



Gestion et ouverture des données de la recherche : pratiques, représentations, besoins

Marie-Liesse Bertram

► To cite this version:

Marie-Liesse Bertram. Gestion et ouverture des données de la recherche : pratiques, représentations, besoins. [Rapport de recherche] Conservation National des Arts et Métiers. 2022. hal-03824487

HAL Id: hal-03824487

<https://cnam.hal.science/hal-03824487>

Submitted on 21 Oct 2022

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 - ShareAlike 4.0 International License



le cnam

Gestion et ouverture des données de la recherche

Pratiques, représentations, besoins

Rapport de l'enquête réalisée en 2022 au
Conservatoire national des arts et métiers.

Direction des bibliothèques et de la documentation

À propos de l'autrice

Marie-Liesse Bertram est médiatrice documentaire à la Direction des bibliothèques et de la documentation (DBD) du Cnam. Elle co-administre le portail HAL de l'établissement. Étudiante en master « Politique des bibliothèques et de la documentation » (PBD) de l'École nationale des sciences de l'information et des bibliothèques (Enssib), elle a mené cette enquête dans le cadre de son apprentissage au Cnam. Elle a rédigé ce rapport.



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/).

Sommaire

À propos de l'autrice	2
Sommaire	3
Remerciements	5
Introduction.....	6
0.1 Contexte	6
0.2 L'enquête du Cnam	8
0.3 Méthodologie	9
1. Profil des répondant·es	12
1.1 Le laboratoire des répondant·es	12
1.2 La discipline des répondant·es	15
1.4 Le genre des répondant·es	17
1.5 L'âge des répondant·es.....	17
1.6 L'expérience des répondant·es dans le milieu de la recherche	18
2. Nature des données produites.....	20
2.1 Les types de données produits.....	20
2.2 Le format des données produites	23
2.3 Le volume des données produites	23
2.4 Les données personnelles	24
3. Pratiques de gestion et d'ouverture des données	26
3.1 Lieu de stockage des données.....	26
3.2 Accès aux données	27
3.3 Partage des données en libre accès	28
3.4 Expérience du PGD	30
3.5 Expérience du data paper	30
3.6 Aide à la gestion des données	31
4. Motivations et freins à l'ouverture des données	34
4.1 Adhésion au principe d'ouverture des données	34
4.2 Raisons d'une mise en libre accès des données.....	35
4.3 Raisons d'une fermeture de l'accès aux données	37
5. Besoins et acteur·rices.....	40
5.1 Connaissance du sujet par les répondant·es.....	40
5.2 Besoins.....	41

5.3 Interlocuteur·rices privilégié·es.....	43
Conclusion	45
Bibliographie.....	46
Annexes	49
Table des illustrations.....	69

Remerciements

L'équipe ayant travaillé sur cette enquête adresse ses remerciements à la Direction de la recherche et à la Direction de la communication du Cnam pour avoir relayé les informations sur l'enquête auprès de la communauté de recherche de l'établissement.

Nous remercions également les chercheur·euses qui nous ont accordé des entretiens exploratoires, ainsi que celles et ceux qui ont relu le questionnaire avant publication.

Merci enfin à toutes les personnes qui ont pris le temps de répondre à cette enquête, et tout particulièrement aux personnes avec qui nous nous sommes entretenues pour les discussions enrichissantes qui ont résulté de notre rencontre.

Introduction

0.1 Contexte

0.1.1 Définition et enjeux des données de la recherche

Plusieurs définitions des données de la recherche coexistent¹, mais la plus communément admise est celle de l'Organisation de coopération et de développement économiques (OCDE) de 2007, qui a été révisée en 2021 pour y ajouter la notion de financements publics :

« Les "données de la recherche" sont définies comme des enregistrements factuels (tels que les chiffres, textes, images et sons) issus de la recherche financée tout ou partie sur fonds publics, qui sont utilisés comme source principale pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. » (OCDE, 2021)

À l'heure du déluge de données et de l'entrée dans un « quatrième paradigme de la science² » (Gray *et al.*, 2005) centré sur la donnée, les données de la recherche posent de nombreux défis quant à leur gestion et à leur ouverture.

« L'ouverture des données de recherche répond à cinq enjeux :

- accélérer les découvertes scientifiques, les innovations et le retour sur investissement en recherche et développement ;
- encourager la collaboration scientifique et les possibilités de recherche interdisciplinaire ;
- éviter la duplication des expériences, favoriser la réutilisation des données et minimiser le risque de perte des données ;
- assurer l'intégrité et la reproductibilité de la recherche (meilleure qualité des résultats, transparence des méthodologies) ;
- accéder librement à une masse de données ouvrant de nouveaux champs d'analyse non envisagés par le producteur des données (gain de temps et de ressources). » (Dedieu et Fily, 2015)

¹ Pour en savoir plus sur les définitions existantes des données de la recherche, nous renvoyons le·la lecteur·rice vers une communication présentée lors du 1^{er} atelier Valorisation et analyse des données de la recherche (VADOR) de l'Inforsid (Schöpfel, Kergosien et Prost, 2017) et vers la thèse de Violaine Rebouillat (Rebouillat, 2019, pp. 23-45).

² Le quatrième paradigme de la science est une hypothèse formulée en 2005 par Jim Gray, qui propose une histoire de la science fondée sur la succession de quatre paradigmes : empirisme, théorie, programmation, exploration de données.

0.1.2 L'actualité du sujet

En France, le premier Plan national pour la science ouverte (PNSO1) (MESRI, 2018), qui portait sur la période 2018-2021, a été renouvelé à l'été 2021 dans un deuxième Plan national pour la science ouverte (PNSO2) (MESRI, 2021). Ce dernier prolonge les axes sur lesquels le travail avait déjà été entamé : généralisation de l'accès ouvert aux publications, structuration et ouverture des données de la recherche, inscription dans une dynamique durable, etc. Le deuxième axe du PNSO2, « Structurer, partager et ouvrir les données de la recherche », occupe une place d'importance dans la politique gouvernementale française en faveur de la science ouverte.

Ainsi, le PNSO2 entend :

- Mettre en œuvre l'obligation de diffusion des données financées sur fonds publics ;
- Créer *Recherche Data Gouv*, la plateforme fédérée des données de la recherche³ ;
- Promouvoir l'adoption d'une politique de données pour les rendre FAIR⁴ ;
- Et plus globalement, transformer les pratiques pour faire de la science ouverte le principe par défaut.

En parallèle du PNSO2 et des directives européennes, toutes les agences de financement françaises et européennes demandent à ce jour la rédaction d'un plan de gestion de données (PGD)⁵, et la plupart d'entre elles⁶ demandent également l'ouverture des données pour celles qui peuvent l'être (en respectant le principe « aussi ouvert que possible, aussi fermé que nécessaire »).

Par ailleurs, le 3 décembre 2021 est paru le décret n° 2021-1572 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique. Les établissements publics concernés⁷ doivent promouvoir « la mise à disposition [...] des données », inciter à « la publication des résultats de recherche dits négatifs » (art. 2), définir « une politique de conservation, de communication et de réutilisation des résultats bruts des travaux scientifiques menés en [leur] sein », veiller « à la mise en œuvre par leur personnel de plans de gestion de données » et contribuer « aux

³ Cette plateforme a été inaugurée le 8 juillet 2022 et proposera à terme cinq modules, parmi lesquels un entrepôt de données généraliste et la labellisation d'ateliers de la donnée destinés à accompagner les chercheur-euses.

⁴ L'acronyme FAIR, proposé en 2016 par un groupe de chercheur-euses issu-es de différents laboratoires (Wilkinson *et al.*, 2016), renvoie aux termes « faciles à trouver, accessibles, interopérables, réutilisables », caractéristiques nécessaires à un jeu de données pour qu'il soit de qualité.

⁵ Un plan de gestion de données (PGD, ou en anglais DMP pour *Data Management Plan*) est un document détaillant les processus permettant de créer, documenter, stocker, partager et protéger les données produites lors d'un projet.

⁶ L'ouverture est demandée par l'ANR et l'Institut national du cancer [INCa], par toutes les agences de financement de la commission européenne, et par quasiment toutes les agences européennes, en dehors d'Initiatives d'excellence [IDEX] en France (Inist-CNRS, 2022).

⁷ « Établissements publics contribuant au service public de la recherche ; fondations reconnues d'utilité publique ayant pour activité principale la recherche publique ; Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES). »

infrastructures qui permettent la conservation, la communication et la réutilisation des données » (art. 6).

Déjà impliqués dans des services liés aux publications ouvertes, de nombreux établissements de recherche français commencent à accompagner leurs chercheur·euses sur la question des données.

0.2 L'enquête du Cnam

0.2.1 Genèse de l'enquête

L'engagement du Cnam en faveur de la science ouverte est porté principalement par la Direction des bibliothèques et de la documentation (DBD) au sein d'une cellule dédiée, en collaboration avec la Direction de la recherche (DR).

Depuis 2019, un portail HAL est administré par Aurélie Puybonnieux (Centre de documentation sur la formation et le travail, DBD) et permet d'offrir une fenêtre sur la production scientifique de l'établissement. Des formations au dépôt sur HAL sont organisées régulièrement à destination des chercheur·euses de l'établissement.

Afin de s'aligner sur les directives ministérielles et de répondre aux besoins des communautés de recherche du Cnam, la cellule science ouverte de la DBD a formulé en 2021 un projet d'enquête sur la question des données de la recherche, afin d'identifier des préconisations pour l'établissement. Les discussions ont donné lieu au recrutement d'une apprentie en master « Politique des bibliothèques et de la documentation » (PBD) à l'Enssib, chargée entre autres missions de conduire cette enquête sur l'année universitaire 2021-2022. Prenant en charge le financement de l'apprentissage, la Direction de la recherche (DR) a soutenu cette initiative.

0.2.2 Quels objectifs ?

Cinq objectifs ont été formalisés pour cette enquête :

- Dresser un état des lieux sur les pratiques concernant les données : types de données, volume, sécurisation, stockage, archivage, ouverture, partage, etc. ;
- Connaître les perceptions des chercheur·euses en matière de données de la recherche ;
- Ancrer l'établissement dans le mouvement plus large de la connaissance des données de la recherche ;
- Sensibiliser par cette enquête et la restitution qui en sera faite les chercheur·euses aux données de la recherche ;
- Établir des préconisations pour l'établissement selon les besoins recueillis.

0.2.3 Déroulé du travail d'enquête

Le travail autour de l'enquête s'est déroulé de novembre 2021 à mai 2022. L'enquête s'est déroulée en deux temps : une phase quantitative, via un questionnaire en ligne ; puis une phase qualitative, à travers des entretiens individuels.

L'alternante a effectué, dans un premier temps, un *benchmark* d'enquêtes similaires menées par d'autres établissements (GTSO données Couperin, 2022). Les questionnaires et les guides d'entretien existants ont été compilés et retravaillés. En lien avec les objectifs formulés et le contexte d'établissement, les questions les plus pertinentes ont été sélectionnées et parfois reformulées pour les expurger des biais repérés. En parallèle, des entretiens exploratoires avec des membres des laboratoires du Cnam ont été menés par deux membres de la cellule science ouverte, en novembre et décembre 2021.

En janvier 2022, le questionnaire a été finalisé, mis en ligne sur la plateforme LimeSurvey et testé par plusieurs enseignant-es-chercheur-euses, qui n'ont pas suggéré de modifications à l'issue de cette phase de test. Il a ensuite été validé par la DBD et transmis à la DR.

Le questionnaire a été ouvert pour une durée d'un mois, à partir de mi-février. La phase quantitative a été suivie d'une phase qualitative. Une série de sept entretiens semi-directifs ont été menés avec huit chercheur-euses différent-es qui avaient accepté d'être recontacté-es à cet effet.

L'analyse des résultats du questionnaire a été menée en parallèle de la conduite des entretiens, sur Excel. La retranscription des entretiens a permis de compléter ces réponses par des exemples et des ressentis individuels, conduisant ainsi à formuler des préconisations pertinentes au regard des aspirations et craintes du public visé.

0.3 Méthodologie

0.3.1 Questionnaire

0.3.1.1 Structure du questionnaire

L'enquête est structurée en sept sections comportant chacune entre deux et huit questions, pour un total de 27 questions. Neuf questions sont obligatoires. L'essentiel de l'enquête est composé de questions fermées : les questions ouvertes sont au nombre de trois.

La section 1 (trois questions) permet d'identifier le laboratoire des répondant-es, leur champ disciplinaire et leur statut, et de filtrer ainsi les répondant-es ne correspondant pas à la cible de l'enquête. Elle est complétée par la section 6 (trois questions) qui vise à recueillir des données sociodémographiques sur les répondant-es.

La section 7 (trois questions) a permis de recueillir le contact des répondant-es ayant accepté de participer à un entretien lors de la deuxième partie de l'enquête.

Le corps de l'enquête est donc composé de 18 questions. La section 2 (quatre questions) porte sur la nature des données produites par les répondant·es. La section 3 (huit questions) permet d'identifier les pratiques des répondant·es quant à la gestion et l'ouverture de leurs données. Elle est complétée par la section 4 (trois questions) qui porte sur les motivations et les freins à l'ouverture des données. Enfin, la section 5 (trois questions) interroge les répondant·es sur leurs besoins et permet de connaître les acteur·ices les mieux identifié·es sur ces questions.

0.3.1.2 Administration du questionnaire

Le questionnaire a été communiqué *via* la plateforme d'enquête en ligne LimeSurvey. Il a été ouvert un mois, du 17 février au 17 mars 2022.

La communication autour de l'enquête s'est faite sur les sites de la recherche et des bibliothèques du Cnam, sur Twitter et *via* un *Savez-vous-que ?*⁸ et l'envoi d'un mail à l'ensemble des chercheur·euses de l'établissement par la DR. Un nouveau *Savez-vous-que ?* reprenant l'information a été publié le 10 mars, une semaine avant la clôture du questionnaire.

0.3.1.3 Nettoyage des données

Le questionnaire a reçu au total 293 réponses, dont 138 réponses complètes (le·la participant·e a cliqué sur « envoyer » à la fin du questionnaire) et 155 réponses incomplètes.

Parmi les réponses incomplètes, 96 d'entre elles (61 % des réponses incomplètes) sont constituées de personnes ayant cliqué sur le lien, consulté la page introductive, mais n'ayant pas lancé le questionnaire.

Les 59 réponses des personnes ayant lancé le questionnaire mais n'étant pas parvenues jusqu'au bout de celui-ci sont réparties comme suit :

- Arrêt sur la première page : 14 réponses (24 %) ;
- Arrêt sur la p. 2 : 21 réponses (36 %) ;
- Arrêt sur la p. 3 : onze réponses (19 %) ;
- Arrêt sur la p. 4 : six réponses (10 %) ;
- Arrêt sur la p. 5 : six réponses (10 %) ;
- Arrêt sur la p. 6 : 0 réponse ;
- Arrêt sur la p. 7 (ayant consulté tout le questionnaire mais n'ayant pas cliqué sur envoyer) : une réponse (2 %).

Après nettoyage des réponses incomplètes inexploitable, nous obtenons 162 réponses exploitables. Les questions sur le laboratoire d'appartenance et le statut des répondant·es ont permis de nettoyer dix réponses supplémentaires (étudiant·es, chercheur·euses extérieur·es au Cnam, etc.). À l'issue du nettoyage des données, 152 réponses ont été conservées car considérées comme exploitables.

⁸ Nom de la *newsletter* interne du Conservatoire national des arts et métiers.

0.3.2 Entretiens

0.3.2.1 Structure des entretiens

Une série d'entretiens semi-directifs a été menée. La structure reprenait les grands thèmes abordés dans l'enquête. Le guide d'entretien⁹ a rarement été respecté, et les questions ont été adaptées aux thèmes abordés spontanément par les enquêté·es.

0.3.2.2 Organisation des entretiens

La dernière question de l'enquête visait à recueillir le contact des chercheur·euses intéressé·es par la participation à un entretien. Un échantillon de vingt personnes a été contacté, en tenant compte du laboratoire et de la discipline, afin de permettre une relative représentativité du panel interrogé. Huit chercheur·euses ont ainsi donné suite. Les entretiens ont été menés par Marie-Liesse Bertram, Aurélie Puybonnieux (responsable de la cellule science ouverte, DBD) et Anaïs Scalla (responsable du pôle politique documentaire et science ouverte, DBD).

Nous souhaitons, par ailleurs, réaliser deux entretiens de groupes avec des chercheur·euses issu·es de laboratoires ayant largement répondu au questionnaire ; cependant, l'organisation de ceux-ci s'est avérée difficile et nous avons privilégié des entretiens individuels.

0.3.2.3 Retranscriptions

Les entretiens ont été enregistrés avec l'accord des répondant·es. Les passages permettant d'illustrer et de compléter les données issues du questionnaire ont été retranscrits afin de servir de base aux préconisations.

Afin de permettre aux répondant·es d'exprimer le plus librement possible leur opinion sur des sujets parfois chargés politiquement, nous avons fait le choix de ne pas partager de retranscriptions complètes. Les répondant·es ont été anonymisé·es : le genre et les fonctions ont pu être modifiés quand ils ne gênaient pas l'appréhension du contexte.

⁹ Reproduit en annexe.

1. Profil des répondant·es

Résumé : tous les laboratoires et les disciplines proposées ont été représentés dans le questionnaire, dans des proportions variables. Les statistiques du questionnaire révèlent une surreprésentation des maître·sses de conférence par rapport à d'autres statuts. La répartition des genres est relativement équilibrée dans les réponses au questionnaire. La tranche d'âge médiane est de 40-49 ans et l'essentiel des répondant·es bénéficie d'une expérience de plus de 20 ans dans le milieu de la recherche.

Les sections 1 et 6 du questionnaire avaient pour but d'obtenir différentes informations sur le profil des répondant·es. On peut ici signaler un biais dans la représentation de la population cible : les personnes n'ayant pas répondu au questionnaire sont peut-être celles qui se sentaient le moins concernés par la question des données ; ainsi, on peut faire l'hypothèse que la connaissance des enjeux liés aux données de la recherche et surtout l'intérêt pour ceux-ci sont plus importants dans les résultats de cette enquête qu'ils ne le sont en réalité.

1.1 Le laboratoire des répondant·es

Les laboratoires¹⁰ sont tous représentés parmi les répondant·es au questionnaire. Sans surprise, les laboratoires comptant le plus de chercheur·euses du Cnam sont généralement ceux ayant fait le plus de réponses. Ainsi, le Lirsa et le Cédric se placent en tête avec respectivement 28 et 18 réponses (*Figure 1 - Nombre de répondant·es par laboratoire d'appartenance*).

Pour autant, la participation des laboratoires s'est faite dans des proportions variables, quand on rapporte le nombre de répondant·es à la taille du laboratoire (*Figure 2 - Pourcentage de l'effectif du laboratoire ayant ou non participé à l'enquête*). Certains laboratoires ont ainsi vu une bonne proportion de leurs membres participer à l'enquête, sans que cela s'explique par la très petite taille de leurs effectifs¹¹ : on peut citer à titre d'exemple le LCM (44 % de participation), le LMSSC (33 % de participation) et SayFood (50 % de participation).

Les dix entretiens complémentaires nous ont permis de rencontrer onze personnes issues de huit laboratoires différents.

¹⁰ Voir liste en annexe.

¹¹ Le laboratoire Metabiot, créé en 2022, a par exemple recueilli 100 % de participation. Mais le fait qu'il ne compte au moment de l'enquête qu'un seul membre issu du Cnam nous empêche de considérer ce résultat comme significatif.

