



**MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR  
ET DE LA RECHERCHE**

*Liberté  
Égalité  
Fraternité*



# Etude sur les métiers de la donnée : méthodologie et premiers retours

Isabelle Blanc – Anne Laurent

ANF : Pratiquer la Science ouverte : des services et des compétences en action

 @AMDAC\_MESR





## Point d'étape sur l'étude

- Objectifs de l'étude
- Modalités de conduite de étude
- Synthèse bibliographique : grands enseignements et cadre de référence
- Enquêtes sur les pratiques et activités au sein des communautés les plus avancées
- Premiers retours

## Activités et métiers de la donnée Constats, Enjeux et Engagement



### Constats

Les métiers de la donnée sont parmi les plus recherchés  
Un manque cruel de formations initiales et continues sur une partie des activités liées aux données

- Activités de gestion-diffusion des données : nettoyage, préparation, description (métiers), partage, circulation, ouverture
- Des activités au sein et orientées vers les métiers qui produisent et utilisent les données
- Ces métiers ou ces activités trop souvent apparentés aux métiers de l'informatique, de la statistique ou du calcul scientifique

### Enjeux pour l'ESR

Développer et reconnaître les compétences en matière de gestion, préservation, circulation et ouverture des données pour les chercheurs, les enseignants-chercheurs, les personnels d'appui à la recherche et administratifs tout au long de leur parcours

### Engagement politique

Au-delà de l'ESR et en tant que ministère en charge de l'Enseignement supérieur :

- Identifier pour la formation initiale : les unités d'enseignement, les filières métiers
- Développer les formations continues et décrire des parcours professionnels
- Soutenir et amplifier les formations continues et initiales existantes



## Contexte

- Des études existantes sur les pratiques, activités et métiers de la donnée
- Des définitions de métiers de la donnée identifiées dans différents contextes et différents pays
- Des ouvrages de référence
- Des cadres de référence pour les métiers de gouvernance de la donnée et du numérique
- Un constat d'hétérogénéité dans la qualification des métiers
- Des besoins exprimés

## Objectifs

*L'étude ne consiste pas à réviser les référentiels de métiers de l'ESR*

- Etudier des activités et métiers qui ont pris forme au sein des communautés les plus avancées (FR et international)
- Ne pas s'arrêter au simple état des lieux mais regarder les dynamiques
- **Qualifier les 3-5 métiers de la gestion, préservation et diffusion des données**
  - Activités
  - Compétences
  - Pré-requis métiers
  - Trajectoires de (re)conversion
  - Formations initiales ou continues



## Démarche

### *Une étude en plusieurs temps :*

Premier focus sur les activités autour de la donnée de recherche

1. **Synthèse bibliographique** sur les activités et métiers clés de la gestion, préservation, diffusion de la donnée (data steward, data librarian, data curator...) => étude bibliographique
2. **Analyse quantitative et qualitative** de l'existant des pratiques et des trajectoires dans les environnements de l'ESR ayant mis en place des modes de fonctionnements et d'organisation performants
  - Identifier les pratiques de la donnée des répondants
  - Identifier le statut professionnel des répondants

Ensuite inventaire des formations initiales et continues

Elargissement des observations et propositions en interministériels

## Modalités

Chargée de l'étude : Anne Laurent, professeur de l'université de Montpellier, Directrice de l'institut de science des données de Montpellier

Groupe de travail de l'étude réunissant des représentants de différentes communautés métiers de l'ESR impliquées dans la gestion des données de la recherche

Prestataire pour accompagner la conduite de l'étude :  
Dataactivist

Pilotage étude : MESR Isabelle Blanc

Planning prévisionnel :

Novembre 2022-Septembre 2023

Isabelle Blanc	MESR
Anne Laurent	Univ. Montpellier, pilote de l'étude
Céline Hernandez	Collège données, CNRS
Christine Hadrossek	Collège données, DDOR, CNRS
Véronique Stoll	Collège données, Observatoire de Paris
Pierre-Yves Arnould	Collège données, CNRS
Joanna Janik	Collège compétences, CNRS
Johann Berti	Collège compétences, AMU, ADBU
Sylvie Rousset	DDOR, CNRS
Zoé Ancion	ANR

Frédéric De Lamotte	IFB, centre ref thématique
Yvan Le Bras	PNDB, MNHN, centre ref thématique
Stéphanie Cheviron	BU, Univ. Strasbourg, atelier de la donnée
Sylvie Damy	Univ. Franche Comté, atelier de la donnée
Sandrine Gropp	GT montée en compétences ateliers de la donnée, Univ. Montpellier
Mariannig Le Behec	GIS URFIST, Univ. Lyon Centre de ressources compétences Recherche Data Gouv
Gabriel Gallezot	GIS URFIST, Univ. Nice Centre de ressources compétences Recherche Data Gouv
Fabien Borget	CEDRE, AMU
Juliette Dibie-Barthelemy	DRH, INRAE



## Périmètre

- Portrait de quelques métiers clé selon ce qu'en dit la littérature internationale
- Pas de recherche bibliographique exhaustive : priorité donnée aux revues de littératures et publications post-2018
- Prise en compte de référentiels existants

## Premiers constats

- Pas d'étonnement
- Beaucoup de littérature sur des métiers et activités bien décrits par ailleurs et ne constituant pas le cœur de notre étude (data scientist/analyste/engineer) +> métiers à composante numérique
- Une distinction à faire entre métiers/activités/compétences
- **Gestion et diffusion des données : activités et métiers "non informatique, statistique ou calcul scientifique"**

## Méthode

- Liste des activités
- Placement sur le cycle de vie de
  - La nature des données (en cours de traitement, validées, citables)
  - L'endroit d'exercice des activités (labos/platformes, structures centrales, ...)
- Besoin de formulations non ambiguës



