



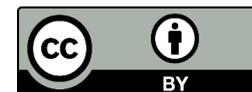
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

*Liberté
Égalité
Fraternité*

Les projets nationaux sur les données de la recherche

Isabelle Blanc [@AMDAC_MESR](https://twitter.com/AMDAC_MESR)
isabelle.blanc@recherche.gouv.fr

Gilles Mathieu gilles.mathieu@recherche.gouv.fr



Données de la recherche : points de repère



Données OCDE 2007

Les « données de la recherche » sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche

Métadonnées

National Information Standards Organization (NISO), 2004

« Une métadonnée est une information structurée qui décrit, explique, localise une ressource informationnelle ou la rend facile à trouver, utiliser ou gérer »
La NISO distingue 3 principaux types de métadonnées :
descriptives, structurelles et administratives.

Borgman, C.L., 2020

Métadonnées scientifiques : description protocole
obtention, traitement, matériel, logiciels...

Métadonnées documentaires : administratives, d'utilisation,
de préservation, d'accès

Ambition en matière d'ouverture de la recherche

posée par la loi pour une république numérique (2016)
rappelée par la science ouverte

Ouvrir l'ensemble des produits et des méthodes de la recherche

- >> Les publications
- >> Les données
- >> Les codes sources et logiciels

Pour une science plus cumulative, plus robuste, plus reproductible

Pour une science pleinement transparente et accessible à toutes et tous

Ouverture, partage/circulation, réutilisation des données & souveraineté

Guide d'application de la loi pour une République numérique pour les données de la recherche

OUVERTURE

OUVERTURE PAR DÉFAUT

Données **achevées**

Données sous-jacentes à une publication présentent un caractère achevé

Pas de droit d'auteur sur données, ne sont pas une création de l'esprit => **producteurs de données**

Etablissements de l'ESR, en tant qu'administrations publiques, **sont les garants, de la mise en œuvre de l'ouverture des données** publiques et non les chercheurs à titre individuel

PARTAGE / CIRCULATION

FERMÉ PAS PLUS QUE NÉCESSAIRE

QUAND LES DONNÉES RELÈVENT D'UNE EXCEPTION LISTÉE PAR LA LOI

Partage en accès restreint

Traitement dans bulles sécurisées

Circulation entre administrations

RÉUTILISATION

RÉUTILISATION LIBRE ET GRATUITE

Licences libres recommandées pour préciser les droits aux utilisateurs :

- **Citation du producteur** (source),
- Mention de la date de la dernière mise à jour des données
- Interdiction d'altérer ou de dénaturer ces dernières
- ...

Données ouvertes sous licence libre contribuent à la **souveraineté**

Concevoir nativement des données réutilisables

Travail scientifique de FAIRisation des données, renseigner :

- Les origines et les transformations des données
- Les guides de codification
- Les modèles et la description détaillée des méthodes de collecte, de nettoyage, d'analyse
- Le code spécifique à un instrument
- Les logiciels conçus pour des applications propres au domaine

Plus la réutilisation est loin du point d'origine, en termes de temps, de problématique de recherche, de discipline... plus il devient difficile de retrouver et d'interpréter des données

Concevoir nativement des données réutilisables

Travail scientifique de FAIRisation des données, renseigner :

- Les origines et les transformations des données
- Les guides de codification
- Les modèles et la description détaillée des méthodes de collecte, de nettoyage, d'analyse
- Le code spécifique à un instrument
- Les logiciels conçus pour des applications propres au domaine

Plus la réutilisation est loin du point d'origine, en termes de temps, de problématique de recherche, de discipline... plus il devient difficile de retrouver et d'interpréter des données

Le premier bénéficiaire d'une donnée bien décrite est celui qui l'a produite !

Retrouver et réutiliser ses données, mieux les gérer, gain de temps et réduction de l'empreinte numérique

Taux de citations des publications dont les données sont partagées : + 25% à + 75%

Les données partagées reconnues comme une production de la recherche



Les engagements

*Deuxième plan national pour la science ouverte
Politique des données, des algorithmes et codes sources*

Soutenir la structuration, la préservation, le partage, l'ouverture,
la découverte des données de la recherche

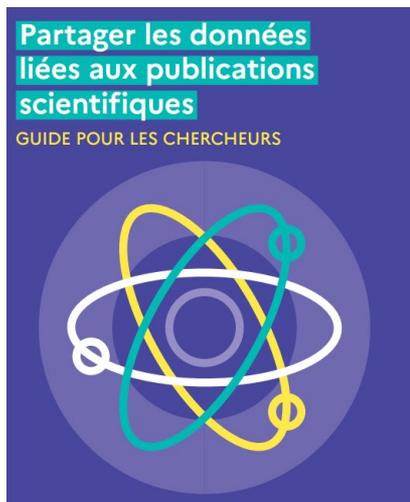
Favoriser les pratiques de gestion raisonnée et écoresponsable des données