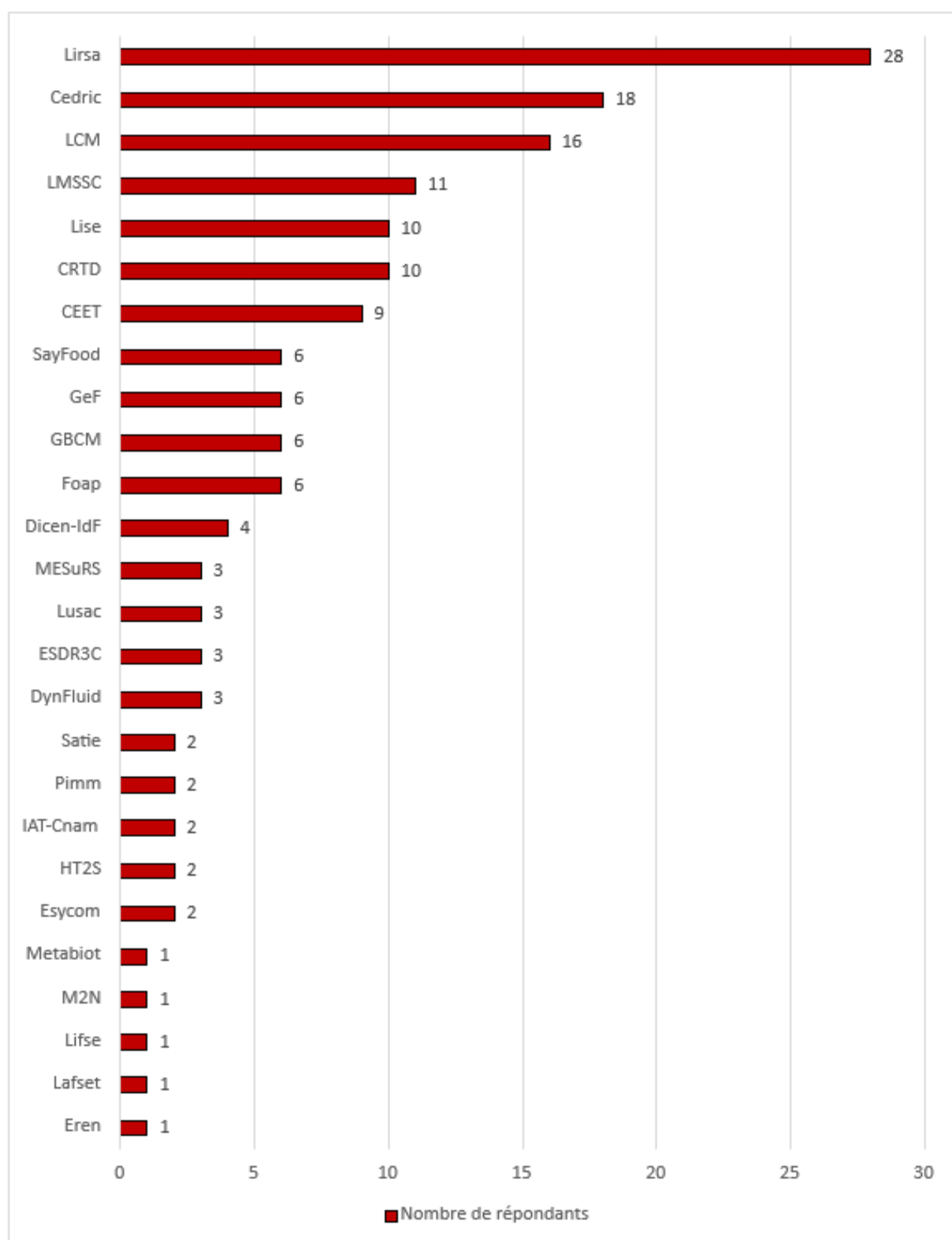


Figure 1 - Nombre de répondant-es par laboratoire d'appartenance

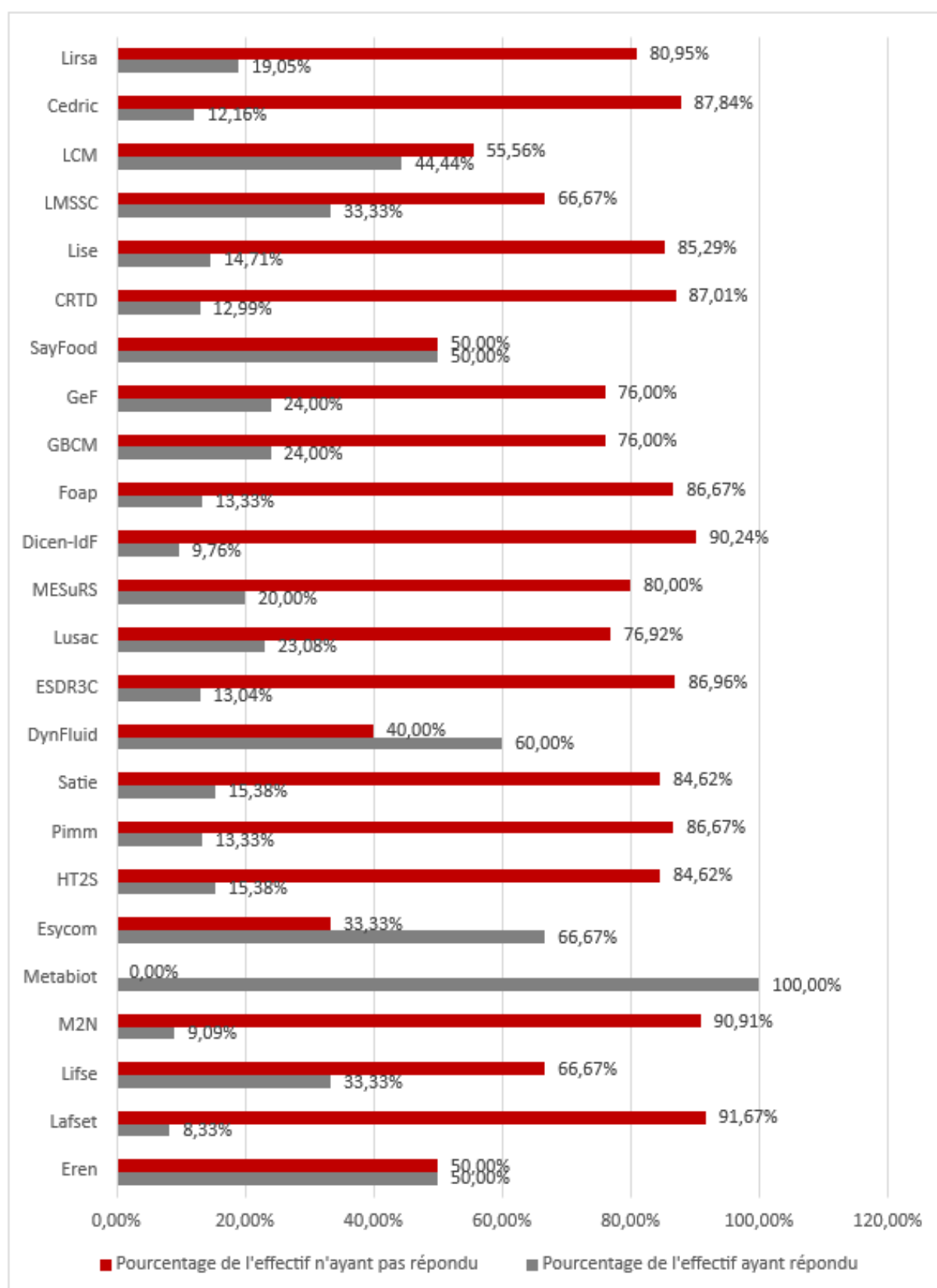


Figure 2 - Pourcentage de l'effectif du laboratoire ayant ou non participé à l'enquête

1.2 La discipline des répondant·es

La typologie des disciplines figurant dans l'enquête proposait des catégories très larges, avec comme objectif de ne pas alourdir le questionnaire par une liste complète des sections du Conseil national des universités (CNU). Cependant, il aurait pu être intéressant de plus la détailler pour permettre une analyse plus fine lors des tris croisés.

Il s'agissait d'une question à choix multiple, permettant aux chercheur·euses à la frontière entre plusieurs disciplines d'en sélectionner plusieurs. Ainsi, si 86 % des répondant·es (131 personnes) n'ont sélectionné qu'un champ disciplinaire, 11 % (17 personnes) en ont sélectionné deux et 3 % (quatre personnes) en ont sélectionné trois.

Toutes les disciplines proposées ont été représentées dans le questionnaire, dans des proportions variables relativement représentatives de la répartition au sein de l'établissement (*Figure 3 - Répartition disciplinaire des répondant·es*). 35 % des répondant·es (53 personnes) ont sélectionné les sciences de l'ingénieur (SI), 24 % (37 personnes) les lettres-sciences humaines et sociales (LSHS), 24 % (37 personnes) le droit-économie-politique-gestion (DEPG), 15 % (23 personnes) les mathématiques-informatique (MI), 8 % (douze personnes) la chimie-matériaux, 7 % (dix personnes) les sciences du vivant-médecine (SVM) et enfin 3 % (cinq personnes) la physique-sciences de la terre et de l'univers (PSTU).

Pour les entretiens, nous nous sommes efforcées d'avoir une certaine représentativité des champs disciplinaires. Si nous n'avons pas pu nous entretenir avec un·e chercheur·euse en sciences du vivant-médecine, nous avons rencontré des personnes issues de toutes les autres disciplines.

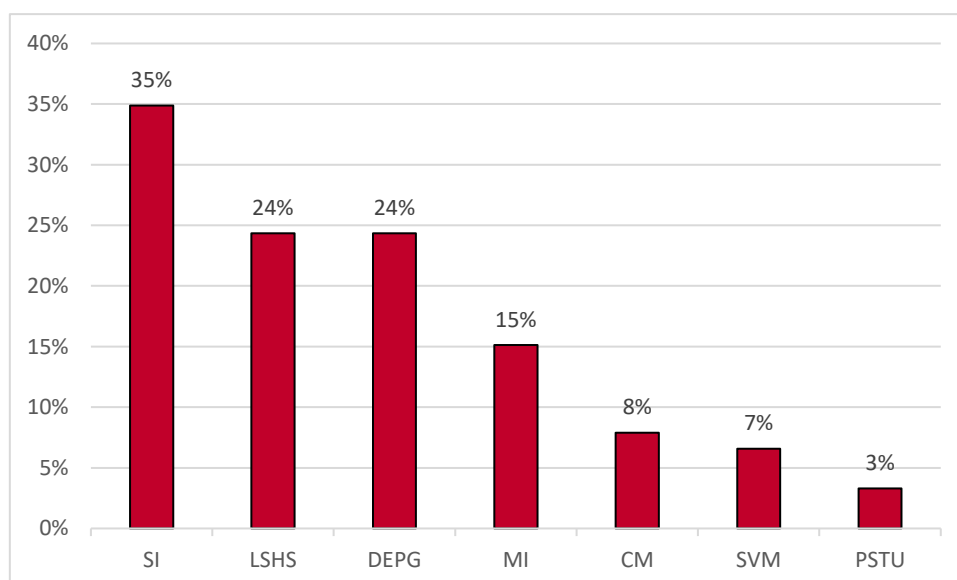


Figure 3 - Répartition disciplinaire des répondant·es

1.3 Le statut des répondant·es

Les statistiques du questionnaire révèlent une surreprésentation des maître·sses de conférence par rapport à d'autres statuts (*Figure 4 - Nombre de répondant·es par statut*). Ceux-ci représentent 46 % des répondant·es (70 personnes), alors que le second statut le plus représenté, celui des personnels ingénieur·es et technicien·nes de recherche et de formation (ITRF), ne représente que 15 % des répondant·es (23 personnes). Viennent ensuite les professeur·es des universités (12 % des répondant·es, 18 personnes), les chargé·es de recherche et les doctorant·es (tous les deux représentant 9 % des répondant·es, 13 personnes), les professeur·es titulaires de chaire (6 % des répondant·es, neuf personnes), les enseignant·es-chercheur·euses associé·es (Past-Mast) (3 % des répondant·es, cinq personnes) et enfin les post-doctorant·es (2 % des répondant·es, trois personnes).

Lors des entretiens, nous n'avons pas pu nous entretenir avec des post-doctorant·es ni des Past-Mast. Les autres statuts ont été représentés. Nous avons rencontré principalement des maître·sses de conférence (la moitié des entretiens). Nous avons aussi rencontré plusieurs directeur·rices d'unités de recherche.

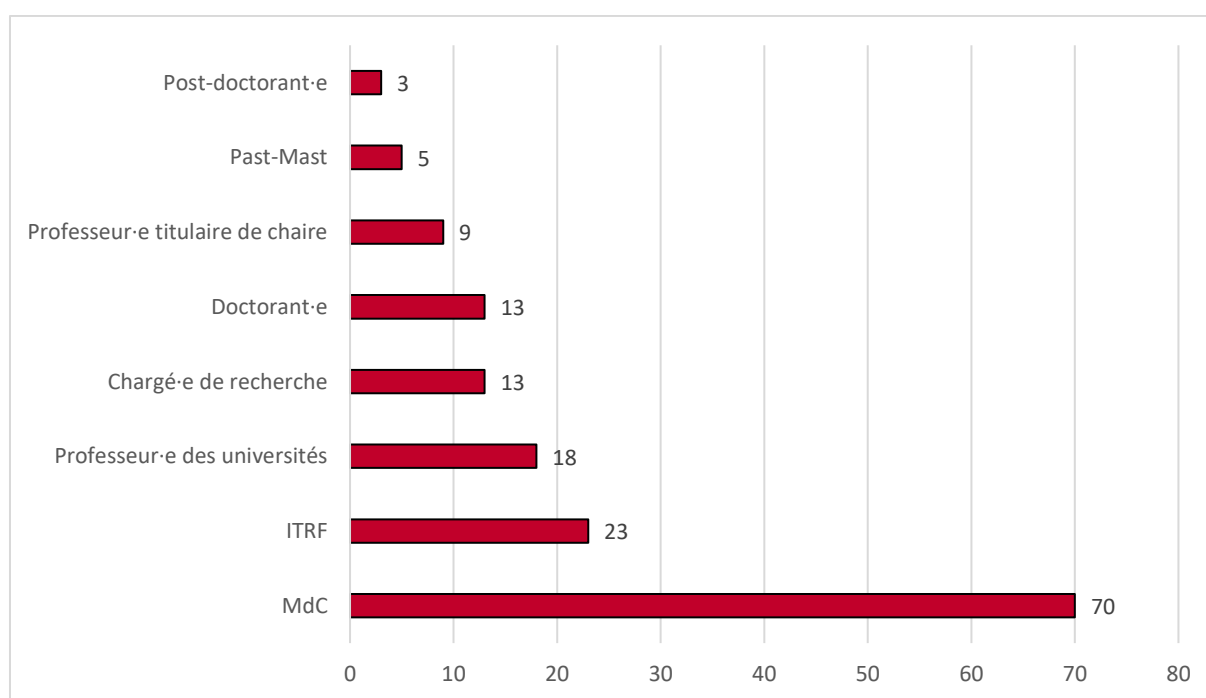


Figure 4 - Nombre de répondant·es par statut

1.4 Le genre des répondant·es

La répartition des genres est relativement équilibrée dans les réponses au questionnaire (*Figure 5 - Genre des répondant·es*). En excluant les 22 % de répondant·es (29 personnes) n'ayant coché aucune réponse, les femmes représentent 49 % des répondant·es (51 personnes) et les hommes 51 % (54 personnes).

Les entretiens sont plus déséquilibrés dans la répartition des genres. Ils ont permis de rencontrer quatre femmes et sept hommes.

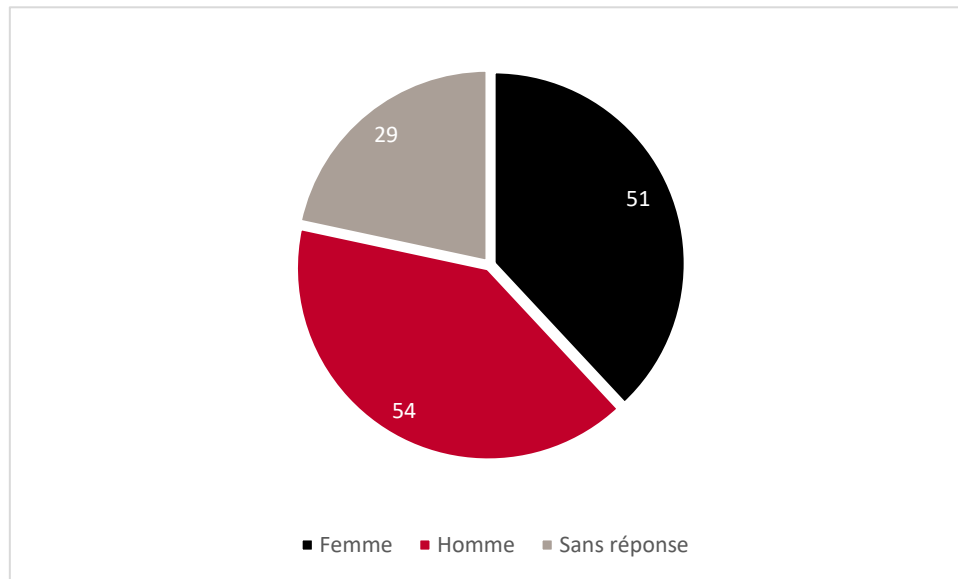


Figure 5 - Genre des répondant·es

1.5 L'âge des répondant·es

Pour les répondant·es au questionnaire, les âges sont répartis comme suit (*Figure 6 - Âge des répondant·es*) : les moins de 30 ans représentent 5 % des réponses (sept personnes), les 30-39 ans 25 % (34 personnes), les 40-49 ans 30 % (40 personnes), de même pour les 50-59 ans, et enfin les plus de 60 ans représentent 10 % des réponses (treize personnes). La tranche d'âge médiane est de 40-49 ans.

Nous n'avons pas recueilli de données de ce type durant les entretiens.

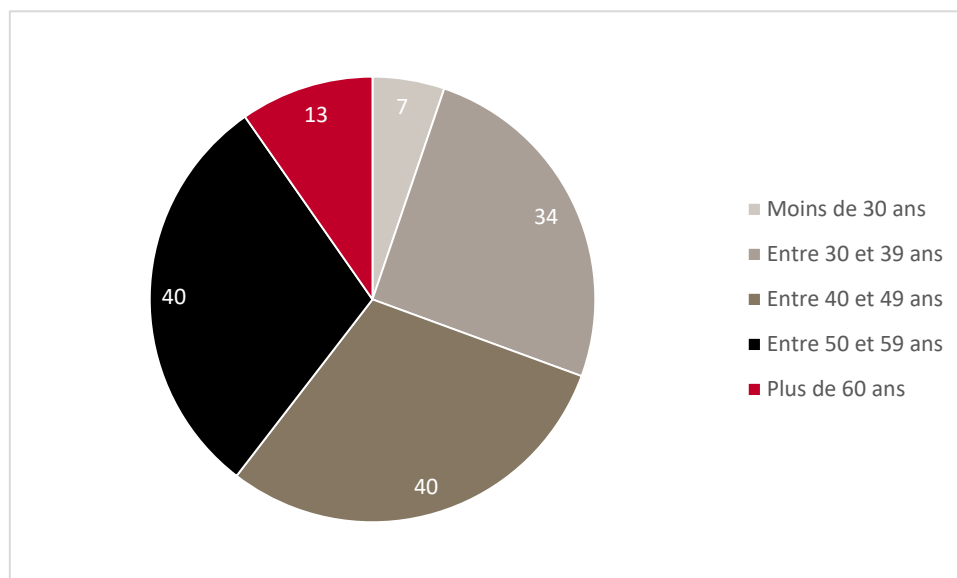


Figure 6 - Âge des répondant·es

1.6 L'expérience des répondant·es dans le milieu de la recherche

L'essentiel des répondant·es bénéficie d'une expérience de plus de 20 ans dans le milieu de la recherche (*Figure 7 - Expérience des répondant·es dans le milieu de la recherche*) : ils-elles représentent 37 % des réponses (49 personnes). Les autres catégories sont mieux réparties entre elles : 14 % des répondant·es (19 personnes) ont une activité de recherche depuis 16 à 20 ans, 21 % des répondant·es (28 personnes) depuis 11 à 15 ans, 15 % (vingt personnes) depuis 5 à 10 ans et 13 % (18 personnes) depuis moins de 5 ans. La médiane se situe dans la tranche de onze à quinze ans d'expérience.