### 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*

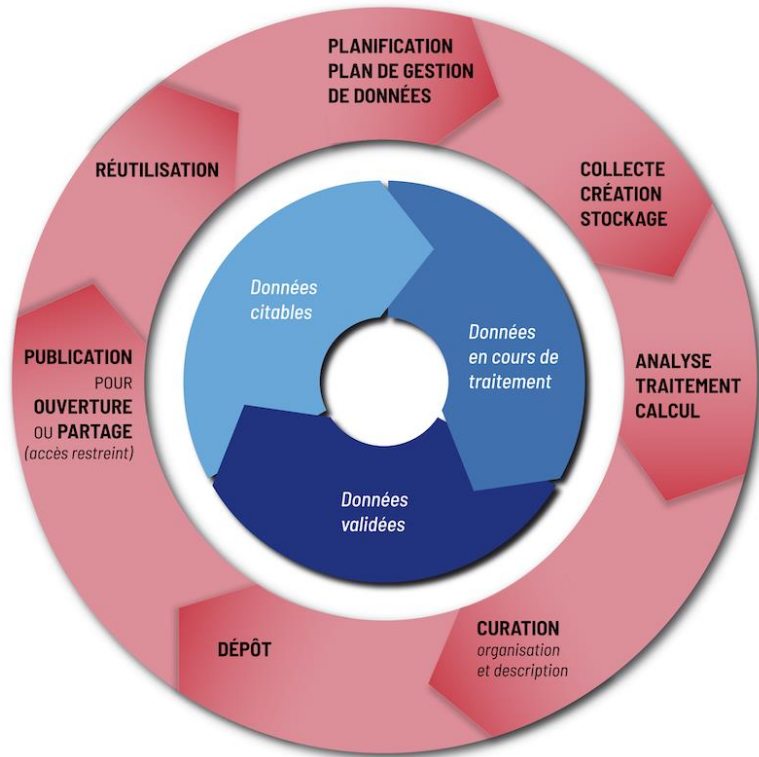
Métadonnées scientifiques : description protocole  
obtention, traitement, matériel, logiciels...

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



## ÉTUDE SUR LES MÉTIERS DE LA DONNÉE

### Cycle de vie de la donnée



Coordonner les activités de gestion et  
diffusion des données

Garantir la qualité des données et leur conformité de  
diffusion par rapport aux cadres législatifs

Superviser l'ensemble du  
cycle de vie

Décrire et mettre en œuvre le protocole d'obtention  
des **données brutes**

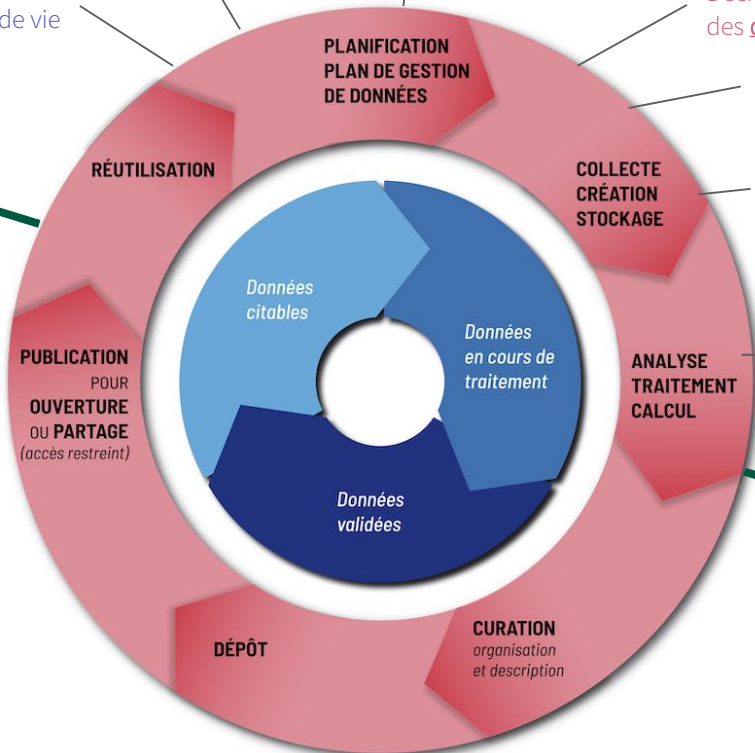
Nettoyer, affiner, convertir et décrire les **données  
brutes** par des **métadonnées scientifiques**

Structurer les données et les outils pour la  
collecte, le stockage et le traitement

Analyser/Représenter/Modéliser/Prédire (stat,  
algo, machine learning) => **données de sortie  
de l'analyse**

Unité / infra de recherche  
Unité / infrastructure de recherche  
Service central

Unité / infra de recherche



Coordonner les activités de gestion et de diffusion des données

Garantir la qualité des données et leur conformité de diffusion par rapport aux cadres législatifs

Superviser l'ensemble du cycle de vie

Décrire et mettre en œuvre le protocole d'obtention des **données brutes**

Nettoyer, affiner, convertir et décrire les **données brutes** par des **métadonnées scientifiques**

Structurer les données et les outils pour la collecte, le stockage et le traitement

Analyser/Représenter/Modéliser/Prédire (stat, algo, machine learning) => **données de sortie de l'analyse**

Créer et formater les **jeux données validées** (ou aide à création et formatage)

Renseigner les métadonnées documentaires (contributeurs, établissements, licences, financeurs projets, ...)

Choisir (ou aider au choix) de l'entrepôt

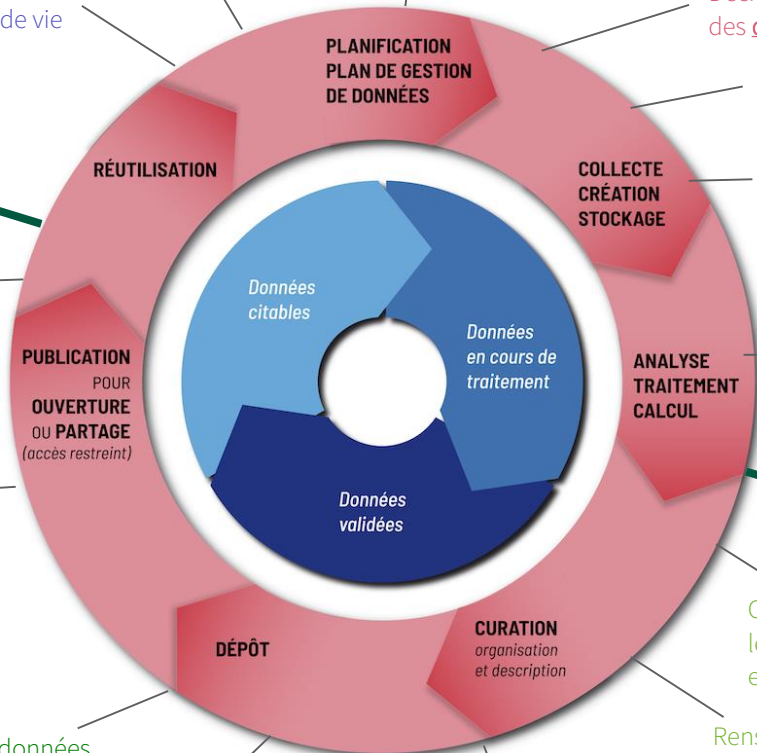
Déposer les jeux de données

Assurer la mise à jour des versions de jeux de données

Publier en ouverture ou en partage restreint

Accompagner la découverte des données

Unité / infra de recherche  
Unité / infrastructure de recherche  
Service central



