

ALIGNER LES DONNÉES DES CHERCHEURS DE MON ÉTABLISSEMENT SUR IDREF



Identifiant pivot et agrégateur de ressources bibliographiques, IdRef permet d'identifier et de valoriser la production des chercheurs. La preuve par l'exemple avec le cas de Nantes Université.

Qui est affilié à mon établissement ? Qui a publié quoi dans HAL ? Pour valoriser la production de ses chercheurs dans HAL ou mesurer l'adoption d'ORCID, chaque établissement se pose ces questions. L'Abes tente d'y répondre en s'appuyant sur IdRef en tant qu'identifiant pivot et agrégateur de ressources bibliographiques, et ce même quand une base tierce n'exploite pas nativement et systématiquement IdRef, ce qui est le cas de HAL.

PRENONS L'EXEMPLE DE L'UNIVERSITÉ DE NANTES¹ ET SUIVONS 5 CHERCHEURS ET 4 PUBLICATIONS PRÉSENTES DANS LE PORTAIL HAL NANTES UNIVERSITÉ CAR ELLES SONT RATTACHÉES À DES UNITÉS DE RECHERCHE NANTAISES.

- V**
 Chercheur junior, travaille à l'université de Nantes depuis 2022 dans le même laboratoire que chercheuse W
- W**
 Chercheuse W, travaille à l'université de Rennes dans le même laboratoire que chercheur V
- X**
 Chercheur X, travaille à l'université de Nantes depuis 2000
- Y**
 Chercheur Y, a travaillé à l'université de Nantes de 2000 à 2020 depuis il est en poste à Sorbonne
- Z**
 Chercheuse Z exerce à Lille

PUBLICATION 1

Chercheur V
Univ Nantes - Labo 1

Chercheuse W
Univ Rennes - Labo 1

PUBLICATION 2

Chercheur V
Univ Nantes - Labo 1

Chercheur Y
Sorbonne Univ - Labo 2

PUBLICATION 3

Chercheur Y
Univ Nantes - Labo 1

Chercheuse Z
Univ Lille - Labo 3

PUBLICATION 4

Chercheur X
Univ Nantes - Labo 4

Chercheuse Z
Univ Lille - Labo 3

[1] Comme d'autres, Nantes a connu une évolution institutionnelle en 2022 passant de Université de Nantes (1962-2021) www.idref.fr/026403447/id à Nantes Université (2022-....) www.idref.fr/258086599/id Les travaux d'alignements évoqués dans cet article concerne les effectifs et publications de l'université de Nantes au sein de l'établissement expérimental Nantes Université.

LA MÉTHODE « ANNUAIRE »

Cette méthode exploite IdRef pour répondre à la question : « Qui [parmi les personnes de l'établissement] a quels identifiants auteur ? »

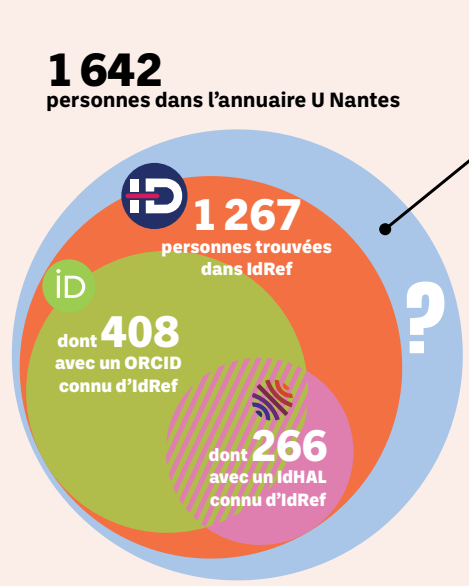
Données initiales fournies par l'établissement : un fichier csv avec une ligne par personne appartenant à l'établissement. A minima : nom | prénom | date de naissance² | laboratoire. Les chercheurs V et X sont dans ce fichier ; W et Y (et évidemment Z) n'y sont pas.

Source des données : extraction du SI RH, LDAP, collecte opérée en amont par le SCD et/ou la direction de la recherche, etc.

Volumétrie dans l'exemple de Nantes : 1 157 maîtres de conférences et professeurs, 86 doctorants, 54 post-doctorants, et 345 personnes aux statuts variés. Le SCD³ a choisi une liste cumulative (les personnels de l'université de Nantes hier et aujourd'hui) avec une interprétation très large de la notion de chercheur.

