

L'UGA au travers du ROR : assurer la qualité des données

Maxence Larrieu

2024-08-18

ROR open data curation identifiers

Notre précédent billet présentait un travail d'identifiant des ORCID de l'UGA à partir de HAL, OpenAlex et ORCID. Par ricochet, ce travail a mis en lumière l'importance, pour un établissement de recherche, d'avoir des données de qualité dans le [Research Organization Registry \(ROR\)](#). En juin de la même année, nous avons donc entrepris un nettoyage des données de l'UGA dans le ROR. Chacune des unités de l'UGA (env. 70) a maintenant son identifiant ROR correctement rattaché à celui de l'université. Cela a nécessité environ 60 demandes de modifications, toutes traitées par le ROR en moins d'un mois.

Lancé en 2019 le ROR succède au Global Research Identifier Database. Il est mondial, "community-led", doté d'une API ouverte et met à disposition les données sous licence CC0 :

We value ourselves as a fully open and public registry of factual information about research organizations. The metadata and curatorial value that we contribute to the ROR registry are not considered the intellectual property of any ROR-related entity. As facts, information stored in the registry, by its nature, cannot be patented. In addition, ROR fully asserts no ownership of the information by making it available to the public under a CC0 1.0 public domain dedication.

Aligning ROR with the Principles of Open Scholarly Infrastructure. ror.org/blog/2020-12-16-aligning-ror-with-posi/

Le ROR est un référentiel ouvert et mondial des structures de recherche. Au niveau français, les référentiels des structures sont par exemple le [RNSR](#) (géré par le MESR), [IdRef](#) (géré par l'ABES) ou encore [AuréHAL structure](#) (géré par le CCSD).

Identifiants externes

 Identifiant HAL : 1043180

 Identifiant HAL : 1041685

 Identifiant HAL : 675

Identifiant RNSR : 200311865N

 Identifiant ROR : 026j45x50

 Identifiant VIAF : <http://viaf.org/viaf/136340959>

 Identifiant Wikidata : Q30262127

 Identifiant Wikipédia : https://fr.wikipedia.org/wiki/Laboratoire_Pacte

 Identifiant ARK BnF : <http://catalogue.bnf.fr/ark:/12148/cb14527532p>

 Identifiant ISNI : 0000000121648234

Listes d'identifiants utilisés dans IdRef pour une structure de recherche. idref.fr/078842697

En France, la curation des données du ROR n'est pas encore rentrée globalement dans les politiques et pratiques. Tout laisse à penser néanmoins que cela changera dans les prochaines années. Les identifiants font en effet partie intégrante de la politique nationale science ouverte et d'autre part les données du ROR font maintenant autorité dans l'écosystème de l'information scientifique : elles sont utilisées pour structurer OpenAlex, avec lequel [le MESR a annoncé un partenariat](#) ; utilisées également dans le récent (2024) [CWTS Leiden Ranking Open Edition](#) ; ou encore au niveau français dans le tout aussi récent [works magnet](#) du MESR qui permet de relier productions scientifiques et établissements.

Sous l'égide de la Commission science ouverte de l'UGA et avec la Cellule Data Grenoble Alpes, nous avons donc réalisé un chantier de curation des données relatives à l'université dans le ROR. L'objectif est que toutes les unités dont l'UGA est tutelle possèdent un identifiant ROR correctement relié à celui de l'université. Le schéma de données du ROR permet en effet de lier les structures : "The relationship types supported are Parent, Child, Related, Successor, and Predecessor." Voir l'excellent billet de blog sur ces relations et les difficultés à dépasser pour un référentiel mondial : ror.org/blog/2023-02-27-parents-children-and-other-relationships-in-ror/.

La curation dans le ROR

Tout le monde peut *demande* une modification des données du ROR à l'aide d'un simple formulaire ; la complétion du formulaire engendre une "issue" Github, laquelle est ensuite traitée par l'équipe du ROR. Ce système permet de suivre finement l'évolution des demandes, cf. github.com/ror-community/ror-updates.

Déroulé

Le chantier s'est déroulé entre juin et mi-juillet 2024, mobilisant 4 personnes, pour un total d'environ 60 "issues", toutes [re-traçables dans Github](#)

Un point fort du ROR est que sa curation est rapide : en quelques jours, on reçoit le retour d'un curateur ; l'issue dans Github change successivement de labels ; et la mise à jour, si elle est acceptée, est effectuée en moins d'un mois.

Ce chantier UGA a été l'occasion de revoir à l'échelle nationale le rattachement des "fédérations de recherche" et autre "groupements de recherche" aux établissements. Pour faire court, ces entités étaient rattachées aux établissements de la même façon que le sont les unités, avec une relation de type *child*. Après notre requête, un échange avec les curateurs du ROR et un avis de l'ABES (cf. [issue github](#)), le type de relation a été changé en *other* (e.g. ror.org/0459fdx51), ce qui permet de les différencier des unités qui composent un établissement.

Bilan

Ce chantier contribue à la visibilité de l'université dans les outils et services utilisant le ROR, citons par exemple lens.org, [OpenAlex](https://openalex.org), [CWTS Leiden Ranking Open Edition](https://www.cwts.com), ou encore le [works-magnet](#) du MESR. Mais l'essentiel est que ces outils et services sont grandissants : le ROR est devenu la solution industrielle pour identifier à échelle mondiale les organisations de recherche.

Au niveau de l'université, l'enjeu pour la suite est de réaliser ce travail régulièrement et surtout de se coordonner entre responsables des référentiels structures (RNSR, AureHAL, IdRef) et autres systèmes d'information internes.

Nous ne pouvons qu'inviter les établissements à réaliser des chantiers de curation similaires. C'est essentiel pour relier chercheurs, productions et établissements. Côté éditeur par exemple, il est possible de relier ces trois entités directement lors de l'enregistrement du DOI ; CrossRef a ainsi récemment ajouté un indicateur sur l'utilisation du ROR dans les métadonnées produites par un éditeur (voir [Re-introducing Participation Reports to encourage best practices in open metadata](#)).

Larrieu, Maxence (2024). L'UGA au travers du ROR : assurer la qualité des données. *Blog GATES Data SHS*. Université Grenoble Alpes.

←

Top

→

