



HAL
open science

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

Gilles Mathieu, Isabelle Blanc

► To cite this version:

Gilles Mathieu, Isabelle Blanc. Un écosystème au service du partage et de l'ouverture des données de recherche. Printemps de la Donnée 2024, INRAE; Université Haute-Alsace; Université de Strasbourg; INSA; PNDB; AgroParisTech; Université de Lille; Sorbonne Université; Data Terra, Jun 2024, Strasbourg, France. hal-04683387

HAL Id: hal-04683387

<https://hal.inrae.fr/hal-04683387v1>

Submitted on 2 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

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

Printemps de la donnée - Juin 2024

Isabelle Blanc isabelle.blanc@recherche.gouv.fr

Gilles Mathieu gilles.mathieu@recherche.gouv.fr



@RechercheDataGv



Ambition

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

Une science plus cumulative, plus robuste, plus reproductible

Une science transparente et accessible à toutes et tous





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



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

recherche.data.gouv.fr

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



@RechercheDataGv



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

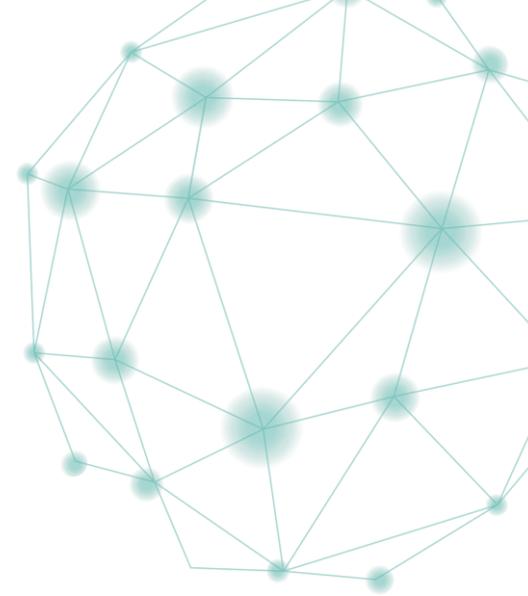
recherche.data.gouv.fr

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



@RechercheDataGv



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

recherche.data.gouv.fr

Objectifs

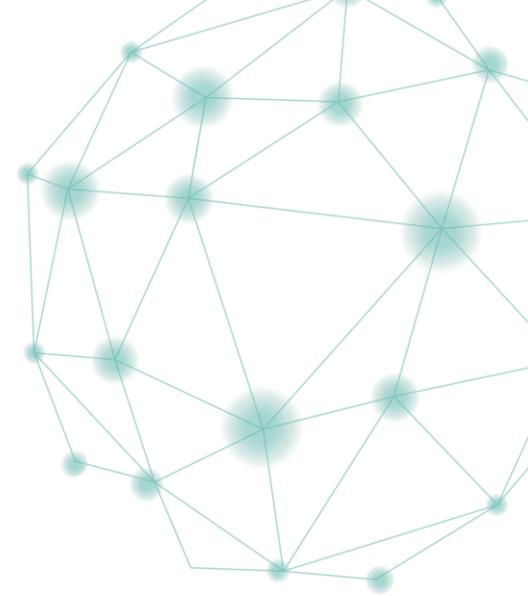
Ne pas laisser d'équipes de recherche sans solution
pour ouvrir ou partager leurs données

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



@RechercheDataGv



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

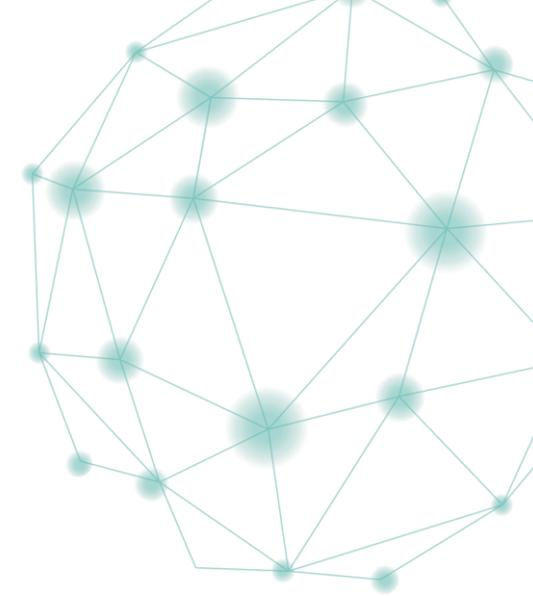
Un maillage d'offres d'accompagnement en proximité de la recherche