Résultat du travail de l'Abes : le même fichier csv enrichi des colonnes suivantes: candidat IdRef | ORCID connu dans IdRef / IdHAL connu dans IdRef.

Quand et à quelle fréquence utiliser cette méthode ? Chaque année pour couvrir les départs et arrivées des personnels ! Après une première fois, le suivi annuel est une routine légère pour l'établissement.

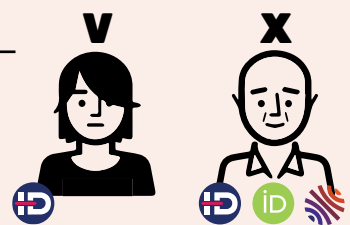


LE TRAVAIL À FAIRE POUR LE CORRESPONDANT AUTORITÉS :

- > **367 personnes pour lesquelles il faut investiguer**
 - soit mettre en attente (les doctorants)
 - soit créer un IdRef
- > **8 personnes à desambiguer**

RÉSULTAT

Chaque personne de l'établissement supposée avoir un IdRef en dispose. L'établissement ne gère plus un annuaire avec des chaînes de caractères mais un annuaire d'identifiants.



[2] Bien que la loi 3 DS du 21 février 2022 et ses décrets d'application posent le principe de l'échange des données entre administrations, l'Abes a constaté des réticences de certains délégués à la protection des données concernant la date de naissance. Nous pouvons vous aider à les lever si besoin.

[3] Notamment Guillaume Godet et Natacha Martin : merci à eux !

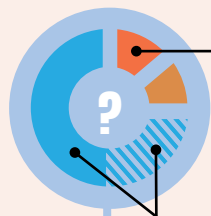
LA MÉTHODE « PORTAIL HAL »

Cette méthode exploite IdRef et HAL pour répondre à la question : « Qui [qualifié par son IdRef] a publié quoi dans HAL ? »

66 535

signatures présentes dans 30 604 publications déposées dans HAL - U Nantes

LE TRAVAIL A FAIRE POUR L'ADMINISTRATEUR PORTAIL ET/OU LE CORRESPONDANT AUTORITÉS :



Des propositions faibles :

644 signatures pour 349 personnes connues d'IdRef

Des propositions fortes :

765 signatures à désambigüiser pour 139 personnes connues d'IdRef

Aucune proposition

2 925 signatures facilement et rapidement étendues à des personnes connues d'IdRef

Environ **10 000** signatures correspondent à **3 987 signatures dédoublonnées** pour lesquelles il faut investiguer

Données initiales et source : l'Abes est autonome pour récupérer les données HAL une fois confirmation par l'établissement de la requête à passer à l'API de HAL pour cibler son périmètre institutionnel.

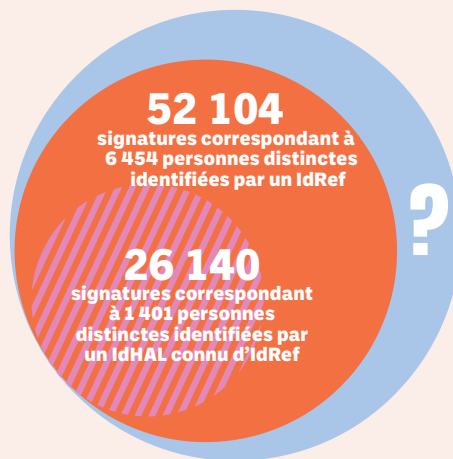
Volumétrie dans l'exemple de Nantes : le portail HAL Nantes Université restreint à la composante de l'université initiale i.e. les métadonnées des 30 604 publications déposées depuis l'origine soit 66 535 signatures.

Résultat du travail de l'Abes : le programme d'alignement exploite le matériau bibliographique brut et privilégie certains critères (coauteurs, titres, discipline...).

Pour HAL Nantes Université, 6 454 personnes distinctes ont été repérées. L'exploitation de l'IdHAL ne permet d'identifier que 1 401 personnes. En faisant un détour par le document, l'Abes brasse plus de données et peut déduire des documents une meilleure identification des auteurs.

