



Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION

scanR, moteur de la recherche et de l'innovation

Un nouvel élan en 2018

CasuHal 2018 - 3 Avril 2018



scanR, aujourd'hui



Ouvert en juillet 2016 en version BETA

Service de profilage recherche et innovation pour :

- 30 000 entreprises
- 5 000 institutions publiques dont 3 700 unités de recherche

Pour

- Connaître pour mieux décider
- Valoriser
- Contribuer à la mise en relation
- Inciter à l'amélioration de la qualité de l'information

Avec

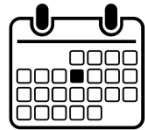
- Des données issues de bases de données
- Des données crawlées sur le web

Derrière l'appli

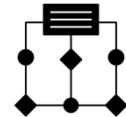
- L'équipe du département des outils d'aide à la décision
- Un partenaire aux compétences complémentaires

<https://scanr.enseignementsup-recherche.gouv.fr/> ou **scanr.fr**

scanR et HAL



Created by Noun Alain Riba
from Noun Project



Created by Alessandro Lodi
from Noun Project



Et aussi

- API scanR
- Open data
Ministériel (moins à
jour)

Chaque jour :

- Les nouvelles
références
- Les
changements

Matching



Created by Adrien Coquet
from Noun Project

- Les limites de l'algorithme de matching
- L'usage du seul RNSR comme référentiel d'alignement
- La qualité des données d'affiliation de HAL

Bilan rapide après 18 mois d'exploitation



- Une approche originale (transversalité, descente au niveau du labo, crawling...)
- Un outil perçu comme performant par ses utilisateurs
- Accueil positif
- Un contexte favorable



- L'accès unique par structure ne permet pas de valoriser l'information disponible
- Une vision plate du système nationale de recherche
- Des listes « infinies » qui ne permettent pas à l'utilisateur de se faire rapidement une idée des orientations d'une structure
- Une couverture inégale des champs

Transformer l'essai mais conserver l' « esprit scanR » :

- Une équipe qui se positionne comme « Product owner »
- Une démarche ouverte sur les utilisateurs (160 réponses à une consultation numérique et 45 participants au 1^{er} barcamp)
- Le travail sur l'outil et celui sur les données sont menés conjointement
- Rendre le code de scanR open source



Framavox utilisateurs
bit.ly/users_scanr

Les chantiers scanR

**DIRECTEMENT AUTOUR DE
L'APPLICATION SCANR**

Orientation 1 – Trouver des structures ... mais également des chercheurs, des projets, des travaux

The screenshot shows the scanR website interface. At the top, there's a navigation bar with 'scanR', 'Recherche', 'À propos', 'Actualités', and 'Emplois'. Below that, a search bar contains '3 358 résultats pour biotechnologie'. A filter bar shows categories: 'TOUT (3 358)', 'Laboratoires, Entreprises (2 500)', 'Chercheurs (158)', 'Projets (372)', and 'Publications, thèses, brevets (328)'. The main content area is divided into four sections, each with a 'Voir toute la liste' button:

- 2 500 laboratoires, entreprises**: Includes 'Incubateur Paris Biotech (PARIS BIOTECH)' and 'ELSALYS BIOTECH'.
- 158 chercheurs**: Includes 'Marie-Laure Quilici' and 'Gérald Spaeth'.
- 372 projets**: Includes 'Biotechnologies rouges' and 'Bioéconomie et biotechnologies'.
- 328 publications, thèses, brevets**: Includes 'Caractérisation de la fraction liquide des digestats...' and 'Caractérisation de la matière organique par spectrofluorimétrie 3D...'.

At the bottom, there's a link 'Quelles sont nos sources de données ?' and a footer with 'Open data - API', 'Contribuer', 'Mentions légales', and 'Contact'.

Alors que la première version de scanR ne proposait qu'un accès par structure, la nouvelle version diversifie les mode d'accès à l'information

➔ Intérêt pour les identifiants chercheurs et la diffusion des identifiants chercheurs dans les sources

➔ Chantier OGP sur les financements sur projet

Orientation 2 – Mieux identifier les thématiques de recherche

scanR Recherche À propos Actualités Emplois connexion FR | EN

scanR Moteur de la Recherche et de l'Innovation

Mots-clés Ex : intelligence artificielle, open government. Périmètre de recherche Tout

Accéder à toute la base

Nos suggestions de recherche SG Biotechnologie Open data PIA

scanR aujourd'hui c'est

34 179 Entités	7 253 Chercheurs	1 247 Projets	420 887 Publications/thèses/brevets
----------------	------------------	---------------	-------------------------------------

Quelles sont nos sources de données ? Découvrir

Tendances

Voici les 10 thèmes les plus actifs dans la recherche française. En nombre de projets, thèses déposées ces douze derniers mois.

