

ORCID, la base internationale d'identifiants de chercheurs, passe le cap des 1 600 000 inscrits



Analyse I/IST-n°4

L'information :

Comme chaque année depuis son lancement en octobre 2012 ORCID (Open Researcher and Contributor Identification Data, www.orcid.org) vient de rendre accessible sous licence Creative Commons (CC0 1.0) son fichier central d'identifiants normalisés de chercheurs. Cette version 2015 compte près de 1 700 000 chercheurs répertoriés : ORCID a désormais atteint la taille critique qui devrait lui permettre de s'imposer rapidement comme le répertoire de métadonnées d'identification des chercheurs au plan mondial. ORCID a dans la période récente multiplié les annonces d'intégration dans des services d'éditeurs (Springer, ProQuest, Biomed Central ...) et signé des accords globaux en Italie, en Grande-Bretagne, pays qui visent un taux de 80% de leurs chercheurs disposant d'un identifiant ORCID dans les deux ans à venir. L'accord passé fin octobre entre ORCID, Crossref et DataCite apparaît particulièrement important : les enregistrements ORCID listant les publications d'un chercheur donné seront mis à jour automatiquement chaque fois que Crossref (attribution de DOI permanents aux publications scientifiques) ou DataCite intégrera une référence enrichie d'un identifiant ORCID. Le chercheur devrait être le premier bénéficiaire de cette possibilité de mettre à jour automatiquement un répertoire de ses publications et contributions.

L'analyse d'Intelligence IST :

Sur le papier l'intérêt pour chaque chercheur d'être désigné par un identifiant numérique non ambigu et persistant dans le temps est évident : en permettant de lever toute incertitude (introduite par des biais d'homonymies) sur l'identité d'un chercheur publiant, un identifiant du type ORCID permet de rendre à César ce qui est à César. Cependant pour qu'un identifiant de ce type soit vraiment utile, il doit être très largement adopté : un « numéro de sécu » qui n'identifierait qu'un ayant-droit sur trois aurait un intérêt limité. La notion de « masse critique » est donc un paramètre essentiel de la crédibilité du projet ORCID. Ce dernier s'inscrivant dans une perspective multidisciplinaire (identifier les chercheurs quelle que soit leur discipline d'appartenance) et globale (identifier les chercheurs quel que soit leur pays d'origine) la barre est placée très haut : il s'agit pour ORCID de se rapprocher le plus possible de la notion de « répertoire universel ». Selon l'OCDE on compte au plan mondial 7,8 millions de chercheurs (Eurostat en dénombre 2,55 dans l'Europe des 28). Il y a donc encore une marge de progression importante pour ORCID. Mais sur la lancée de la progression récente, il n'est pas utopique de penser que le répertoire ORCID puisse compter 5 millions d'identifiants de chercheur à échéance de trois ans. L'année 2015 a vu le nombre d'annonces autour de l'adoption d'ORCID se multiplier : Springer vient d'annoncer (le 3 novembre) qu'il introduira chaque fois qu'ils sont disponibles les identifiants ORCID dans les métadonnées relatives aux livres scientifiques qu'il publie, et ce en distinguant les rôles d'« editor » (en charge de la cohérence éditoriale et scientifique d'un ouvrage) et les rôles d'auteurs de chapitres. Cette diversification des types de publications intégrant un identifiant ORCID concerne de nombreux éditeurs. ProQuest l'introduira dans les métadonnées descriptives des thèses et dissertations qu'il met en ligne (annonce du 13/10). Biomed Central (annonce du 5/10) va lancer un projet pilote sur sa revue GigaScience (dédiée aux applications du « Big Data » à tous les champs de recherche scientifique) qui permettra dans la liste de contributeurs à un article de distinguer les différents rôles de ces contributeurs : rédaction des articles, vérification des résultats, édition des jeux de données, etc. L'objectif est de prendre en compte des contributions à la publication scientifique qui ne sont pas forcément reconnues dans le champ auteurs. Dans le même ordre d'idée Publons, une start-up qui développe des bases de données d'identification de « peer reviewers » introduira chaque fois qu'ils sont disponibles les identifiants ORCID de façon à ce que les tâches de reviewing soient prises en compte dans les curriculum scientifiques des chercheurs. Ces quatre annonces illustrent l'un des avantages potentiels d'un identifiant comme ORCID : garder trace de tous types de contributions (et pas seulement des articles publiés) afin de construire au fil du temps un curriculum qui reflète plus fidèlement les différents apports du chercheur dans le processus de la communication scientifique.

Un autre levier de croissance d'ORCID est la signature d'accords à couverture nationale. En juin dernier le JISC britannique, qui mutualise outre-Manche les achats de ressources documentaires numériques a conclu avec ORCID un accord pour le compte d'une cinquantaine d'universités britanniques. A la même date c'est la conférence des recteurs d'universités italiennes qui a conclu avec ORCID un accord qui devrait assurer qu'à la fin 2016 80% des chercheurs italiens disposeront d'un identifiant ORCID, un effort d'enrichissement rétrospectif permettant que les publications depuis 2006 de ces chercheurs soient enrichies par cet identifiant. L'exemple britannique ou italien montre que ce sont de plus en plus les entités impliquées dans la recherche qui sont intéressées par l'adoption d'un identifiant « universel ». Les avantages qu'elles en escomptent ne se limitent pas au fait de garder la trace des publications : les logiques d'évaluation, de reconstitution de carrière, peuvent bénéficier du recours à un identifiant des chercheurs stable dans le temps et non ambigu.

L'affiliation à ORCID est toutefois in fine gérée par le chercheur lui-même (et cette inscription est pour lui gratuite). Il est donc essentiel que le chercheur perçoive clairement les bénéfices qu'il peut tirer de cette affiliation. L'accord signé le 27 octobre entre ORCID et Crossref d'une part, DataCite d'autre part devrait contribuer puissamment à cette motivation. Crossref est on le sait le consortium attribuant aux articles scientifiques un DOI unique et stable dans le temps, DataCite fait de même pour identifier les jeux de données issus de la recherche. Si l'auteur d'un article référencé dans Crossref où le contributeur d'un jeu de données référencé dans DataCite dispose d'un identifiant ORCID, sa fiche personnelle ORCID sera automatiquement mise à jour et enrichie des références Crossref et DataCite. La possibilité de construire ainsi de façon automatique et incrémentale un répertoire de ses contributions scientifiques sera pour le chercheur un atout appréciable.

Le consortium ORCID reste une structure légère (mobilisant une vingtaine de personne). Son modèle économique repose sur les contributions des entités (éditeurs, agences de financement de la recherche, universités, organismes de recherche) qui soutiennent le projet. La production de grands référentiels d'identification (des chercheurs, des organismes de recherche, des labos...) est une condition préalable au développement de l'« e-science ». Ces référentiels doivent être des référentiels ouverts et non propriétaires pour satisfaire aux objectifs d'une science ouverte (par exemple pour qu'ils soient interopérables avec les outils de workflow éditorial mis en œuvre par les éditeurs). En quelques années ORCID est en train de faire la démonstration que dans une logique collaborative il n'est en rien utopique de construire un « répertoire universel ».

ORCID s'acclimatera-t-il dans le contexte français, au-delà de l'adhésion spontanée des chercheurs ? Fin août la direction Europe d'ORCID présentait à l'ABES les grandes lignes des accords conclus en Grande-Bretagne, en Italie, mais aussi aux Pays-Bas, au Danemark et au Portugal. Un prélude à une réflexion sur l'adoption consortiale d'ORCID dans le contexte français ?