- 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*

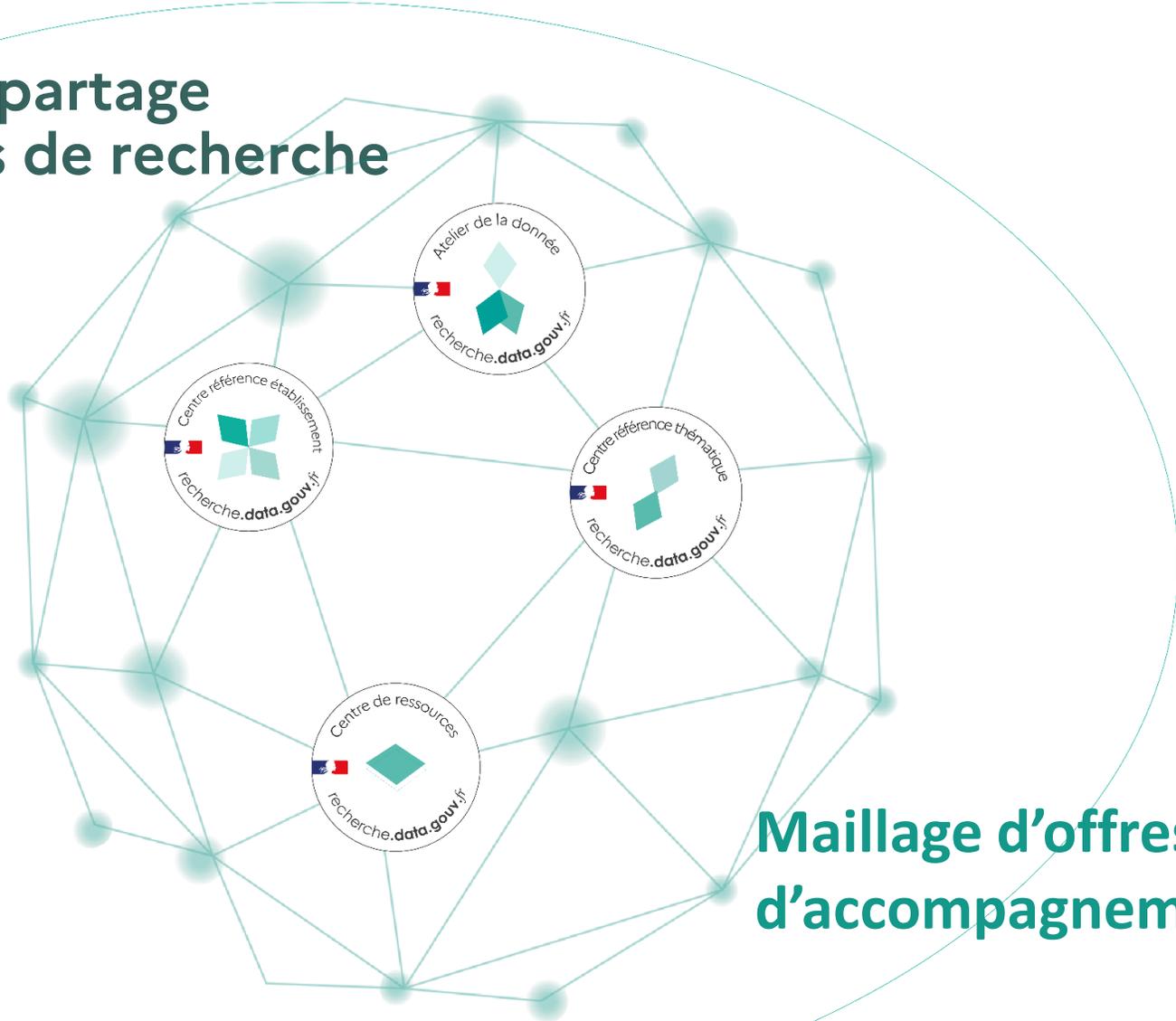




Un écosystème au service du partage et de l'ouverture des données de recherche



Entrepôt-catalogue



Maillage d'offres d'accompagnement



Premier point de contact : Les ateliers de la donnée



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

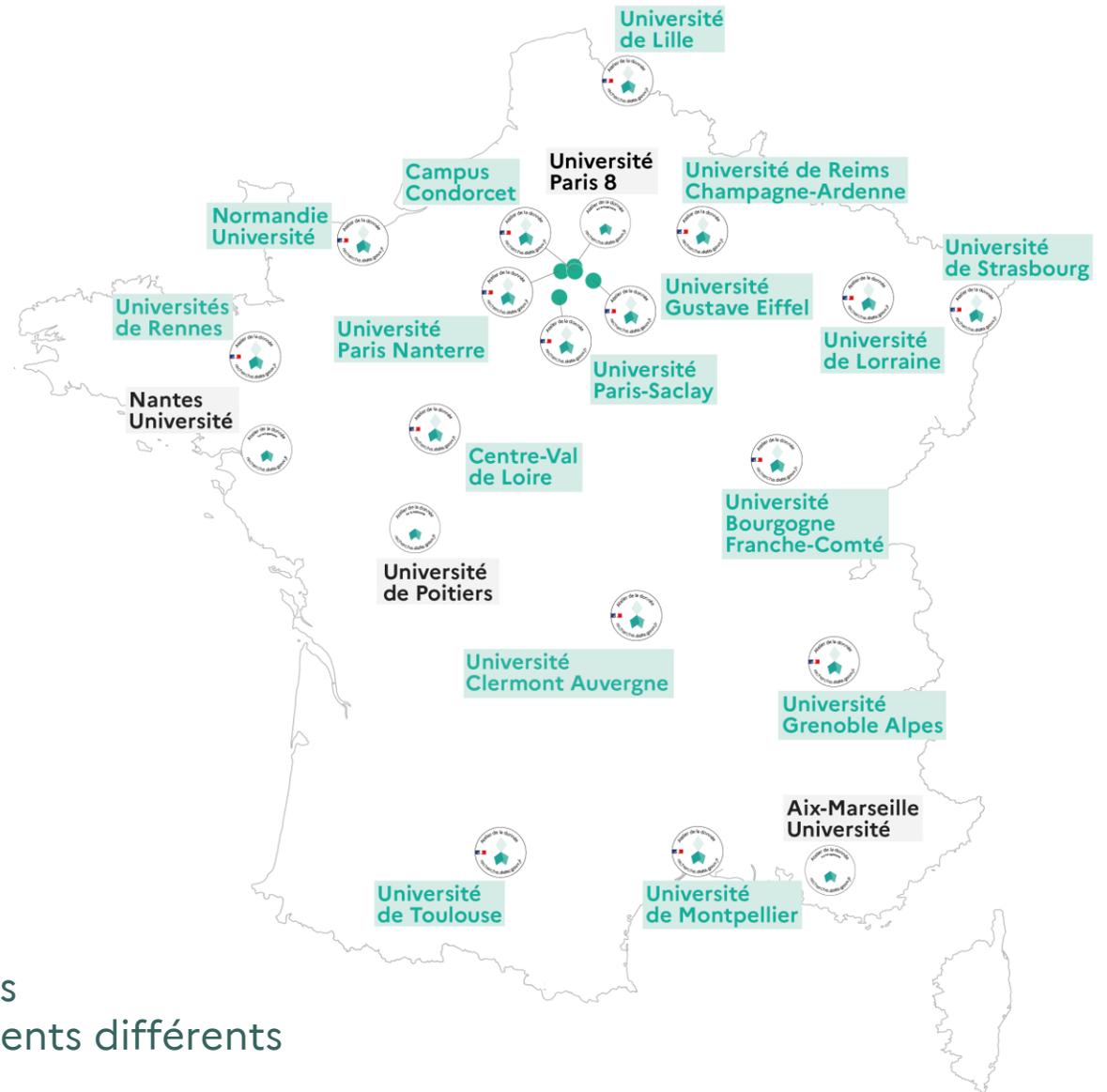
Regroupement d'expertises complémentaires de différents établissements

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

20 ateliers de la donnée opérationnels

Plusieurs ateliers en projet

95+ ETP
370+ personnes
80+ établissements différents



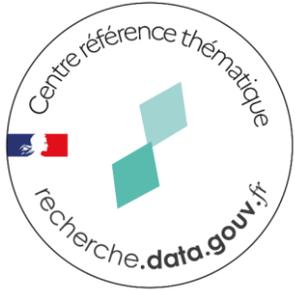


RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

recherche.data.gouv.fr

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 pour un élargissement de la couverture disciplinaire





Les centres de référence établissement



Offre d'accompagnement propre aux orientations de recherche d'un établissement

Etablissements nationaux de l'Enseignement supérieur et/ou de la Recherche

Proposent depuis plusieurs années une offre d'accompagnement propre à leurs orientations de recherche et à destination de leurs équipes de recherche





RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

recherche.data.gouv.fr

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



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

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



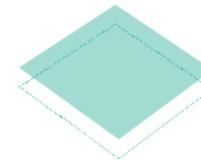
Centre de ressources
compétences
 Recherche Data Gouv