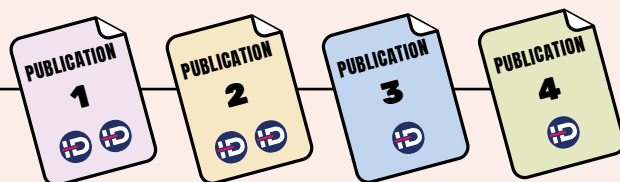
Quand et à quelle fréquence utiliser cette méthode ?

une seule fois car l'Abes traite depuis le printemps 2023 tous les nouveaux dépôts dans HAL.



RÉSULTAT

l'Abes sait que tel IdRef est auteur de tel document déposé dans HAL même en l'absence d'IdHAL.



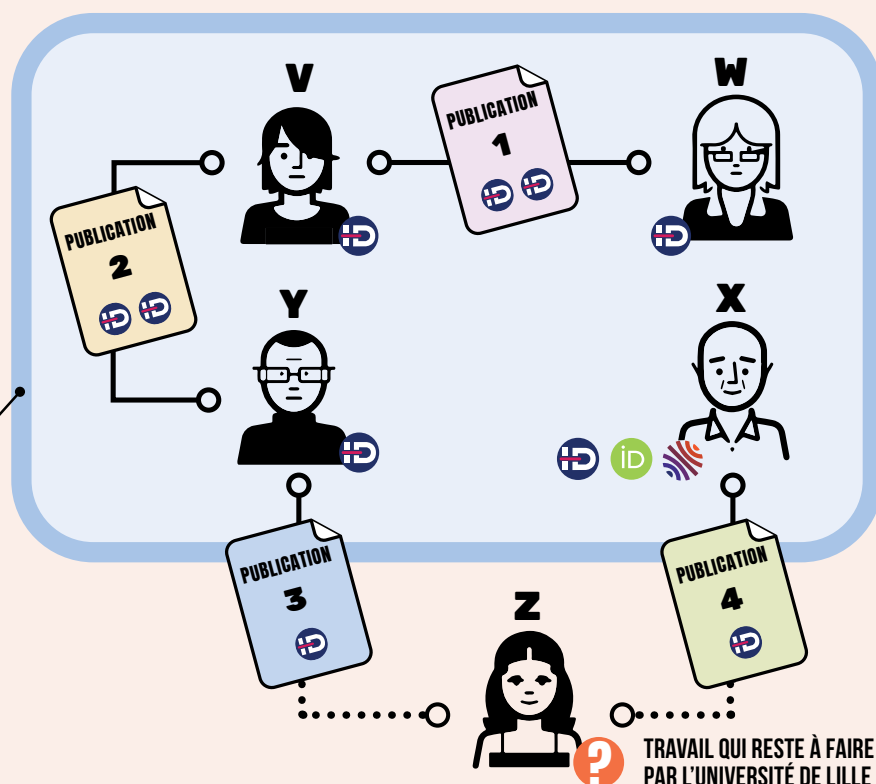
COMPLÉMENTARITÉ DES DEUX MÉTHODES

Lorsque les deux méthodes sont combinées, un graphe de connaissances très complet est généré et exploité dans IdRef et dans data.idref.fr.

Dans notre exemple, le travail mené par les collègues de Nantes a permis d'inclure dans data.idref.fr les triplets qui concernent les chercheurs V, W, X, Y et leurs publications respectives. Le chercheur Z, qui travaille à l'université de Lille, n'a évidemment pas été traité par la méthode « annuaire ». Ses publications ont en revanche été traitées par la méthode « portail HAL » : un candidat IdRef a peut-être été détecté par le programme de l'Abes mais les collègues nantais n'ont pas vérifié, car ils avaient suffisamment à faire : ce serait aux collègues lillois de prendre le relais.

RÉSULTAT

Ensemble des triplets désormais dans data.idref.fr grâce à l'utilisation des deux méthodes par l'Université de Nantes



SI VOTRE ÉTABLISSEMENT N'A PAS EU RECOURS À CE SERVICE, N'HÉSITEZ PAS À PRENDRE CONTACT VIA LE GUICHET ABESSTP > IDREF > DONNÉES !

article établi par Benjamin Bober, Isabelle Mauger Perez, Carole Melzac, François Mistral idref@abes.fr

Infographie : Anne Ladevie

TRAVAIL QUI RESTE À FAIRE PAR L'UNIVERSITÉ DE LILLE