Coordonner les activités de gestion et de diffusion des données

Garantir la qualité des données et leur conformité de diffusion par rapport aux cadres législatifs

Data scientist  
Data analyst

Superviser l'ensemble du cycle de vie

Décrire et mettre en œuvre le protocole d'obtention des données brutes

Nettoyer, affiner, convertir et décrire les données brutes (par des métadonnées scientifiques)

Structurer les données et les outils pour la collecte, le stockage et le traitement

Analyser/Représenter/Modéliser/Prédire (stat, algo, machine learning) => données de sortie de l'analyse

Créer et formater les jeux données validées (ou aide à création et formatage)

Renseigner les métadonnées documentaires (contributeurs, établissements, licences, financeurs projets, ...)

Choisir (ou aider au choix) de l'entrepôt

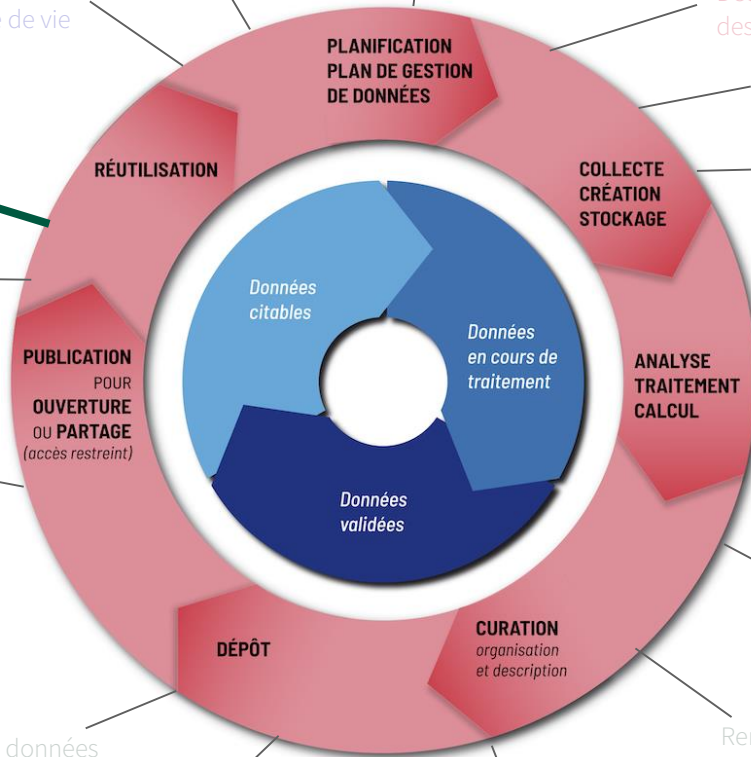
Déposer les jeux de données

Assurer la mise à jour des versions de jeux de données

Publier en ouverture ou en partage restreint

Accompagner la découverte des données

Unité / infra de recherche  
Unité / infrastructure de recherche  
Service central



Coordonner les activités de gestion et de diffusion des données

Garantir la qualité des données et leur conformité de diffusion par rapport aux cadres législatifs

Superviser l'ensemble du cycle de vie

Décrire et mettre en œuvre le protocole d'obtention des données brutes

Nettoyer, affiner, convertir et décrire les données brutes (par des métadonnées scientifiques)

Structurer les données et les outils pour la collecte, le stockage et le traitement

Analyser/Représenter/Modéliser/Prédire (stat, algo, machine learning) => données de sortie de l'analyse

Créer et formater les jeux données validées (ou aide à création et formatage)

Renseigner les métadonnées documentaires (contributeurs, établissements, licences, financeurs projets, ...)

Choisir (ou aider au choix) de l'entrepôt

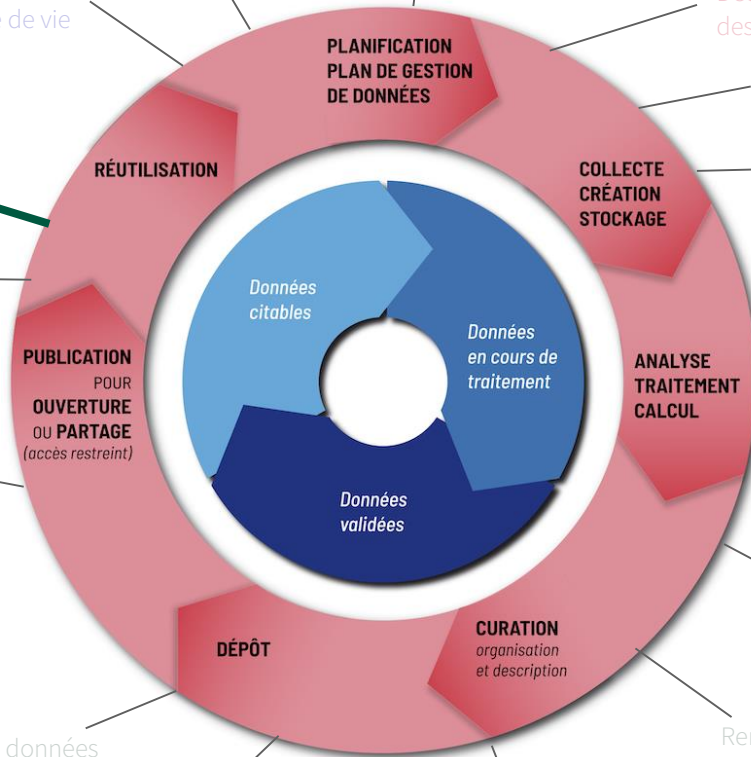
Déposer les jeux de données

Assurer la mise à jour des versions de jeux de données

Publier en ouverture ou en partage restreint

Accompagner la découverte des données

Unité / infra de recherche  
Unité / infrastructure de recherche  
Service central



Coordonner les activités de gestion et de diffusion des données

Garantir la qualité des données et leur conformité de diffusion par rapport aux cadres législatifs

Superviser l'ensemble du cycle de vie

Décrire et mettre en œuvre le protocole d'obtention des données brutes

Nettoyer, affiner, convertir et décrire les données brutes (par des métadonnées scientifiques)

Structurer les données et les outils pour la collecte, le stockage et le traitement

Analyser/Représenter/Modéliser/Prédire (stat, algo, machine learning) => données de sortie de l'analyse

Créer et formater les jeux de données validées (ou aide à création et formatage)

Renseigner les métadonnées documentaires (contributeurs, établissements, licences, financeurs projets, ...)