Brain to computer	Anthropologie évolutive	biotechnologie
Réalité virtuelle	Open source	Fusion nucléaire
Brain content	Cryptographie	Materiaux
Intelligence artificielle		

Suggestion de recherche avec possibilité de faire des mises en avant

Utiliser les données de navigation pour faire ressortir des tendances de recherche

Orientation 3 - Internationaliser



	Effectif
Docid avec au moins 1 Grid	137 578
Docid avec au moins 1 RNSR	21 244
Docid avec 1 Siret	10 273
Docid avec au mieux un Siren (sans Siret)	6 899



- Priorité au codage des structures étrangères et intégration du Grid
- Intégration SIRENE, RNSR, Wikidata, UAI, OpenCorporates
- Près de 50 % de docid identifiés avec un identifiant interopérable



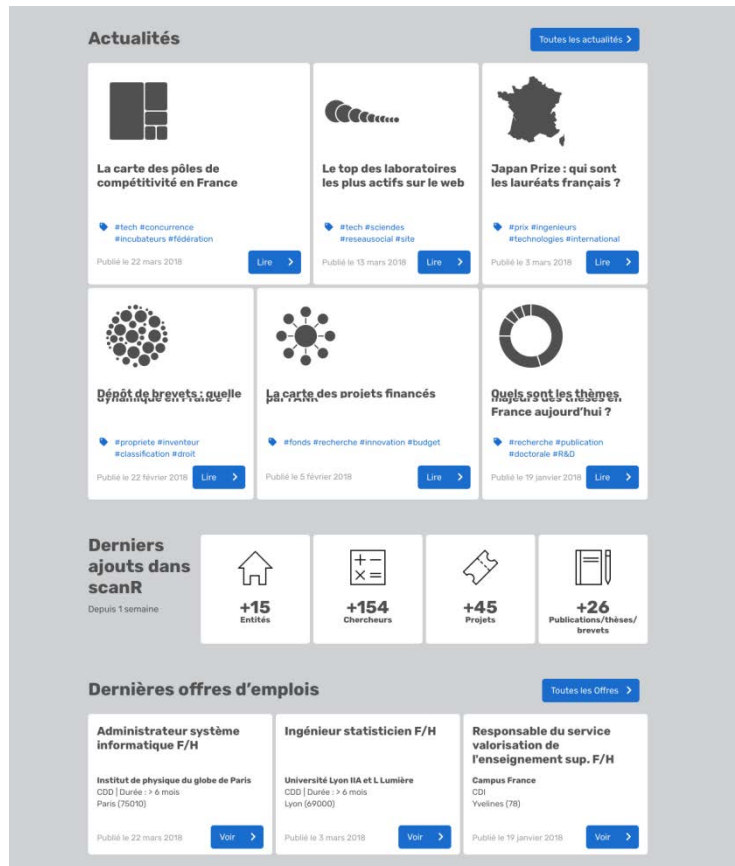
- Poursuite des recodages
- Retour vers l'équipe HAL
- Intégration à scanR

The screenshot shows the HAL website interface in English. At the top, there is a navigation bar with 'connexion' and 'FR | EN' language options. Below this is a blue header with the text 'VOUS NE RETROUVEZ PAS TOUTES VOS PRODUCTIONS ?' and a search bar. The main content area is divided into several sections: 'Ecosystème' with a circular network diagram, 'Certifications & Prix' with a list of awards, and 'Dernières offres d'emplois' with a list of job openings. The interface is clean and modern, with a focus on navigation and search.

Une interface en anglais

Des partenaires étrangers identifiés

Orientation 4 – Rendre visible la richesse de scanR



Une éditorialisation automatisée du contenu de scanR pour proposer les « visualisations du moment »

Le dynamisme du contenu de scanR mis en lumière

De nouveaux contenus pour élargir l'audience de scanR (et améliorer la connaissance des structures)

Les chantiers scanR

ET AUSSI AUTOUR DE SCANR

La machine à données



Un service générique (et ouvert) d'aide à la mise en cohérence des données avec des référentiels pour les développeurs (API) et les non développeurs (interface graphique)

Objectifs :

- Travailler plus de sources à moindre cout (interne)
- Rendre service
- Eliminer l'excuse de « défaut de qualité » comme frein à la circulation de l'information

LA MACHINE A DONNEES

Application web permettant le **RAPPROCHEMENT** générique de fichiers CSV

S'enregistrer & Commencer un projet

Référentiels disponibles

Afin de faciliter vos mises en correspondance, un ensemble de référentiels est disponible et mis à jour régulièrement

Insee

rnSr
répertoire national des structures de recherche

GRID

GitHub

Merge-Machine

Le projet se trouve sur GitHub sous licence MIT. Vous pouvez suivre et/ou participer à son amélioration.