Nous n'avons pas systématiquement recueilli de données de ce type durant les entretiens, mais les deux extrêmes ont pu être rencontrés lors de nos discussions : nous avons ainsi rencontré au moins deux personnes ayant une expérience de moins de cinq ans dans le milieu de la recherche, et au moins deux personnes ayant une expérience de plus de vingt ans.

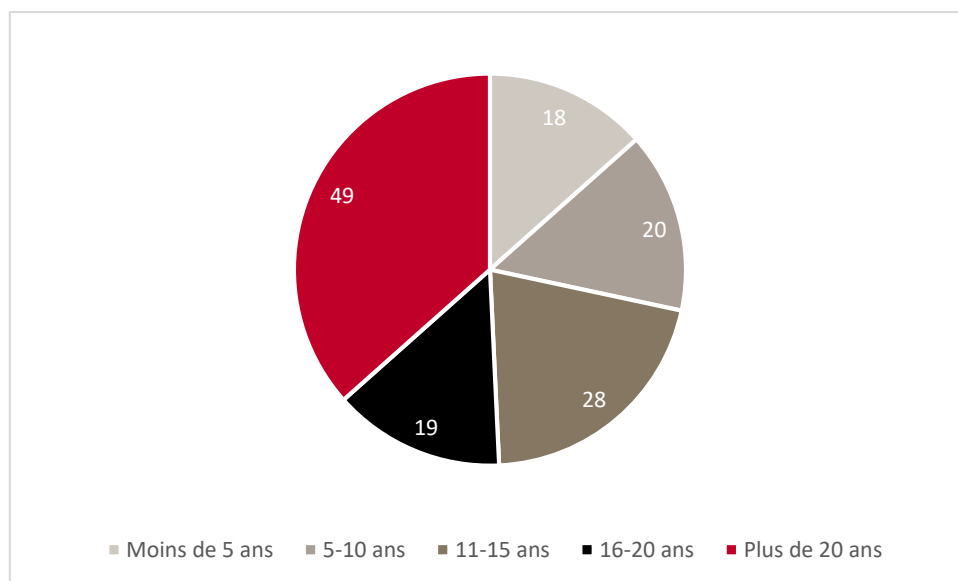


Figure 7 - Expérience des répondant-es dans le milieu de la recherche

2. Nature des données produites

Résumé : concernant la nature des données produites, on observe une grande variété des types, formats et volumes de données, à toutes les échelles (au sein de l'établissement, d'un champ disciplinaire, d'un même laboratoire). La problématique spécifique posée par les données personnelles concerne 39 % des répondant·es à l'enquête et ce type de données est manipulé (dans des mesures variables) dans tous les laboratoires.

2.1 Les types de données produits

Cette question à choix multiple avait pour but de connaître les différents types de données produits par les répondant·es. Le choix a été fait de proposer une typologie de données assez large, afin d'éviter le risque de non-exhaustivité. Le code informatique, même s'il n'est pas à proprement parler une donnée, a été inclus en raison de sa proximité avec les données et de sa mise en avant dans les plans nationaux pour la science ouverte.

Les types de données ont été définis comme suit¹² :

- Données d'observation (données non reproductibles, par exemple : relevés de température ou de précipitations, données de télédétection, photos d'un événement, données d'enquêtes, etc.) ;
- Données expérimentales (données censées être reproductibles, par exemple : générées par un équipement en laboratoire, mesures de performance d'une machine, etc.) ;
- Données de simulation numérique (données reproductibles produites par des logiciels, par exemple : modèles climatiques ou économiques, etc.) ;
- Données dérivées ou compilées (données brutes qui ont fait l'objet de traitements et d'analyses successifs, par exemple : données obtenues par de la fouille de textes ou de données d'une base, etc.).

2.1.1 Principaux types de données produits

Les quatre types de données retenus pour cette question et le code informatique sont représentés, à plus de 20 % chacun (*Figure 8 - Types de données produits*). Les données dérivées ou compilées (49 %) et données d'observation (47 %) arrivent en tête des réponses : 70 % des répondant·es indiquent produire au moins un de ces deux types de données.

¹² En reprenant la typologie proposée sur l'espace chercheurs de l'École nationale des ponts et chaussées (Bordignon).

Les entretiens et le champ « Autre » permettent de détailler certains types de données produits et manipulés par les répondant·es : « lever topographique », « lever géophysique », « données géoréférencées », « coordonnées de points », « nuages de points », « mesures physiques », « données satellitaires », « questionnaire », « entretiens », « textes officiels », « données personnelles », « données bathymétriques », « données météo », « données issues du gouvernement », « données d'autres chercheurs », « données publiques en quanti », « travail qualitatif avec des entretiens », « enregistrements », « vidéo », « gravimétrie quantique », « données comportementales », « corpus de textes », « données produites par un code de calcul », « modèle numérique CAO », « statistiques ». Leur variété témoigne de la diversité des usages et des données manipulées selon le champ disciplinaire mais aussi au sein de ces champs disciplinaires.

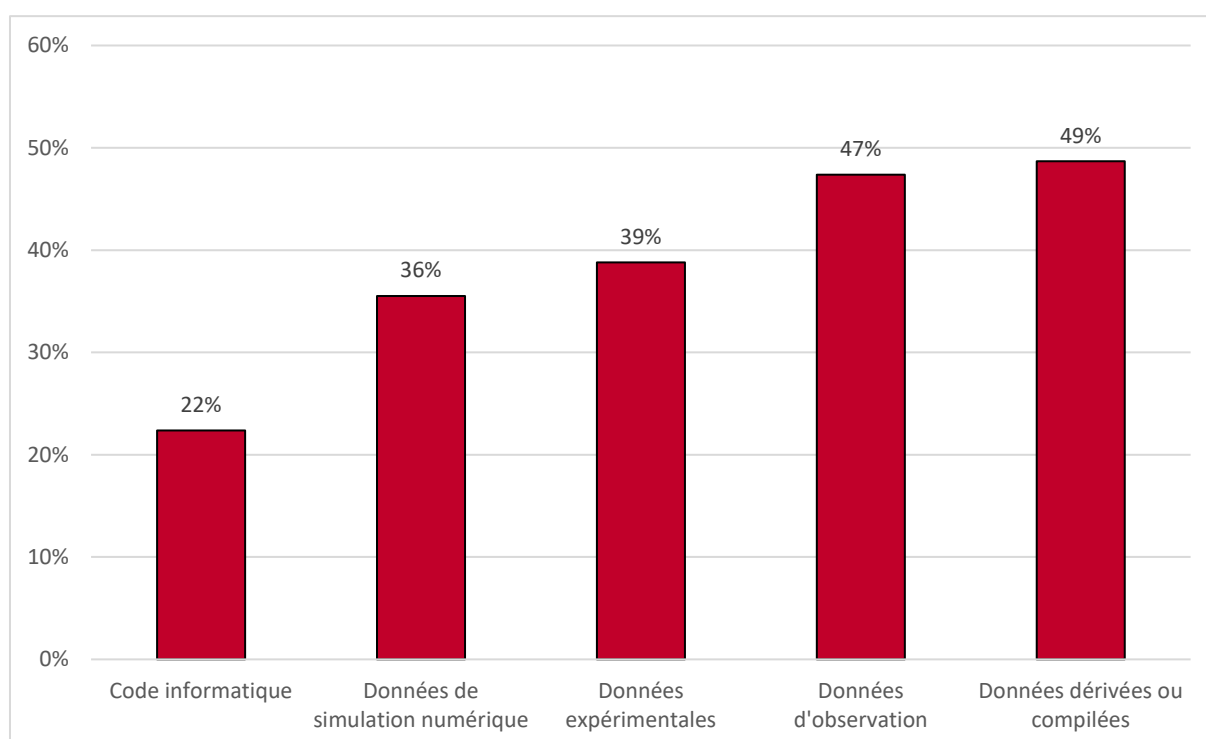


Figure 8 - Types de données produits

2.1.2 Types de données produits par champ disciplinaire

Pour chaque champ disciplinaire représenté tous les types de données ont été cités au moins une fois, mais des tendances se dessinent (Figure 9 - Répartition des types de données produits au sein des champs disciplinaires). Certaines disciplines se spécialisent fortement dans un type de données (CM dans les données expérimentales) ou deux (LSHS et DEPG dans les données d'observation et les données dérivées ou compilées) ; d'autres ont une activité plus polyvalente (SI).

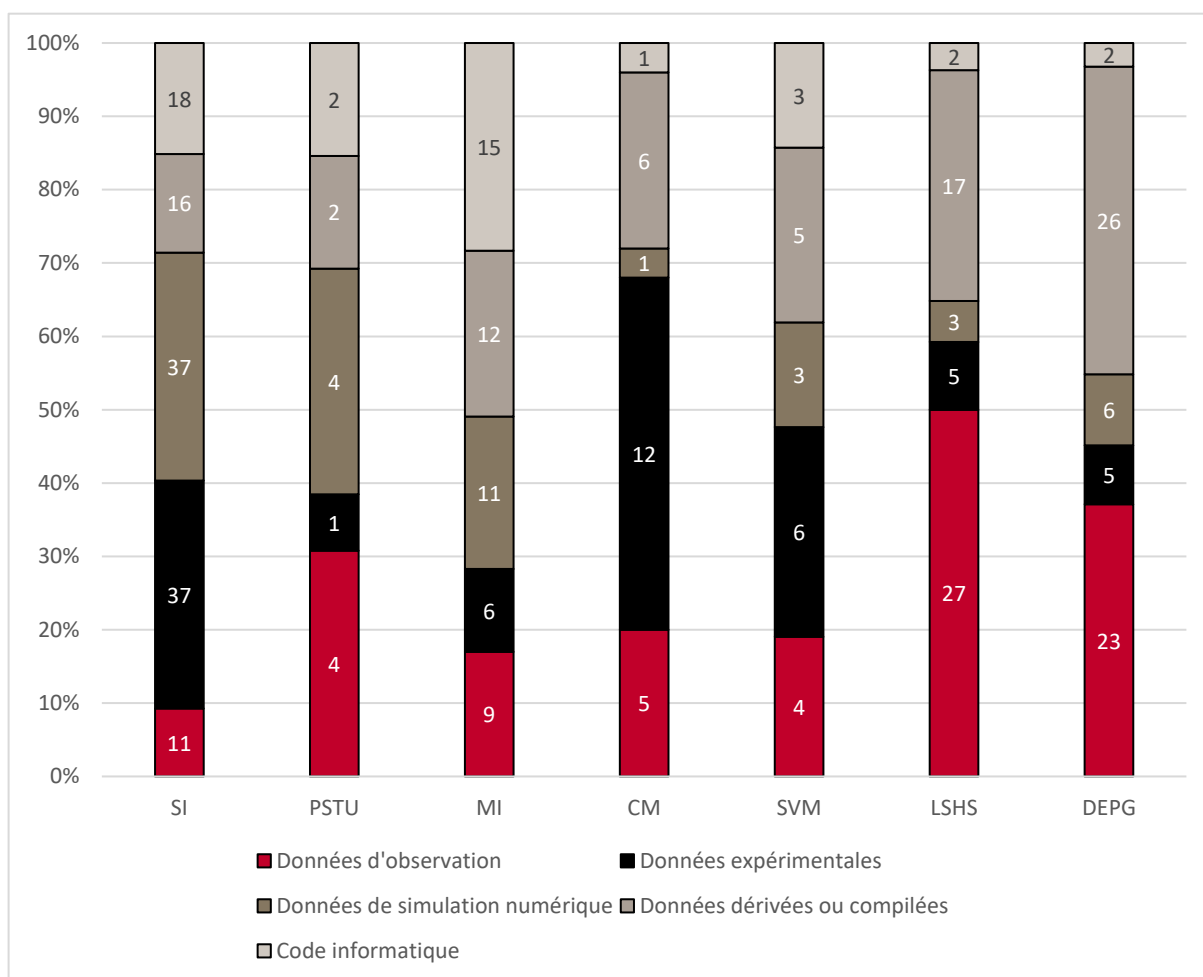


Figure 9 - Répartition des types de données produits au sein des champs disciplinaires

2.1.3 Absence de réponse à la question

Le champ « Autre » a permis de nettoyer ces données et de rattacher les données annoncées dans celui-ci à un des champs proposés, quand cela était possible.

Quatre des sept réponses indiquées dans le champ « autre » – et où la personne n'avait indiqué aucun type de données – désignaient sans équivoque des données d'observation dans le champ des sciences humaines et sociales (SHS). Par ailleurs, une réponse laissée dans le champ « Autre » indique : « Ces catégories ne correspondent à rien d'équivalent en SHS donc j'ignore si mes réponses sont pertinentes ». Cela dit quelque chose de la représentation des données de la recherche en SHS, et du possible sentiment de ne pas être concerné par la question et/ou d'une spécificité qui demanderait à aborder la question sous un angle nécessairement différent de celui des disciplines issues des sciences, techniques et médecine (STM)¹³.

¹³ Ces spécificités disciplinaires expliquent d'ailleurs la multiplication des travaux sur l'ouverture des données en SHS. En 2019, la revue *Tracés* publiait ainsi un hors-série sur le sujet (Galonier, 2019) ; la même année, le Comité pour la science ouverte (CoSO) publiait un guide à destination des chercheur-euses de SHS concernant les données à caractère personnel.

2.2 Le format des données produites

La question portant sur le format des données produites (*Figure 10 - Répartition des types de données au sein de la production des chercheur-euses du Cnam*) permettait de préciser la question précédente. La typologie proposée, reprise de précédentes enquêtes axées SHS (Serres *et al.*, 2017) n'était cependant pas assez complète pour notre cas et demanderait l'ajout de différentes catégories. Certaines ont été proposées dans les réponses dans le champ « Autre » : « logiciel LaTeX », « nuages de mots Ircanteq », « fichiers propres aux logiciels utilisés (fichiers universels pour l'expérimental, fichiers d'export de logiciels de simulation, LaTeX pour les rapports) », « shapefile », « textes structurés (xml et al.) », « fichiers QGIS », « dumps de réseaux de neurones », « format 3D (.stl .xyz, etc.) », « R », « fichiers spécifiques pour logiciels de simulation électrique », « Shiny », « Text (source) et binaire », « fichiers binaires .bin », « Jupyter », « fichiers vtk », « raster », « fichiers SAS, stata », « format libre », « formats compressés et dédiés ».

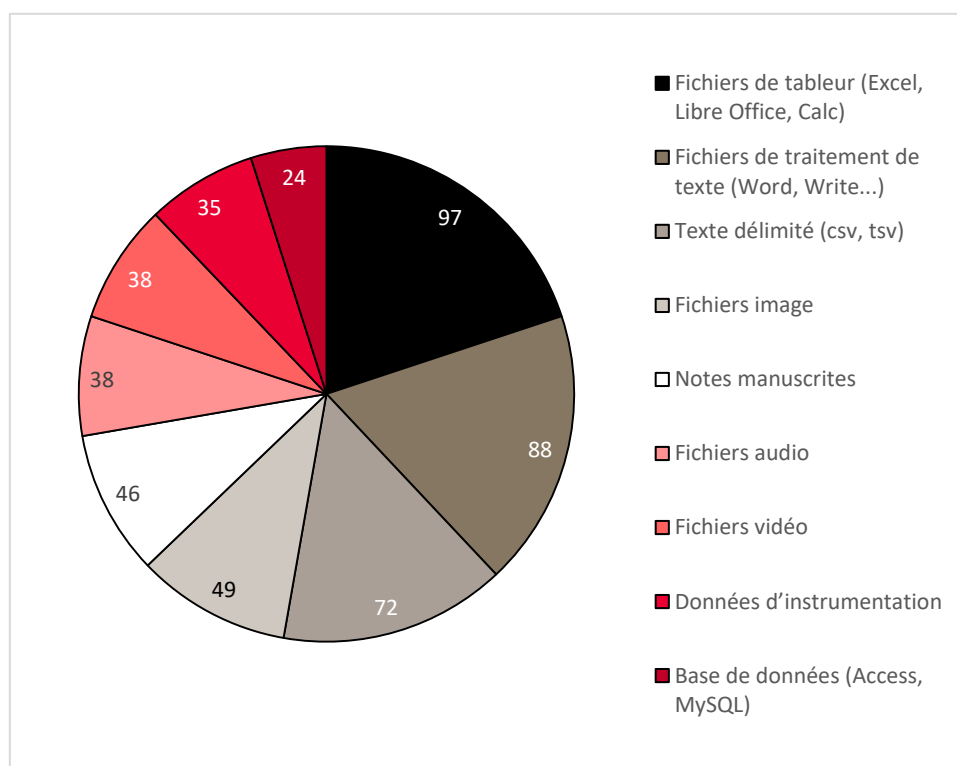


Figure 10 - Répartition des types de données au sein de la production des chercheur-euses du Cnam

2.3 Le volume des données produites

Sur la question des volumes de données, les réponses ont été très variables et un croisement par laboratoire est assez parlant.

On constate qu'au sein d'un même laboratoire, les situations peuvent être extrêmement contrastées. Dans trois unités de recherche (Cédric, CEET et Lirsa), les répondant-es ont à eux tous sélectionné les cinq réponses possibles. Ainsi, sur les 18 répondant-es du Cédric, trois

indiquent produire un volume de données inférieur à 1 Go, trois autres un volume compris en 1 Go et 20 Go, quatre autres entre 20 Go et 100 Go, puis encore trois entre 100 Go et 1 To, et enfin trois autres disent produire un volume de données supérieur à 1 To.

D'ailleurs, si on se penche sur les laboratoires dont le nombre de réponses a été le plus significatif, donc qui dépassent dix réponses à cette question (c'est-à-dire le Cédric, le Lirsa et le LMSSC), au moins quatre volumes sont représentés à chaque fois (*Figure 11 - Répartition des volumes de données au sein de trois laboratoires*).

Plusieurs entretiens témoignent de leur travail sur de « *gros volumes de données* », nécessitant des infrastructures et un budget spécifiques.