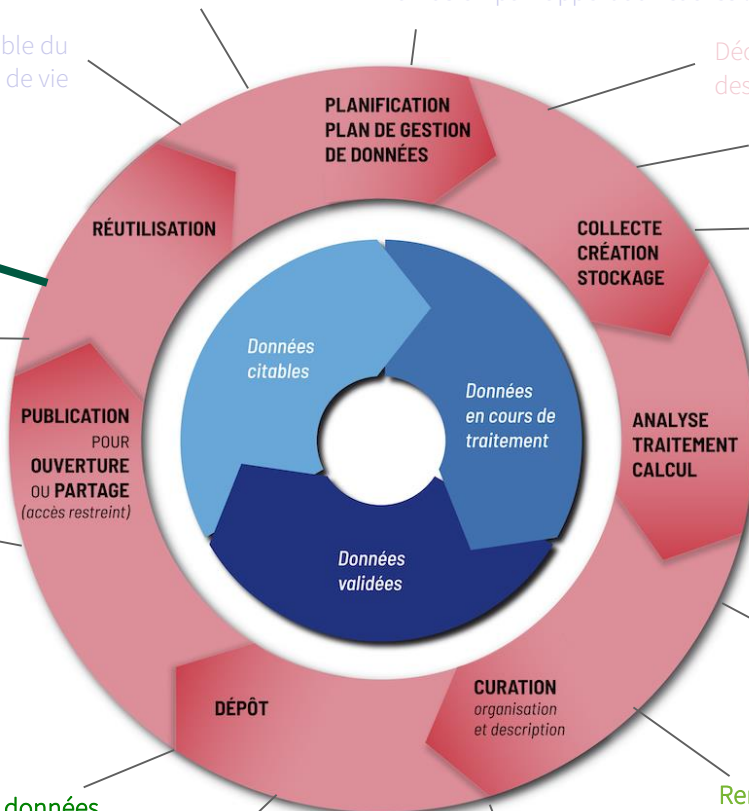
Déposer les jeux de données Choisir (ou aider au choix) de l'entrepôt

Assurer la mise à jour des versions de jeux de données

Publier en ouverture ou en partage restreint

Accompagner la découverte des données

Unité / infra de recherche  
 Unité / infrastructure de recherche  
 Service central



Data librarian

Superviser l'ensemble du cycle de vie

Décrire et mettre en œuvre le protocole d'obtention des données brutes

Nettoyer, affiner, convertir et décrire les données brutes (par des métadonnées scientifiques)

Structurer les données et les outils pour la collecte, le stockage et le traitement

Analyser/Représenter/Modéliser/Prédire (stat, algo, machine learning) => données de sortie de l'analyse

Créer et formater les jeux de données validées (ou aide à création et formatage)

Renseigner les métadonnées documentaires (contributeurs, établissements, licences, financeurs projets, ...)

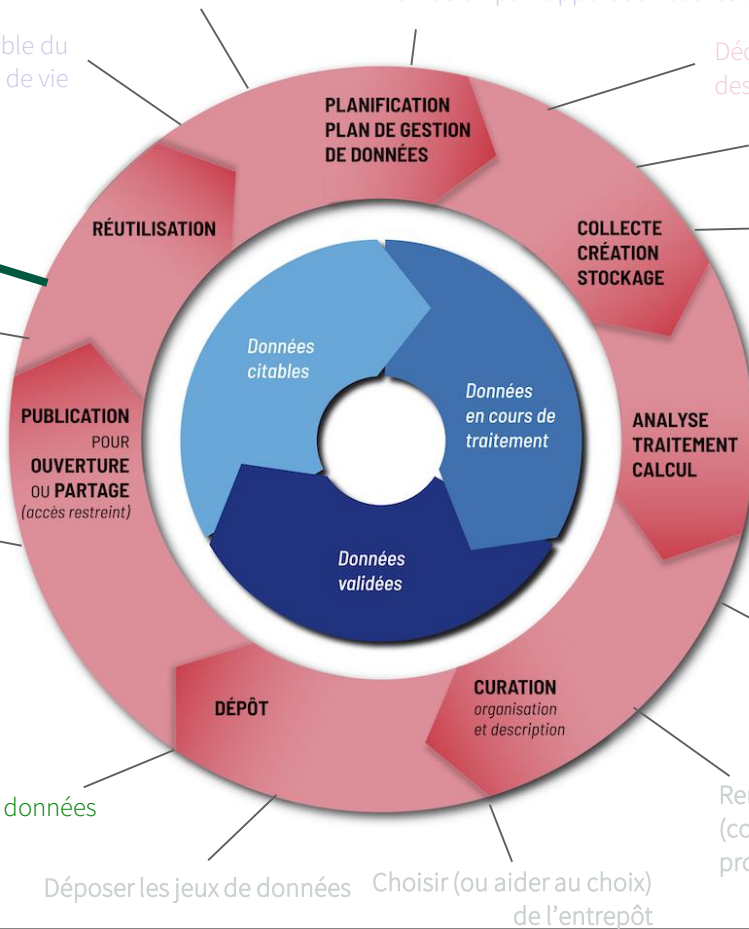
Déposer les jeux de données Choisir (ou aider au choix) de l'entrepôt

Assurer la mise à jour des versions de jeux de données

Publier en ouverture ou en partage restreint

Accompagner la découverte des données

Unité / infra de recherche  
Unité / infrastructure de recherche  
Service central



Coordonner les activités de gestion et de diffusion des données

Garantir la qualité des données et leur conformité de diffusion par rapport aux cadres législatifs

MetaData- curator

Superviser l'ensemble du cycle de vie

Décrire et mettre en œuvre le protocole d'obtention des données brutes

Nettoyer, affiner, convertir et **décrire les données brutes (par des métadonnées scientifiques)**

Structurer les données et les outils pour la collecte, le stockage et le traitement

Analyser/Représenter/Modéliser/Prédire (stat, algo, machine learning) => données de sortie de l'analyse

Créer et formater les jeux données validées (ou aide à création et formatage)

Renseigner les métadonnées documentaires (contributeurs, établissements, licences, financeurs projets, ...)

