



HAL
open science

Gestion et ouverture des données de recherche en SHS : une enquête auprès des communautés de Lyon 2, Lyon 3 et la MSH Lyon St-Étienne

Marie-Emilia Herbet, Aurore Montébran, Alexandra Dugué, Raphaëlle Billy,
Mariana Maire, Christophe Honet, Aurore Cartier, Marie-Lise Krumenacker,
Ling Li, Camille Espiau-Bechetoille, et al.

► To cite this version:

Marie-Emilia Herbet, Aurore Montébran, Alexandra Dugué, Raphaëlle Billy, Mariana Maire, et al..
Gestion et ouverture des données de recherche en SHS : une enquête auprès des communautés de Lyon
2, Lyon 3 et la MSH Lyon St-Étienne. Université Jean Moulin Lyon 3; Université Lumière Lyon 2;
MSH Lyon - Saint-Étienne. 2023. hal-04087568

HAL Id: hal-04087568

<https://univ-lyon3.hal.science/hal-04087568v1>

Submitted on 15 May 2023

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

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



Distributed under a Creative Commons Attribution 4.0 International License

GESTION ET OUVERTURE DES DONNÉES DE RECHERCHE EN SHS : UNE ENQUÊTE AUPRÈS DES COMMUNAUTÉS DE LYON 2, LYON 3 ET LA MSH LYON ST-ÉTIENNE

Françoise BARRE¹, Raphaëlle BILLY¹, Aurore CARTIER³, Alexandra DUGUE², Camille ESPIAU-
BECHETOILLE¹, Marie-Emilia HERBET³, Christophe HONET⁴, Marie-Lise KRUMENACKER¹,
Ling LI¹, Mariana MAIRE¹, Aurore MONTEBRAN³

(1) SCD Université Lyon 2, (2) Maison des sciences de l'homme Lyon St-Étienne, (3) SCD Université Lyon 3,

(4) Enssib, élève conservateur



Table des matières

Introduction	4
Contexte.....	4
Objectifs de l'enquête.....	5
Méthodologie.....	6
Limites.....	7
Choix de présentation des résultats.....	8
Profils des répondants.....	8
Caractérisation des données produites et stockées	11
Caractéristiques générales.....	11
Formats informatiques.....	13
Volumétrie et stockage.....	14
Pratiques en matière de gestion des données	18
Le plan de gestion de données : un exercice encore peu répandu	18
Données sensibles ou personnelles : des doutes sur la conduite à mener	20
Recours aux métadonnées : un usage équivoque.....	22
Diffusion des données : un intérêt non-identifié	23
Des pratiques d'archivage quasi-inexistantes	25
Des besoins traduisant le positionnement inconfortable des SHS.....	26
Enquête qualitative : des résultats clivants	28
Des politiques de gestion et d'ouverture des données mal articulées avec certaines disciplines et pratiques en SHS	28
Un cadre normatif hérité des sciences exactes.....	28
Gestion et documentation des données : de nouvelles obligations jugées lourdes ou inadéquates	29
Des chercheurs démunis face aux règles encadrant le traitement des données.....	30
Les données sensibles	30
Propriété intellectuelle.....	31
Vers une meilleure utilisation des données, une opportunité en SHS ?.....	32
Le défi de la conservation et de l'archivage	32
Un gisement scientifique à l'épreuve de compétences en devenir.....	33
Conclusion et perspectives.....	35

TABLE DES FIGURES	37
GLOSSAIRE	38
ANNEXES	39
Annexe 1 : questionnaire en ligne.....	39
Annexe 2 : grille d'entretien.....	60
Annexe 3 : services institutionnels de stockage en date (février 2023).....	62
Annexe 4 : liste des unités de recherche des répondants.....	63

