

Adoption de l'identifiant chercheur ORCID : le cas des universités toulousaines

Marie Dominique Heusse, Guillaume Cabanac

▶ **To cite this version:**

Marie Dominique Heusse, Guillaume Cabanac. Adoption de l'identifiant chercheur ORCID : le cas des universités toulousaines. 38e congrès INFORSID (informatique des organisations et systèmes d'information et de décision), Jun 2020, Dijon, France. pp.19-34. hal-02862519

HAL Id: hal-02862519
https://hal.science/hal-02862519

Submitted on 9 Jun 2020

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.

Adoption de l'identifiant chercheur ORCID : le cas des universités toulousaines

Marie-Dominique Heusse, Guillaume Cabanac

IRIT UMR 5505 CNRS, Université Paul Sabatier, Toulouse, France
{marie-dominique.heusse,guillaume.cabanac}@irit.fr

RÉSUMÉ. *Les systèmes d'information de la recherche collectent et mettent en visibilité la production scientifique des chercheurs. Leur désambiguïsation est capitale pour ne pas fusionner les productions de plusieurs personnes (cas des homonymes). Or, l'initiative ORCID offre un identifiant à chaque chercheur, pointant vers ses affiliations et sa bibliographie. Les agences de financement (ANR et ERC) et les revues savantes encouragent l'adoption d'ORCID. Nous présentons une méthode pour quantifier cette adoption selon la discipline et de la catégorie d'emploi des publiants d'un établissement. La preuve de concept est réalisée sur les données des 6 471 personnels rattachés aux 150 laboratoires du site toulousain. Nous confrontons avec une validation manuelle leur identité aux 7,3 de millions profils d'orcid.org. Nous observons une adoption croissante d'ORCID avec une disparité d'adoption selon les disciplines. Étonnement, des profils sont uniquement créés pour obtenir un ORCID, sans renseigner ni affiliation ni bibliographie. Ces profils « vides » ont peu d'intérêt pour la tâche de désambiguïsation des identités. À notre connaissance, aucune autre étude de cette ampleur n'a été publiée concernant l'adoption d'ORCID sur un site universitaire multidisciplinaire. La méthode proposée est réplicable et de futures études pourront chercher à confronter les situations et les dynamiques d'évolution.*

ABSTRACT. *Research-focused information systems harvest and promote researchers' scientific output. Disambiguating their identities is key not to merge several persons' records (case of homonyms). ORCID offers an identifier to link one's identity, affiliations, and bibliography. Funding agencies (e.g., ANR, ERC) and scholarly journals promote ORCID. We introduce a method to quantify its adoption according to researchers' discipline and occupation in a higher-education organisation. We semi-automatically matched the 6,471 staff members affiliated to the 150 labs of the Toulouse scientific area with the 7.3 million profiles at orcid.org. The increasing ORCID adoption comes with discipline-wise disparities. Unexpectedly, many profiles are void of information and might have been created only to get an identifier. Those empty profiles are of little interest for the entity disambiguation task. To our knowledge, this is the first study of ORCID adoption at the scale of a multidisciplinary scientific metropole. This method is replicable and future studies can target other cases to contrast the dynamics of ORCID adoption worldwide.*

MOTS-CLÉS : *bibliographie, désambiguïsation, ORCID, identifiant chercheur, multidisciplinarité.*

KEYWORDS: *bibliography, disambiguation, ORCID, researcher identifier, multidisciplinarity.*

1. Introduction

La question des identifiants est au cœur de tout système d'information gérant des individus au sein d'une organisation. Les établissements d'enseignement supérieur et de recherche (ESR) n'échappent pas à la règle, compliquée cependant par la diversité même des activités de leurs membres :

– fonctions administratives, d'enseignement, de recherche ;

– et pour la recherche, multiplicité des acteurs/opérateurs avec lesquels les enseignants-chercheurs et chercheurs sont en interaction : tutelles locales et nationales, agences de financement, éditeurs publics et privés, plateformes d'archives ouvertes, réseaux sociaux académiques, etc.

Dans les faits, deux pratiques coexistent :

– les systèmes de gestion des établissements organisent la communication des informations au sein de modules spécifiques : gestion des ressources humaines (SIRH), gestion des formations, gestion financière, en particulier. Des actions sont menées au niveau national pour améliorer la qualité des données dans ces outils, voir par exemple le projet Sinaps [1] ;

– pour l'activité recherche, le paysage est beaucoup plus éclaté. Si de très rares établissements français commencent à s'intéresser à des outils de type *Current research information system* ou CRIS [2], il n'y a pas pour le moment de réalisations à l'instar des universités d'Europe du Nord ou d'Amérique du Nord. Le repérage des productions de la recherche en particulier s'effectue au travers d'une série de bases de données, revues en ligne, archives ouvertes, qui ont chacune leur propre système d'identification des auteurs, spécifique à la base, comme ResearcherId (Web of Science), AuthorId (Scopus), IdHal (HAL) et IdRef (Sudoc). En parallèle se sont développées des initiatives internationales pour créer des normes d'identifiants univoques et pérennes : ISNI [3] et VIAF [4] notamment.