Choisir (ou aider au choix) de l'entrepôt

Déposer les jeux de données

Assurer la mise à jour des versions de jeux de données

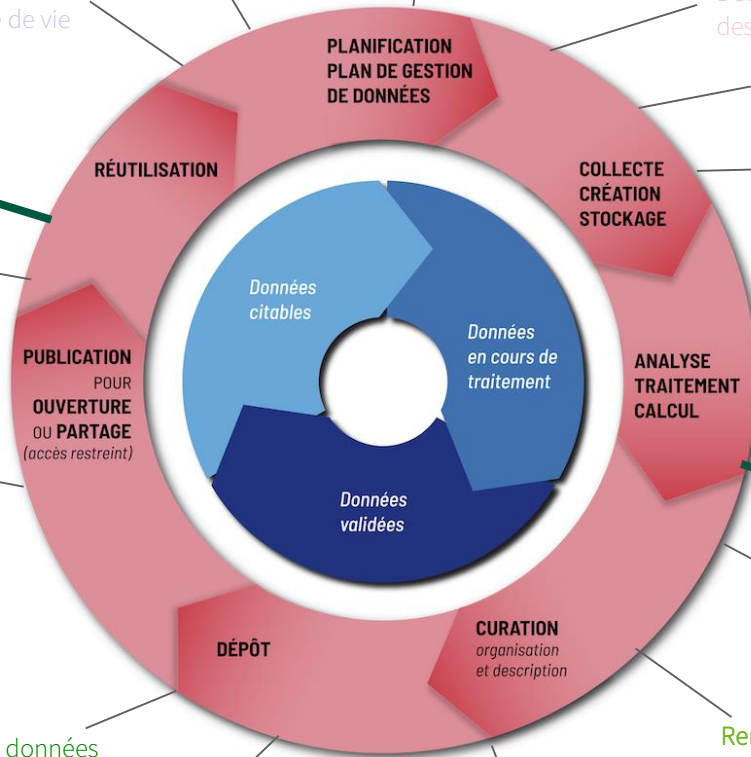
Publier en ouverture ou en partage restreint

Accompagner la découverte des données

Unité / infrastructure de recherche  
Service central

Unité de recherche

Unité de recherche



MetaData curator



Coordonner les activités de gestion et de diffusion des données

Garantir la qualité des données et leur conformité de diffusion par rapport aux cadres législatifs

Superviser l'ensemble du cycle de vie

Décrire et mettre en œuvre le protocole d'obtention des données brutes

Nettoyer, affiner, convertir et décrire les données brutes (par des métadonnées scientifiques)

Structurer les données et les outils pour la collecte, le stockage et le traitement

Analyse stat, algo, représentation des données, modélisation, machine learning, analyse prédictive => données de sortie de l'analyse

Créer et formater les jeux données validées (ou aide à création et formatage)

Renseigner les métadonnées documentaires (contributeurs, établissements, licences, financeurs projets, ...)

Choisir (ou aider au choix) de l'entrepôt

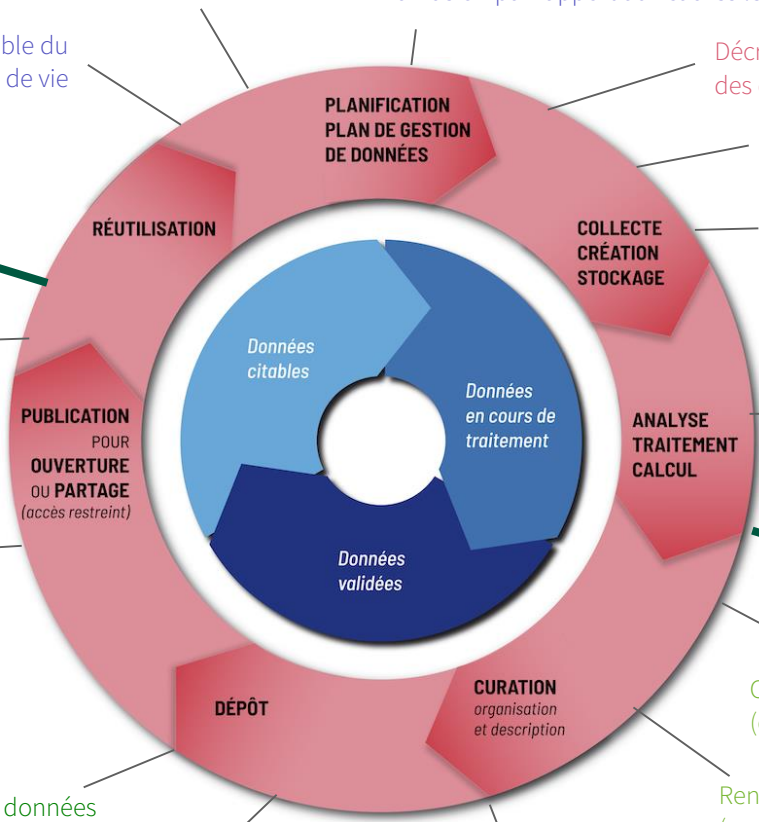
Déposer les jeux de données

Assurer la mise à jour des versions de jeux de données

Publier en ouverture ou en partage restreint

Accompagner la découverte des données

Unité / infra de recherche  
 Unité / infrastructure de recherche  
 Service central





