



recherche.**data.gouv.fr**

Vers la création d'une plateforme nationale fédérée des données de recherche



1. Une stratégie nationale pour les données de la recherche

Commencer par les données associées aux publications

L'enjeu du lien entre données et publications est inscrit comme priorité dans le Plan national pour la science ouverte

DEUXIÈME AXE : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE



Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes. Lors de l'annonce du plan « Intelligence artificielle » au Collège de France, le 29 mars 2018, le Président de la République a annoncé la mise en place d'un principe d'ouverture par défaut pour toutes les données publiées dans le cadre d'appels à projet sur fonds publics. Cette obligation sera limitée par les exceptions légitimes encadrées par la loi, par exemple en ce qui concerne le secret professionnel, les secrets industriels et commerciaux, les données personnelles ou les contenus protégés par le droit d'auteur. Elle sera par ailleurs encadrée par les bonnes pratiques définies par chaque communauté scientifique, par exemple pour définir des durées d'embargo.

D'autre part, le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation se dotera d'un administrateur des données de la recherche. Il travaillera avec l'administrateur général des données de la France et animera un réseau des administrateurs de données de la recherche dans les établissements concernés. Un appel « flash » de l'ANR permettra d'accélérer la structuration de la communauté scientifique afin de promouvoir les principes « FAIR » et de développer l'ouverture des données. D'une façon générale, les dépenses de traitement des données seront éligibles dans les appels à projets.

Les chercheurs seront invités à déposer les données dans des entrepôts de données certifiés, dont la gouvernance et les règles de propriété intellectuelle seront conformes aux bonnes pratiques. À ce titre, les infrastructures nationales et européennes de recherche seront privilégiées, notamment via des centres de données thématiques et disciplinaires. Les plans de gestion des données, instrument de définition des règles de construction, conservation et diffusion des données, seront généralisés. Un prix des données de la recherche sera mis en place afin de mettre en valeur et récompenser les équipes qui réalisent un travail exemplaire dans ce domaine.

La France apportera son soutien à la **Research Data Alliance (RDA)**, réseau international définissant les bonnes pratiques dans le domaine des données de la recherche. Elle soutiendra aussi le développement et la conservation des logiciels, support indissociable des connaissances techniques et scientifiques de l'humanité tout entière. Dans cette optique, elle apportera son soutien à **Software heritage**.

Dans le cadre du soutien public aux revues, la France recommandera l'adoption d'une politique de données ouvertes associées aux articles et le développement des **data papers**. Une politique similaire concernant les thèses sera également mise en place.

« Les données de la recherche sont la matière première de la connaissance. Les partager, c'est ouvrir de nouvelles perspectives scientifiques. »

MESURES

- 4 Rendre obligatoire la diffusion ouverte de données de recherche issues de programmes financés par appels à projets sur fonds publics.
- 5 Créer la fonction d'administrateur des données et le réseau associé au sein des établissements.
- 6 Créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs.



Accélérer

- Proposer un appel ANR Flash destiné à accélérer l'adoption des principes FAIR et l'ouverture des données de la recherche en France.
- Créer un prix des données de la recherche récompensant les équipes et projets exemplaires dans ce domaine.

Coordonner

- Construire autour de l'administrateur des données un réseau de correspondants dans les établissements, pour répondre aux questions que se posent les chercheurs sur les données de la recherche.
- Dans le cadre du soutien public aux revues, recommander l'adoption d'une politique de données ouvertes associées aux articles, le développement des articles de données et des revues de données.

Structurer

- Généraliser la mise en place de plans de gestion des données dans les appels à projets de recherche
- Développer des centres de données thématiques et disciplinaires.
- Développer un service générique d'accueil et de diffusion des données simples.
- Engager un processus de certification des infrastructures de données.