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



Centre de ressources
outils communs aux ateliers
 Recherche Data Gouv

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



Centre de ressources
entrepôt-catalogue
 Recherche Data Gouv

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



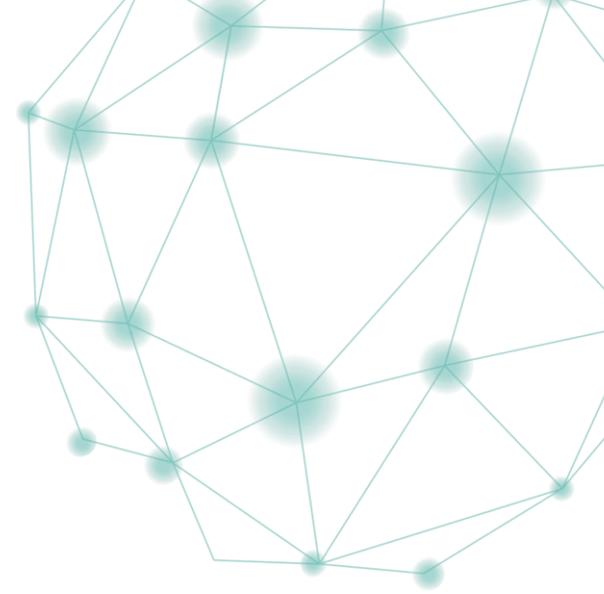
@RechercheDataGv



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

recherche.data.gouv.fr



Un entrepôt de données mutualisé

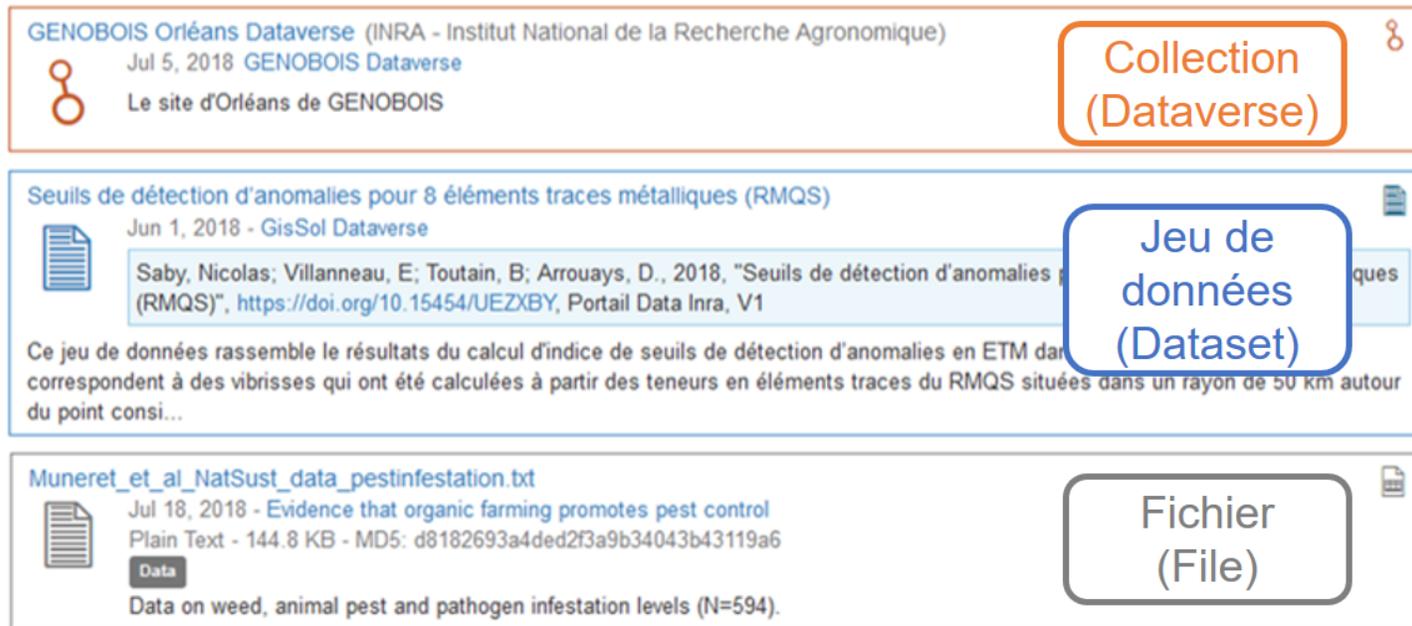
entre les établissements d'enseignement supérieur et de la recherche