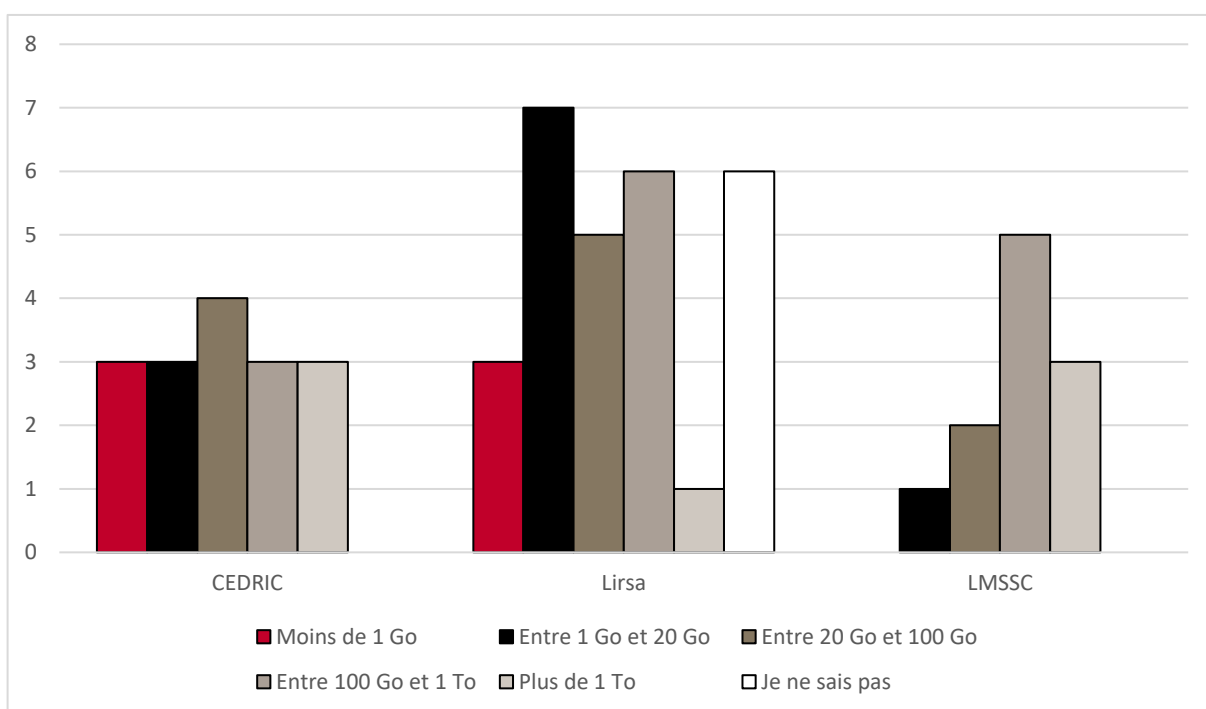


Figure 11 - Répartition des volumes de données au sein de trois laboratoires

2.4 Les données personnelles

Toute information se rapportant à une personne physique identifiée ou identifiable est une donnée à caractère personnel. La publication de données personnelles est soumise à des conditions juridiques spécifiques.

39 % des répondant·es (56 personnes) déclarent gérer des données personnelles, 51 % (73 personnes) indiquent ne pas en gérer, et 9 % (13 personnes) ne savent pas si elles en gèrent ou non (*Figure 12 - Répartition des réponses à la question « Gérez-vous des données personnelles ? »*).

Sans surprise, c'est au sein des « Lettres, sciences humaines et sociales » et du « Droit, économie, gestion » que la collecte de données personnelles est la plus fréquente

(respectivement 45 % et 36 % des réponses). Il faut noter cependant que la collecte de ce type de données est présente dans tous les champs disciplinaires (*Figure 13 - Pratiques de gestion de données personnelles selon le champ disciplinaire*).

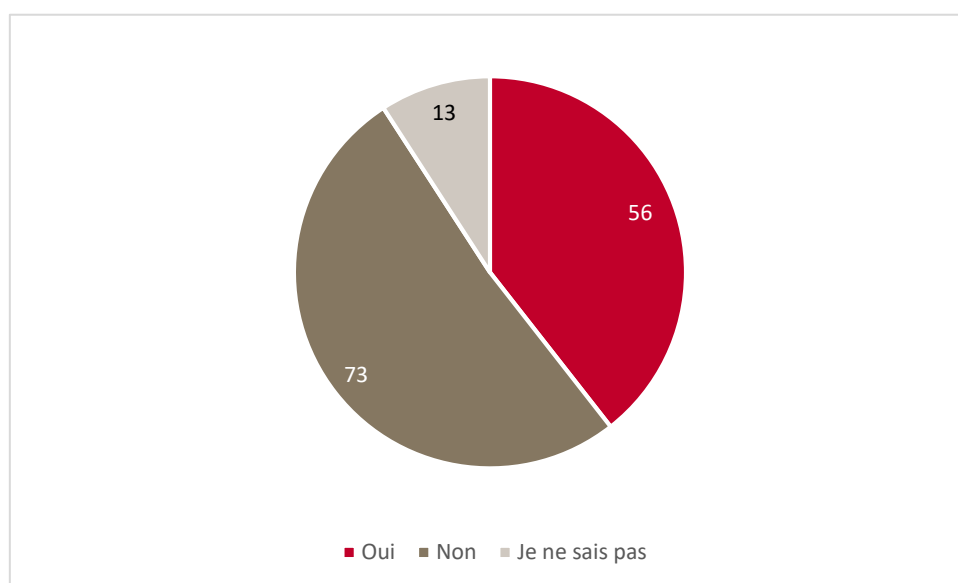


Figure 12 - Répartition des réponses à la question « Gérez-vous des données personnelles ? »

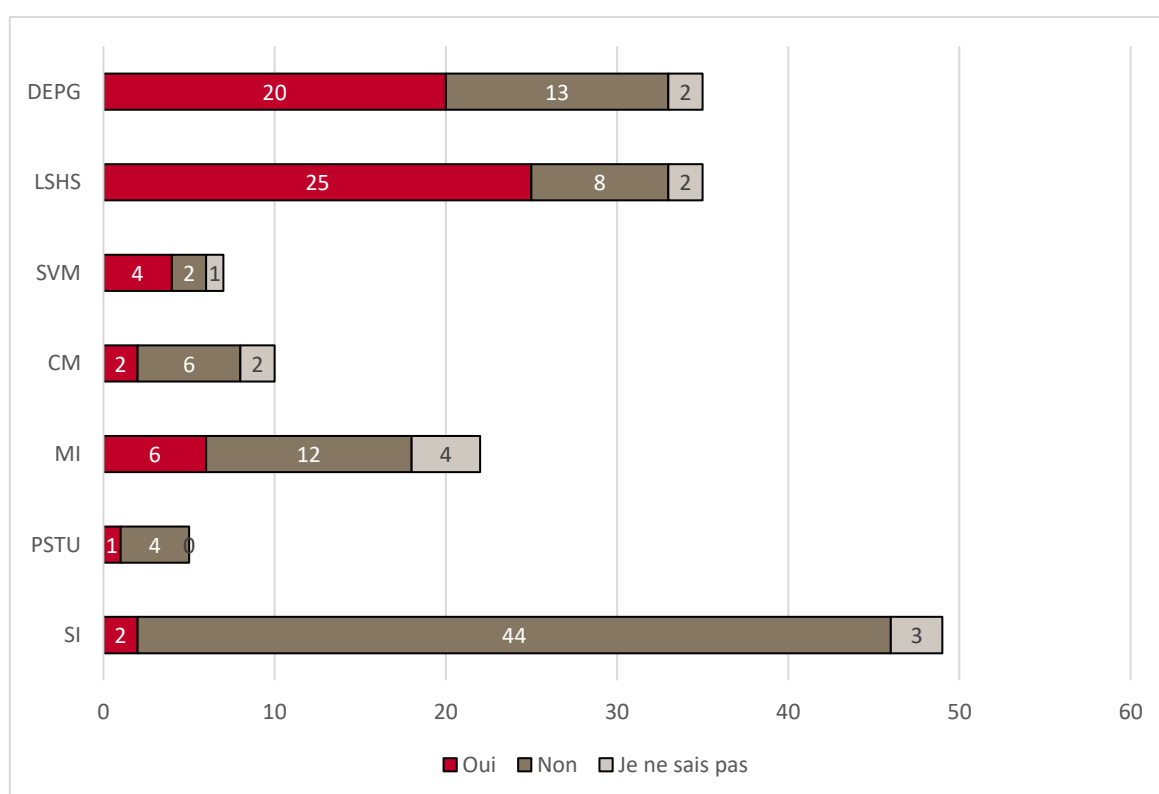


Figure 13 - Pratiques de gestion de données personnelles selon le champ disciplinaire

3. Pratiques de gestion et d'ouverture des données

Résumé : 83 % des répondant·es cumulent plusieurs lieux de stockage. Si la moitié des répondant·es indiquent que leur groupe de travail a accès à leurs données, ils·elles sont aussi 46 % (66 personnes) à indiquer que personne n'a accès à celles-ci. Les répondant·es étaient peu nombreux·ses à déclarer partager des données en libre accès, à avoir déjà participé à la rédaction d'un plan de gestion de données ou à avoir publié un *data paper*.

3.1 Lieu de stockage des données

Nous avons choisi d'inclure les dossiers papier dans la liste des lieux de stockage pour tenter de se rapprocher de l'exhaustivité. Le champ « Autre » permet aussi d'identifier des lieux de stockage que nous n'avions pas pris en compte : « ordinateurs pilotant les instruments scientifiques », « pc des appareils ».

Le stockage en local est le plus représenté (*Figure 15 - Lieux de stockage des données des répondant·es*). 91 des 144 répondant·es (63 %) stockent leurs données « en local, sur un disque dur externe ou un autre support (clé USB, CD, DVD, etc.) ».

83 % des répondant·es cumulent plusieurs lieux de stockage (*Figure 14 - Cumul de lieux de stockages par les répondant·es*). 17 % en ont un, 24 % en ont deux, 28 % en ont trois, 23 % en ont quatre, 5 % en ont cinq et 4 % en ont six.

Nombre de lieux de stockage sélectionnés	Nombre de réponses	Pourcentage des réponses
1	24	17%
2	33	24%
3	39	28%
4	32	23%
5	7	5%
6	6	4%

Figure 14 - Cumul de lieux de stockages par les répondant·es

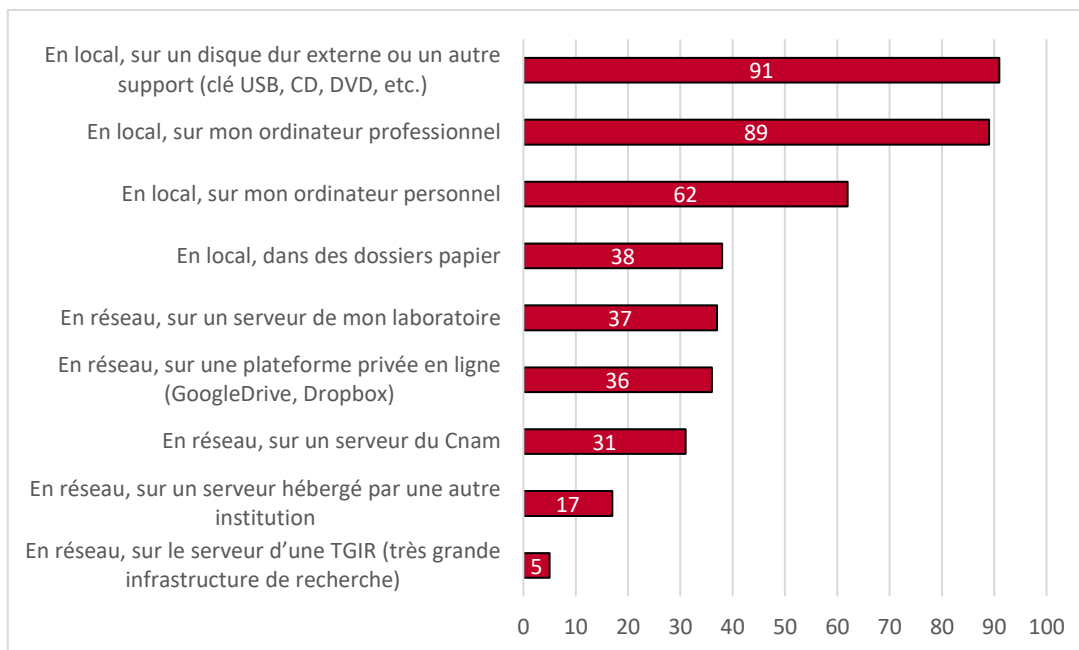


Figure 15 - Lieux de stockage des données des répondant-es

3.2 Accès aux données

La question suivante concernait les pratiques de partage de données des répondant-es. Si la moitié des répondant-es indiquent que leur groupe de travail a accès à leurs données, ils-elles sont aussi 46 % (66 personnes) à indiquer que personne n'a accès à celles-ci. 23 % (27 personnes) indiquent les partager avec leur laboratoire. Le croisement par laboratoire ne permet que difficilement de distinguer une tendance. Seuls 4 % (six personnes) indiquent les partager avec tout le monde.

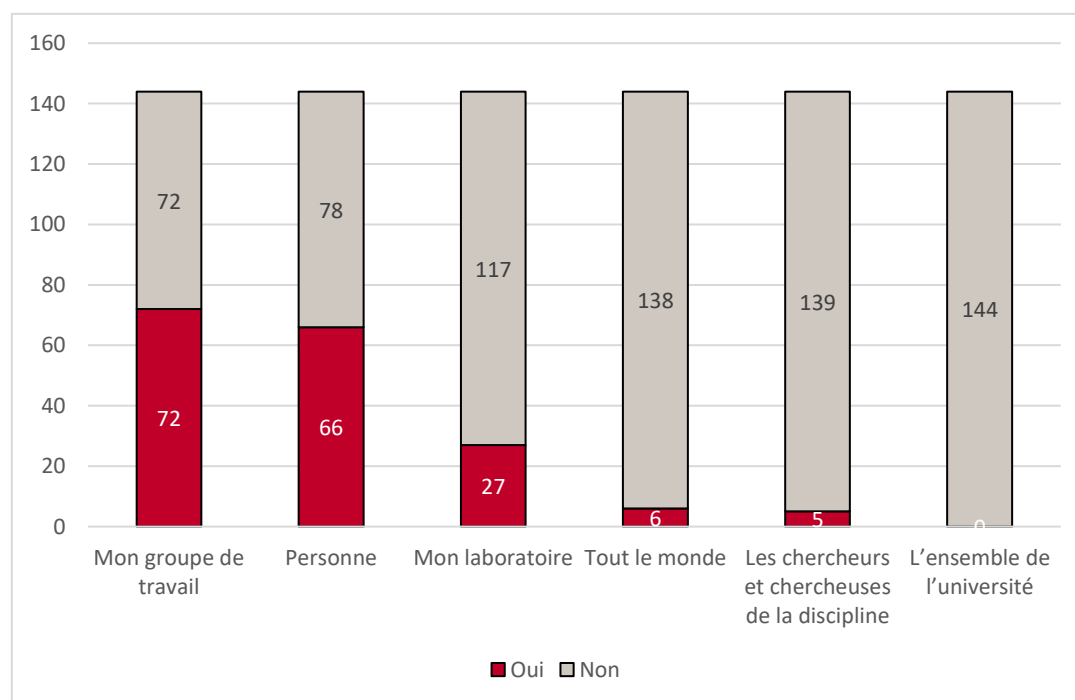


Figure 16 - Personnes ayant accès ou non aux données des répondant-es

3.3 Partage des données en libre accès

Les répondant-es étaient peu nombreux-ses à déclarer partager des données en libre accès. 85 % (111 personnes) ne l'ont jamais fait, contre 15 % (19 personnes) à l'avoir fait au moins une fois (*Figure 18 - Pourcentage de l'effectif ayant déjà partagé ses données en libre accès selon l'âge*).

Par ailleurs, on constate que parmi les personnes ayant déjà mis leurs données en libre accès, 84 % (seize personnes) l'ont fait plus d'une fois.

La variable de l'âge (*Figure 17- Réponses à la question « Avez-vous déjà partagé des données en libre accès sur le web ? »*) n'est pas très significative. Si l'on observe une courbe descendante entre les tranches d'âge « 30-39 ans » (24 % des 30-39 ans ont déjà partagé leurs données en libre accès) et « 40-49 ans » (5 %), la courbe remonte par la suite (13 % pour les « 50-59 ans », 15 % pour les « plus de 60 ans »). L'âge ne semble donc pas être un critère déterminant quant au partage des données. On peut toutefois souligner que le « pic de partage » se situe dans la tranche d'âge « 30-39 ans ».

Le constat est le même concernant la durée de l'expérience dans le milieu de la recherche (*Figure 19 - Pourcentage de l'effectif ayant déjà partagé ses données en libre accès selon la durée de l'expérience dans le milieu de la recherche*) : la courbe, ascendante jusqu'à la tranche « 11 à 15 ans » (25 % des personnes ayant une expérience de 11 à 15 ans dans le milieu de la recherche ont déjà partagé leurs données en libre accès, connaît une brutale chute pour la tranche « 16 à 20 ans » (5 %) avant de remonter légèrement pour la tranche « plus de 20 ans » (12 %).

La progression relativement similaire des deux courbes (âge et expérience) était attendue.

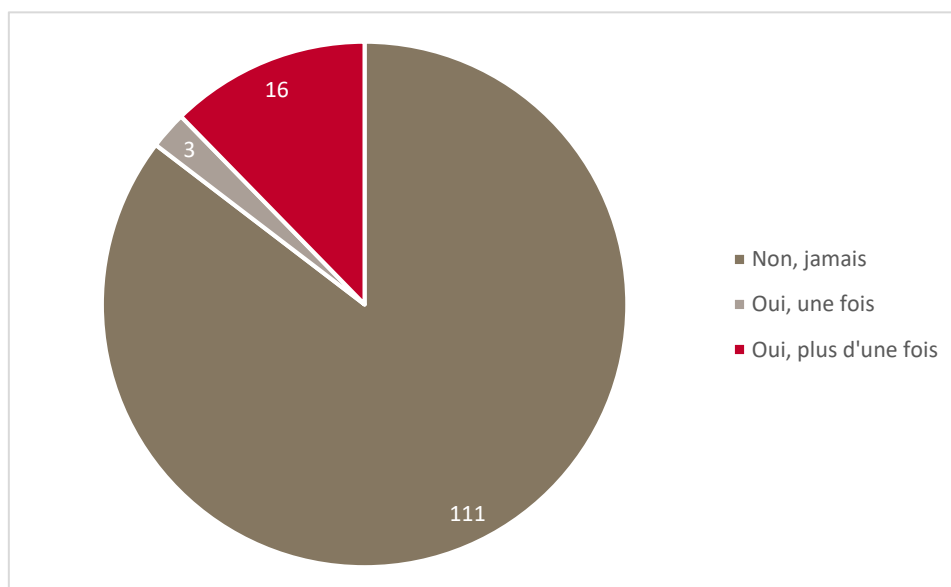


Figure 17- Réponses à la question « Avez-vous déjà partagé des données en libre accès sur le web ? »

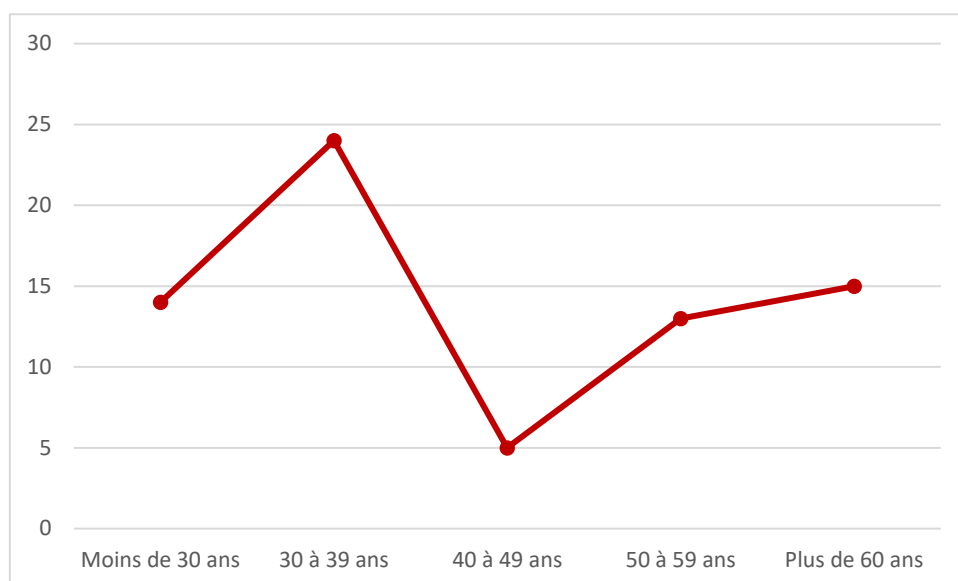


Figure 18 - Pourcentage de l'effectif ayant déjà partagé ses données en libre accès selon l'âge

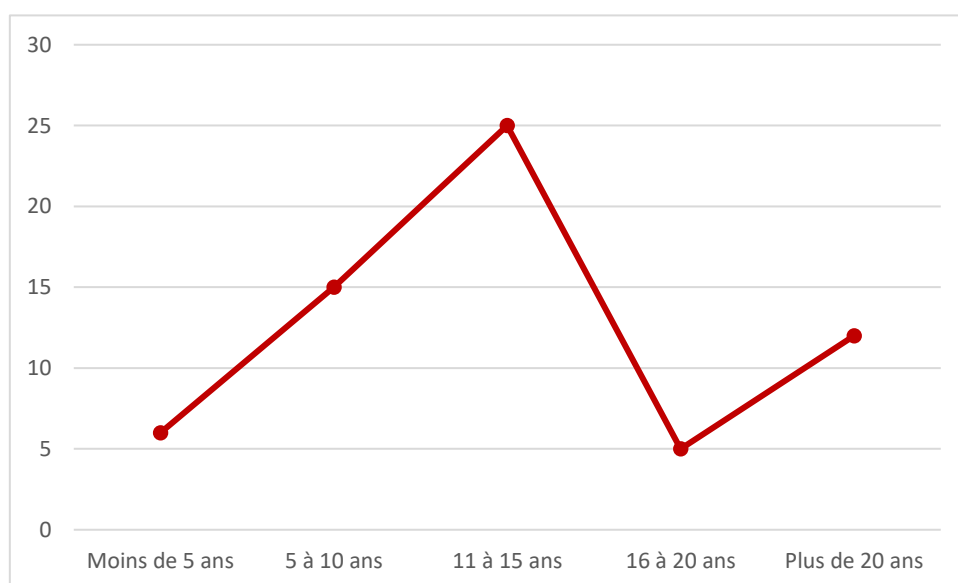


Figure 19 - Pourcentage de l'effectif ayant déjà partagé ses données en libre accès selon la durée de l'expérience dans le milieu de la recherche

3.4 Expérience du PGD

La question portant sur les plans de gestion de données était précédée d'une définition (voir note 5) : en effet, les entretiens exploratoires nous avaient déjà montré que le niveau de connaissance sur le sujet pouvait être très variable d'un·e chercheur·euse à l'autre.