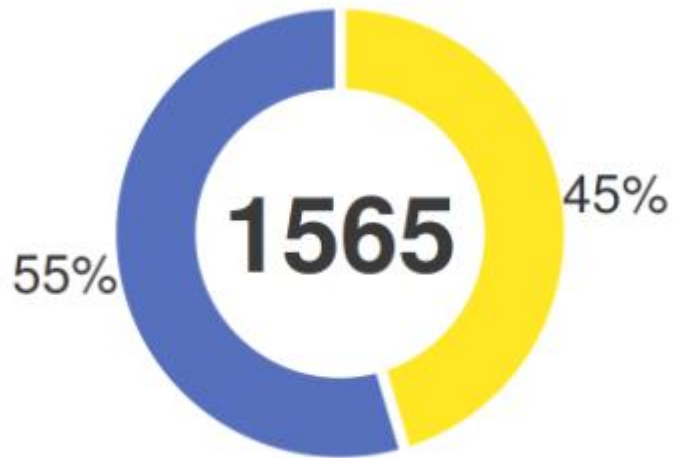
## Objectifs



- Destinée à toutes les personnes ayant des activités en lien avec les données de la recherche
- Vise à identifier
  - Les pratiques de la donnée des répondants
  - Lieu de réalisation des activités
  - Liste des activités
  - Rôle dans les activités (« je réalise » / « je coordonne »)
  - Poids des activités
  - Le statut professionnel des répondants
  - Les parcours et les dynamiques



## Premiers résultats

1565 répondants

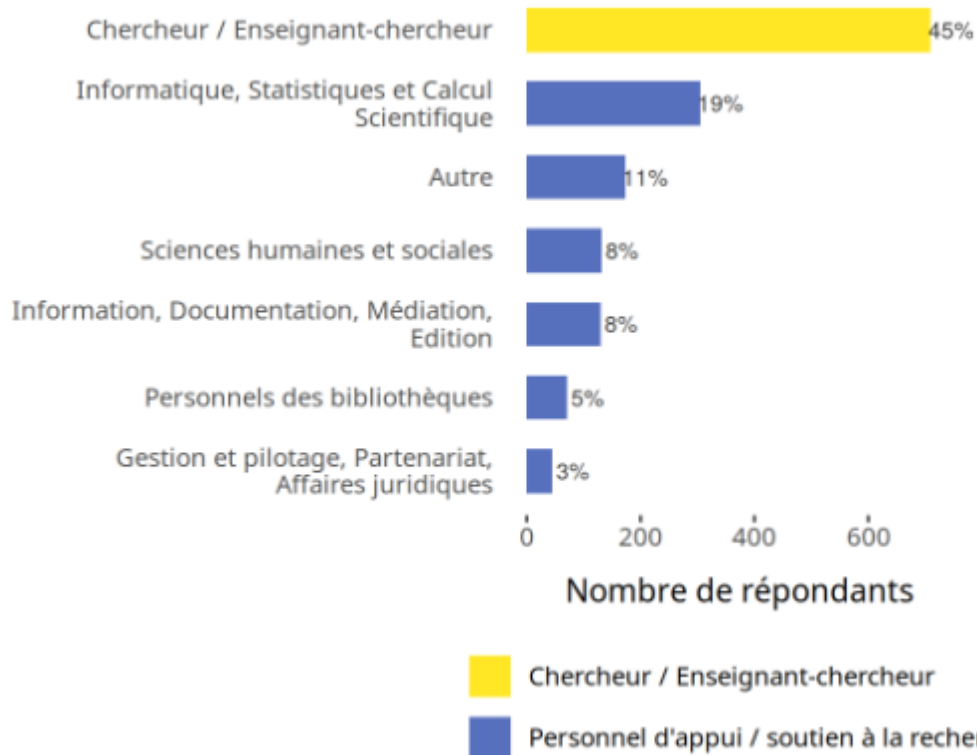


-  Chercheur / Enseignant-chercheur
-  Personnel d'appui / soutien à la recherche