ORCID, acronyme de *Open Researcher and Contributor ID*, est un système international créé en 2012 et construit sur la norme ISNI dans le but d'identifier de manière unique les auteurs de publications scientifiques et académiques (Haak *et al.*, 2012). Porté par une organisation à but non lucratif – mais à laquelle participent les plus grands éditeurs privés – il a été rapidement adopté par les chercheurs eux-mêmes : au 17 octobre 2019 [5] on dénombre 7 314 172 profils créés tandis que l'UNESCO estime

1. http://www.amue.fr/pilotage/logiciels/sinaps/
2. https://www.eurocris.org/
3. http://www.isni.org/
4. http://viaf.org/
5. Export annuel des profils ORCID en XML sous licence CC0 : https://doi.org/10.23640/07243.9988322.v2, voir aussi https://orcid.org/statistics

à 7,8 millions le nombre de scientifiques (Soete *et al*., 2015, p. 32). Les institutions et les revues ont également adopté ORCID : les agences de financement nationales ou européennes ainsi que nombre d'éditeurs exigent désormais cet identifiant pour les soumissions : dépôts de dossiers de demande de financement, soumission d'articles, revues par les pairs (Hanson *et al*., 2016 ; Dunford, Rosenblum, 2018). Par ailleurs, les universités adhèrent à l'organisation ORCID et mettent en œuvre des actions d'incitation pour leurs chercheurs, dans le contexte de la mise en œuvre d'un CRIS [6] ainsi que pour améliorer leurs dispositifs d'archives ouvertes (Brown *et al*., 2016) ou d'évaluation (Haak *et al*., 2018).

Si les avantages de ce dispositif ne font aucun doute, on peut s'interroger sur le degré d'adoption réel par les chercheurs eux-mêmes. Les freins à l'utilisation ne sont en effet pas négligeables : des alternatives existent en matière de mise en visibilité d'un chercheur, plus connues, populaires, ou (mais pas toujours) plus faciles à mettre en œuvre : profil Google Scholar, réseaux sociaux académiques comme Mendeley, ResearchGate, Academia.edu, voire LinkedIn (Tran, Lyon, 2017 ; French, Fagan, 2019). L'argument d'efficacité du référencement (pérennité garantie de l'identifiant ORCID, assurance de l'élimination des doublons) peut aussi entraîner des réticences chez des universitaires attachés à leur liberté académique et hostiles à des dispositifs de repérage de leur activité. Nous avons donc cherché à mesurer le taux d'adoption d'ORCID dans un cadre circonscrit et connu : les enseignants-chercheurs et chercheurs des établissements constituant l'université de Toulouse, sur la période 2013–2017. Au-delà d'une réponse binaire (adoption : oui ou non), nous avons voulu affiner l'analyse en repérant l'évolution dans le temps, le lien éventuel avec la discipline, ainsi que l'utilisation qui était faite des fonctionnalités d'ORCID : renseignement des données d'affiliation et de biographie, alimentation de la notice avec les publications, ou à l'inverse profil « vide ».

La section 2 présente les principales caractéristiques d'ORCID, et notamment celles qui peuvent influer sur les pratiques des adoptants, au moment de l'inscription et ensuite, pour mettre à jour et alimenter les profils. La section 3 présente notre méthode de collecte, de validation et d'analyse des données relatives aux 6 471 personnels du site toulousain. La section 4 présente la dynamique de l'adoption de l'identifiant ORCID par cette communauté universitaire au regard des catégories d'emploi et des pôles disciplinaires. La section 5 discute les résultats inattendus que nos analyses révèlent, avant la conclusion en section 6.

2. Profil ORCID : création, alimentation et visibilité des profils

La création de profil ORCID peut se faire soit de manière individuelle à l'initiative d'un chercheur ou d'une personne mandatée par lui, soit dans le cadre d'une politique mise en œuvre par une institution qui a contractualisé avec ORCID et bénéficie à ce

6. L'université du Colorado est un des cas emblématiques : https://www.colorado.edu/fis/orcid.

titre d'outils permettant d'automatiser un certain nombre de processus [7]. Fin 2019, ORCID comptait 7,8 millions de profils créés de par le monde, et 1 107 institutions partenaires dont 79 % d'universités et organismes de recherche. Plus de la moitié de ces institutions partenaires sont en Europe, cependant la France n'en compte que six : 2 maisons d'édition, 4 organismes de recherche, et donc aucune université. Un accord de consortium entre ORCID et la France a néanmoins été signé en 2019 dans le but de développer le réseau d'établissements affiliés [8].

Au moment de la création d'un profil ORCID, le chercheur choisit une option de visibilité (figure 1). Le profil sera consultable par quiconque (visibilité publique), ou par des tiers de confiance [9] seulement (visibilité restreinte), ou uniquement par son créateur (visibilité privée). Puis il renseigne les données le concernant dans un certain nombre de champs (figure 2) : *Person identifiers*, *Employment*, *Education and qualification*, *Invited positions and distinctions*, *Membership and services*, *Funding*, et *Works*. De nouveaux champs ont été rajoutés récemment : *Peer-review* et *Research resources*.

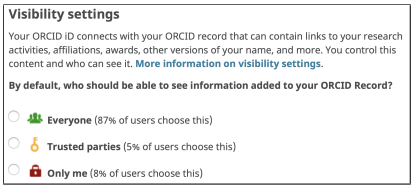


Figure 1. Paramètres de visibilité disponibles lors de la création d'un profil ORCID.

La saisie et notamment la complétude des informations entrées n'étant pas contrôlée, les profils créés peuvent être extrêmement lacunaires. Nous avons rencontré dans les résultats de notre travail un très grand nombre de profils « vides » (section 5.1), sans pouvoir déterminer si c'était parce qu'aucune information n'avait été saisie, ou parce que le chercheur avait fait le choix d'un accès fermé. Cependant, l'information donnée par ORCID (8 % des utilisateurs choisissent la visibilité privée) est bien plus faible que le pourcentage observé à Toulouse, où il semble s'agir surtout de profils non renseignés.