- 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



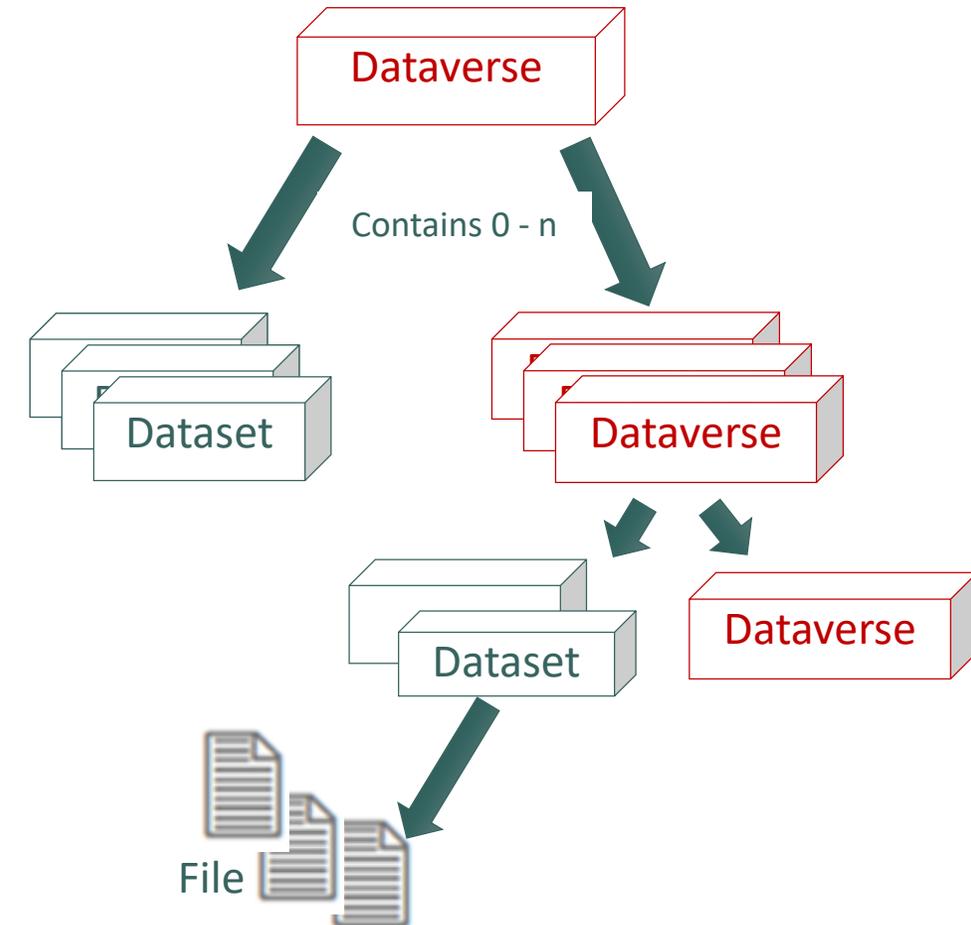
@RechercheDataGv

Entrepôt Recherche Data Gouv : structure

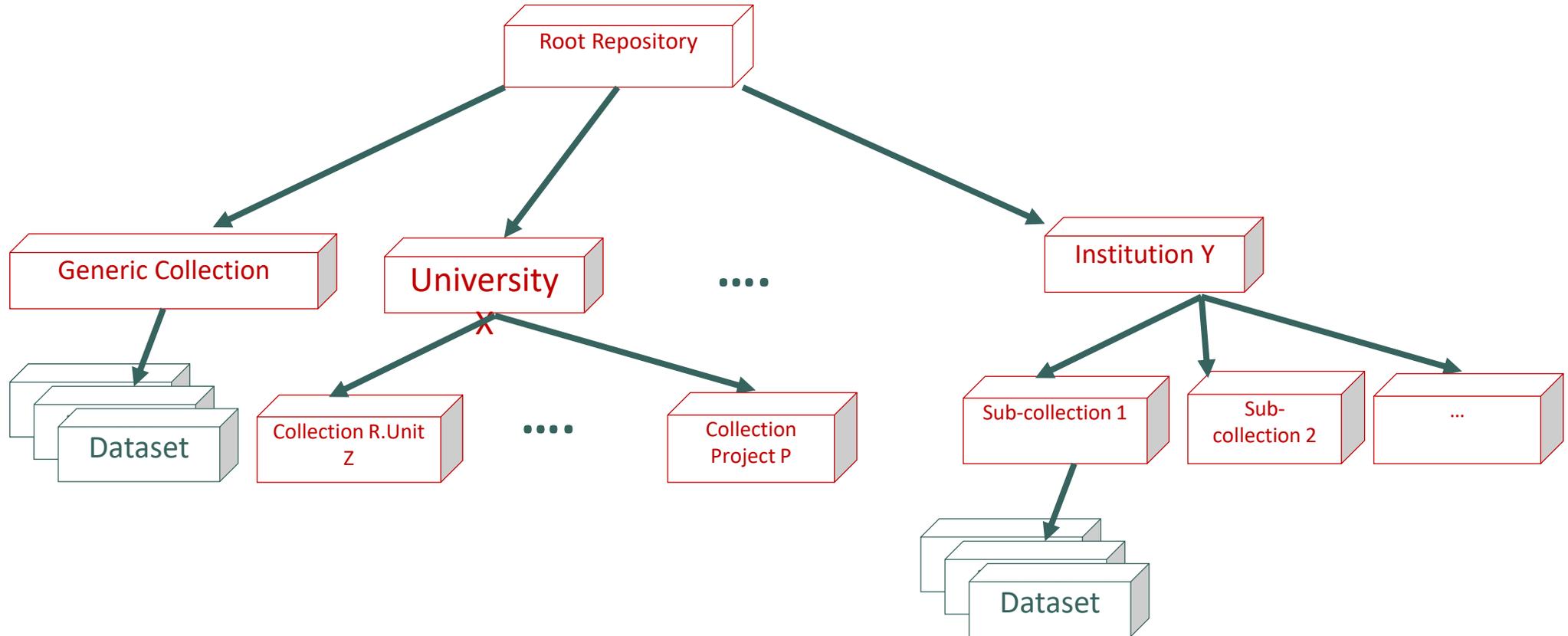


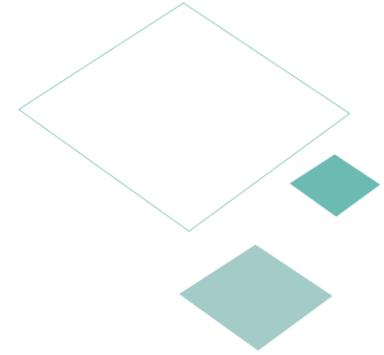
The screenshot illustrates the structure of the Data Gouv repository with three examples:

- Collection (Dataverse):** GENOBOIS Orléans Dataverse (INRA - Institut National de la Recherche Agronomique), Jul 5, 2018. Description: "Le site d'Orléans de GENOBOIS".
- Jeu de données (Dataset):** Seuils de détection d'anomalies pour 8 éléments traces métalliques (RMQS), Jun 1, 2018. Description: "Ce jeu de données rassemble le résultats du calcul d'indice de seuils de détection d'anomalies en ETM dans les zones de culture correspondant à des vibrisses qui ont été calculées à partir des teneurs en éléments traces du RMQS situées dans un rayon de 50 km autour du point consi..."
- Fichier (File):** Muneret_et_al_NatSust_data_pestinfestation.txt, Jul 18, 2018. Description: "Data on weed, animal pest and pathogen infestation levels (N=594)".