Organiser

- Soutenir la **Research data alliance (RDA)** et créer le chapitre français de l'alliance (RDA France).
- Soutenir **Software heritage**, la bibliothèque des codes sources

« 6. Une politique de données ouvertes associées aux articles publiés par les chercheurs. »

Comment avons-nous travaillé

Etude de faisabilité

- 3 rapports pilotés par Jean-Christophe Desconnets (IRD) et le Comité pour la science ouverte

<https://www.ouvrirlascience.fr/service-generique-daccueil-et-de-diffusion-de-donnees-simples/>

- *Étude de faisabilité d'un entrepôt générique : besoins des usagers*
- *Étude comparative des services nationaux de données.
Australie, Norvège, Pays-Bas, Canada, Royaume-Uni, Allemagne, Royaume-Uni.*
- *Ambitions du service et scénarios de mise en œuvre*

- 48 personnes membres du groupe de travail ou auditionnées
- Dont les responsables de 7 entrepôts nationaux.

Groupe de préfiguration

- 24 personnes membres du groupe de travail ou auditionnées

Total

- 72 personnes mobilisées, que nous remercions !

2. Recherche Data Gouv

Une plateforme à 5 modules

Une plateforme à cinq modules

RECHERCHE DATA GOUV



ATELIERS DE LA DONNÉE

Experts de la donnée
en charge de
l'accompagnement
des chercheurs

Une plateforme à cinq modules

RECHERCHE DATA GOUV



ATELIERS DE LA DONNÉE

Experts de la donnée
en charge de
l'accompagnement
des chercheurs



ENTREPÔT

Interface Web
de dépôt de ses données
par le chercheur
+ espace de modération

Une plateforme à cinq modules

RECHERCHE DATA GOUV



ATELIERS DE LA DONNÉE

Experts de la donnée
en charge de
l'accompagnement
des chercheurs



ENTREPÔT

Interface Web
de dépôt de ses données
par le chercheur
+ espace de modération



CATALOGUE

Repérage et moissonnage
des données
des entrepôts externes

Une plateforme à cinq modules

RECHERCHE DATA GOUV



ATELIERS DE LA DONNÉE

Experts de la donnée
en charge de
l'accompagnement
des chercheurs



ENTREPÔT

Interface Web
de dépôt de ses données
par le chercheur
+ espace de modération



CATALOGUE

Repérage et moissonnage
des données
des entrepôts externes



CENTRES DE RÉFÉRENCE THÉMATIQUES

Experts disciplinaires de la donnée

Une plateforme à cinq modules

RECHERCHE DATA GOUV



ATELIERS DE LA DONNÉE

Experts de la donnée
en charge de
l'accompagnement
des chercheurs



ENTREPÔT

Interface Web
de dépôt de ses données
par le chercheur
+ espace de modération



CATALOGUE

Repérage et moissonnage
des données
des entrepôts externes



CENTRES DE RÉFÉRENCE THÉMATIQUES

Experts disciplinaires de la donnée



CENTRES DE RESSOURCES RATTACHÉS À RECHERCHE DATA GOUV

Les chercheurs déposent dans l'entrepôt le plus adapté à leurs yeux





ENTREPÔTS NATIONAUX THÉMATIQUES

*Ex : Data Terra, Nakala,
PNDB, France Cohortes, ...*



Les chercheurs déposent
dans l'entrepôt le plus adapté à leurs yeux



ENTREPÔTS NATIONAUX THÉMATIQUES

*Ex : Data Terra, Nakala,
PNDB, France Cohortes, ...*

Les chercheurs déposent
dans l'entrepôt le plus adapté à leurs yeux



ENTREPÔTS INTERNATIONAUX THÉMATIQUES

*Ex : PANGAEA, GBIF,
EBI, CERN, GEO...*





**ENTREPÔTS
NATIONAUX
THÉMATIQUES**

*Ex : Data Terra, Nakala,
PNDB, France Cohortes, ...*



**RECHERCHE
DATA GOUV**



**ENTREPÔTS
INTERNATIONAUX
THÉMATIQUES**

*Ex : PANGAEA, GBIF,
EBI, CERN, GEO...*



Subsidiarité, confiance, mutualisation et visibilité



Principe de subsidiarité avec les entrepôts disciplinaires existants, nationaux ou internationaux



Plateforme de qualité : une plateforme modérée



Plateforme de confiance (certifications)



Mutualisation des moyens logiciels et de l'infrastructure



Principe de subsidiarité dans l'accompagnement des chercheurs et la modération des dépôts



Visibilité de la production des établissements (logo, ...), des structures de recherche, des chercheurs et indicateurs de suivi et *reporting* (nombre de jeux de données déposés, téléchargés...)