Favoriser les pratiques de réutilisation des données de la recherche

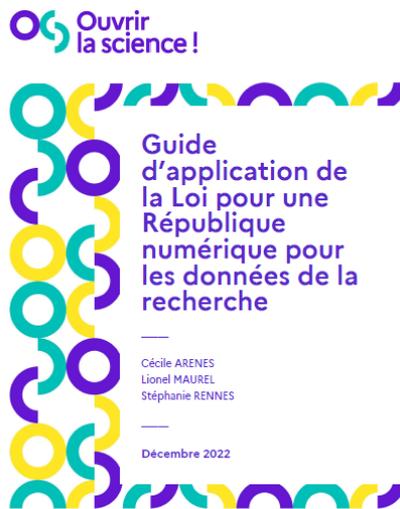
Fédérer les initiatives locales et thématiques



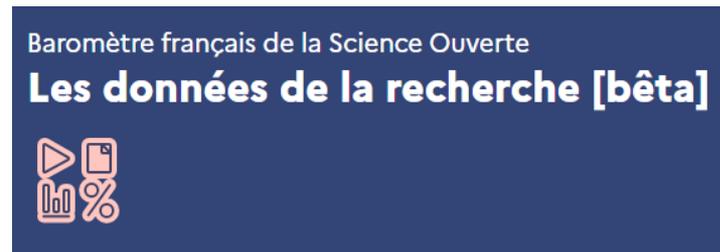
Données de la recherche : guides, actions pour soutenir les équipes de recherche



<https://www.ouvri.la-science.fr/comment-partager-les-donnees-liees-a-vos-publications-scientifiques-un-guide-vous-accompagne/>



<https://hal-lara.archives-ouvertes.fr/OUVRIR-LA-SCIENCE/hal-03968218v1>



<https://barometredelascienceouverte.esr.gouv.fr/donnees-de-la-recherche/general/>



La certification des entrepôts
et services de données

Reconnaître les métiers de la donnée

Identifier et décrire les
activités de gestion et
diffusion des données

Décembre 2023

Liste des entrepôts de données de confiance

Guider les équipes
de recherche vers l'entrepôt le
plus adapté à leur domaine

Octobre 2023

Science ouverte - Données de la recherche

Edition livret – Fin 2023

Harmonisation de l'information juridique

Création d'une plateforme de
ressources sur le droit des
données de la recherche

2024



RÉPUBLIQUE
FRANÇAISE recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

Un écosystème au service du partage et de l'ouverture des données de recherche



LICENCE OUVERTE
OPEN LICENSE



Objectifs

Ne pas laisser d'équipes de recherche sans solution
pour ouvrir ou partager leurs données

Faire de l'accompagnement un élément central du dispositif

Reconnaître, fédérer et amplifier les initiatives existantes





Objectifs

Ne pas laisser d'équipes de recherche sans solution
pour ouvrir ou partager leurs données

Faire de l'accompagnement un élément central du dispositif

Reconnaître, fédérer et amplifier les initiatives existantes

Un maillage d'offres d'accompagnement en proximité de la recherche

- Sensibiliser, former ou accompagner les équipes de recherche





Objectifs

Ne pas laisser d'équipes de recherche sans solution pour ouvrir ou partager leurs données

Faire de l'accompagnement un élément central du dispositif

Reconnaître, fédérer et amplifier les initiatives existantes

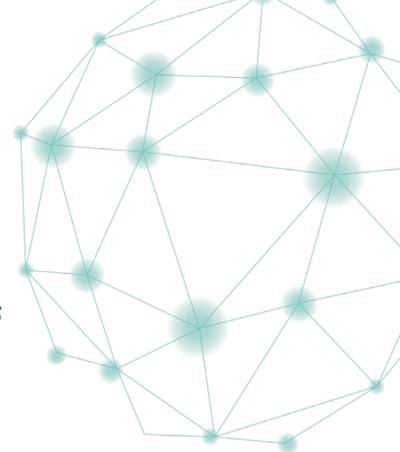
Un maillage d'offres d'accompagnement en proximité de la recherche

- Sensibiliser, former ou accompagner les équipes de recherche

Une offre souveraine de dépôt, publication et de signalement des données

Plateforme nationale pluridisciplinaire et fédérée des données de la recherche

- Complémentaire des entrepôts thématiques
- **Une alternative à la publication sur des plateformes commerciales ou ne répondant pas à des critères de confiance**
- Devenir un des services de *l'European open science cloud*