Organisation globale de l'entrepôt





Collection générique et espaces institutionnels

- Collection générique
 - Administration et curation réalisée par l'équipe du centre de ressources « entrepôt & catalogue »
 - Ouverte à tous, en particulier aux équipes d'établissements ne disposant pas d'un espace institutionnel dédié
- Espaces institutionnels
 - Administration et curation réalisée par l'établissement porteur
 - Réseau des administrateurs/curateurs



Un catalogue des données de la recherche française

Données de la recherche

Produite par une activité de recherche

Données de la recherche française

Produites ou coproduites par au moins un auteur affilié à un établissement de recherche ou d'enseignement français ; ou dans des **entrepôts de données gérés par des établissements français de la recherche, sans faire de distinction sur l'affiliation des auteurs des données (français ou étrangers)**

Données hébergées dans

Potentiellement tout type d'entrepôt dans la mesure où elles sont signalées manuellement

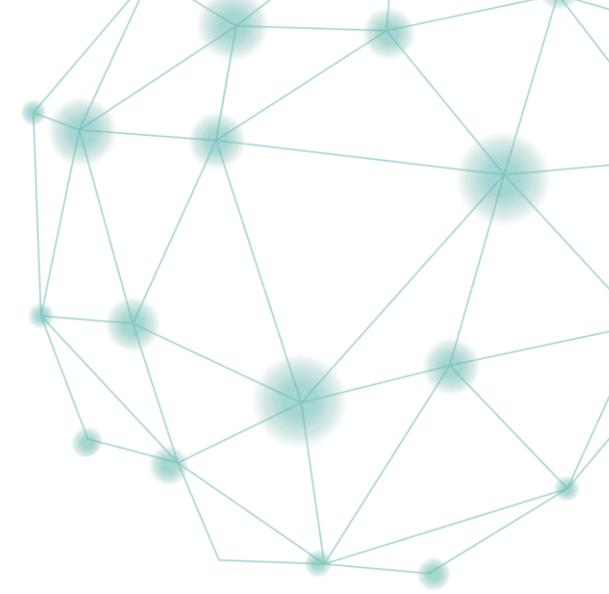
—> Uniquement les entrepôts priorisés par les acteurs de Recherche Data Gouv concernant le référencement automatisé

Travail en cours



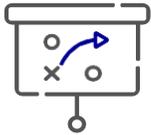


Entre le 8 juillet 2022 et le 25 juin 2024



Chiffres audience du portail (visiteurs acceptant cookies)

- 59 119 visiteurs uniques
- 208 740 pages vues



Chiffres activité support

- 47 formations réalisées à distance
- 838 participants à ces formations
- 1 603 demandes traitées par e-mails pour diverses sollicitations