Principe de non cession de la propriété intellectuelle sur les données



Visibilité internationale

Le projet entrepôt et catalogue de données confié à INRAE

- Le projet est lancé pour 3 ans (dispositif bêta) avant de se doter d'une structure définitive
- Il est physiquement hébergé dans un Datacenter national
- Il est financé par le Fonds national de la science ouverte
- Il se dote d'un calendrier à 3 ans :
 - Début 2022 (PFUE) : sortie de l'entrepôt/catalogue et lancement progressif des ateliers de la donnée
 - 2022-2023 : campagnes d'alimentation avec les ateliers de la donnée, avec des acteurs de l'édition et des archives ouverte
 - 2023-2024 : catalogue des données externes
- Des **établissements volontaires sont invités à contribuer en *in kind*** au module d'entrepôt pendant la phase bêta du dispositif
Ils sont associés à la gouvernance.



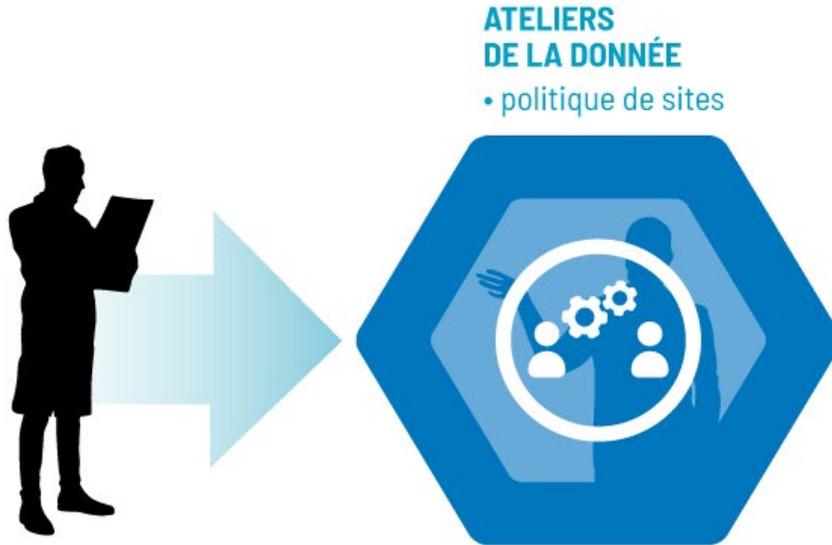


3. Accompagnement des chercheurs

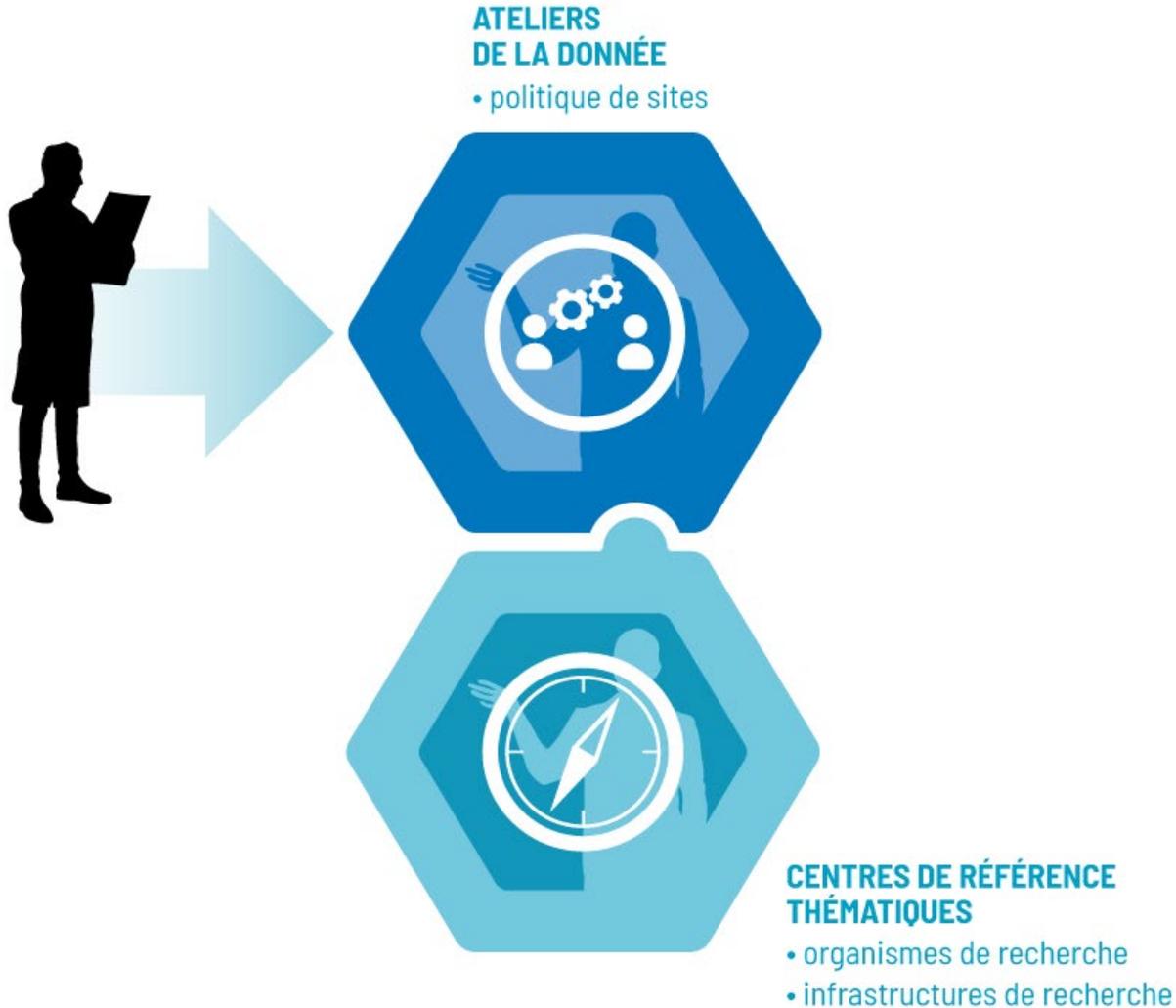
Investir avant tout dans l'accompagnement
et la réduction du « fardeau de la donnée » pour les chercheurs