Dans les premières années d'ORCID, la création de doublons (plusieurs profils – donc identifiants – pour un même auteur) était également possible, et nous en avons rencontré dans notre corpus. Cet inconvénient a été repéré et depuis 2017 ce risque de doublons est efficacement détecté au moment de la création d'un nouveau profil [10].

ORCID a assez tôt développé des outils pour faciliter l'alimentation des données sur les publications, et plus récemment sur l'activité d'évaluation par les pairs. Selon le rapport 2019 de l'organisation (Haak *et al*., 2019, p. 15), près de 67 % des 49 millions de notices de publications qui sont rattachées à des profils l'ont été via une API (Scopus, ResearcherId, Pubmed Central, CrossRef, pour l'essentiel).

7. https://members.orcid.org/api/integrate
8. https://www.couperin.org/services-et-prospective/adhesions-consortiales/orcid
9. À qui le créateur octroie ou révoque ce privilège (cas des plateformes de soumission d'article, notamment).
10. https://orcid.org/blog/2014/01/09/managing-duplicate-iDs

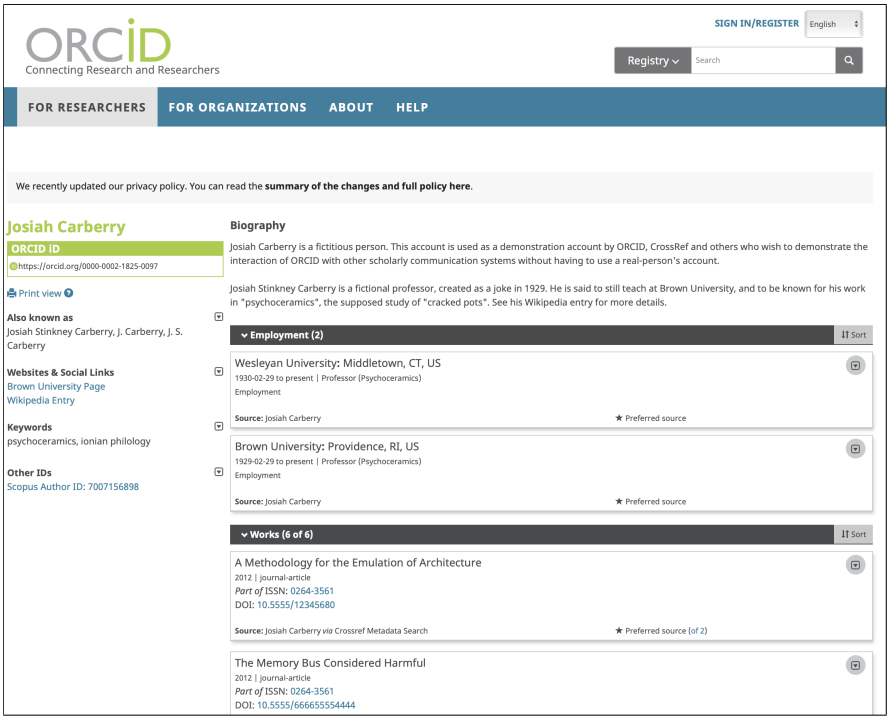


Figure 2. Profil ORCID « modèle » pour le chercheur factice Josiah Carberry.

3. Méthodes et données

Cette section décrit le protocole de collecte des données provenant des établissements ainsi que d'ORCID. Nous détaillons l'appariement automatique de chaque identité de personnel avec les profils ORCID correspondants, puis la démarche de validation manuelle de ces appariements. L'analyse de la base de données originale ainsi constituée permet de révéler la dynamique d'adoption d'ORCID à l'échelle d'une métropole scientifique. C'est à notre connaissance la première étude à cette échelle.

3.1. Collecte des données

Les données sur les 6 471 personnels des laboratoires ont été rassemblées dans le cadre d'un premier travail en 2014–2016 sur la caractérisation de l'activité de recherche du site toulousain (Heusse, 2016). Ce registre des personnels est un tableau comprenant les champs suivants : identité (prénom et nom), établissement, catégorie d'emploi, corps et grade, sexe, année de naissance, laboratoire de rattachement, domaine scientifique. Dans un premier temps, l'ensemble des personnels a été pris en compte, y

compris les personnels BIATSS [11] des catégories A, B et C, car des premiers sondages avaient permis de repérer que certains d'entre eux co-signent des publications et ont un identifiant ORCID. Après rapprochement avec les données ORCID, les personnels de catégorie B et C ont été retirés de la base (130 personnes), à l'exception de ceux identifiés comme publiants, soit 18 personnes. L'ensemble des personnels se répartit en trois grandes catégories d'emploi [12] :

– enseignants-chercheurs ;

– chercheurs : chercheurs des organismes et post-doctorants ;

– autres personnels : ingénieurs d'étude, ingénieurs de recherche, enseignants du second degré, praticiens hospitaliers, notamment.

Les informations, qui sont une agrégation de données fournies par les laboratoires, sont de qualité hétérogène : la date de naissance et le domaine scientifique notamment sont remplis de manière lacunaire, ce qui nous empêche par exemple de repérer pleinement si l'adoption d'ORCID est plutôt le fait de chercheurs jeunes, ou au contraire de ceux qui, montés en responsabilité, l'utiliseraient pour des demandes de financement. Les lacunes sur le domaine scientifique sont compensées par le regroupement des quelques 150 laboratoires du site en six grands pôles disciplinaires, et notre analyse est donc effectuée au niveau de ces ensembles :

– IMI : *Ingénierie, mathématiques et informatique* ;

– SM : *Sciences de la matière* ;

– BSA : *Biologie, santé, agronomie* ;

– SHS : *Sciences humaines et sociales* ;

– DEG : *Droit, économie, gestion* ;

– STU : *Sciences de la Terre et de l'univers*.