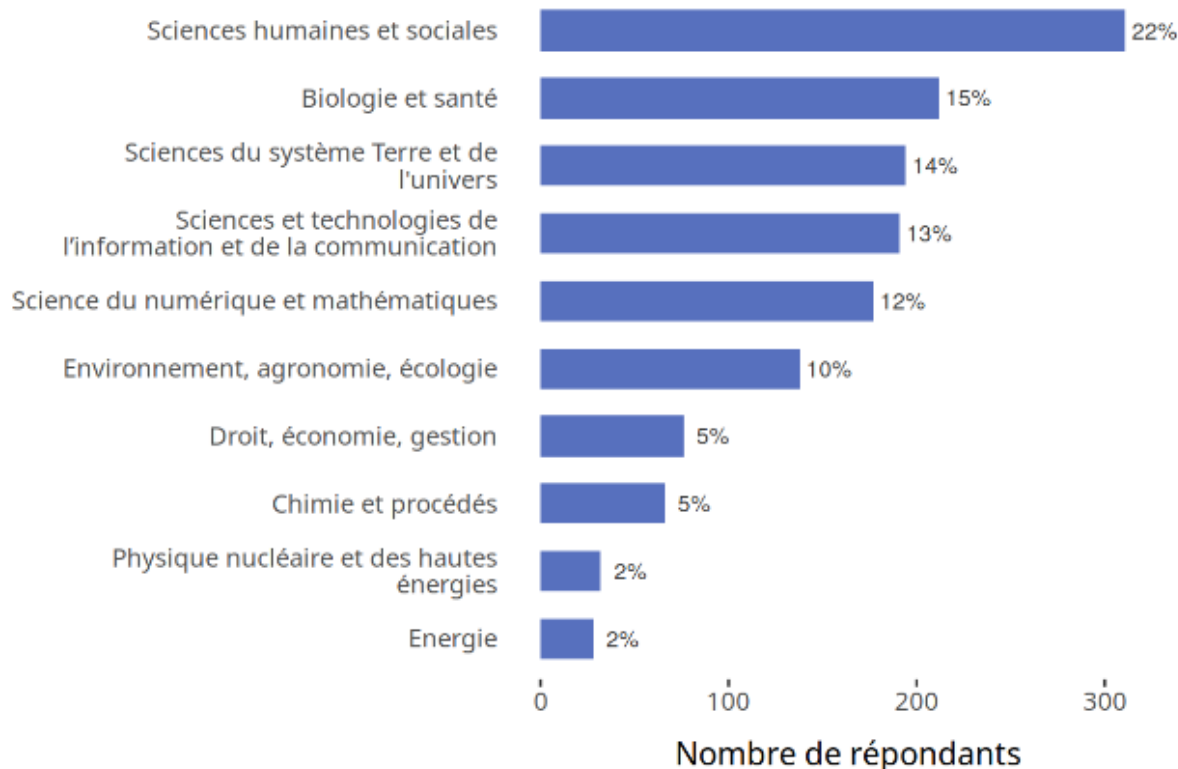
## Premiers résultats

1565 répondants



## Quel est le domaine de votre formation initiale ?

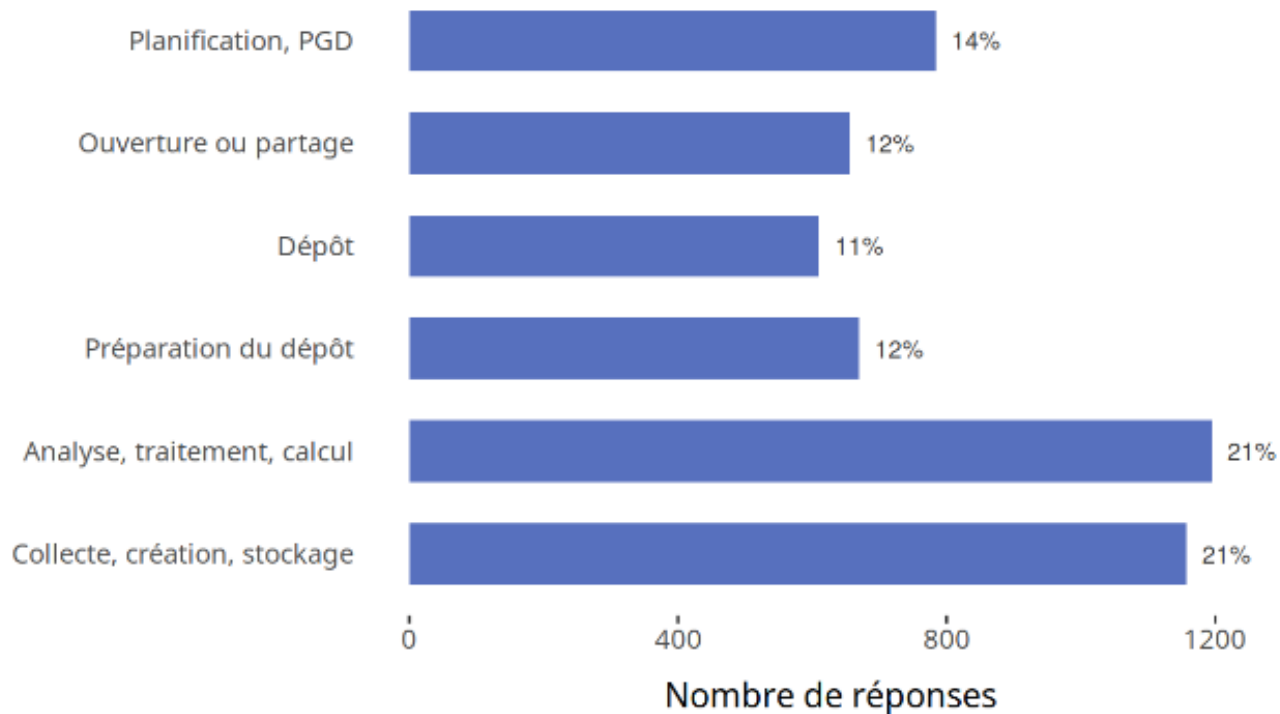
1565 réponses, soit 100% des répondants





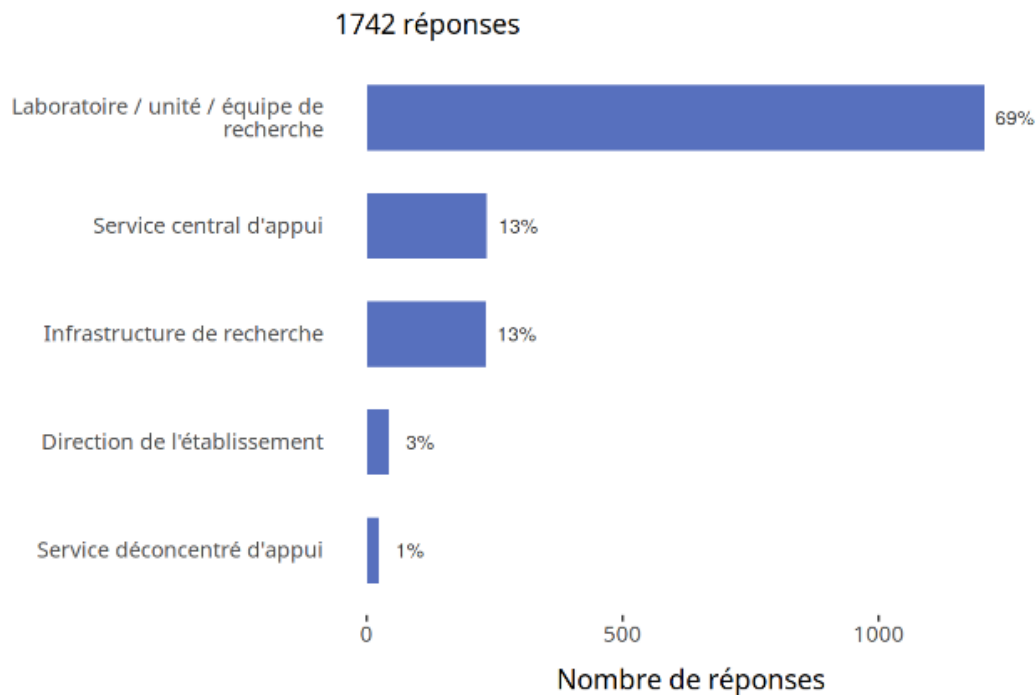
## Sur quelle(s) partie(s) du cycle de vie des données intervenez-vous ?