PARTENAIRES

Programme EXO

MESRS

Conditions générales d'utilisation

API

Contact

<http://machine-a-donnees.fr>


- Version BETA opérationnelle
- Code sous licence MIT disponible sur GitHub
- Hébergement sur serveurs ministère en cours
- Référentiels proposés actuellement
 - SIRENE (10 millions d'enregistrements)
 - RNSR
 - Grid (ES et R à l'étranger)
 - Répertoire national des associations
 - Référentiel géographique
 - Codes postaux villes > communes INSEE
- En projet :
 - Opencorporates

dataESR [EIG2018] – Gestion des référentiels

#dataESR

Retour Cet établissement est ouvert depuis le May 17, 1980

Nom de l'établissement Paris IV - Université Paris Sorbonne **Actif**



Adresse de l'établissement **Sorbonne** **Active**
1 Rue des Ecoles, 75005 Paris
01 40 46 22 11

Identifiants référentiels

UAI **Actif**
0752756N

+ Ajouter un code

Evolution: Aucune évolution enregistrée actuellement...
+ Ajouter une évolution

Caractérisations de l'établissement
secteur public

Présence sur le web

http://www.paris-sorbonne.fr/ **x**

https://fr.wikipedia.org/wiki/Sorbonne_Universit% **x**

+ Ajouter un lien

Rattachements

	Mère	Fille
siège_composant	Université de Paris	

#dataESR

Alignement des sources sur les référentiels-clé



Ecoles doctorales

	Effectif
IdRef avec NUMED	384
IdRef non identifiés	15
IdRef étrangers	12
Total fichier ABES	411

- 93 % de numéro d'ED ré-affectés



- Publication d'un jeu de données ouvert avec les ED en historique enrichi avec les EdRef associés à chaque ED (dans quelques jours)
- Intégration d'une entrée ED dans scanR avec affichage thèses (dans quelques semaines)



Unités de recherche

	Effectif
IdRef avec RNSR	1384
IdRef Etrangers	175
IdRef avec SIREN/SIRET	12
IdRef orphelins	957
IdRef reste à qualifier	261
Total fichier ABES	2614

- 50 % IdRef labo alignés sur RNSR

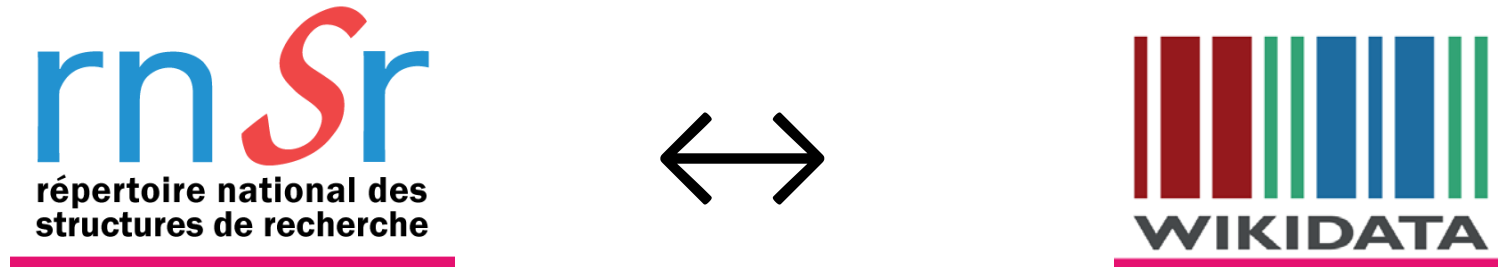
- Poursuite des travaux de recodage
- Intégration des références bibliographiques associées aux laboratoires présents dans scanR (dans quelques mois)

Personnels de direction des structures de recherche du RNSR

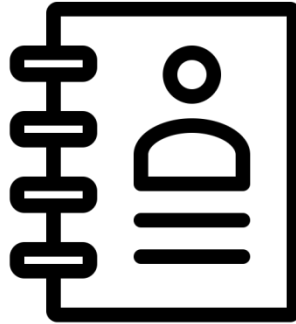
- Affectation d'un IdRef à plus de 80 % des personnels de direction des structures de recherche du RNSR

- Intégration prochaine IdRef/Orcid au RNSR

Alignement des sources sur les référentiels-clé



- Numéro national de structure (NNS), propriété de wikidata (initiative scanR)
- Intégration du RNSR à Wikidata (intégration des structures de recherche manquantes et intégration de leur NNS) (initiative d'Antonin Delpeuch)
- Intégration du Wikidata au RNSR (Automne 2018)



Contacts :

Yann Caradec et Emmanuel Weisenburger

[yann.caradec\[at\]recherche.gouv.fr](mailto:yann.caradec[at]recherche.gouv.fr) et [emmanuel.weisenburger\[at\]recherche.gouv.fr](mailto:emmanuel.weisenburger[at]recherche.gouv.fr)

Sur la Machine à données

[leo.bou\[at\]gmail.com](mailto:leo.bou[at]gmail.com)

[jeremy.peglion\[at\]recherche.gouv.fr](mailto:jeremy.peglion[at]recherche.gouv.fr)

Sur les projets EIG 2018

[ambre.arnaud\[at\]recherche.gouv.fr](mailto:ambre.arnaud[at]recherche.gouv.fr)

[maxime.pierrot\[at\]recherche.gouv.fr](mailto:maxime.pierrot[at]recherche.gouv.fr)

[sami.moustachir\[at\]recherche.gouv.fr](mailto:sami.moustachir[at]recherche.gouv.fr)