Cette question était une question à choix multiple. Cependant, on constate que l'ensemble des répondant·es n'ont choisi qu'une seule réponse parmi les quatre possibles (*Figure 20 - Réponses à la question « Avez-vous déjà participé à la rédaction d'un plan de gestion de données ? »*). Ainsi, les neuf répondant·es (6 % des réponses) qui affirment avoir déjà rédigé un PGD sans qu'il soit exigé par le financeur indiquent aussi n'avoir jamais participé à la rédaction de ce document dans le cadre de projets où le financeur l'exigeait. De même, les 18 répondant·es (13 % des réponses) ayant été impliqué·es dans ce travail en raison des exigences de financeurs n'ont pas effectué de travail similaire sur d'autres projets.

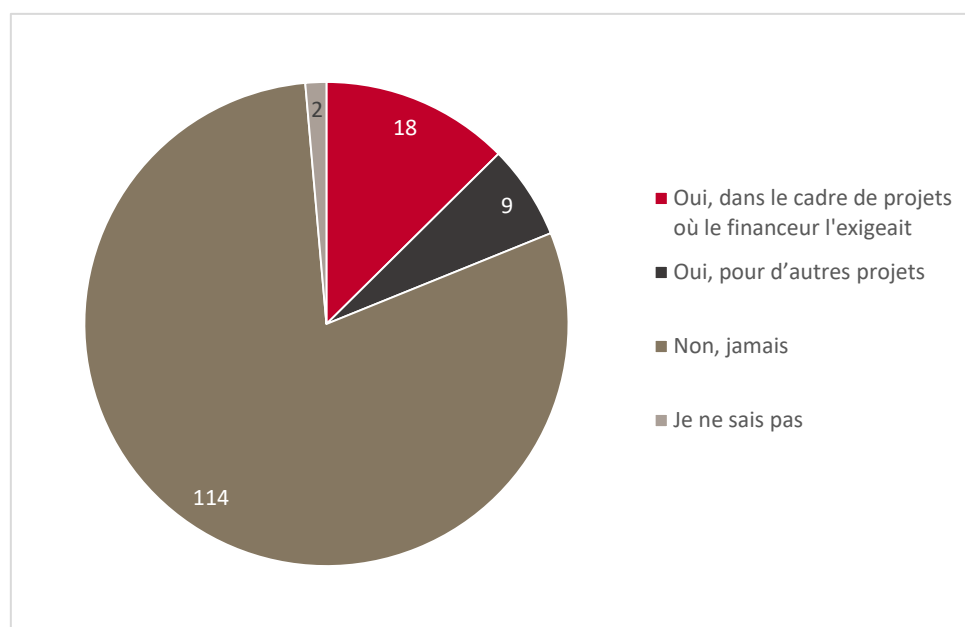


Figure 20 - Réponses à la question « Avez-vous déjà participé à la rédaction d'un plan de gestion de données ? »

3.5 Expérience du data paper

Comme celle sur les PGD, la question sur les data papers débutait par une définition :

« Un data paper est une publication scientifique, soumise à la validation des pairs, dont le but principal est de décrire un ou plusieurs jeux de données, afin de permettre leur réutilisation. » (Dataacc)

Peu en ont publié (19 %) et les réponses « Oui » sont principalement le fait des sciences de l'ingénieur (56 % des réponses positives) et des mathématiques-informatique (28 %). Plusieurs entretiens, notamment au sein de disciplines axées sur le recueil de données de

mesures géomatiques ou bathymétriques, ont mis en avant l'intérêt porté au sujet de ce type de publication : « j'ai des collègues qui ont commencé à s'intéresser aussi aux data papers », « c'est quelque chose que les gens feraient volontiers. La contrepartie, c'est évidemment que derrière il y ait une valorisation en termes de publication reconnue par les pairs », « Donc, oui, je pense que passer par un "data paper" puis un "paper", ça serait l'idéal, dans ce sens-là », « Décrire les données de toute façon, on le fait dans les articles, qui ne sont pas des data papers, on le fait dans la méthodologie au préalable ».

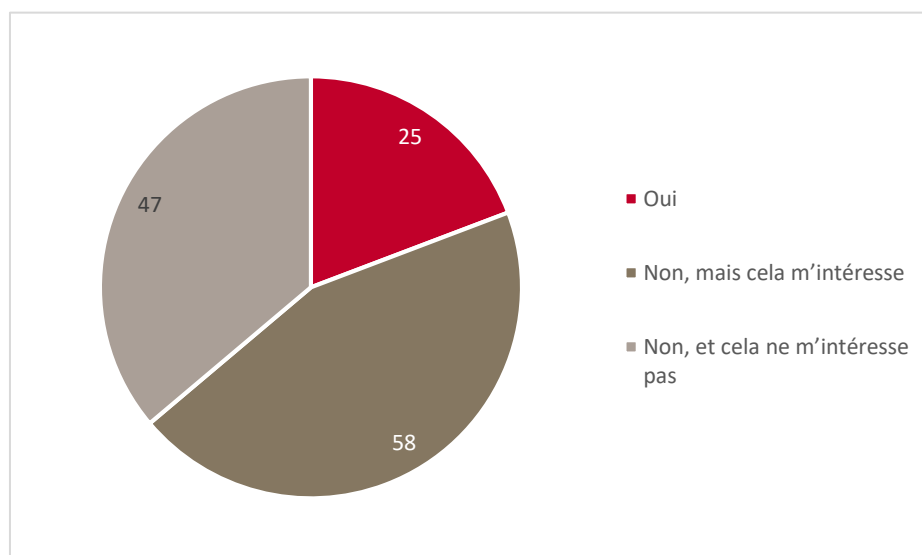


Figure 21 - Réponses à la question « Avez-vous déjà publié un data paper ? »

3.6 Aide à la gestion des données

Peu d'aide à la gestion des données a été rapportée de manière générale. 85 % des répondant·es (121 personnes) indiquent que dans leur laboratoire, chacun gère ses propres données (Figure 22 - Réponses à la question « Bénéficiez-vous d'une aide spécifique sur la question des données ? »)

Dans les entretiens, certaines phrases ressortent : « *on fait ça un peu artisanalement* ». Il a fallu, selon les répondant·es ayant déjà rédigé un PGD ou ouvert des données, consacrer du temps à se renseigner sur les attentes du PGD, sur le fait de trouver un entrepôt adapté, etc.

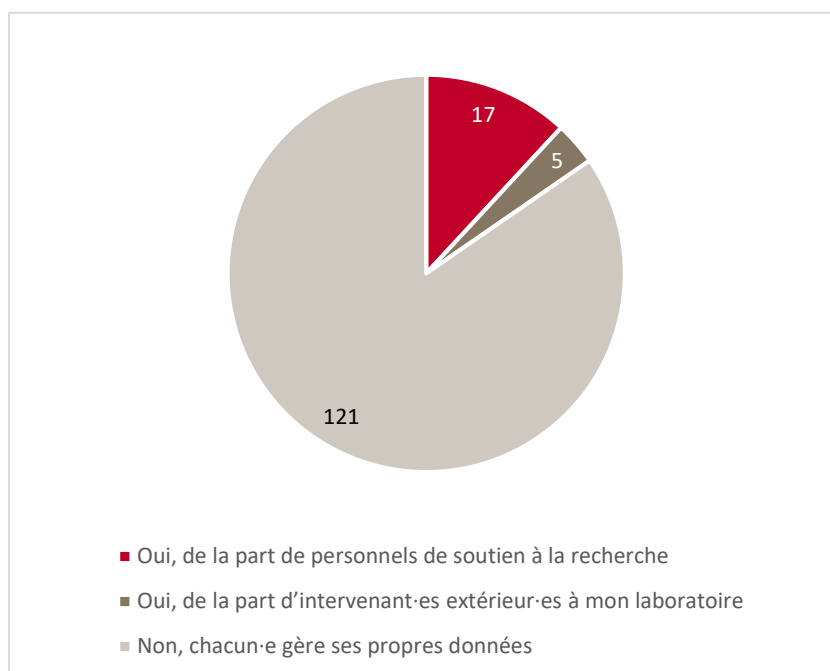


Figure 22 - Réponses à la question « Bénéficiez-vous d'une aide spécifique sur la question des données ? »

3.7 Politique du laboratoire

La question 3.7 « Votre laboratoire a-t-il mis en place des règles spécifiques sur la gestion des données (stockage, archivage, diffusion) ? » a reçu des réponses positives pour huit unités de recherche sur 26 : Dynfluid, ESDR3C, FoAP, GeF, LCM, Lirsa, LMSSC et Sayfood (Figure 23 - Présence de règles spécifiques sur la gestion des données (stockage, archivage, diffusion) au sein des laboratoires).

Pourtant, il faut noter que ces réponses positives ne font jamais l'unanimité : elles sont presque systématiquement mitigées par des réponses négatives (sauf dans le cas de Dynfluid, où la deuxième réponse est « C'est en projet ou en cours »). Cela pourrait s'expliquer par des interprétations multiples de la question, qui englobait stockage, archivage et diffusion en un tout. Si certain-es répondant-es ont pu estimer que la présence de règles de stockage uniquement ne permettait pas de répondre « Oui », d'autres ont pu considérer cela comme un cadre suffisant pour le faire. Une formulation plus explicite de la question aurait pu faire ressortir des réponses plus marquées.

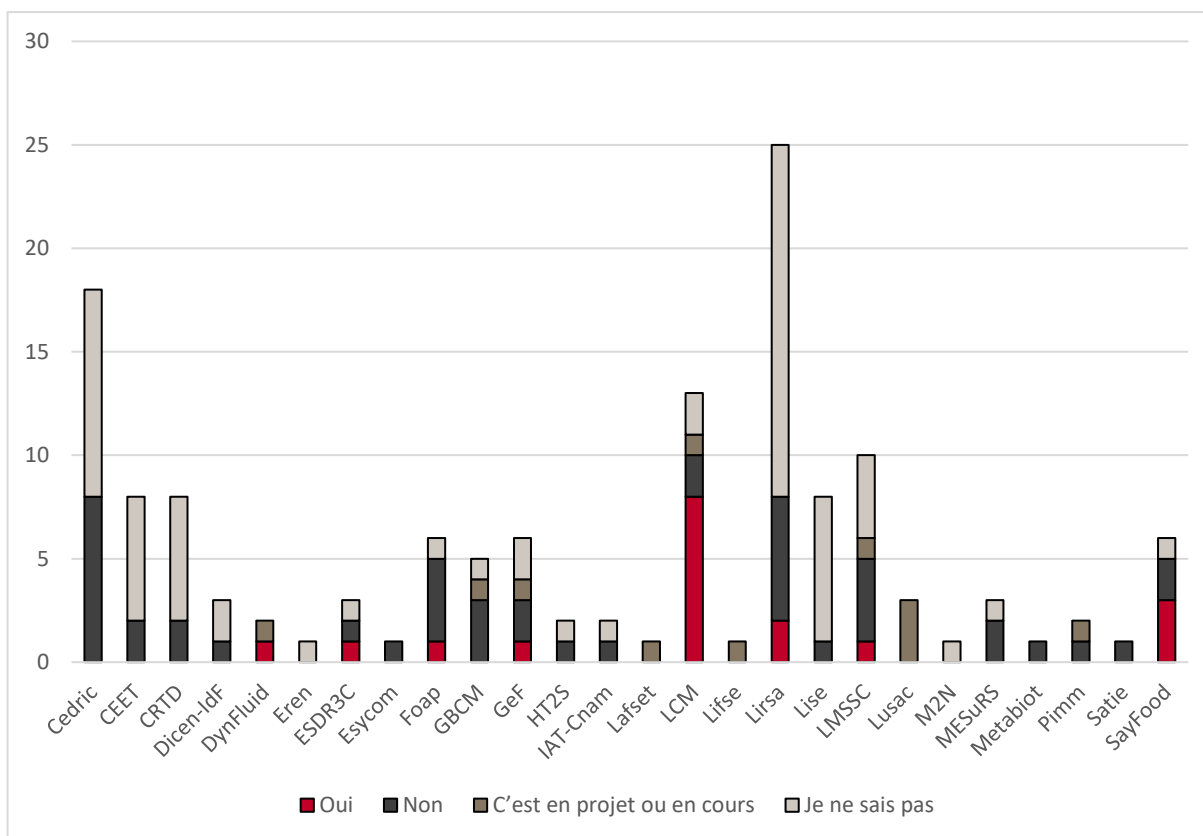


Figure 23 - Présence de règles spécifiques sur la gestion des données (stockage, archivage, diffusion) au sein des laboratoires

4. Motivations et freins à l'ouverture des données

Résumé : les réponses aux questions liées à l'adhésion au principe d'ouverture des données montrent un décalage entre les pratiques (peu de partage des données) et les représentations des répondant·es (généralement favorables à leur diffusion). Les principales raisons d'une mise en libre accès des données sont le sentiment d'aider la communauté scientifique et la visibilité offerte aux travaux ; les principales raisons de ne pas le faire sont les freins juridiques et le manque de temps.

4.1 Adhésion au principe d'ouverture des données

Les réponses aux questions liées à l'adhésion au principe d'ouverture des données montrent un décalage entre les pratiques et les représentations des répondant·es (*Figure 24 - Adhésion au principe d'ouverture des données*). 68 % des répondant·es se déclarent en effet prêt·es à partager leurs données de recherche avec d'autres (total des « tout à fait d'accord » et « plutôt d'accord »). Il faut toutefois prendre en compte un biais, le terme « partage » étant généralement connoté positivement. Le rapport de l'enquête de Rennes 2 le signale : « *il est toujours délicat de se prononcer contre le partage* » (Serres *et al.*, 2017, p. 63).

La question du besoin exprimé de pouvoir réutiliser les données d'autres chercheur·euses (exprimé par 63 % des répondant·es) peut être croisée avec le champ disciplinaire : la demande est particulièrement forte dans les sciences dites « dures » (56 % des répondant·es en SVM, 50 % des PSTU, 43 % des SI, 33 % des CM, 23 % des MI), et moins importantes pour ce qui concerne les sciences humaines (respectivement 18 % et 12 % des LSHS et des DEPG).

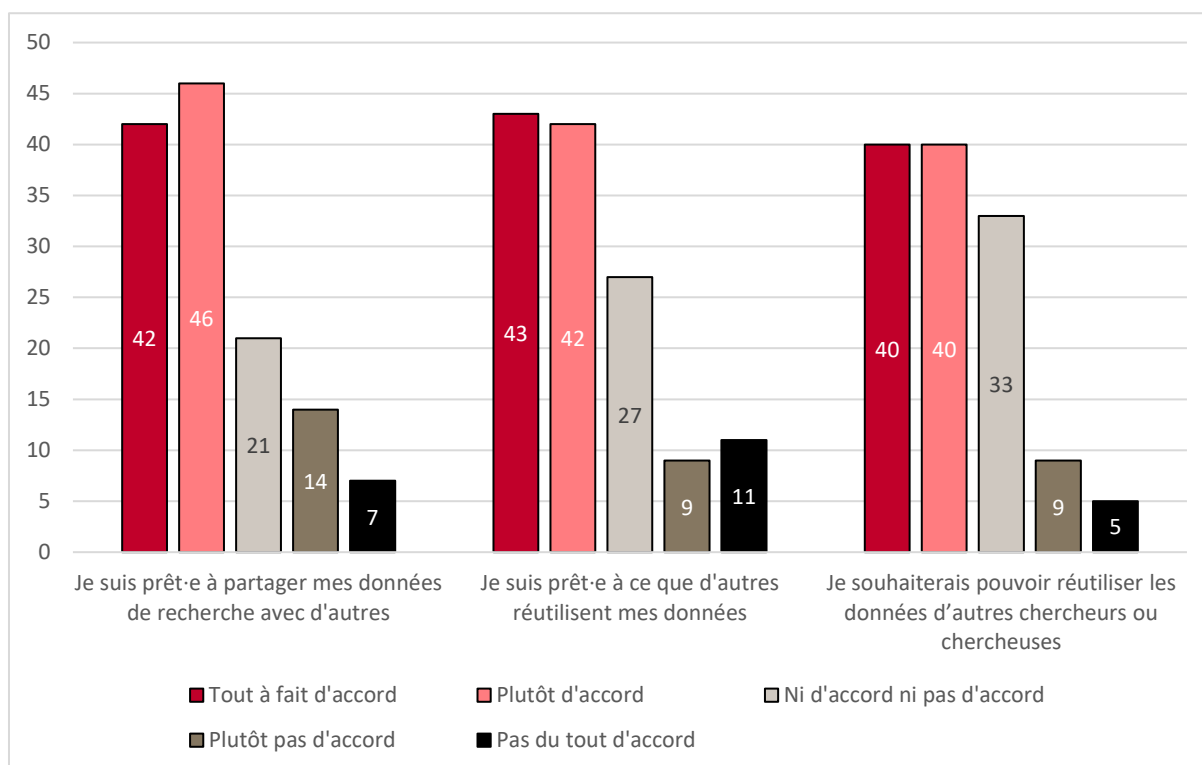


Figure 24 - Adhésion au principe d'ouverture des données

4.2 Raisons d'une mise en libre accès des données

La question 4.2 (« Pour quelle(s) raison(s) choisiriez-vous de mettre vos données de recherche en libre-accès ? ») et la suivante (« Pour quelle(s) raison(s) choisiriez-vous de ne pas mettre vos données de recherche en libre accès ? ») sont directement inspirées des questions 25 et 26 de l'enquête de Rennes 2. Elles ont été reformulées pour ne plus faire apparaître la notion d'« incitation » et de « frein », qui nous paraissaient susceptibles de porter un biais (Figure 25 - Principales raisons d'une ouverture de l'accès aux données).