La figure 3 montre la diversité des configurations, avec un poids différent de chaque catégorie selon les pôles : les personnels relevant de *Sciences humaines et sociales* et *Droit, économie, gestion* sont caractérisées par la très faible part de chercheurs (organismes et post-doctorants) et de personnels BIATSS en leur sein. Les chercheurs sont à l'inverse légèrement majoritaires en *Biologie, santé, agronomie* et en *Sciences de la Terre et de l'univers*. Enfin, l'importance des autres personnels en BSA s'explique partiellement par le poids de la recherche clinique et le rôle des personnels du CHU dans celle-ci.

Pour chaque identité des 6 471 personnels, nous avons interrogé l'API publique [13] d'ORCID pour obtenir le ou les profils correspondants à partir de la requête composée du prénom et du nom du personnel en question. Certains résultats sont pléthoriques : c'est le cas de la requête « Philippe Durand » restituant 4 534 profils dont 5 profils

11. Personnels ingénieurs, administratifs, techniques, sociaux et de santé et des bibliothèques.

12. Dans la suite de l'article, le terme générique « chercheur » sera utilisé pour désigner l'ensemble des personnels, sauf lorsqu'il s'agit de repérer les situations propres à une catégorie spécifique.

13. https://orcid.org/organizations/integrators/API

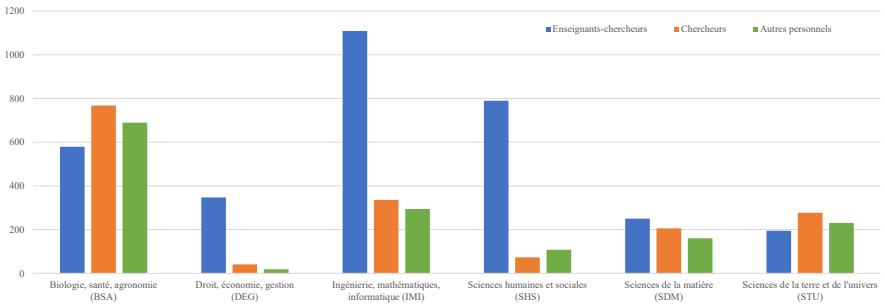


Figure 3. Répartition des 6 471 personnels par pôle et par catégorie d'emploi.

d'homonymes parfaits en tête de liste puis des appariements partiels, tels que « Philippe S. Durand » ou même « Romain Durand ». [14] Nous avons par la suite traité les 20 premiers profils restitués au plus. Chaque profil est un fichier XML [15] contenant, notamment, la date de création du profil, sa date de dernière mise à jour, l'identité déclarée, une biographie, des affiliations et des publications. Le contenu du fichier XML reflète les informations présentées sur le profil consultable en ligne (figure 2).

Nous avons cherché dans les profils une évocation du site toulousain pour désambiguïser les homonymes : ses lieux (Toulouse, Albi, Castres, Rodez et Tarbes) et le nom de ses universités (Capitole, Mirail devenue Jean Jaurès et Paul Sabatier). Cet indicateur était ensuite mobilisé en phase de validation manuelle détaillée dans la section suivante.

3.2. Validation manuelle de la jointure entre registre des personnels et ORCID

Une fois l'interrogation via l'API d'ORCID réalisée, la base de données constituée proposait, pour chaque individu, un des cas de figure suivants :

- aucune correspondance dans la base d'identifiants d'ORCID,
- une seule correspondance trouvée,
- de 2 à 20 correspondances trouvées.

Une validation a été effectuée quand au moins une correspondance a été détectée. La validation était facilitée par l'extraction de données biographiques et d'affiliation lorsqu'elles étaient présentes, ou par le repérage des affiliations sur les articles quand ceux-ci étaient indiqués. La difficulté a été plus grande avec les profils « vides » : une quasi-certitude a pu être établie dans le cas d'un groupe prénom et nom très discriminant, avec au besoin vérification de cohérence de la discipline et de l'absence de deux chercheurs portant le même nom via une interrogation de Google Scholar.

14. Nous avons observé que certaines identités fournies par les laboratoires diffèrent parfois des identités de signature des articles : cas des noms de naissance *versus* noms d'usage, notamment. De fait, l'interrogation par appariement exact (avec guillemets, c.-à-d. "Philippe Durand") n'est pas judicieuse pour ces cas.

15. https://members.orcid.org/api/tutorial/reading-xml

Cette quasi-certitude a été écartée dans le cas de noms très répandus. L'étiquetage de l'appariement a utilisé les codes suivants :

– Pas de numéro ORCID correspondant à l'identité du personnel recherchée,

– `0` : il ne s'agit pas de la même personne (cas des différences entre le groupe prénom et nom recherché *versus* restitué par l'interrogation d'ORCID),

– `1` : le profil ORCID permet de valider l'appariement,

– `?` : profil ORCID vide, mais un seul chercheur correspond après vérification dans Google Scholar,

– `??` : le groupe prénom et nom restitué correspond à l'identité recherchée, mais plusieurs chercheurs correspondent et le profil vide ne permet pas de trancher.