Les ateliers de la donnée s'installent progressivement sur tout le territoire



Expertise généraliste en proximité des équipes de recherche pour toute question relative à la donnée

Regroupement d'expertises complémentaires de différents établissements

Un réseau d'entraide, de mutualisation et de capitalisation des services actifs

19 ateliers de la donnée opérationnels

10 ateliers en projet

90+ ETP
350+ personnes
80+ établissements différents

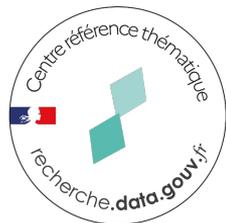




RÉPUBLIQUE
FRANÇAISE recherche.data.gouv.fr

Liberté
Égalité
Fraternité

Les centres de référence thématiques



Expertise par domaine scientifique

Définition et diffusion des **bonnes pratiques et des standards internationaux de gestion, traitement et diffusion des données** par domaine scientifique

Premier paysage :
infrastructures de recherche ayant une activité structurante de gestion et diffusion de données pour leur communauté scientifique

Réflexion en cours entre toutes les infrastructures nationales en biologie et santé



INSTITUT FRANÇAIS DE BIOINFORMATIQUE



DATA
INFRASTRUCTURE



Pôle National
de Données de Biodiversité



@RechercheDataGv



Les centres de ressources au service de tous les autres modules de Recherche Data Gov



Centre de ressources
supports pédagogiques, e-Formations
Recherche Data Gov

Mise en commun et certification de
supports pédagogiques et élaboration et
diffusion de **e-Formations généralistes**



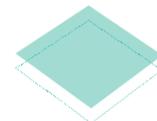
Centre de ressources
compétences
Recherche Data Gov

Dispositif de **montée en compétences**
pour les ateliers de la donnée



Centre de ressources
outils communs aux ateliers
Recherche Data Gov

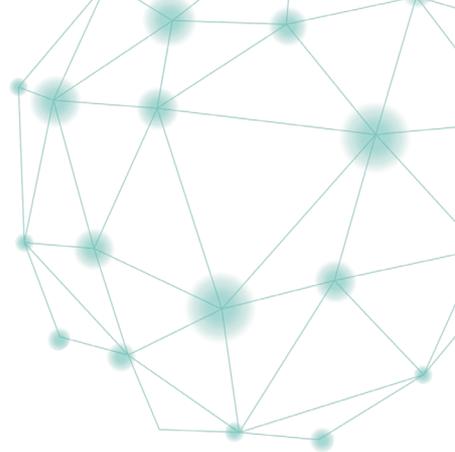
Élaboration et maintien d'**outils communs**
aux ateliers



Centre de ressources
entrepôt-catalogue
Recherche Data Gov

Développe et maintien l'**entrepôt-catalogue**
et accompagne les utilisateurs





Un entrepôt de données mutualisé

entre les établissements d'enseignement supérieur et de la recherche

- Une offre nationale développée et maintenue par des établissements au service de toute la communauté de recherche
- Les équipes de recherche déposent et publient directement leurs données
- Des espaces pour les établissements qui assurent la gestion et la modération de leurs données
- Décharge des établissements du coût et du maintien d'une offre technique
- Une offre maintenue à l'état de l'art
- Masse critique pour atteindre une dimension internationale



Créer un espace institutionnel

Gérer l'ouverture et le partage des données de son établissement

http://recherche.data.gouv.fr/uploads/general/Documents/Demande_creation_espace_institutionnel.pdf



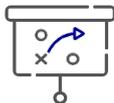


Depuis le 8 juillet 2022



Chiffres audience du portail (visiteurs acceptant cookies)

- 28 333 visiteurs uniques
- 115 790 pages vues



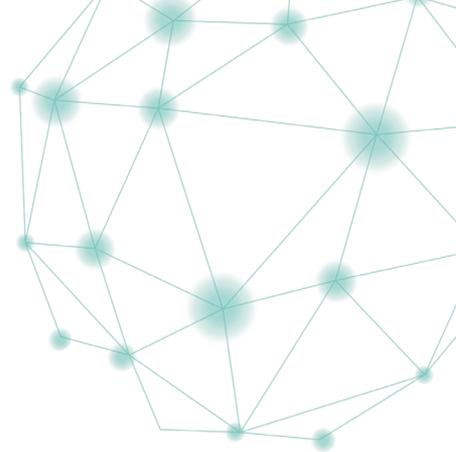
Chiffres activité support

- 31 formations réalisées à distance
- 617 participants à ces formations
- 958 demandes traitées par e-mails pour diverses sollicitations