POINTS CLÉS À RETENIR

- 75 % des répondants n'ont jamais partagé leurs données en accès ouvert
- Seuls 6,5 % disent publier leurs données en accès ouvert régulièrement
- Pour justifier ce non-partage, les répondants invoquent surtout le fait qu'ils n'en voient pas l'intérêt ou qu'ils n'y ont pas pensé
- Le support de stockage des données le plus fréquemment utilisé est le disque dur personnel
- Les répondants souhaitent en priorité recevoir des conseils d'ordre général et une aide à la gestion informatique des données

Introduction

Contexte

D'abord articulées autour de l'ouverture des publications, les politiques de science ouverte accordent désormais un intérêt croissant à l'enjeu des données de recherche, dont la réalité tangible diffère fortement d'une discipline à l'autre. Dans l'acceptation officielle proposée par la directive européenne de 2019 sur l'open data, les données de recherche correspondent à des « documents numériques » constituant des « éléments probants dans le processus de recherche » et jugés « nécessaires pour valider des conclusions »¹.

Cette approche, adossée au principe de « répliquabilité et de vérification des données »², induit qu'il « n'est pas de science (y compris sociale) sans possibilité de vérifier ses résultats »³. Comment les disciplines relevant des SHS s'inscrivent-elles dans cette « injonction à faire preuve »⁴ ?

Si certains champs des SHS mobilisent des outils statistiques ou des bases de données (géographie, archéologie, économie, psychologie, linguistique...), d'autres peuvent se percevoir comme davantage comme utilisateurs de sources que producteurs de données (droit, philosophie, littérature, histoire...) voire tirent leur légitimité d'enquêtes de terrain parfois

¹ Article 2 de la directive européenne 2019/1024 concernant les données ouvertes et la réutilisation des informations du secteur public. Disponible sur : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019L1024&from=ES>

² Le procès des données, 2022. *Genèses*. Vol. 129, n° 4, p. 3-6. Disponible sur : <https://doi.org/10.3917/gen.129.0003>

³ *Ibid.*

⁴ Siméant-Germanos, Johanna. Qui protéger, consentir à quoi, enquêter comment ? Les sciences sociales face à la bureaucratisation de la vertu scientifique. *Genèses*. 2022. Vol. 129, n° 4, p. 66-87. Disponible sur : <https://doi.org/10.3917/gen.129.0066>

sensibles, qui n'ont pas vocation à être ouvertes intégralement, même pour faire preuve (sociologie, ethnologie...).

Or, les données en tant que résultats susceptibles d'être partagés à grande échelle font désormais partie des attendus des financeurs de la recherche dans le cadre des appels à projet de l'Agence nationale de la recherche ou Horizon Europe. D'où la nécessité d'interroger les communautés en SHS pour mieux cerner leur positionnement vis-à-vis des politiques de gestion et d'ouverture des données, ainsi que les besoins d'accompagnement exprimés dans ce paysage très évolutif modelé par plusieurs initiatives nationales :

- Introduction des plans de gestion de données par l'ANR⁵ en 2019 ;
- Incitation à la structuration et au partage des données de recherche dans l'axe 2 des deux plans nationaux pour la science ouverte en 2018 et 2021, qui insistent sur la transformation des pratiques afin d'aboutir à des données dites FAIR (faciles à trouver, accessibles, interopérables et réutilisables) ;
- Adoption du décret sur l'intégrité scientifique en 2021 visant à généraliser les plans de gestion de données dans les établissements de l'ESR ;
- Publication de la feuille de route 2021-2024 du ministère de l'Enseignement supérieur et de la Recherche sur la politique des données, des algorithmes et des codes sources ;
- Lancement de la plateforme nationale fédérée Recherche data gouv en 2022 ;

C'est dans cet entour opportun que s'est structurée la coopération entre les SCD des universités Lyon 2, Lyon 3 et la MSH Lyon Saint-Étienne en vue de la réalisation d'une enquête.