On a donc considéré comme adoptants d'ORCID les personnels étiquetés avec un code `1` ou `?`, soit 2 580 personnes. Il faut noter par conséquent que le corpus retenu représente une fourchette basse de l'adoption d'ORCID : des identifiants ORCID du groupe `??` peuvent correspondre de fait à des chercheurs du site.

4. Résultats quantitatifs

Le taux d'adoption d'ORCID pour l'ensemble de la population concernée, mesuré début mars 2020, est de 39,87 %. En excluant les « autres personnels », ce taux est de 44,6 %. Ces deux taux recouvrent des disparités importantes que nous avons souhaité mettre en lumière dans les sections suivantes.

4.1. Adoption de l'identifiant ORCID par catégorie d'emploi

Comme il apparaît dans la figure 4, les chercheurs (non enseignants) sont sans grande surprise les adoptants majoritaires (à 55 %). Le taux plus faible des enseignants-chercheurs (39,2 %) s'explique sans doute en partie par leur surreprésentation dans les disciplines de *Sciences humaines et sociales* et de *Droit, économie, gestion* où les exigences des revues et des tutelles vis-à-vis d'ORCID sont plus faibles. Pour les « autres personnels » (24,2 %), s'ils prennent part à la recherche menée dans les équipes, peu d'entre eux sont associés aux publications. Rappelons qu'il s'agit ici, pour les BIATSS, des catégorie A et des 18 publiants de catégorie B et de catégorie C, ainsi que des praticiens hospitaliers et des professeurs du second degré, principalement.

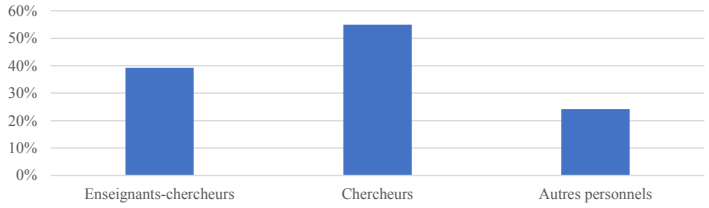


Figure 4. Pourcentage d'adoption d'ORCID par catégorie d'emploi.

4.2. Adoption de l'identifiant ORCID par pôle disciplinaire

Le taux d'adoption par pôle disciplinaire (figure 5) est en cohérence avec ce qu'on sait des pratiques de publication, des politiques des revues dans les différents grands champs disciplinaires, et de l'usage des identifiants plus ancien et plus fort en sciences exactes. L'analyse est à compléter cependant par la figure 8, qui représente les courbes respectives d'adoption dans le temps pour chacun des pôles.

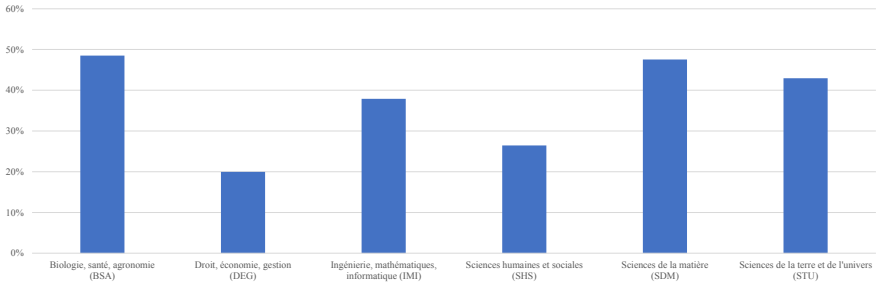


Figure 5. Pourcentage d'adoption d'ORCID par pôle disciplinaire.

La figure 6 combine les taux d'adoption par pôle et par catégorie d'emploi. Elle permet de constater que le taux d'adoption plus élevé pour les chercheurs se retrouve de façon assez homogène dans les différents pôles disciplinaires, même s'il est plus faible que la moyenne en *Droit, économie, gestion* : il faut sans doute y voir l'impact des consignes des organismes de recherche en la matière [16]. Pour les enseignants-chercheurs, elle montre nettement le plus faible pourcentage en SHS et DEG, déjà évoqué plus haut. Les autres personnels ont des taux significativement plus élevés en BSA et SDM, et dans une moindre mesure en STU : dans ces pôles, plusieurs laboratoires ont été repérés comme associant des personnels BIATSS comme cosignataires des articles de recherche (section 5).

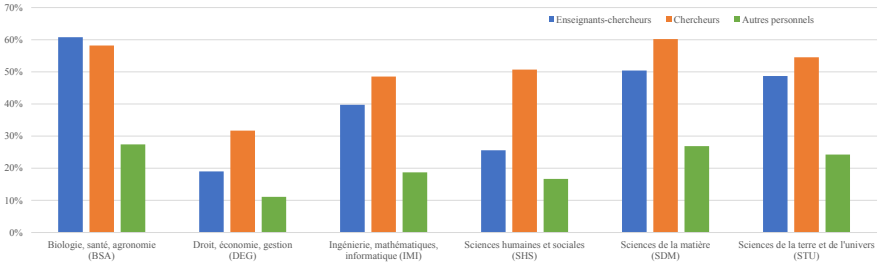


Figure 6. Pourcentage d'adoption d'ORCID par pôle disciplinaire et par catégorie d'emploi.

16. Fiche DORANUM https://doranum.fr/wp-content/uploads/Fiche-ORCID.pdf

4.3. Adoption progressive de l'identifiant ORCID : analyse longitudinale

4.3.1. Évolution globale

La figure 7 montre que loin d'être linéaires, les créations s'articulent sur le calendrier universitaire avec les échéances de dépôts de projets ANR [17], et que cette tendance paraît se renforcer avec les années, les mois d'octobre constituant des pics de création de manière de plus en plus visible. Cette tendance est probablement à rapprocher de la pratique de création de profils « vides », la motivation du chercheur étant principalement d'obtenir un identifiant pour compléter le dossier de demande de financement.