Chiffres entrepôt de données

- 31 espaces institutionnels actifs
- 14 demandes en cours d'instruction
- 946 jeux de données déposés
- 593 jeux de données moissonnés
- ~30 000 fichiers de données ouverts ou partagés en accès restreints
- 335 812 fichiers téléchargés

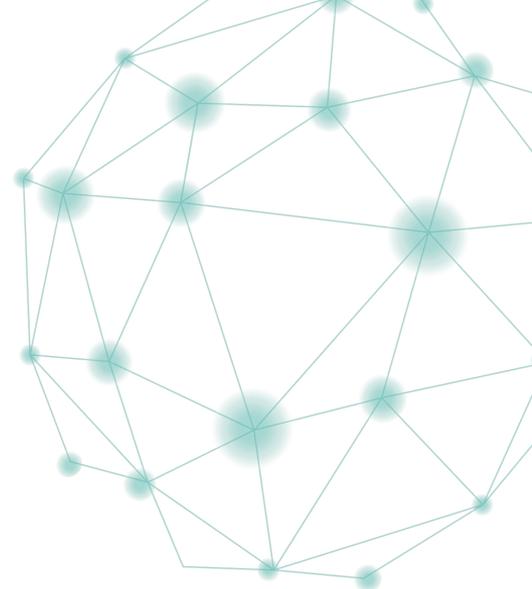


Une construction collective Une gouvernance collective

- Partenaires du projet



- Représentants de la communauté ESR





Un écosystème au service du partage et de l'ouverture des données de recherche

19 + ATELIERS DE LA DONNÉE

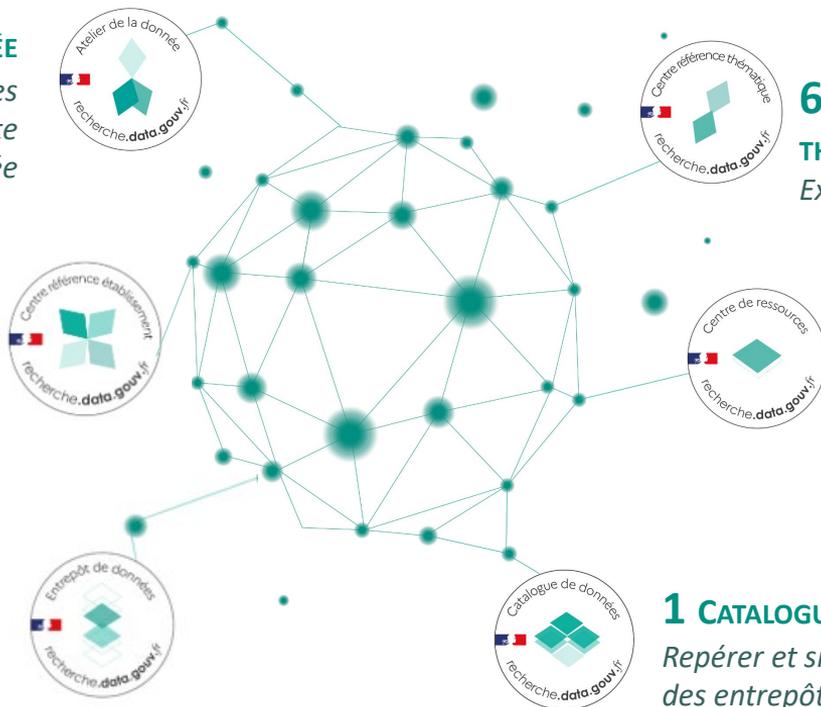
Expertise généraliste en proximité des équipes de recherche pour toute question relative à la donnée

4 CENTRES DE RÉFÉRENCE

Accompagnement aux orientations de l'établissement

1 ENTREPÔT DE DONNÉES

Offre mutualisée pour tous les établissements pour le dépôt et la publication des données



6 + CENTRES DE RÉFÉRENCE THÉMATIQUES

Expertise par domaine scientifique

4 CENTRES DE RESSOURCES

Pour soutenir les ateliers et capitaliser leurs pratiques

1 CATALOGUE DES DONNÉES

Repérer et signaler les données des entrepôts externes de confiance





MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

*Liberté
Égalité
Fraternité*

Merci de votre attention !

Isabelle Blanc

@AMDAC_MESR
isabelle.blanc@recherche.gouv.fr

Gilles Mathieu

gilles.mathieu@recherche.gouv.fr