L'appel à manifestation d'intérêt sur les ateliers de la donnée, porté par Recherche data gouv, a par ailleurs permis de légitimer cette enquête, qui s'inscrit dans une logique de coopération interétablissement, étant avéré que les conclusions de l'enquête constitueraient un premier socle fédérateur pour faciliter la participation à l'AMI* sur les données.

Objectifs de l'enquête

Dans ce contexte, l'enquête portée conjointement par les SCD Lyon 2, Lyon 3 et la Maison des sciences de l'homme (MSH), Lyon Saint-Étienne a pour objectif premier de dresser un état des lieux, même partiel :

- Des pratiques de gestion de données locales des communautés SHS du site ;
- Des principaux types, natures et formats de données en SHS produites au sein des structures de recherche ;
- Des expertises et compétences ainsi que des initiatives existantes au sein des unités de recherche ;
- Des besoins en accompagnement, sensibilisation et formation des communautés de recherche tout au long du cycle de vie des données ;
- Des besoins complémentaires au regard des infrastructures et des feuilles de routes nationales.

Cette enquête constitue également une première base de travail en vue d'établir de futures recommandations permettant le développement de politiques d'établissement ou de site.

⁵ La définition des acronymes marqués d'une * est disponible dans le glossaire p. 38

L'objectif final est d'améliorer la qualité globale des solutions proposées pour accompagner la gestion, la diffusion et l'archivage des données de recherche en SHS.

Méthodologie

L'enquête s'appuie sur une étude en deux volets conduits entre septembre 2021 et juillet 2022.

Le premier volet quantitatif a pris la forme d'une enquête en ligne réalisée avec le logiciel LimeSurvey, destinée à l'ensemble des chercheurs, enseignants-chercheurs, doctorants et personnels de la recherche des trois établissements. Le questionnaire disponible en annexe⁶ de ce rapport comportait au total 77 questions, dont 16 obligatoires et 61 conditionnées par le choix de réponses antérieures. L'hébergement des données a été assuré par l'université Lyon 3.

La phase d'élaboration et de diffusion de ce questionnaire, d'une durée de deux mois et demi, a bénéficié du renfort d'un conservateur stagiaire des bibliothèques de l'Ensib en stage au sein de la BU Lyon 3 de septembre à décembre 2021. Avant diffusion, le questionnaire a fait l'objet de plusieurs soumissions pour relecture auprès des délégués à la protection des données personnelles (DPD*) de Lyon 2 et Lyon 3 et de la référente RGPD* de la MSH, du Service des affaires juridiques, générales et des archives (SAJGA) et du Service général à la recherche (SGR) de Lyon 3. Une version bêta du questionnaire a également été visée par trois enseignants-chercheurs de Lyon 3, qui ont formulé des remarques, des questions ou des propositions de modifications.

L'enquête a été mise en ligne et diffusée par chacun des partenaires du 19 novembre au 10 décembre 2021 via ses réseaux de diffusion :

- Pour Lyon 2 : *via* son site web et les réseaux professionnels internes de l'université.
- Pour Lyon 3 : *via* les listes de diffusion institutionnelles.
- Pour la MSH : à l'ensemble des personnels des laboratoires associés, par mail.

L'enquête comprend 6 blocs de questions visant à cerner le profil des répondants et les pratiques de ces derniers au regard des grandes étapes du cycle de vie des données :

1. Profil des répondants (13 questions)
2. Caractérisation des données produites (24 questions)
3. Modalités de gestion des données produites (12 questions)
4. Modalités de partage et de diffusion des données (7 questions)
5. Documentation et description des données (7 questions)
6. Besoins d'accompagnement en matière de gestion des données (14 questions)

Parallèlement à la phase de dépouillement et d'analyse commune de l'enquête quantitative amorcée à compter du mois de février 2022, une étude qualitative complémentaire a été menée, sur la base d'entretiens en présentiel ou en distanciel avec différents participants à l'enquête ayant volontairement communiqué leurs coordonnées.