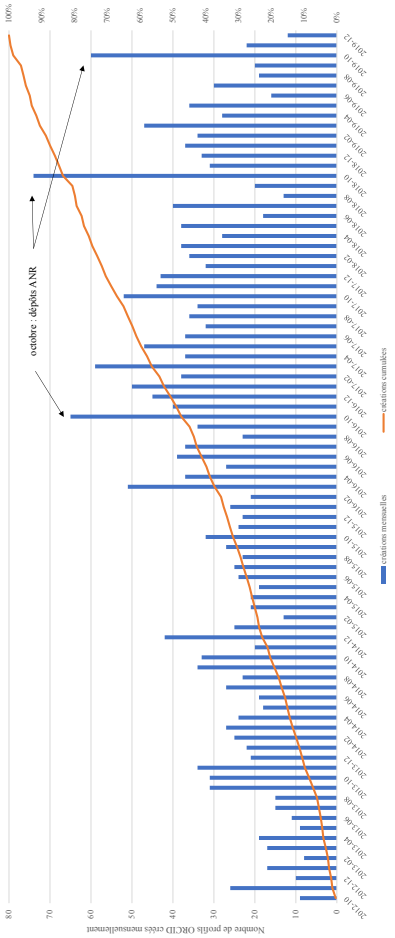


Figure 7. Créations de profils ORCID au fil du temps.

17. cf. page 20 du guide : https://anr.fr/fileadmin/aap/2019/aapg-anr-2019-Guide.pdf

4.3.2. Évolution par pôle disciplinaire

La figure 8 représente la courbe d'évolution des créations de profils ORCID pour chaque pôle disciplinaire. Les formes respectives des courbes sont dans l'ensemble homogènes, mais permettent cependant de repérer certaines caractéristiques :

– les disciplines STU et BSA ont été au début (et sur presque toute la période pour STU) les plus actives dans les créations,

– d'autres, comme SDM, sont parties moins vite, mais ont connu une forte accélération dans les dernières années. C'est aussi le cas de SHS, marqué par des taux faibles au départ, mais qui a fini par combler ce décalage.

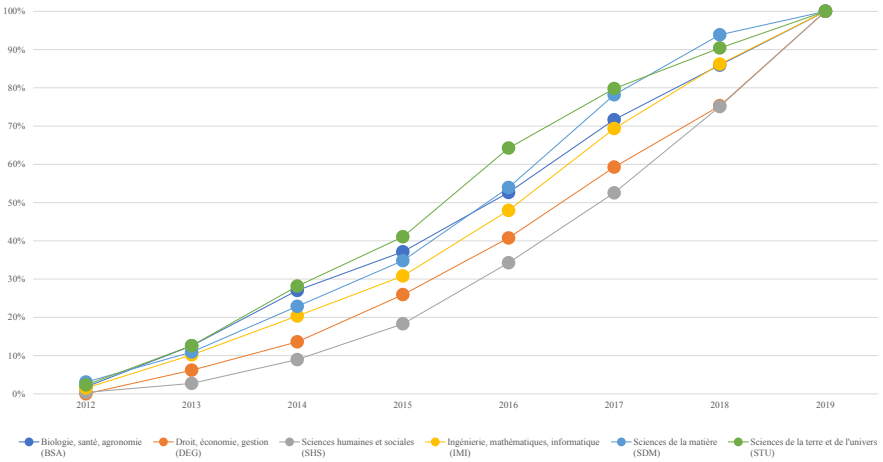


Figure 8. Part cumulée d'adoption d'ORCID par pôle disciplinaire. La part des profils correspondant à $y = 100\,\%$, pour un pôle disciplinaire donné, est présentée en figure 5.

5. Observations qualitatives

Cette section complète l'étude quantitative des données par des observations qualitatives liées à la diversité des usages des profils ORCID. Nous commentons également quelques cas de mésusages notables puis réalisons un focus sur la catégorie d'emploi des « autres personnels ».

5.1. Diversité des (més)usages des profils ORCID

On constate une très grande variété dans les modes d'appropriation et d'usage d'ORCID. Rappelons que l'inscription se fait à l'initiative du chercheur, et qu'il a la possibilité de renseigner les champs énumérés à la section 2. Des profils ORCID détaillés comprenant à la fois des données biographiques et de publications ne sont cependant pas la majorité : certains indiquent seulement les éléments d'affiliation,

d'autres uniquement les publications. Surtout, nombre de profils sont complètement vides [18] (figure 9).



Figure 9. Exemple d'un profil ORCID « vide ».

Comme indiqué précédemment, les profils vides ont deux origines : profil non renseigné ou profil dont la visibilité est privée. Répartir les profils de notre corpus dans ces deux catégories n'est pas possible techniquement. Indiquons seulement que la fréquence des profils vides dans notre corpus s'élève à 40 %, alors qu'ORCID indique qu'il y a 8 % de profils privés (figure 1). Il serait intéressant de repérer la variété des pratiques, du profil complet au profil vide, et les priorités éventuellement données par les chercheurs aux données d'affiliation ou de publications. À titre d'exemple, la figure 10 présente le pourcentage de profils ORCID listant au moins une publication, par pôle disciplinaire.

