.fr/hal-03640303v1>

[13] <https://researchdata.jiscinvolve.org/wp/2018/09/13/organisation-identifiers-recs>

[14] <https://info.orcid.org/orcid-support-for-ringgold-organization-ids-ending>

[15] <https://github.com/ror-community/ror-updates/issues/4930>

[16] <https://www.data.gouv.fr/en/datasets/repertoire-national-des-structures-de-recherche-rnsr>



Crédit Adobe stock, par byrdyak