Cette deuxième étape a donné lieu à la mise au point d'un formulaire d'entretien permettant de questionner le rapport que les chercheurs entretiennent vis-à-vis des données qu'ils produisent éventuellement ou réutilisent (enjeu de propriété intellectuelle, difficultés rencontrées, etc.) ainsi que les pratiques des doctorants s'ils sont amenés à en encadrer. Les enregistrements sonores, lorsqu'ils ont eu lieu, ont toujours fait l'objet d'une demande de consentement auprès des participants concernés. Ce second volet de l'enquête a bénéficié d'un renfort de 6 mois d'une chargée de projet contractuelle en poste à Lyon 3 ayant un profil de chercheuse en sciences du langage et bénéficiant d'une expérience avérée dans la conduite d'entretiens qualitatifs directifs ou semi-directifs.

⁶ Annexe n° 1 p. 39

	Lyon 2	Lyon 3	MSH
Entretiens en présentiel	6	8	4
Entretiens en distanciel	3	2	6
Profil des répondants	8 enseignants-chercheurs 1 doctorant	7 enseignants-chercheurs 3 doctorants	8 enseignants-chercheurs 2 ingénieurs
Total	9	10	10

Les entretiens, d'une durée moyenne de 45 minutes, ont été menés à partir d'une trame disponible en annexe⁷.

Limites

L'équipe chargée de réaliser l'enquête et d'en analyser les résultats a évolué à plusieurs reprises au cours du déroulement du projet, notamment pendant la phase d'exploitation des réponses au questionnaire. Certaines formulations de questions et des options de réponses du questionnaire ont pu par ailleurs conduire à des ambiguïtés.

La difficulté à définir un périmètre exact de la population cible à l'échelle des trois établissements constitue un biais sur la représentativité. Les différents établissements n'ont pas tous utilisé les mêmes moyens de diffusion du questionnaire, les populations cibles n'ont donc pas été touchées de manière équivalente.

Par ailleurs, le choix de réponse proposé pour identifier la discipline de rattachement des enseignants-chercheurs, fondé essentiellement sur la section du Conseil national des universités (CNU) et son équivalent CNRS a parfois donné lieu à des réponses difficilement exploitables (plusieurs sections cochées ou absence de réponse).

L'enquête permet néanmoins de dégager des tendances sur le positionnement des chercheurs en SHS sur l'enjeu des données de recherche, le type de données produites et les besoins d'accompagnement.

⁷ Annexe n° 2, p. 60

Choix de présentation des résultats

Le questionnaire a recueilli 264 réponses complètes. L'équipe a fait le choix de ne retenir que les résultats les plus pertinents et exploitables, aboutissant à la production de 28 graphiques reprenant le nombre de répondants (exprimé par n) ou de réponses et couvrant de manière équilibrée les six rubriques du questionnaire.

Profils des répondants

Les répondants issus des universités de Lyon 2 et Lyon 3 représentent 81 % des répondants (**fig. 1**), en cohérence avec les effectifs de ces établissements. Avec 181 réponses, soit 70 % du total des répondants, les enseignants-chercheurs constituent la catégorie la mieux représentée (**fig. 2**), en cohérence avec la cible de cette enquête qui leur était adressée en priorité.