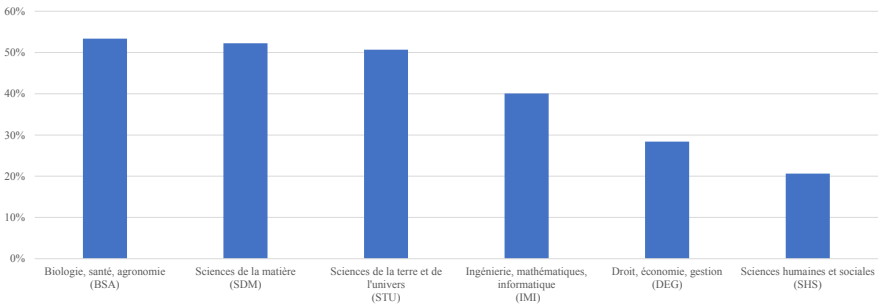


Figure 10. Part des profils ORCID listant au moins une publication.

18. Il y a, par exemple, cinq profils distincts pour l'identité « Philippe Durand » dans ORCID, tous vides…

Nous avons aussi observé des pratiques étonnantes, aussi bien en matière d'affiliation que de signalement des publications, sans toutefois que ces pratiques aient été quantifiées :

– les sections *Biography* et *Employment* du profil, qui permettent de repérer les mobilités dans la carrière d'un chercheur et donc l'évolution de ses affiliations, sont parfois détournées pour mentionner les universités étrangères avec lesquelles il a des liens d'échange mais qui ne sont pas ses « employeurs » à proprement parler. Il arrive même que la rubrique énumère seulement ces différents établissements prestigieux (cas des universités américaines, notamment) mais sans indiquer celui de rattachement (c'est-à-dire l'employeur). Pour les chercheurs rattachés à des organismes nationaux, d'autre part, il n'est pas rare de trouver « CNRS Paris » ou « INSERM Paris », sans aucune référence à leur implantation régionale ni à leur laboratoire sur le site – que l'on trouve seulement en consultant les articles, lorsqu'ils sont présents.

– le nombre de publications que les chercheurs ajoutent à leur profil ORCID sous la section *Works* varie d'une seule à plusieurs centaines. À l'évidence, certains ne maîtrisent pas les processus d'alimentation automatique de cette section, et la qualité des données en pâtit (pas de date de publication, pas de DOI…). Nous avons réalisé des sondages pour mesurer l'écart entre la production réelle d'un auteur et le nombre d'articles signalés dans ORCID, et cet écart peut atteindre, lui aussi, plusieurs centaines d'articles.

La création de plusieurs profils ORCID par la même personne, non contrôlée jusqu'en 2017 [19], a été également repérée. Nous avons quantifié cette pratique, qui reste cependant très minoritaire : il y a quinze profils ORCID correspondant à des doublons certains dans notre corpus : 8 cas de doublons entre en un profil renseigné et un profil vide, 7 cas entre deux profils vides. Plus surprenante est la découverte de 3 autres doublons où les deux profils sont à chaque fois renseignés pour la même personne – avec une quantité de données inégale, cependant.

Une hypothèse d'interprétation de la diversité de ces pratiques serait la « concurrence » entre les différents dispositifs de mise en visibilité de la production des chercheurs, et le fait pour eux d'en être plus familiers ou plus convaincus de leur efficacité. Pour explorer cette hypothèse sur une échelle très réduite, on a mis en regard utilisation d'ORCID *versus* création d'un profil Google Scholar pour 15 chercheurs du site distingués au cours des années récentes comme *Highly Cited Researchers* par Clarivate Analytics [20] (tableau 1).

On le voit : aucune corrélation n'apparaît avec certitude, même pour des chercheurs familiers des consécrations comme ceux du palmarès de Clarivate Analytics.

19. Depuis, le contrôle est réalisée sur l'adresse email, voir note 10.

20. https://recognition.webofsciencegroup.com/awards/highly-cited/2019/ et années antérieures.

Tableau 1. Présence en ligne sur ORCID et Google Scholar de 15 chercheurs listés comme Highly Cited Researchers *(HCR) par Clarivate Analytics. La complétude des profils ORCID est indiquée : existence d'un profil, renseignement de la bibliographie, de l'affiliation et des publications.*

n°	discipline	ORCID				Google Scholar
		profil	biographie	affiliation	publications	
1	agronomie	×		×	×	×
2	biologie	×		×		
3	écologie					×
4	écologie	×		×		×
5	économie					×
6	ingénierie	×		×	×	×
7	santé	×				×
8	santé	×			×	×
9	santé	×	×		×	
10	santé					
11	santé	×				
12	santé	×				
13	santé	×		×	×	
14	santé	×			×	
15	terre et espace	×				×

5.2. Focus sur les créations de profils ORCID par les « autres personnels »

Moins de 25 % des individus de la catégorie d'emploi « autres personnels » a créé un ORCID au cours de la période (section 4.1). Ce taux global recouvre des disparités, et celles-ci sont très clairement liées aux pratiques de publication, et d'association – ou pas – de l'ensemble des personnels ayant pris part à la recherche en qualité de cosignataires de l'article. Nos données permettent de repérer les disciplines qui semblent considérer cette association comme normale : en tête vient le pôle *Biologie, santé, agronomie* où cette pratique se rencontre dans 7 laboratoires sur les 30 qui constituent le pôle, et représente visiblement une caractéristique d'un certain nombre de recherches dans le domaine de la santé ; on la trouve ensuite dans les différents laboratoires du pôle *Sciences de la Terre et de l'univers*, dans plusieurs laboratoires du pôle *Sciences de la matière* (chimie, matériaux…) et enfin dans le pôle *Ingénierie, mathématiques et informatique* (robotique, mécanique des fluides). La création d'un ORCID par ces

personnels est un corollaire logique de la reconnaissance qui leur est accordée par les autres chercheurs du laboratoire.