Les principales raisons qui ressortent sont le sentiment d'aider la communauté scientifique (85 réponses, soit 62 % des répondant-es) et la visibilité offerte aux travaux (78 personnes, 57 % des répondant-es). Viennent ensuite les opportunités de contacts et de coopération avec d'autres (75 personnes, 54 % des répondant-es), le souhait de voir se développer de nouvelles recherches à partir de jeux de données mis à disposition (71 personnes, 51 %), l'adhésion aux valeurs de la science ouverte (68 personnes, 49 %), les obligations des financeurs (50 personnes, 36 %) et les enjeux d'intégrité scientifique avec le souci de faire valider ses résultats via la mise à disposition d'un jeu de données (43 personnes, 31 %).

Le champ « Autre » et les entretiens viennent compléter les réponses proposées.

Dans les entretiens apparaît l'idée que la recherche, étant financée sur fonds publics, devrait conduire à la mise à disposition des données : « *c'est plutôt philosophique, mais aussi, on est payés par des fonds publics* », « *Il faut être clair que dans la mesure où on utilise des moyens publics, [...] les données doivent devenir publiques à terme. Donc le fait qu'un propriétaire, 10 ans après, dise : "c'est ma donnée, je ne veux pas la publier", ça n'est pas acceptable de mon point de vue.* »

On relève aussi la volonté que ces données ne soient pas perdues au moment de quitter un organisme : « *il me paraissait essentiel de faire ça en quittant ce travail, parce que si je ne l'avais pas fait, clairement, ces données auraient été pour moi complètement perdues* », « *ça repose sur la disponibilité des gens qui ont encore la compétence et le savoir pour répondre à la question. C'est aussi une des motivations qui m'a amené à publier toutes ces données sur la base [nom d'une base]* ».

Le champ « Autre » vient aussi nuancer :

« Pour les retravailler avec d'autres, mais je n'envisage pas un accès totalement libre. Les données sont construites dans un contexte que seules les personnes qui ont participé au recueil connaissent. La réutilisation hors contexte n'aurait pas de sens. »

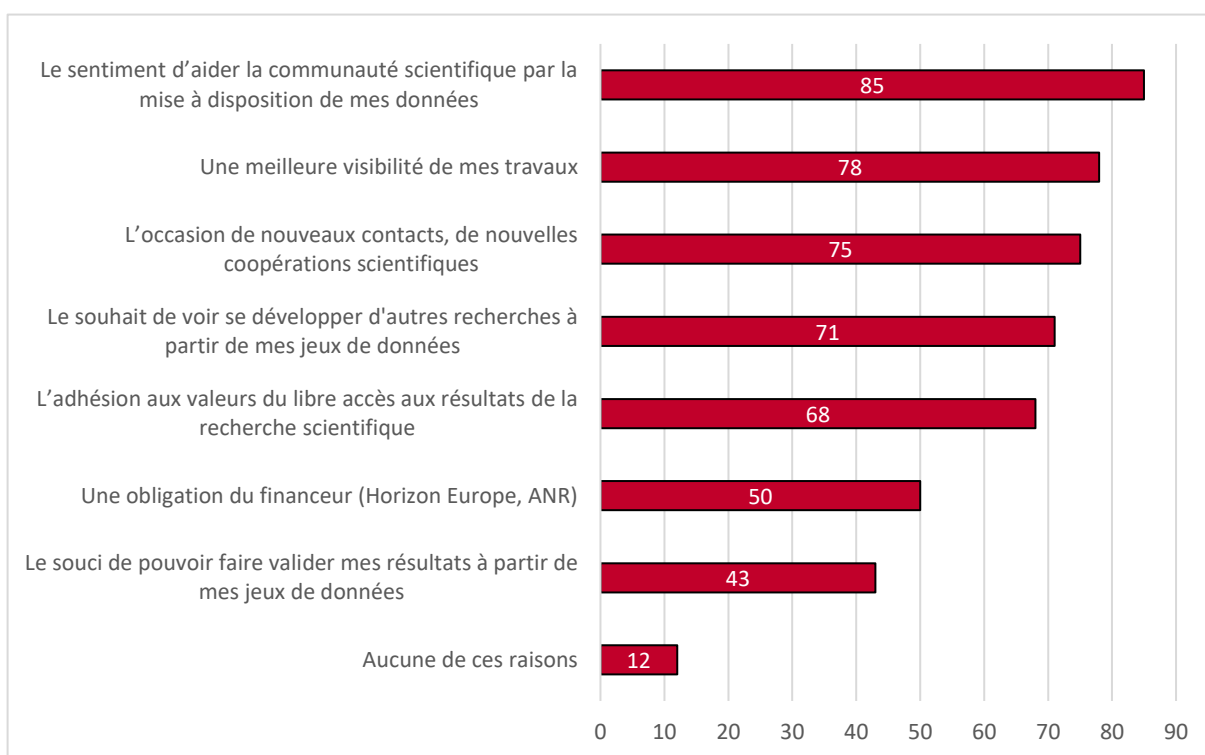


Figure 25 - Principales raisons d'une ouverture de l'accès aux données

4.3 Raisons d'une fermeture de l'accès aux données

La question facultative 4.3 – « Pour quelle(s) raison(s) choisiriez-vous de ne pas mettre vos données en libre-accès ? » – a réuni 131 réponses sur les 138 personnes parvenues jusqu'à ce point du questionnaire, soit un taux de réponse de 95 % (*Figure 26 - Principales raisons d'une fermeture de l'accès aux données*).

La question des freins juridiques occupe la première place (53 % des répondant·es), suivie par le manque de temps (44 %). Le manque de connaissance et la crainte du plagiat ont tous deux été sélectionnés par 36 % des répondant·es. Les raisons économiques ou commerciales arrivent en dernier (15 %). Enfin, 8 % des répondant·es ont jugé qu'au sein de la liste proposée par la question, aucune réponse ne constituait pour eux une raison pour ne pas mettre leurs données en libre-accès.

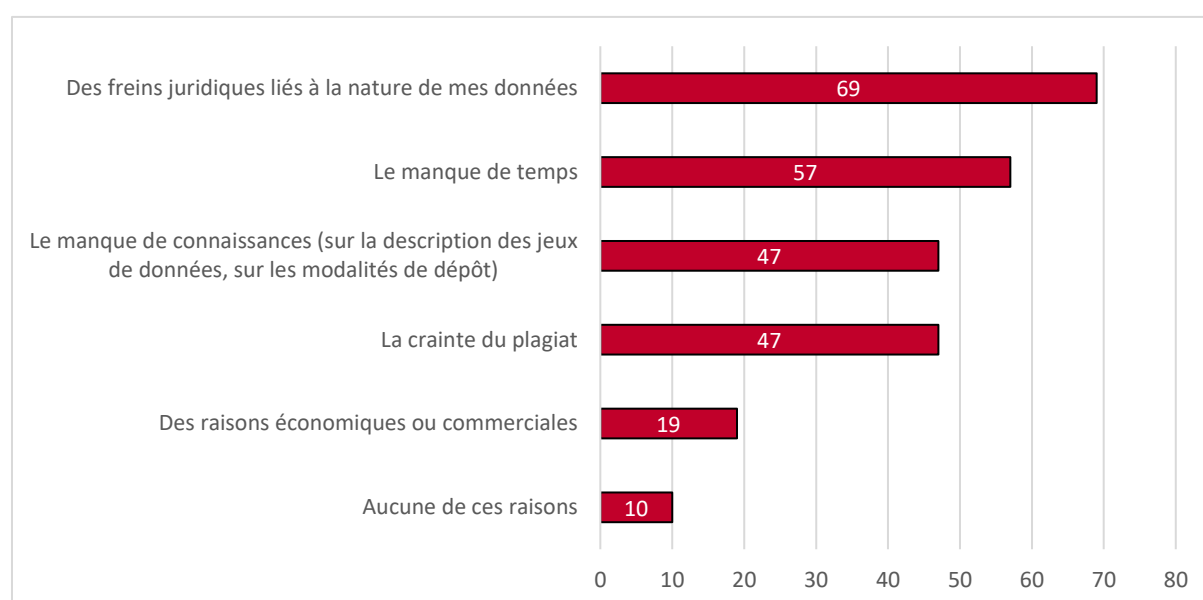


Figure 26 - Principales raisons d'une fermeture de l'accès aux données

Le champ « Autre », dans lequel ont été laissés 18 commentaires, nous permet de compléter la liste incomplète des raisons proposées dans la question.

La question du manque de sens de la démarche pour certaines données revient beaucoup : « *pour des raisons de sens et de contextualisation du recueil* », « *les données dépendent de la problématique* », « *Pas pertinent (recherches qualitatives)* », « *elles ne s'appliquent qu'à ma recherche et beaucoup de données sont personnelles. Il faudrait beaucoup de temps et de travail pour les rendre "partageables", et par respect pour mes enquêté·es je ne le ferais pas avec n'importe qui. Cela risque que leur propos soit dénaturé si le contexte de recueil de données n'est pas pris en compte pour l'analyse.* », « *Éthique interprétation* ».

Une insistance est également portée dans certains commentaires sur la notion de concurrence au sein de la recherche et de volonté d'être la principale personne à publier sur ses données : « *Garder de l'avance sur mes travaux* », « *La culture globalement concurrentielle*

et non collaborative de la communauté scientifique à ce jour », « l'utilisation de mes données difficilement collectées alors que je n'ai pas encore pu les exploiter complètement », « tant que les données n'ont pas été analysées et traitées et fait l'objet d'une publication ».

Un certain nombre de commentaires viennent préciser l'obstacle juridique : « *non-respect de la confidentialité* », « *habilitation secret-défense* », « *confidentialité des propos* », « *données personnelles non partageables* ».

Enfin, un commentaire vient indiquer une préférence pour d'autres modalités de partage : « *Je préfère une collaboration, un échange : données contre données* ».

Les entretiens viennent aussi préciser l'aspect chronophage de cette ouverture : « *Parce que ça relève souvent du volontariat : consacrer son temps de chercheur pour rendre la donnée publique, ce n'est pas valorisant, ça ne rentre pas dans les papiers.* »

Il est intéressant de croiser cette question avec la question 1.1, « À quelle unité de recherche du Cnam appartenez-vous ? », obligatoire (*Figure 27 - Principales raisons de ne pas mettre ses données en libre accès selon le laboratoire des répondant-es*). Comme à Rennes 2, nous avons sélectionné les laboratoires comptabilisant dix répondant-es ou plus, dont on peut espérer qu'ils soient les plus significatifs.

Selon le laboratoire, ces raisons ne sont pas les mêmes : au Lirsa, au Cédric et au CRTD, ce sont d'abord les freins juridiques qui font obstacle, puis le manque de temps. Ce dernier apparaît en bonne première place au LCM et au LMSSC. Enfin, au Lise, les freins juridiques passent en deuxième, la crainte du plagiat représentant la première source de réticences des répondant-es.

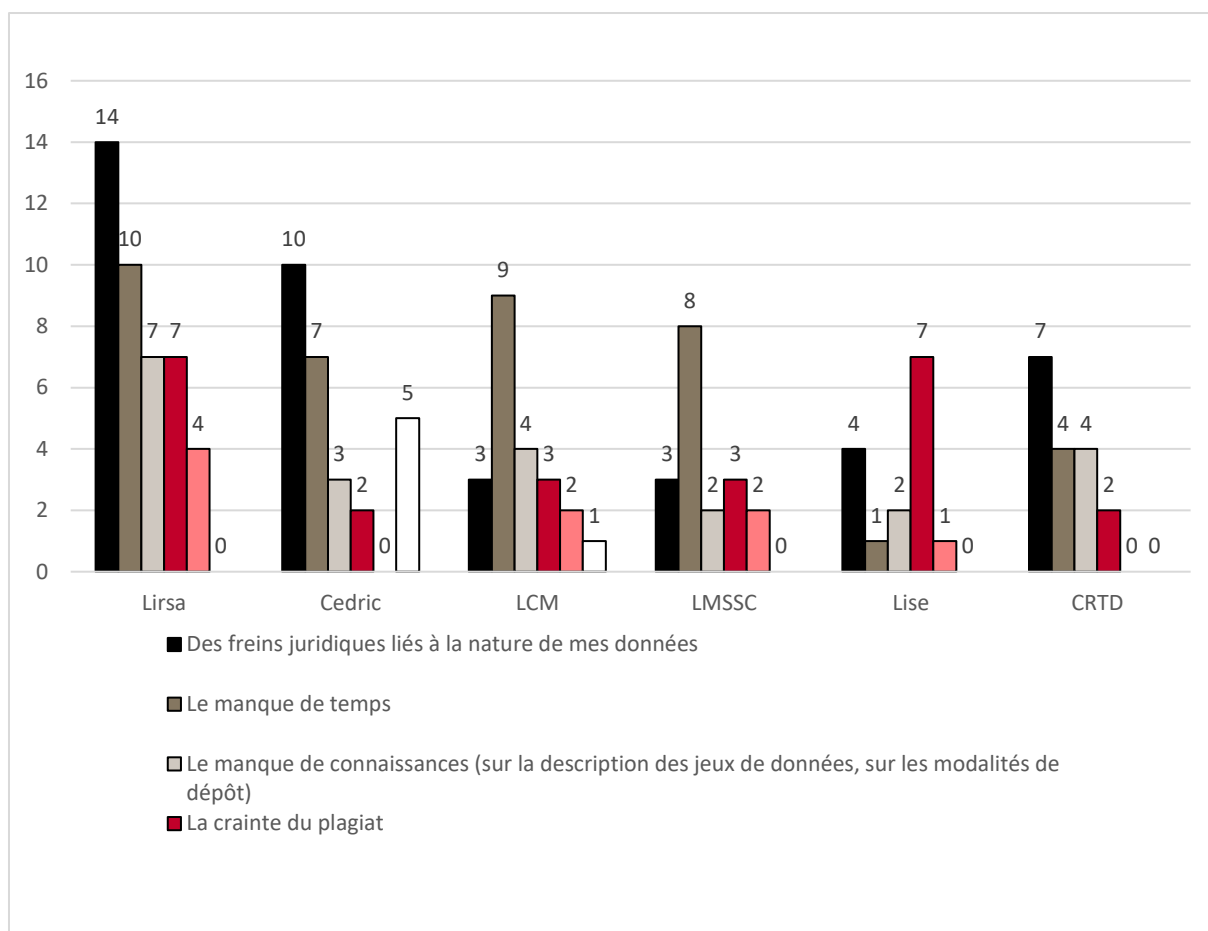


Figure 27 - Principales raisons de ne pas mettre ses données en libre accès selon le laboratoire des répondant-es

5. Besoins et acteur·rices

Résumé : la connaissance du sujet des données de la recherche par les répondant·es est globalement imprécise et mal maîtrisée. Les besoins sont assez mal identifiés et la demande d'une formation générale sur le sujet ressort. Les interlocutrices principales identifiées sur le sujet sont principalement la direction du laboratoire et la direction de la recherche, suivis de la direction des bibliothèques et de la documentation.

5.1 Connaissance du sujet par les répondant·es

Cette question avait pour but d'évaluer la connaissance des répondant·es sur quatre aspects importants de la gestion et de l'ouverture des données (*Figure 28 - Évaluation des compétences des chercheur·euses sur quatre aspects de la gestion des données*. « Je sais... » ou « Je connais... »). Le partage de jeux de données en accès ouvert est celui qui recueille le plus de réponses positives (33 réponses « tout à fait d'accord » ou « plutôt d'accord », soit 26 %), suivi de près par la capacité à identifier des plateformes où déposer un jeu de données en accès ouvert (29 réponses, 23 %). La connaissance des principes FAIR vient après (vingt réponses, 17 %) et enfin celle des techniques de rédaction d'un plan de gestion de données (huit réponses, 7 %).

Il faut noter que l'immense majorité des réponses est négative : 44 % de « pas du tout d'accord » pour le partage de jeu de données en accès ouvert, 48 % pour les plateformes sur lesquelles déposer un jeu de données en accès ouvert, 63 % pour la connaissance des principes FAIR et enfin 71 % pour la rédaction d'un plan de gestion de données. Soit un total de 56 % de réponses « pas du tout d'accord » ; en cumulant à ces dernières les réponses « plutôt pas d'accord », on atteint 74 % de réponses négatives.

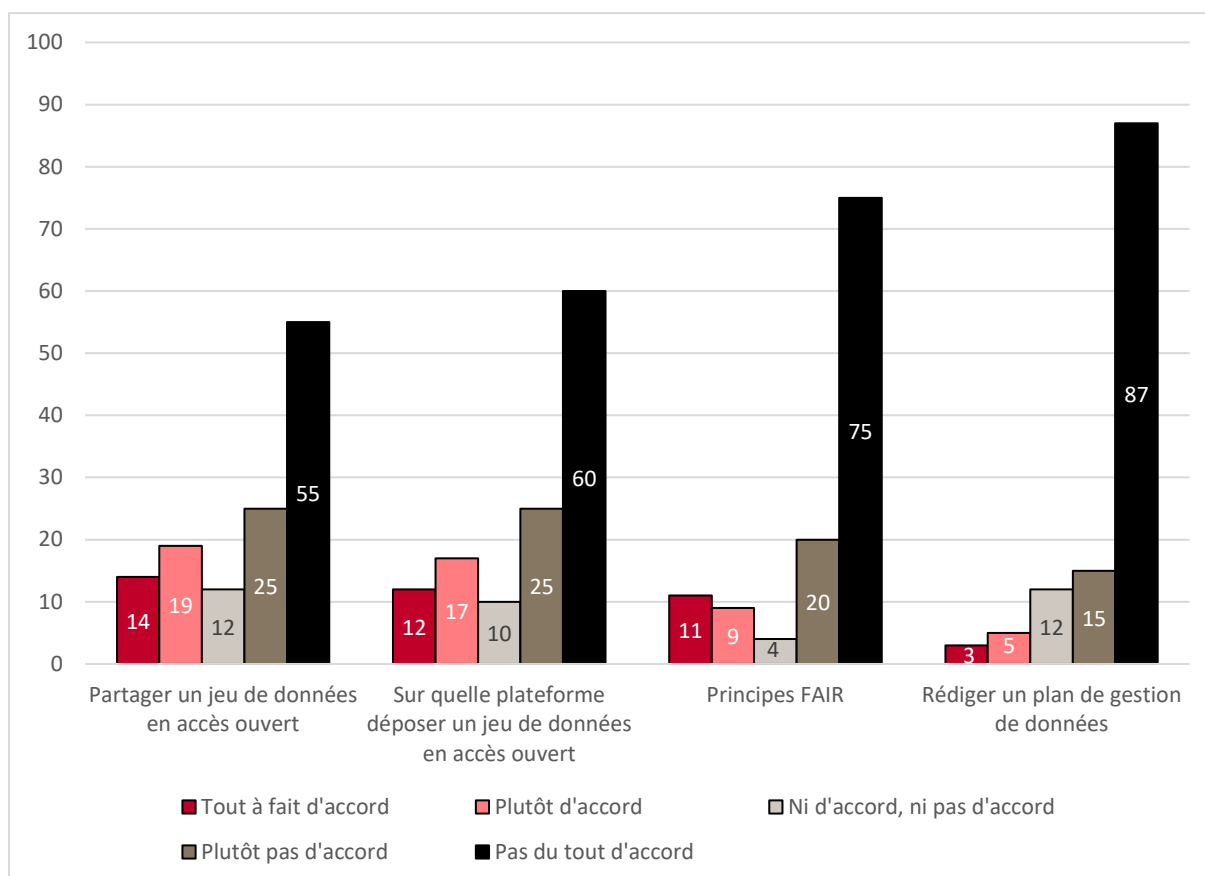


Figure 28 - Évaluation des compétences des chercheur·euses sur quatre aspects de la gestion des données. « Je sais... » ou « Je connais... »

5.2 Besoins

La question sur les besoins ne fait pas particulièrement ressortir de besoins phare (Figure 29 - Principaux besoins des répondant·es).