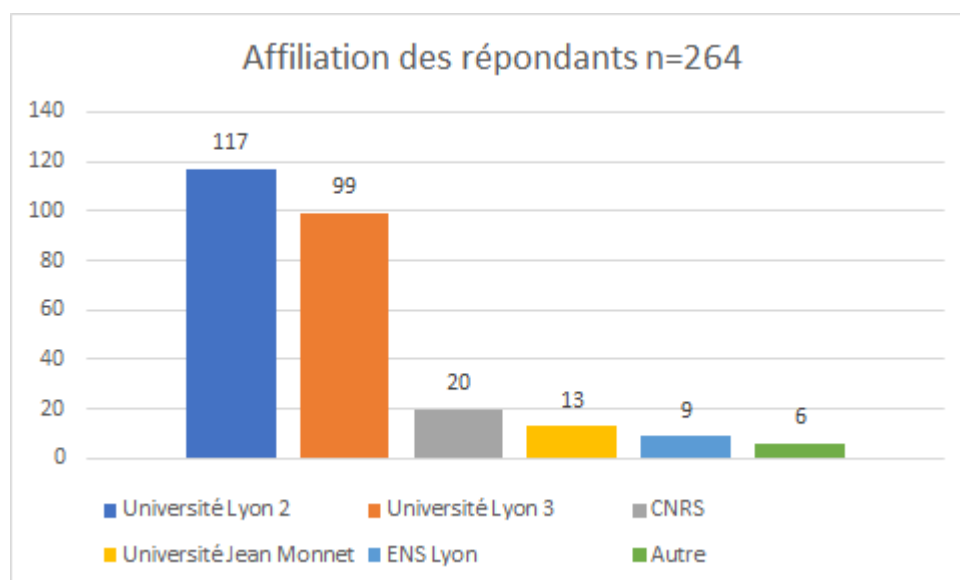


Fig. 1

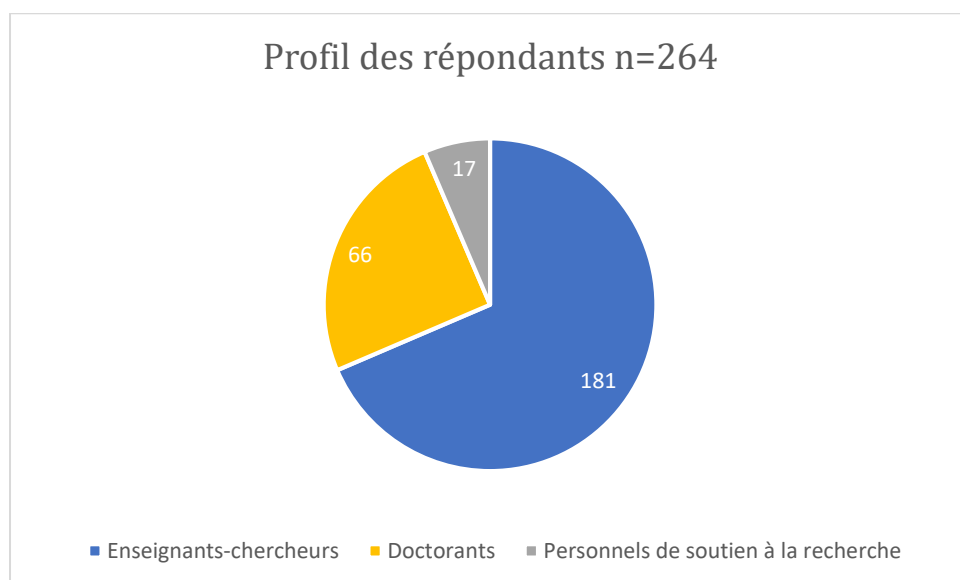


Fig.2

La répartition disciplinaire des participants révèle de forts contrastes (**fig. 3**). La communauté des historiens et archéologues arrive en tête, avec 34 répondants, alors que les sciences politiques ferment la marche. La psychologie, discipline uniquement représentée à Lyon 2, occupe le quatrième rang, juste devant les sciences de gestion et du droit, où 22 des 24 répondants sont issus de Lyon 3. Certaines disciplines sont essentiellement représentées par un laboratoire. C'est le cas en géographie, dont la totalité des réponses émane de l'UMR Environnement, ville, société, principal laboratoire du site à se consacrer à cette discipline. En sciences de gestion, 16 répondants sur 26, soit 60 %, proviennent du laboratoire Magellan (**fig. 4**). Les réponses issues des chercheurs en linguistique sont quant à elles réparties de manière équitable entre trois laboratoires (DDL, ICAR, et CEL).