6. Conclusion et perspectives

L'identifiant ORCID introduit en 2012 pour aider à la désambiguïsation des identités de chercheurs s'est progressivement inscrit dans le paysage de l'enseignement supérieur et de la recherche. Cet article est la première étude de l'adoption d'ORCID sur le périmètre d'une métropole scientifique majeure à l'échelle nationale. Analysant les profils des 6 471 personnels du site toulousain, notre étude montre une dynamique d'adoption croissante quelle que soit la discipline des chercheurs. À l'heure actuelle près de 40 % de la population étudiée possède un profil ORCID. L'analyse par pôle disciplinaire montre des disparités d'adoption : entre 60 % des enseignants-chercheurs en *Biologie, santé, agronomie* et 20 % des enseignants-chercheurs en *Droit, économie et gestion*. Plusieurs raisons peuvent expliquer ces différences, dont les recommandations voire injonctions inégales des établissements, des sociétés savantes, des agences de financement et le degré d'internationalisation des champs de recherche. Une analyse qualitative de l'adoption d'ORCID, pour les 60 % des profils renseignés, montre qu'ils sont partiellement remplis : les sections disponibles pour décrire la biographie, la trajectoire institutionnelle, les autres identifiants, les financements obtenus… sont très souvent vides et il manque bien souvent de nombreuses publications. Par ailleurs, nous avons observé des mésusages dans la façon d'utiliser et de renseigner les différents champs de la notice ORCID, l'une des omissions les plus critiques étant l'information sur *Employment*, c'est-à-dire l'affiliation du chercheur. Ces observations questionnent quant à la perception de la finalité de l'identifiant ORCID par une partie de la communauté savante.

Il est envisagé de mener des travaux plus qualitatifs pour compléter ce premier travail : explorer via des enquêtes les pratiques et les comportements des usagers, estimer le taux de complétude des publications signalées dans ORCID en comparant avec le Web of Science par exemple ou avec le CV exhaustif fourni par le chercheur lui-même (Youtie *et al*., 2017), repérer l'impact des stratégies d'établissement (par ex., dépôt des publications de l'INRA dans son archive ouverte institutionnelle ProdINRA [21]).

Étant donnée la politique actuelle menée par les opérateurs de l'enseignement supérieur et de la recherche visant à favoriser l'adoption généralisée d'ORCID au sein des établissements, cette étude permet d'estimer l'état actuel de l'adoption tout en soulignant les écueils rencontrés à ce jour, et permet donc de dégager des orientations pour les actions à conduire.

21. http://wiki.inra.fr/wiki/prodinra/Qualite/ORCID+ResearcherId

Bibliographie

Brown J., Demeranville T., Meadows A. (2016). Open access in context: Connecting authors, publications and workflows using ORCID identifiers. *Publications*, vol. 4, n° 4, p. 30. doi:10.3390/publications4040030

Dunford R., Rosenblum B. (2018). Keeping it authentic: Reconciling ORCID iDs gathered at submission with the author manuscript. *Learned Publishing*, vol. 31, n° 3, p. 236–240. doi:10.1002/leap.1159

French R. B., Fagan J. C. (2019). The visibility of authority records, researcher identifiers, academic social networking profiles, and related faculty publications in search engine results. *Journal of Web Librarianship*, vol. 13, n° 2, p. 156–197. doi:10.1080/19322909.2019.1591324

Haak L. L., Fenner M., Paglione L., Pentz E., Ratner H. (2012). ORCID: A system to uniquely identify researchers. *Learned Publishing*, vol. 25, n° 4, p. 259–264. doi:10.1087/20120404

Haak L. L., Meadows A., Brown J. (2018). Using ORCID, DOI, and other open identifiers in research evaluation. *Frontiers in Research Metrics and Analytics*, vol. 3. doi:10.3389/frma.2018.00028

Haak L. L., Petro J. A., Simpson W., Demeranville T., Wijnbergen I., Hershberger S. *et al.* (2019). *ORCID 2019 annual report*. Rapport technique. ORCID. Consulté sur https://doi.org/10.23640/07243.12009153.v1

Hanson B., Lawrence R., Meadows A., Paglione L. (2016). Early adopters of ORCID functionality enabling recognition of peer review: Two brief case studies. *Learned Publishing*, vol. 29, n° 1, p. 60–63. doi:10.1002/leap.1004

Heusse M.-D. (2016). Faire parler les données sur la recherche grâce au web sémantique : le projet VIVO. In *Actes de l'atelier Valorisation et Analyse des Données de la Recherche (VADOR) organisé dans le cadre du congrès INFORSID*, p. 19–25.

Soete L., Schneegans S., Eröcal D., Angathevar B., Rasiah R. (2015). A world in search of an effective growth strategy. In S. Schneegans (Ed.), *UNESCO Science Report: Towards 2030*, p. 20–55. Paris. Consulté sur http://unesdoc.unesco.org/images/0023/002354/235406e.pdf

Tran C. Y., Lyon J. A. (2017). Faculty use of author identifiers and researcher networking tools. *College & Research Libraries*, vol. 78, n° 2, p. 171–182. doi:10.5860/crl.78.2.171

Youtie J., Carley S., Porter A. L., Shapira P. (2017). Tracking researchers and their outputs: New insights from ORCIDs. *Scientometrics*, vol. 113, n° 1, p. 437–453. doi:10.1007/s11192-017-2473-0