31 % des répondant·es souhaitent bénéficier d'une offre de formation générale « en priorité » ; 51 % « éventuellement » ; 18 % ne sont pas intéressé·es. Les entretiens ont confirmé que l'offre de formation générale était une priorité :

« Comme je n'y connais pas grand-chose, je ne sais pas si j'ai des besoins de formation. Je ne peux pas appréhender ce qu'il y a à apprendre. J'ai besoin d'une sensibilisation générale : qu'est-ce que la science ouverte peut-être déjà, qu'est-ce que les données personnelles, qu'est-ce qu'il y a à protéger, pourquoi ? Pourquoi la politique de science ouverte aussi, qu'est-ce qui se cache derrière, est-ce que c'est parce qu'il y a cette idée de "contribuons tous, mettons tous à contribution pour progresser ensemble" ou est-ce que c'est qu'il y a plein de recherches publiques qui bénéficient ensuite au secteur privé ? Ce serait intéressant aussi de voir quels enjeux il y a derrière. »

Des chercheur·euses en sciences humaines ont par ailleurs exprimé un intérêt marqué pour une assistance juridique, au cours des entretiens : « *Je gère des données et je ne suis pas très au courant des cadres que je dois appliquer, juridiquement* », « *Ça ne suffit pas de lire. Le RGPD, je l'ai lu, je le connais, et après ?* »

Se posent aussi souvent des questions sur les études longitudinales, sur la réutilisation des données sur la durée ou après un certain temps : les chercheur·euses constatent que cela soulève des enjeux juridiques mais aussi des questions de stockage, d'archivage.

Sur les PGD, les répondant·es ont indiqué lors des entretiens que ce document avait surtout été perçu comme un « *livrable en plus* » parmi d'autres, une tâche administrative plus qu'une aide à la gestion de leurs données :

« Je dirais que ça fait une couche administrative en plus. On est complètement noyés. Dans un financement, on arrive à un moment où il y a à peine 50 % du financement qui sert à la recherche, le reste ça n'est que de la paperasse. Que ce soit en termes de rédaction, de publication... C'est vraiment un boulet très fort. »

Un chercheur nous indique ne pas avoir trouvé la démarche « utile », tandis que d'autres seraient curieux·ses d'en savoir plus : « *Je ressens un besoin de formation* ».

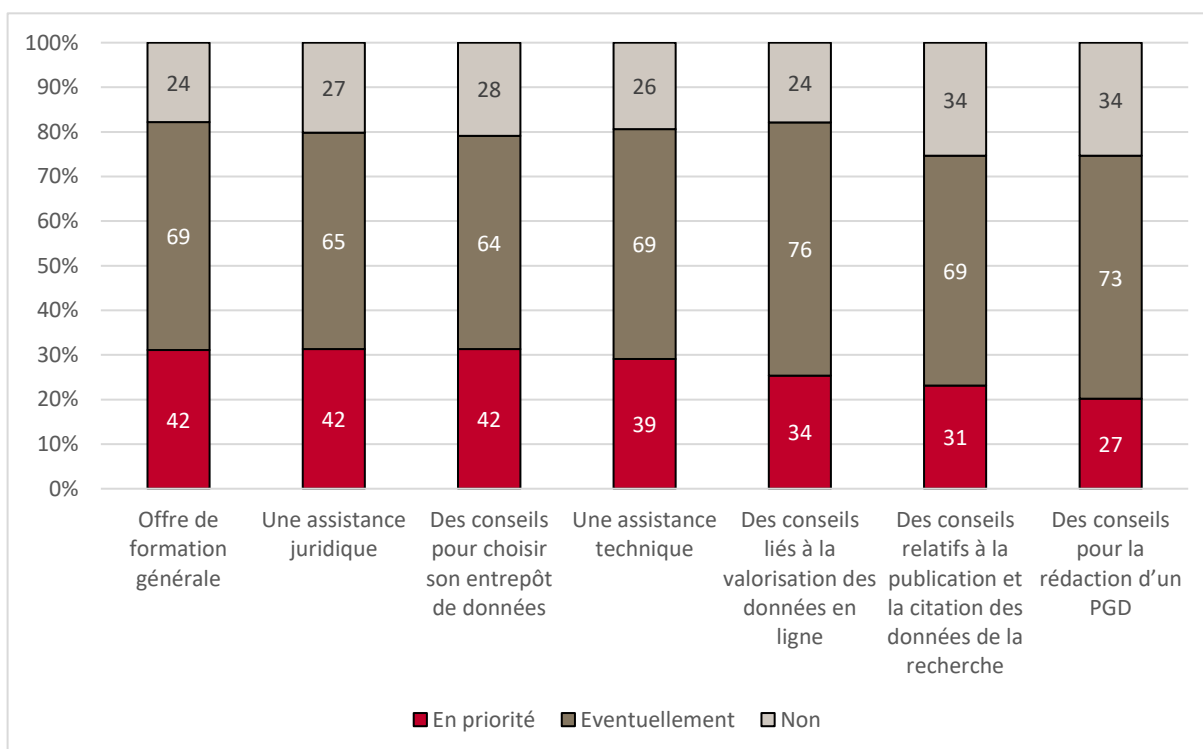


Figure 29 - Principaux besoins des répondant·es

5.3 Interlocuteur·rices privilégié·es

La question sur les interlocuteur·rices privilégié·es par les chercheur·euses pour répondre à leurs questions sur la gestion et l'ouverture des données avaient pour objectif d'identifier les acteur·rices les plus à même d'être impliqué·es dans les futurs services de l'établissement. Dans l'ordre, ce sont d'abord la direction du laboratoire puis la Direction de la recherche (DR) qui sont identifiés comme interlocutrices, suivies de la Direction des bibliothèques et de la documentation (DBD). (*Figure 30 - Interlocuteur·rices privilégié·es par les répondant·es sur la question des données*). Les autres interlocuteur·rices proposé·es dans la question (Service juridique et Direction des systèmes d'information), sont plus perçu·es comme des interlocuteur·rices ponctuel·les, en cas de besoin sur des thématiques spécifiques. La Direction de la communication (Dircom) est peu identifiée.

Une seconde question portait sur ces interlocuteur·rices potentiel·les. On y demandait aux répondant·es d'évaluer les affirmations suivantes : « Je sais à qui m'adresser pour obtenir des informations théoriques sur la gestion et l'ouverture des données » et « Je sais à qui m'adresser pour obtenir une assistance technique sur la gestion et l'ouverture des données ». On y constate que les interlocuteur·rices sont plutôt mal identifié·es. Les résultats ne divergent que peu selon qu'il s'agisse d'obtenir des informations théoriques ou une aide technique. Les réponses positives (« tout à fait d'accord » ou « plutôt d'accord ») représentent respectivement 21 et 20 réponses (soit 17 % des réponses dans les deux cas). Les réponses négatives (« pas du tout d'accord » ou « plutôt pas d'accord ») représentent 86 (71 %) et 89 réponses (74 %).

Il faut noter un lien fort entre l'offre de service sur HAL et la question des données de la recherche. Les entretiens abordent très régulièrement le sujet, et le CDFT (Centre de documentation sur la formation et le travail, auquel est à ce jour rattachée l'administratrice HAL) est notamment identifié comme acteur sur les sujets de science ouverte par les laboratoires dont les objets de recherche touchent à la formation et au travail.

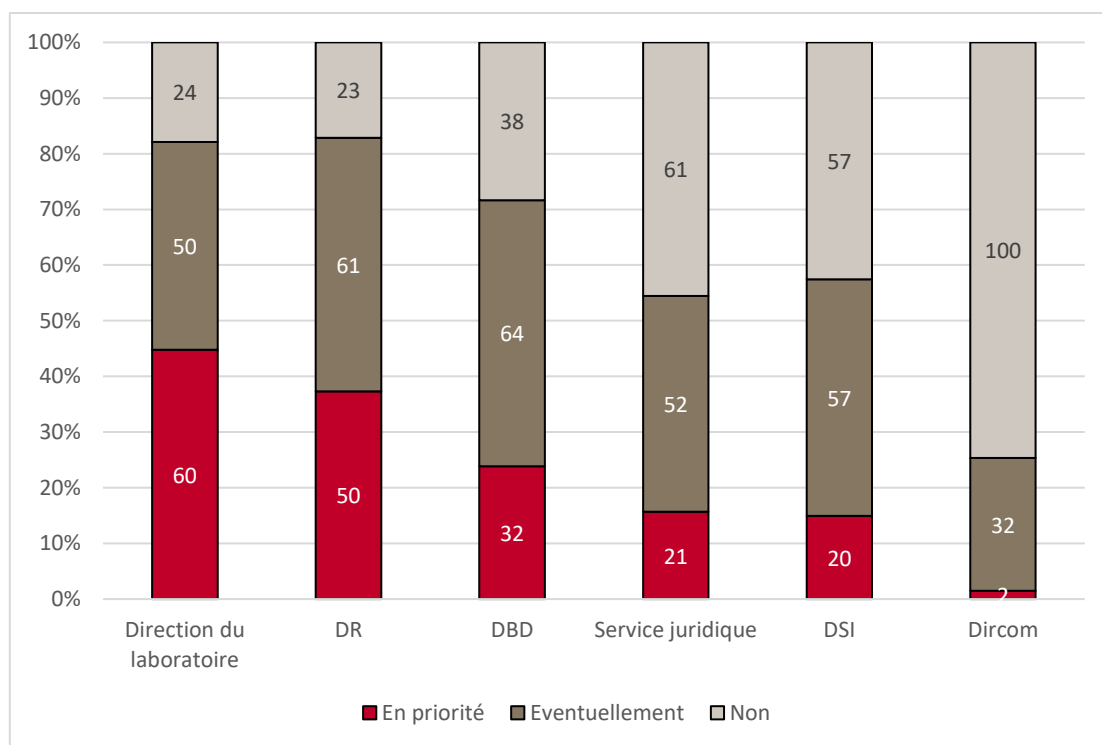


Figure 30 - Interlocuteur·rices privilégié·es par les répondant·es sur la question des données

Conclusion

Les 152 réponses au questionnaire considérées comme exploitables et les 10 entretiens complémentaires ont permis de tirer plusieurs conclusions de cette enquête, résumées au début de chaque chapitre.

Concernant la nature des données produites, on observe une grande variété dans les types, formats et volumes de données, à toutes les échelles (au sein de l'établissement, d'un champ disciplinaire, d'un même laboratoire). La problématique spécifique posée par les données personnelles concerne 39 % des répondant·es à l'enquête et ce type de données est manipulé (dans des mesures variables) dans tous les laboratoires.

83 % des répondant·es cumulent plusieurs lieux de stockage. Si la moitié des répondant·es indiquent que leur groupe de travail a accès à leurs données, ils·elles sont aussi 46 % (66 personnes) à indiquer que personne n'a accès à celles-ci. Les répondant·es étaient peu nombreux·ses à déclarer partager des données en libre accès, à avoir déjà participé à la rédaction d'un plan de gestion de données ou à avoir publié un *data paper*.

Les réponses aux questions liées à l'adhésion au principe d'ouverture des données montrent un décalage entre les pratiques (peu de partage des données) et les représentations des répondant·es (généralement favorables à leur diffusion). Les principales raisons d'une mise en libre accès des données sont le sentiment d'aider la communauté scientifique et la visibilité offerte aux travaux ; les principales raisons de ne pas le faire sont les freins juridiques et le manque de temps.

La connaissance du sujet des données de la recherche par les répondant·es est globalement imprécise et mal maîtrisée. Les besoins sont assez mal identifiés et la demande d'une formation générale sur le sujet ressort. Les interlocutrices principales identifiées sur le sujet sont principalement la direction du laboratoire et la Direction de la recherche, suivies de la Direction des bibliothèques et de la documentation.

Les résultats de l'enquête ont été diffusés à l'occasion d'un webinaire introduit par Stéphane Lefebvre, adjoint de l'Administrateur général en charge de la recherche. La publication de ce rapport lors de l'*open access week* 2022 a été accompagnée d'une infographie interactive¹⁴. Cette enquête a permis de définir plusieurs pistes de travail internes à l'établissement, et notamment au sein de la Direction des bibliothèques et de la documentation.

¹⁴ BERTRAM Marie-Liesse, 2022. Les données de la recherche au Cnam [en ligne]. Disponible sur URL : <https://view.genial.ly/629b4ccf4e5e8e0018ed9e50/interactive-content-les-donnees-de-la-recherche-au-cnam>

Bibliographie

Enquêtes précédentes

DEBOIN, Marie-Claude, 2020. Bilan d'enquête sur la pratique de plan de gestion de données (PGD) au Cirad [en ligne]. Rapport d'enquête. Cirad. Disponible sur URL : https://agritrop.cirad.fr/597251/1/Bilan_enquete_PGD_Cirad_20201214.pdf [consulté le 23 avril 2022].

GTSO DONNEES COUPERIN, 2022. Enquêtes. Science ouverte France [en ligne]. Disponible sur URL : <https://scienceouverte.couperin.org/donnees-recherche-enquetes/> [consulté le 20 août 2022].

HERBET, Marie-Emilia, LEONARD, Jérémie, SANTANGELO, Maria Grazia, et ALBARET, Lucie, 2021. Dissimuler ou disséminer ? Une étude sur le sort réservé aux résultats négatifs [en ligne]. DOI : <https://doi.org/10.5281/zenodo.5541589> [consulté le 3 avril 2022].

REBOUILLAT, Violaine, 2019. Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs [en ligne]. Thèse. Sciences de l'information et de la communication. Paris : Conservatoire national des arts et métiers - Cnam. Disponible sur URL : <https://tel.archives-ouvertes.fr/tel-02447653> [consulté le 25 novembre 2021].

SERRES, Alexandre, MALINGRE, Marie-Laure, MIGNON, Morgane, PIERRE, Cécile, COLLET, Didier, 2017. Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2 [en ligne]. Rapport de recherche. Université Rennes 2. Disponible sur URL : <https://hal.archives-ouvertes.fr/hal-01635186v2/document> [consulté le 18 août 2022].

Données de la recherche : généralités

BORDIGNON, Frédérique. *Espace chercheurs ENPC* [en ligne]. Disponible sur URL : <https://espacechercheurs.enpc.fr/fr/donnees-recherche-intro> [consulté le 9 septembre 2022].

DATAACC. Publier un data paper : où et comment ? *Dataacc.org* [en ligne]. Disponible sur URL : <https://www.dataacc.org/vos-besoins/valoriser-ses-donnees/data-papers/> [consulté le 9 septembre 2022].

DEDIEU, Laurence et FILY, Marie-Françoise, 2015. Rendre publics ses jeux de données scientifiques, en 6 points. Montpellier (FRA) : CIRAD, 6 p. DOI : <https://doi.org/10.18167/coopist/0059> [consulté le 15 août 2022].

GRAY, Jim, LIU, David T., NIETO-SANTISTEBAN, Maria, SZALAY, Alexander S., DEWITT, David, HEBER, Gerd, 2005. Scientific Data Management in the Coming Decade [en ligne]. ACM

Sigmod Record. Vol. 34, n° 4, pp. 34-41. DOI : <https://doi.org/10.1145/1107499.1107503> [consulté le 23 mars 2022].

GALONNIER, Juliette, *et al.*, 2019. Les sciences humaines et sociales au travail (ii) : que faire des données de la recherche ?. Tracés [en ligne]. N° 19. DOI : <https://doi.org/10.4000/traces.10518> [consulté le 15 août 2022].

LEHOUX, Elise, 2021. Les données de recherche. Questions à Christine L. Borgman. Bulletin des bibliothèques de France (BBF) [en ligne]. Disponible sur URL : https://bbf.enssib.fr/consulter/Contribution_2021-032.pdf [consulté le 16 mars 2022].

OCDE, 2007. Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics [en ligne]. Disponible sur URL : <https://www.oecd.org/fr/science/inno/38500823.pdf> [consulté le 9 mars 2022].

OCDE, 2021. Recommendation of the Council concerning Access to Research Data from Public Funding [en ligne]. Disponible sur URL : <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347> [consulté le 22 août 2022].

SCHÖPFEL, Joachim, KERGOSIEN, Eric et PROST, Hélène, 2017. « Pour commencer, pourriez-vous définir 'données de la recherche' ? » Une tentative de réponse. *Atelier VADOR : Valorisation et Analyse des Données de la Recherche ; INFORSID 2017*, mai 2017, Toulouse, France. [\(hal-01530937\)](#) [consulté le 13 juillet 2022].

WILKINSON, Mark D., DUMONTIER, Michel, AALBERSBERG IJsbrand Jan et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data [en ligne]. N°3. DOI : <https://doi.org/10.1038/sdata.2016.18> [consulté le 17 mai 2022].

Données de la recherche : cadre légal et politique

ANR, 2018. Plan d'action 2019 [en ligne]. Disponible sur URL : <https://anr.fr/fileadmin/documents/2018/Plan-d-action-ANR-2019.pdf> [consulté le 20 août 2022].

Inist-CNRS, 2022. Les exigences des financeurs. Doranum [en ligne]. DOI : <https://doi.org/10.13143/37ss-m608> [consulté le 20 août 2022].

MESRI, 2018. Plan national pour la science ouverte [en ligne]. Disponible sur URL : <https://www.ouvrirlascience.fr/plan-national-pour-la-science-ouverte/> [Consulté le 9 mars 2022].

MESRI, 2020. Passeport pour la science ouverte [en ligne]. Disponible sur URL : https://www.ouvrirlascience.fr/wp-content/uploads/2021/10/SO_21-10-14-WEB-FR.pdf [consulté le 20 août 2022]

MESRI, 2022. Stratégie nationale des infrastructures de recherche. Edition 2021 [en ligne]. Disponible sur URL : <https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2022->

[03/feuille-de-route-nationale-des-infrastructures-de-recherche---2021-v2--17318.pdf](#) [consulté le 30 mars 2022].

Premier ministre, 2021. Décret n° 2021-1572 du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique [en ligne]. JORF n° 0283 du 5 décembre 2021. Disponible sur URL : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360> [consulté le 31 janvier 2022].

Annexes

Guide d'entretien

Thème	Directeur de laboratoire	SHS	STM
	Quels types de données produisez-vous ?		
Nature des données produites	Y a-t-il des types de données variés au sein du laboratoire ? Ou au contraire un type de données particulièrement représenté ?		
	Ces types de données ont-ils des spécificités impactant leur gestion ?		
	Quelle taille font ces données ?		
		Gérez-vous des données personnelles ? Que pensez-vous d'une éventuelle spécificité des SHS quant à la question des données ?	
Stockage	Où stockez-vous vos données ?		
Pratiques de gestion et d'ouverture des données	Avez-vous été impliqué dans des projets ANR, ERC ? Avez-vous déjà rédigé un plan de gestion de données ? Que pouvez-vous dire sur cette expérience ? Était-ce facile ? Quelles étaient les difficultés ? Avez-vous trouvé cela utile ? Auriez-vous eu besoin d'aide ?		
	Si non : définition du PGD, présentation de son utilité : qu'en pensez-vous ?		
Représentations de l'ouverture des données	Que pensez-vous du principe « aussi ouvert que possible, aussi fermé que nécessaire » ? Y voyez-vous des limites ?		
	Quelles est votre position en tant que directeur de laboratoire, sur une éventuelle politique des données de recherche : est-ce que c'est une question qui a déjà été débattue dans votre laboratoire ? Si non, pensez-vous que votre laboratoire devrait avoir une politique commune,		

	cohérente ? Et sur quoi devrait-elle porter ?		
Besoins	Quelles sont les principales difficultés que vous rencontrez quant à la gestion des données de la recherche ?		
	Et quant à leur ouverture ?		
	De quel genre d'accompagnement auriez-vous besoin ? Quelle forme cela pourrait-il prendre ? Sur quel sujet privilégié (aucun n'étant particulièrement ressorti dans l'enquête) ?		
	Quels seraient les principaux besoins selon vous en termes de services, d'accompagnement, de formation, d'outils techniques, d'infrastructures de stockage ?		
Réutilisation	Réutilisez-vous les données d'autres chercheurs ? Pourquoi ? Où les trouvez-vous ?		
Politique	Que pensez-vous de l'éventualité d'une mise en place d'une politique générale sur le sujet ? Politique d'établissement, de laboratoire, aide ponctuelle ? Qu'est-ce qui serait le plus pertinent à vos yeux : une réflexion collective sur les données ou une approche ponctuelle et individuelle ?		