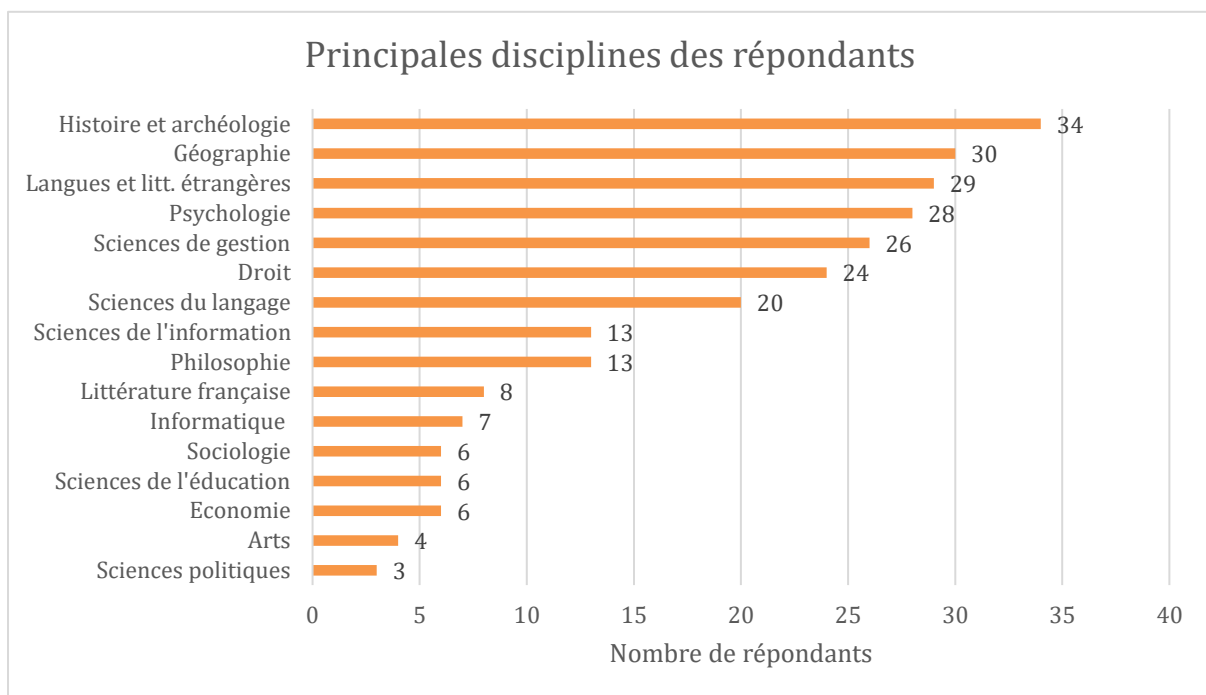


Fig. 3

Affiliation des répondants par unité de recherche

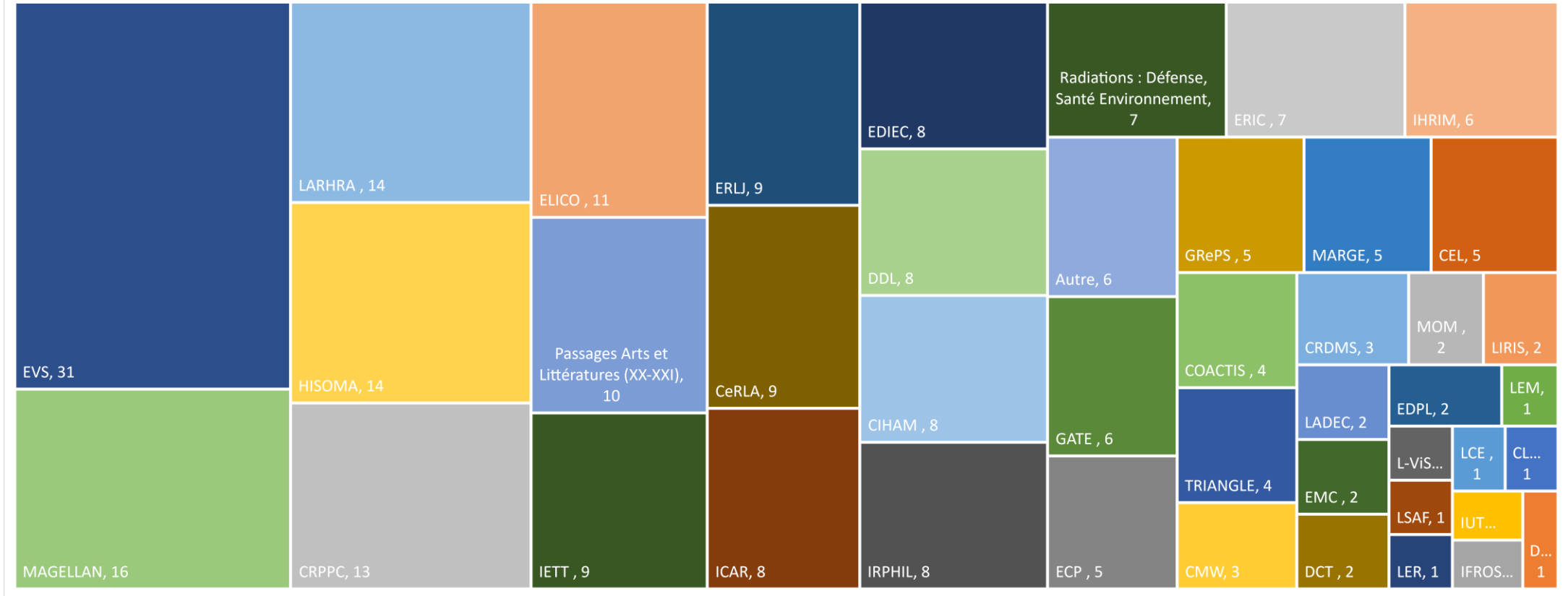


Fig. 4 (voir annexe n° 4 p. 62 pour les acronymes)

Caractérisation des données produites et stockées

La connaissance des types de données et leurs caractéristiques techniques (volume, format, stockage, etc.) permettent d'identifier les besoins en termes d'infrastructure et d'accompagnement.

Caractéristiques générales

Au sein de l'éventail d'options proposées, les résultats indiquent que les chercheurs produisent plusieurs types de données puisque cette question a suscité 456 réponses. Seul un quart des réponses proviennent de répondants ne produisant qu'un seul type de données.

Trois types de données dominent largement (**fig. 5**), indépendamment des questions de disciplines :

- **Les données canoniques ou de référence** collectées : pour les répondants, elles correspondent surtout à des corpus de texte, des bases de données, des recueils statistiques, des catalogues, des cartes et plans.
- **Les données d'observation** proviennent en majorité d'entretiens, questionnaires et remontées de terrain.
- **Les données compilées ou dérivées** découlent majoritairement de tableaux de données et statistiques.