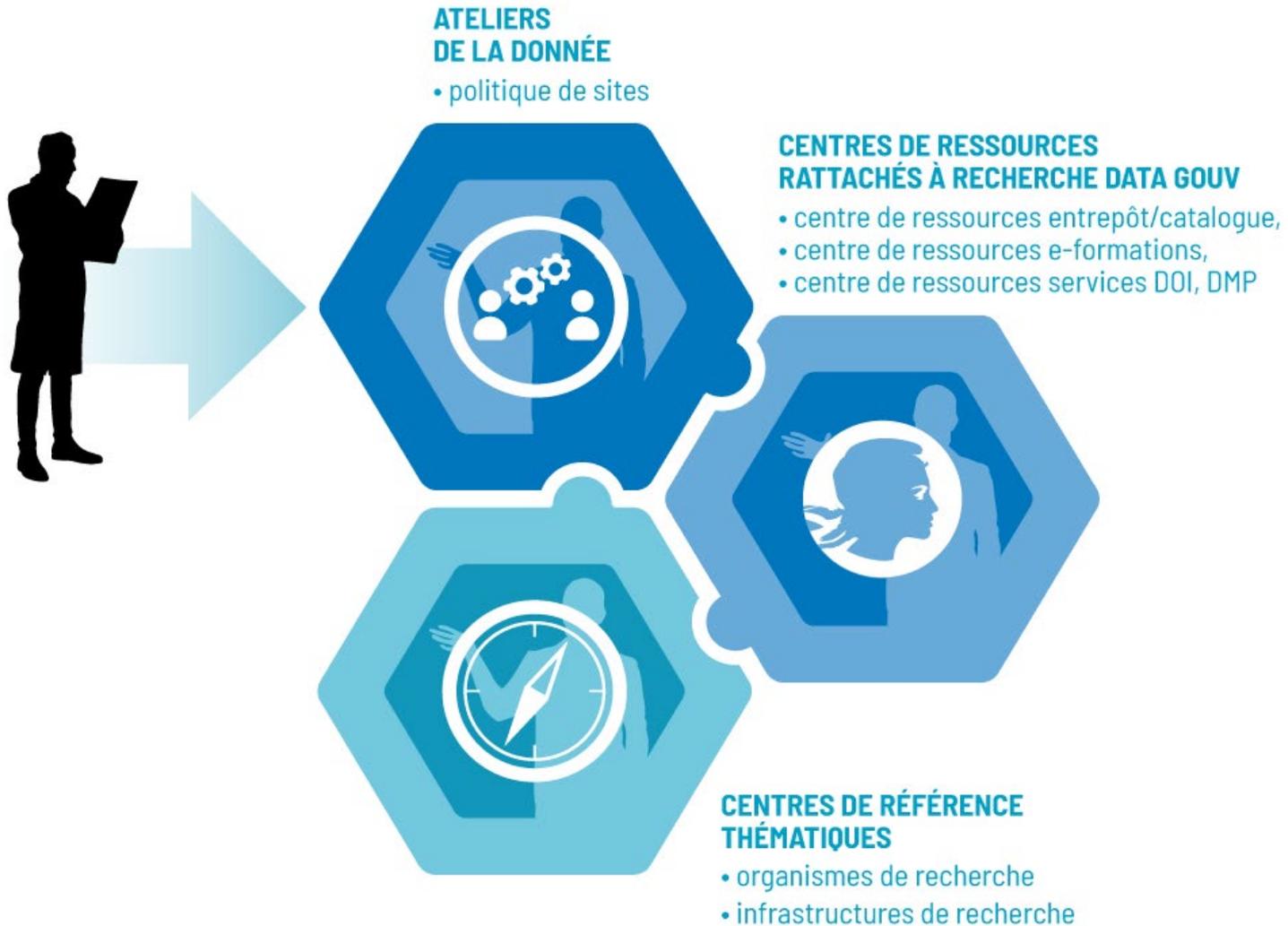
Accompagnement des chercheurs



Accompagnement des chercheurs



Accompagnement des chercheurs



Construction des Ateliers de la donnée



Ateliers de la donnée (politique de sites)

- Point d'entrée en proximité des chercheurs sur toute nature de besoin relatif à la donnée
- Regroupent en leur sein une palette de compétences qui constitue une alliance interprofessionnelle
- Regroupent des ressources de différents établissements
- Apportent un premier niveau d'expertise et de services
- S'appuient sur l'expertise des services juridiques, partenariales, SI de proximité
- S'appuient sur des expertises scientifiques nationales (centres de référence thématiques)
- Animés en réseau pour mutualiser leurs pratiques, formations...

Appel à manifestation d'intérêt :

- Labelliser les initiatives existantes
- Labelliser les nouvelles initiatives
- Accompagner la maturation et la préfiguration de nouvelles propositions
- Structurer au plan territorial le réseau des ateliers de la donnée

Modalités de l'appel

- 3 appels par an - 3 millions d'euros
- Soutien financier pour le démarrage/renforcement des initiatives 50 à 200 k€ / projets
- Suivi des projets labellisés à 18 et 36 mois
- Lancement du premier appel septembre 2021



Conclusion

Conclusion

- Accompagnement des chercheurs est primordial :
La « *charge de la donnée* » est lourde et nous ne pouvons pas nous permettre de laisser les chercheurs seuls face à cette nouvelle ambition
- Nous proposons une **architecture centralisée** (entrepôt et catalogue) **/distribuée** (accompagnement des chercheurs en proximité), qui a pour objectif de s'appuyer sur les forces et dynamiques existantes.
- Vous pouvez contribuer au projet entrepôt-catalogue et soumettre vos initiatives aux ateliers de la donnée !