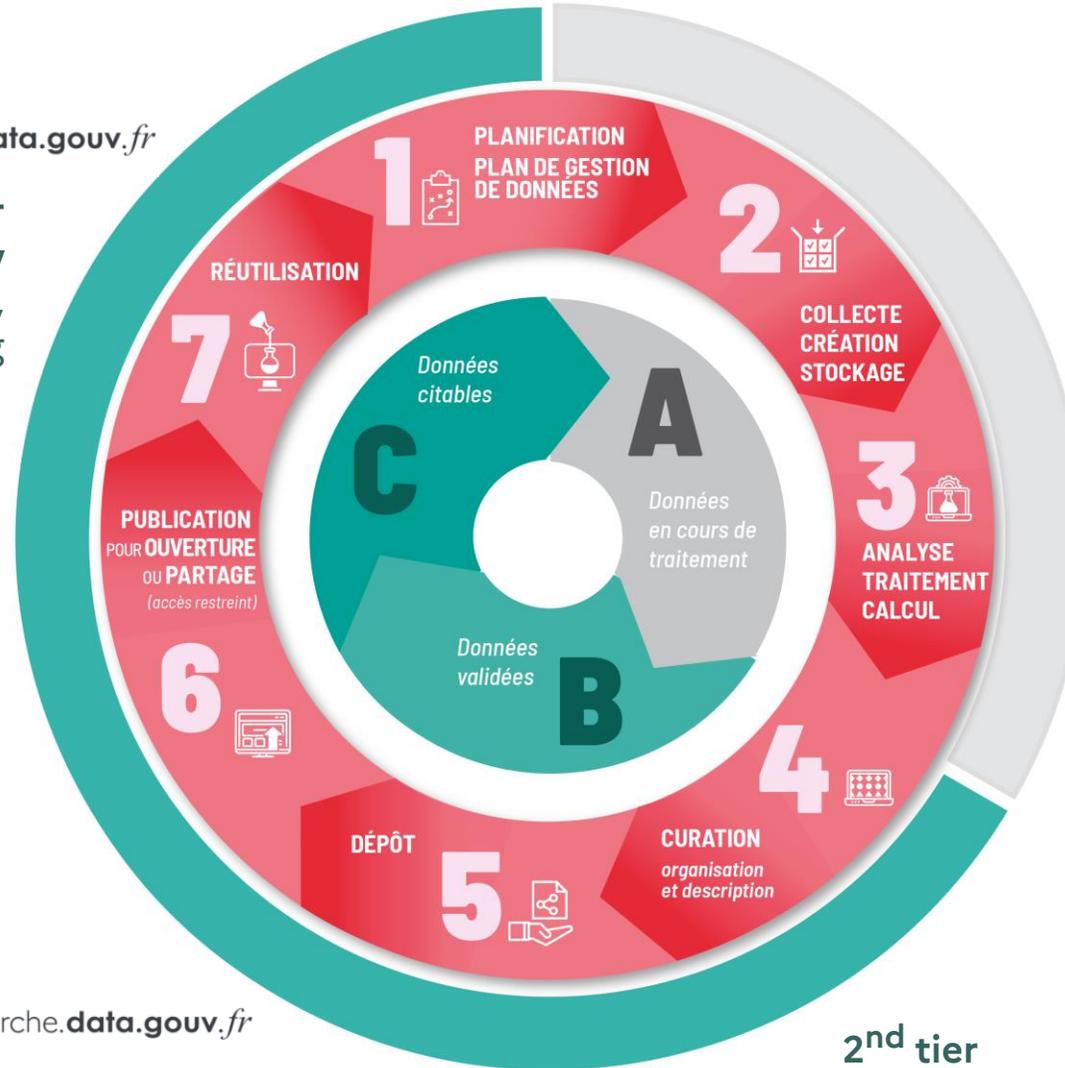
Chiffres entrepôt de données

- 50 espaces institutionnels actifs
- 4060 jeux de données déposés ou moissonnés
- ~51 564 fichiers de données ouverts ou partagés en accès restreints
- 1 240 000 téléchargements de fichiers



3rd tier

Recherche Data Gouv exposes citable data, through opening or sharing



1st tier

Collect, create, store and compute scientific data. This is done within the laboratories and in collaboration with storage and computing service providers



2nd tier

Recherche Data Gouv groups curation, deposit and publication services for validated scientific data



Recherche Data Gouv à l'international

Du côté de la plateforme nationale des données de la recherche

- Rejoindre et co-construire le réseau européen des entrepôts de confiance
- Travailler au-delà des frontières sur les sujets transdisciplinaires

Du côté de l'accompagnement

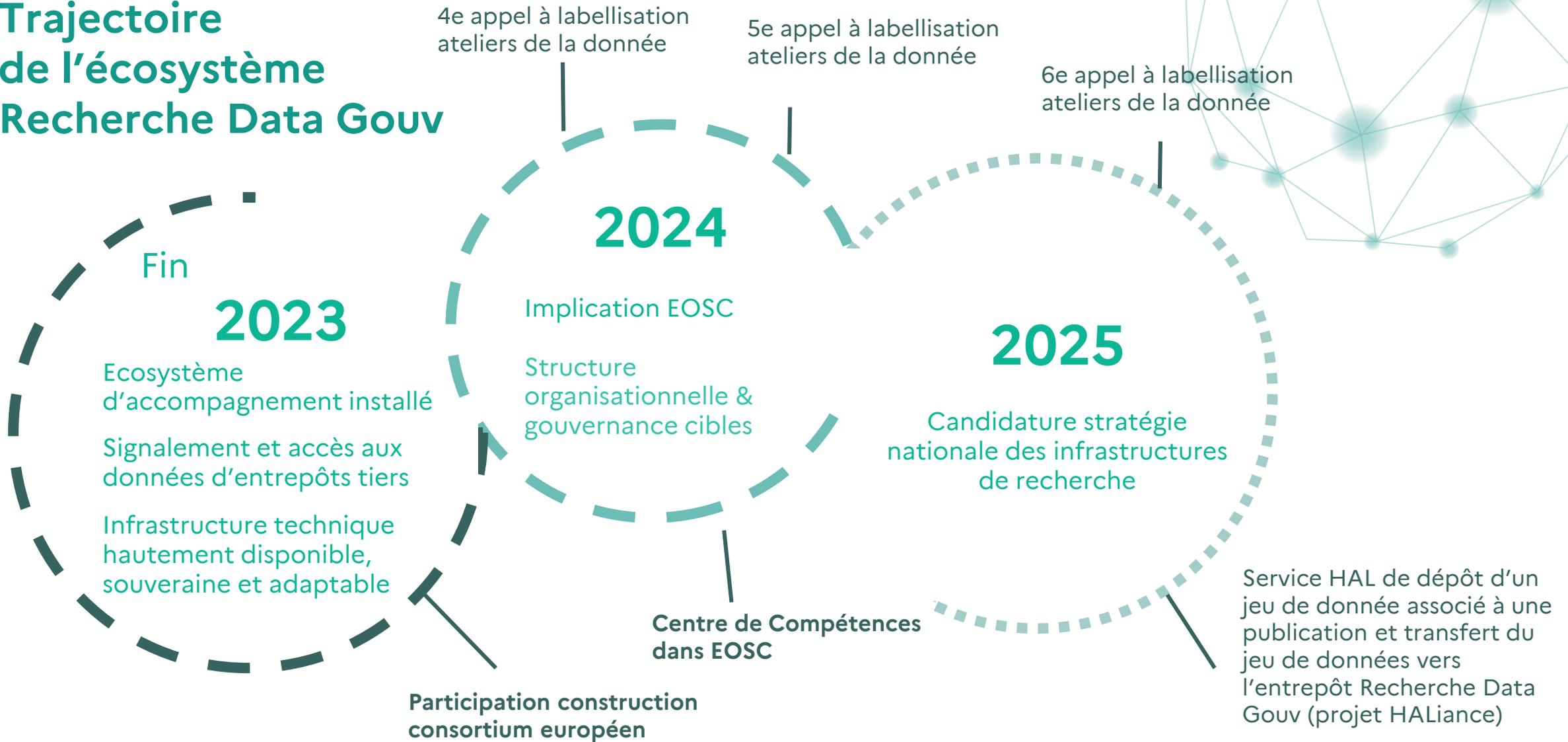
- Premier périmètre du « Centre de Compétences » Français via Skills4EOSC
- Partager les pratiques de support et formation au niveau européen