5642 réponses





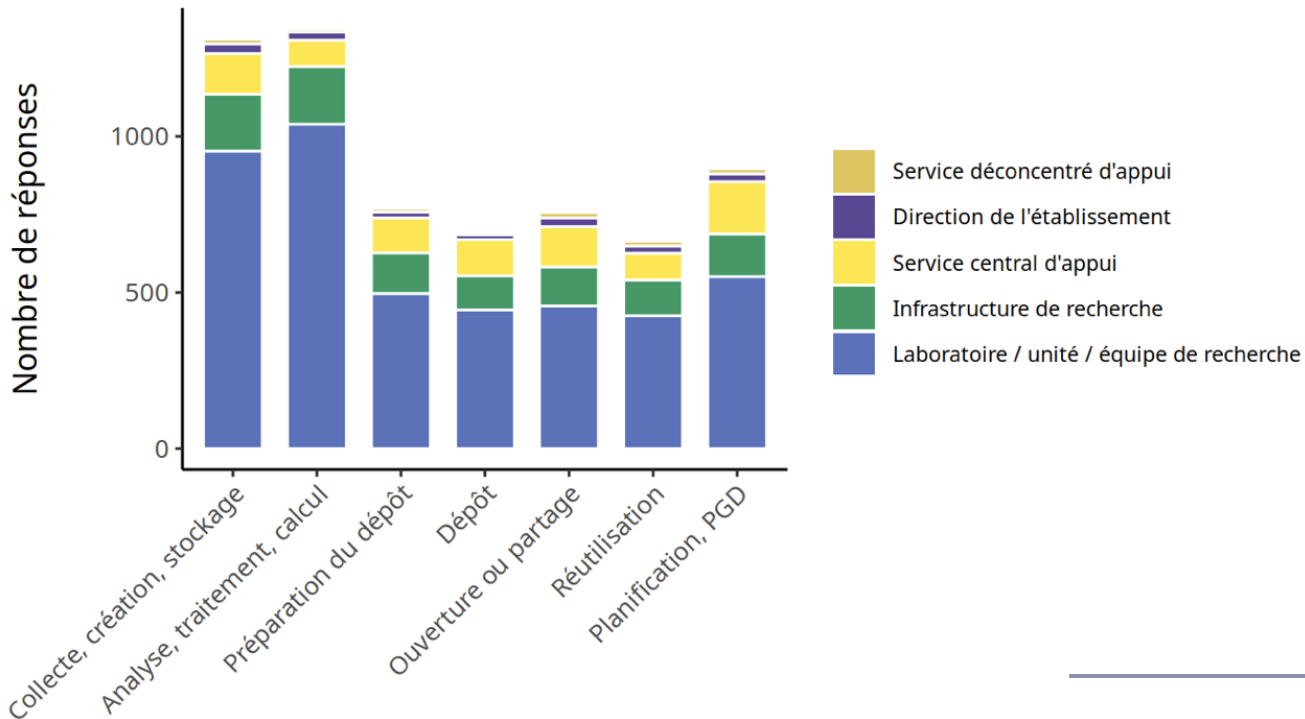
## Où réalisez-vous votre activité en lien avec les données de recherche ?



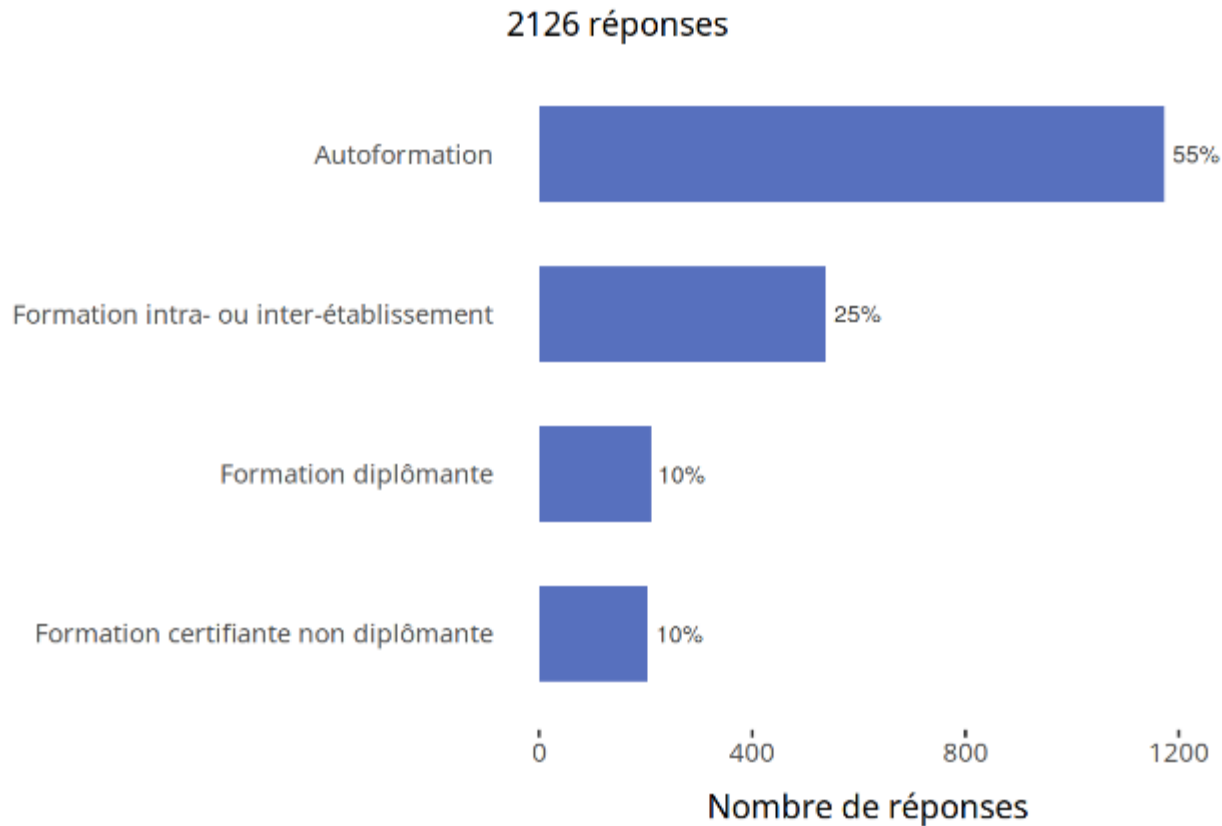


## Lieux d'activités selon l'intervention dans le cycle de vie de la donnée

6434 réponses









## Objectifs

Approfondir la réalité des activités et des parcours

- Position dans l'organisation
- Tâches et compétences liées aux données
- Parcours professionnel (y compris parcours de formation/auto-formation)

## Modalités

- Entretien avec un panel large de 30 acteurs représentatifs, notamment les volontaires de l'étude par questionnaire

## Prochaines étapes

- Finalisation du travail de bibliographie sous la forme d'un rendu mettant en lumière les activités et leur déclinaison le long du cycle de vie de la donnée
- Approfondissement de l'étude statistique du questionnaire
- Analyse et synthèse des entretiens
- Travail sur les formations et les parcours



**MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR  
ET DE LA RECHERCHE**

*Liberté  
Égalité  
Fraternité*



# Rapport complet de l'étude sur les métiers de la donnée

**Octobre 2023**



@AMDAC\_MESR