Questionnaire

Enquête sur la gestion et l'ouverture des données de recherche au Cnam

La Direction des Bibliothèques et de la Documentation et la Direction de la Recherche du Cnam lancent une enquête sur les **pratiques de gestion et d'ouverture des données de la recherche** des chercheurs et chercheuses du Cnam.

Objectifs de l'enquête :

- Connaître les pratiques de stockage, d'archivage et de diffusion des données de la recherche au sein du Cnam ;
- Comprendre les représentations des données de la recherche ;
- Identifier les besoins afin de proposer des services d'accompagnement pertinents et adaptés.

Que sont les données de la recherche ?

Les données de recherche sont 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 » (OCDE).

Le questionnaire est ouvert jusqu'au 17/03/2022. La durée estimée de réponse est de 10 minutes.

Merci beaucoup de l'intérêt que vous lui accorderez.

Traitement des données

Les réponses à ce questionnaire seront traitées de manière anonyme. Elles seront analysées par les membres de la cellule science ouverte (https://bibliotheques.cnam.fr/opac/article/la-science-ouverte-au-cnam/5_so_cnam) de la DBD. Les résultats pourront donner lieu à une diffusion anonymisée et à des publications associées.

Les informations recueillies dans ce questionnaire sont enregistrées dans un fichier informatisé par le Cnam-Etablissement public, ce traitement étant nécessaire aux fins des intérêts légitimes poursuivis par l'établissement.

Les données collectées seront communiquées exclusivement à la Direction des bibliothèques et de la documentation du Cnam-Etablissement public et seront conservées pour une durée de 6 mois.

Vous pouvez accéder aux données vous concernant, les rectifier, demander leur effacement ou exercer votre

droit à la limitation du traitement de vos données.

Pour exercer ces droits ou pour toute autre question sur le traitement de vos données dans ce dispositif, vous pouvez contacter la déléguée à la protection des données du Cnam-Etablissement public à l'adresse mail suivante : ep_dpo@lecnam.net (mailto:ep_dpo@lecnam.net) ; ou la cellule science ouverte à l'adresse mail suivante : science.ouverte@cnam.fr (<mailto:science.ouverte@cnam.fr>)

Une réclamation peut être adressée à la Cnil.

Il y a 27 questions dans ce questionnaire.

1. Votre statut au sein du Cnam

1.1 A quelle(s) unité(s) de recherche du Cnam appartenez-vous ? *

❗ Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ CEDRIC
- ☐ CEET
- ☐ CRTD
- ☐ Diden-IDF
- ☐ DynFluid
- ☐ Eren
- ☐ ESDR3C
- ☐ Esycom
- ☐ FoAP
- ☐ GBCM
- ☐ GeF
- ☐ HT2S
- ☐ IAT-Cnam
- ☐ Lafset
- ☐ LCM
- ☐ Lifse
- ☐ Lirsa
- ☐ Lise
- ☐ LMSSC
- ☐ Lusac
- ☐ M2N
- ☐ MESuRS
- ☐ Metabiot
- ☐ Pimm
- ☐ Satie
- ☐ SayFood
- ☐ Autre unité de recherche:

1.2 À quel(s) champ(s) disciplinaire(s) êtes-vous rattaché-e ? *

❗ Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Sciences de l'ingénieur (SI)
- ☐ Physique-Sciences de la Terre et de l'univers (PSTU)
- ☐ Mathématiques-Informatique (MI)
- ☐ Chimie-Matériaux (CM)
- ☐ Sciences du vivant-Médecine (SVM)
- ☐ Lettres et sciences humaines et sociales (LSHS)
- ☐ Droit, économie, politique, gestion (DEPG)

1.3 Quel est votre statut ? *

❗ Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Chargé-e de recherche
- ☐ Directeur·rice de recherche
- ☐ Doctorant·e
- ☐ Enseignant·e et chercheur·se associé·e ou invité·e (PAST-MAST)
- ☐ Ingénieur·e ou personnel technique de recherche et de formation (ITRF)
- ☐ Maître·sse de conférence
- ☐ Professeur·e des universités
- ☐ Professeur·e titulaire de chaire
- ☐ Post-doctorant·e
- ☐ Autre statut:

2. Nature des données produites

2.1 Quel(s) type(s) de données produisez-vous dans le cadre de vos recherches ?

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Données d'observation (données non reproductibles, par exemple : relevés de température ou de précipitations, données de télédétection, photos d'un événement, données d'enquêtes, etc.)
- ☐ Données expérimentales (données censées être reproductibles, par exemple : générées par un équipement en laboratoire, mesures de performance d'une machine, etc.)
- ☐ Données de simulation numérique (données reproductibles produites par des logiciels, par exemple : modèles climatiques ou économiques, etc.)
- ☐ Données dérivées ou compilées (données brutes qui ont fait l'objet de traitements et d'analyses successifs, par exemple : données obtenues par de la fouille de textes ou de données d'une base, etc.)
- ☐ Code informatique
- ☐ Autre:

2.2 Quel est leur format ?

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Texte délimité (csv, tsv)
- ☐ Fichiers de tableur (Excel, Libre Office, Calc)
- ☐ Fichiers de traitement de texte (Word, Write...)
- ☐ Fichiers audio
- ☐ Fichiers image
- ☐ Fichiers vidéo
- ☐ Données d'instrumentation
- ☐ Base de données (Access, MySQL)
- ☐ Notes manuscrites
- ☐ Autre:

2.3 À combien estimez-vous l'espace nécessaire au stockage de vos données de recherche ?

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Moins de 1 Go
- ☐ Entre 1 Go et 20 Go
- ☐ Entre 20 Go et 100 Go
- ☐ Entre 100 Go et 1 To
- ☐ Plus de 1 To
- ☐ Je ne sais pas

2.4 Avez-vous déjà collecté au moins une fois des données à caractère personnel ?

Toute information se rapportant à une personne physique identifiée ou identifiable est une donnée à caractère personnel.

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
- ☐ Non
- ☐ Je ne sais pas

3. Pratiques de gestion et d'ouverture des données

3.1 Où stockez-vous vos données de recherche ?

Veuillez choisir toutes les réponses qui conviennent :

- ☐ En local, dans des dossiers papier
- ☐ En local, sur mon ordinateur personnel
- ☐ En local, sur mon ordinateur professionnel
- ☐ En local, sur un disque dur externe ou un autre support (clé USB, CD, DVD, etc.)
- ☐ En réseau, sur un serveur de mon laboratoire
- ☐ En réseau, sur un serveur du Cnam
- ☐ En réseau, sur une plateforme privée en ligne (GoogleDrive, Dropbox)
- ☐ En réseau, sur un serveur hébergé par une autre institution
- ☐ En réseau, sur le serveur d'une TGIR (très grande infrastructure de recherche)
- ☐ Sur un autre support:

3.2 En dehors de vous, qui peut accéder directement à vos données ?

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Personne
- ☐ Mon groupe de travail
- ☐ Mon laboratoire
- ☐ L'ensemble de l'université
- ☐ Les chercheurs et chercheuses de la discipline
- ☐ Tout le monde
- ☐ Un autre groupe de personnes:

3.3 Avez-vous déjà stocké vos données en libre accès sur le web ?

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui, plus d'une fois
- ☐ Oui, une fois
- ☐ Non, jamais

Si oui, sur quel(s) site(s) / entrepôt(s) de données ?

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui, plus d'une fois' ou 'Oui, une fois' à la question '10 [G3Q00003]' (3.3 Avez-vous déjà stocké vos données en libre accès sur le web ?)

Veuillez écrire votre réponse ici :

3.4 Avez-vous déjà participé à la rédaction d'un Plan de Gestion de Données ?

Un plan de gestion de données (PGD, ou en anglais DMP pour Data Management Plan) est un document détaillant les processus permettant de créer, documenter, stocker, partager et protéger les données produites lors d'un projet.

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Oui, dans le cadre de projets où le financeur l'exigeait
- ☐ Oui, pour d'autres projets
- ☐ Non, jamais
- ☐ Je ne sais pas

3.5 Avez-vous déjà publié un *data paper* ?

Un data paper est une publication scientifique, soumise à la validation des pairs, dont le but principal est de décrire un ou plusieurs jeux de données, afin de permettre leur réutilisation.

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
- ☐ Non, mais cela m'intéresse
- ☐ Non, et cela ne m'intéresse pas

3.6 Bénéficiez-vous d'une aide spécifique sur la question des données (gestion, traitement, archivage...) ?

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Oui, de la part de personnels de soutien à la recherche
- ☐ Oui, de la part d'intervenant·es extérieur·es à mon laboratoire
- ☐ Non, chacun·e gère ses propres données
- ☐ Oui, de la part d'autres intervenants:

3.7 Votre laboratoire a-t-il mis en place des règles spécifiques sur la gestion des données (stockage, archivage, diffusion) ?

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
- ☐ Non
- ☐ Je ne sais pas
- ☐ C'est en projet ou en cours

4. Représentations de l'ouverture des données

4.1 Dans quelle mesure êtes-vous d'accord ou pas d'accord avec ce qui suit ?

Choisissez la réponse appropriée pour chaque élément :

	Tout à fait d'accord	Plutôt d'accord	Ni d'accord, ni pas d'accord	Plutôt pas d'accord	Pas du tout d'accord
Je suis prêt-e à partager mes données de recherche avec d'autres	☐	☐	☐	☐	☐
Je suis prêt-e à ce que d'autres réutilisent mes données	☐	☐	☐	☐	☐
Je souhaiterais pouvoir réutiliser les données d'autres chercheurs ou chercheuses	☐	☐	☐	☐	☐

4.2 Pour quelle(s) raison(s) choisiriez-vous de mettre vos données de recherche en libre accès ?

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Une obligation du financeur (Horizon Europe, ANR)
- ☐ Le sentiment d'aider la communauté scientifique par la mise à disposition de mes données
- ☐ Le souhait de voir se développer d'autres recherches à partir de mes jeux de données
- ☐ Une meilleure visibilité de mes travaux
- ☐ L'occasion de nouveaux contacts, de nouvelles coopérations scientifiques
- ☐ Le souci de pouvoir faire valider mes résultats à partir de mes jeux de données
- ☐ L'adhésion aux valeurs du libre accès aux résultats de la recherche scientifique
- ☐ Aucune de ces raisons
- ☐ Autre:

4.3 Pour quelle(s) raison(s) choisiriez-vous de ne pas mettre vos données de recherche en libre accès ?

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Des freins juridiques liés à la nature de mes données
- ☐ Des raisons économiques ou commerciales
- ☐ La crainte du plagiat
- ☐ Le manque de temps
- ☐ Le manque de connaissances (sur la description des jeux de données, sur les modalités de dépôt)
- ☐ Aucune de ces raisons
- ☐ Autre:

5. Besoins

5.1 Dans quelle mesure êtes-vous d'accord ou pas d'accord avec ce qui suit ?

Choisissez la réponse appropriée pour chaque élément :

	Tout à fait d'accord	Plutôt d'accord	Ni d'accord, ni pas d'accord	Plutôt pas d'accord	Pas du tout d'accord
Je sais comment partager un jeu de données en accès ouvert	☐	☐	☐	☐	☐
Je sais sur quelle plateforme déposer un jeu de données en accès ouvert	☐	☐	☐	☐	☐
Je dispose d'assez de temps pour ouvrir mes données de recherche	☐	☐	☐	☐	☐
Je sais à qui m'adresser pour obtenir des informations théoriques sur l'ouverture des données de recherche	☐	☐	☐	☐	☐
Je sais à qui m'adresser pour obtenir une aide technique sur l'ouverture des données de recherche	☐	☐	☐	☐	☐
Je connais les principes FAIR	☐	☐	☐	☐	☐
Je sais comment rédiger un plan de gestion de données	☐	☐	☐	☐	☐

5.2 Sur quel(s) point(s) souhaiteriez-vous bénéficier d'une aide ? *

Choisissez la réponse appropriée pour chaque élément :

	Oui, en priorité	Éventuellement	Non
Une offre de formation générale sur la gestion des données de la recherche	☐	☐	☐
Des conseils pour la rédaction d'un plan de gestion de données	☐	☐	☐
Une assistance technique (par exemple sur les métadonnées, les normes, l'archivage à long terme)	☐	☐	☐
Une assistance juridique (par exemple sur les restrictions d'accès, le traitement des données sensibles, l'utilisation de licences, le RGPD)	☐	☐	☐
Des conseils pour choisir son entrepôt de données	☐	☐	☐
Des conseils relatifs à la publication et la citation des données de la recherche	☐	☐	☐

	Oui, en priorité	Éventuellement	Non
Des conseils liés à la valorisation des données en ligne (par exemple via les data papers)	☐	☐	☐

5.3 Sur la question de la gestion et de l'ouverture des données de la recherche, à qui vous adresseriez-vous en priorité au sein du Cnam ? *

Choisissez la réponse appropriée pour chaque élément :

	En priorité	Éventuellement	Non
La direction de votre laboratoire	☐	☐	☐
La direction de la recherche	☐	☐	☐
La direction de la communication	☐	☐	☐
La direction des services informatiques	☐	☐	☐
Le service juridique	☐	☐	☐
La direction des bibliothèques et de la documentation	☐	☐	☐

6. Qui êtes-vous ?

6.1 Quel est votre genre ?

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Féminin
- ☐ Masculin

6.2 Quel âge avez-vous ? *

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Moins de 30 ans
- ☐ Entre 30 et 39 ans
- ☐ Entre 40 et 49 ans
- ☐ Entre 50 et 59 ans
- ☐ Plus de 60 ans

6.3 Depuis combien de temps exercez-vous une activité de recherche ? *

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Moins de 5 ans
- ☐ 5-10 ans
- ☐ 11-15 ans
- ☐ 16-20 ans
- ☐ Plus de 20 ans

7. Contact

7.1 Acceptez-vous d'être recontacté ultérieurement pour de futurs entretiens dans le cadre de la deuxième partie de l'enquête ? *

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

Si oui, merci de nous laisser votre adresse email afin que nous puissions vous recontacter. *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '25 [G7Q00001]' (7.1 Acceptez-vous d'être recontacté ultérieurement pour de futurs entretiens dans le cadre de la deuxième partie de l'enquête ?)

Veuillez écrire votre réponse ici :

7.2 Souhaitez-vous ajouter quelque chose ?

Veuillez écrire votre réponse ici :

Merci de votre contribution !

Contact : science.ouverte@cnam.fr (mailto:science.ouverte@cnam.fr)

En savoir plus sur la science ouverte au Cnam

(https://bibliotheques.cnam.fr/opac/article/la-science-ouverte-au-cnam/5_so_cnam)

Liste des unités de recherche, programmes transverses et instituts du Cnam

Sciences humaines et sociales

CEET

Centre d'études de l'emploi et du travail

CRTD

Centre de recherche sur le travail et le développement

Dicen-IDF

Dispositifs d'information et de communication à l'ère numérique - Paris, Île-de-France

ESDR3C

Équipe Sécurité & Défense - Renseignement, criminologie, crises, cybermenaces

Foap

Formation et apprentissages professionnels

HT2S

Histoire des technosciences en société

Lirsa

Laboratoire interdisciplinaire de recherche en sciences de l'action

Lise

Laboratoire interdisciplinaire pour la sociologie économique

Sciences de l'ingénieur

Cédric

Centre d'études et de recherche en informatique et communications

DynFluid

Dynamique des fluides

Eren

Équipe de recherche en épidémiologie nutritionnelle

Esycom

Équipe systèmes de communication et microsystèmes

GBCM

Génomique, bioinformatique et chimie moléculaire

GeF

Géomatique et foncier

Lafset

Laboratoire du froid et des systèmes énergétiques et thermiques

LCM

Laboratoire commun de métrologie LNE-Cnam

Lifse

Laboratoire d'ingénierie des fluides et des systèmes énergétiques

LMSSC

Laboratoire de mécanique des structures et des systèmes couplés

Lusac

Institut national des sciences et techniques de la mer (Intechmer)

M2N

Modélisation mathématique et numérique

MESuRS

Modélisation, épidémiologie et surveillance des risques sanitaires

Metabiot

Metabiot

Pimm

Procédés et ingénierie en mécanique et matériaux

Satie

Systèmes et applications des technologies de l'information et de l'énergie

SayFood

Ingénierie des aliments et bioproduits

IAT

Institut aérotechnique

Table des illustrations

Figure 1 - Nombre de répondant·es par laboratoire d'appartenance	13
Figure 2 - Pourcentage de l'effectif du laboratoire ayant ou non participé à l'enquête.....	14
Figure 3 - Répartition disciplinaire des répondant·es	15
Figure 4 - Nombre de répondant·es par statut	16
Figure 5 - Genre des répondant·es.....	17
Figure 6 - Âge des répondant·es.....	18
Figure 7 - Expérience des répondant·es dans le milieu de la recherche.....	19
Figure 8 - Types de données produits	21
Figure 9 - Répartition des types de données produits au sein des champs disciplinaires	22
Figure 10 - Répartition des types de données au sein de la production des chercheur·euses du Cnam	23
Figure 11 - Répartition des volumes de données au sein de trois laboratoires.....	24
Figure 12 - Répartition des réponses à la question « Gérez-vous des données personnelles ? »	25
Figure 13 - Pratiques de gestion de données personnelles selon le champ disciplinaire	25
Figure 14 - Cumul de lieux de stockages par les répondant·es	26
Figure 15 - Lieux de stockage des données des répondant·es	27
Figure 16 - Personnes ayant accès ou non aux données des répondant·es.....	27
Figure 17- Réponses à la question « Avez-vous déjà partagé des données en libre accès sur le web ? »	28
Figure 18 - Pourcentage de l'effectif ayant déjà partagé ses données en libre accès selon l'âge	29
Figure 19 - Pourcentage de l'effectif ayant déjà partagé ses données en libre accès selon la durée de l'expérience dans le milieu de la recherche	29
Figure 20 - Réponses à la question « Avez-vous déjà participé à la rédaction d'un plan de gestion de données ? »	30
Figure 21 - Réponses à la question « Avez-vous déjà publié un data paper ? »	31
Figure 22 - Réponses à la question « Bénéficiez-vous d'une aide spécifique sur la question des données ? »	32
Figure 23 - Présence de règles spécifiques sur la gestion des données (stockage, archivage, diffusion) au sein des laboratoires	33
Figure 24 - Adhésion au principe d'ouverture des données	35
Figure 25 - Principales raisons d'une ouverture de l'accès aux données.....	36
Figure 26 - Principales raisons d'une fermeture de l'accès aux données	37
Figure 27 - Principales raisons de ne pas mettre ses données en libre accès selon le laboratoire des répondant·es	39
Figure 28 - Évaluation des compétences des chercheur·euses sur quatre aspects de la gestion des données. « Je sais... » ou « Je connais... ».....	41
Figure 29 - Principaux besoins des répondant·es.....	42
Figure 30 - Interlocuteur·rices privilégié·es par les répondant·es sur la question des données	44