En tant que service national de données

- Partage de bonnes pratiques au sein de RDA
- Lien et interactions avec d'autres groupes d'intérêt



Trajectoire de l'écosystème Recherche Data Gouv



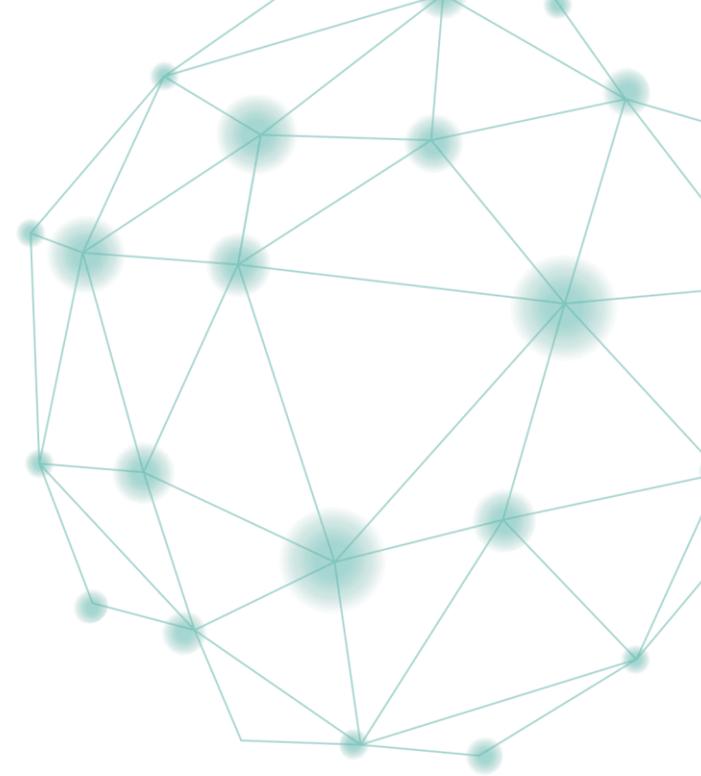


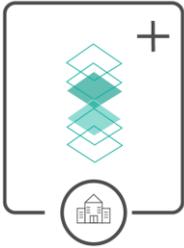
Une construction collective Une gouvernance collective

- Partenaires du projet



- Représentants de la communauté ESR

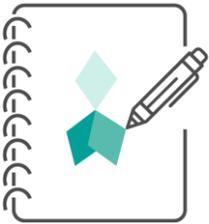




Créer un espace institutionnel

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

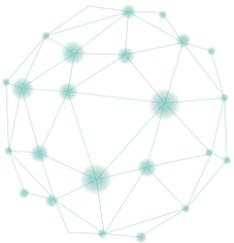
http://recherche.data.gouv.fr/uploads/general/Documents/Demande_creation_espace_institutionnel.pdf



Devenir atelier de la donnée

Pour accompagner en proximité les équipes de recherche dans la gestion, la structuration et la diffusion de leurs données

<https://recherche.data.gouv.fr/fr/page/rejoindre-lecosysteme>



Soutenir le dispositif et participer à la gouvernance

Mobilisation de ressources d'établissements au sein d'une unité mixte de service qui maintient et fait évoluer l'entrepôt mutualisé des données

**Un écosystème
ouvert à toute la
communauté de
recherche et
construit par les
établissements
partenaires**



RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

Pour suivre toutes les actualités

<https://recherche.data.gouv.fr/fr>



@RechercheDataGv



<https://www.linkedin.com/company/recherche-data-gouv/>



@RechercheDataGv