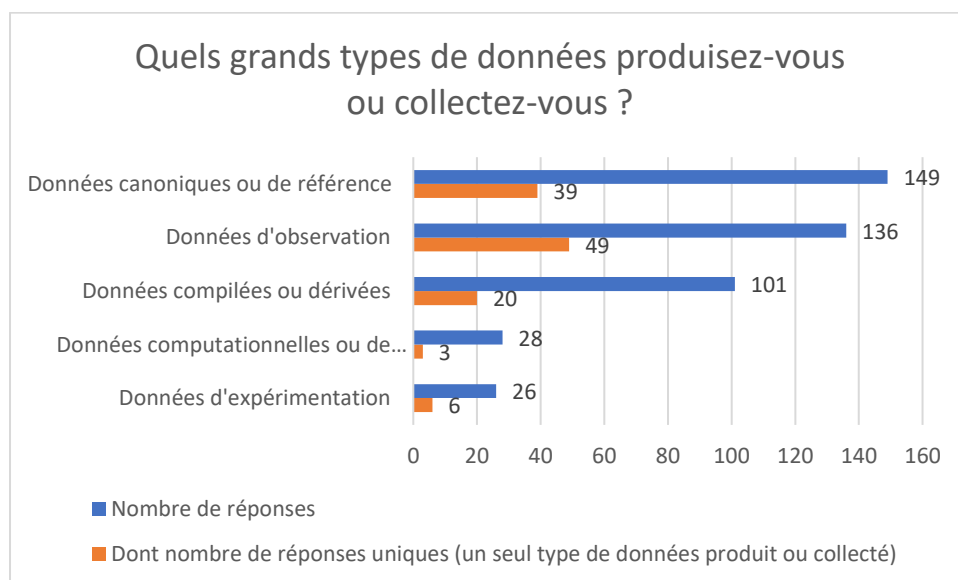


Fig. 5

Une analyse plus fine permet néanmoins de dégager des spécificités disciplinaires (**fig. 5.1**) : la psychologie, la géographie et les sciences de gestion semblent être les plus grandes pourvoyeuses de données d'observation. En psychologie, ce type de données est prépondérant puisqu'il est cité 23 fois, représentant 62 % des réponses des chercheurs de cette discipline. Parmi les exemples fournis par les répondants issus de la psychologie et des sciences de gestion, les résultats émanant de questionnaires, d'entretiens et de remontées de terrain, sont mentionnés en priorité devant les séquences vidéos, les photographies et les données provenant des réseaux sociaux. Au sein de l'éventail des données d'observation, les géographes se distinguent par l'utilisation d'images satellites, de données de géolocalisation, de photographies, mais aussi de carnets de terrain ou de données plus techniques récoltées *in situ* (sédiments, charbons, macro-restes végétaux...).

Les données canoniques ou de référence sont quant à elles citées de manière prépondérante par les chercheurs en langues (22 réponses sur 38 soit 57 %), qui s'appuient essentiellement sur des corpus de textes.

En histoire et archéologie, les données canoniques sont mentionnées quasiment à parts égales avec les données compilées ou dérivées. Les premières correspondent essentiellement à des corpus de texte et dans une moindre mesure à des cartes. Les secondes reposent essentiellement sur des bases de données.

Les données computationnelles et les données d'expérimentation, caractéristiques des sciences dures, sont logiquement les deux catégories les moins représentées dans cette enquête. Néanmoins, l'analyse des réponses révèle la présence de données issues d'expériences en sciences du langage, en géographie, en psychologie ainsi qu'en gestion. Pour cette dernière discipline, elles renvoient exclusivement à des mises en situation. La psychologie et les sciences du langage citent les mises en situation, aux côtés des tests et des modélisations. La géographie s'appuie sur ces deux types de données expérimentales, mentionnant également des analyses de sédiment et de datation.

Les données computationnelles (simulations, calculs) apparaissent quant à elle dans la plupart des disciplines couvertes par l'enquête, mais de manière très marginale.

Des disciplines comme le droit et la philosophie, pour lesquelles l'identification des données est source de questionnements (voir infra partie IV-enquête qualitative), s'appuieraient surtout sur des données canoniques et compilées. Les réponses des chercheurs traduisent une confusion entre les données et les publications. Parmi les exemples de données citées dans le champ de réponse libre, les juristes mentionnent par exemple les « articles », « chroniques », « notes d'arrêts », « commentaires de jurisprudence et de lois ». Un chercheur en philosophie cite les « publications scientifiques ». Ces réponses traduisent le peu de réalité tangible de la notion de données dans ces disciplines, où la démonstration repose sur d'autres prémisses (sources, principes fondamentaux, règles de droit, etc.) qui ne correspondent pas à proprement parler à des données de recherche.

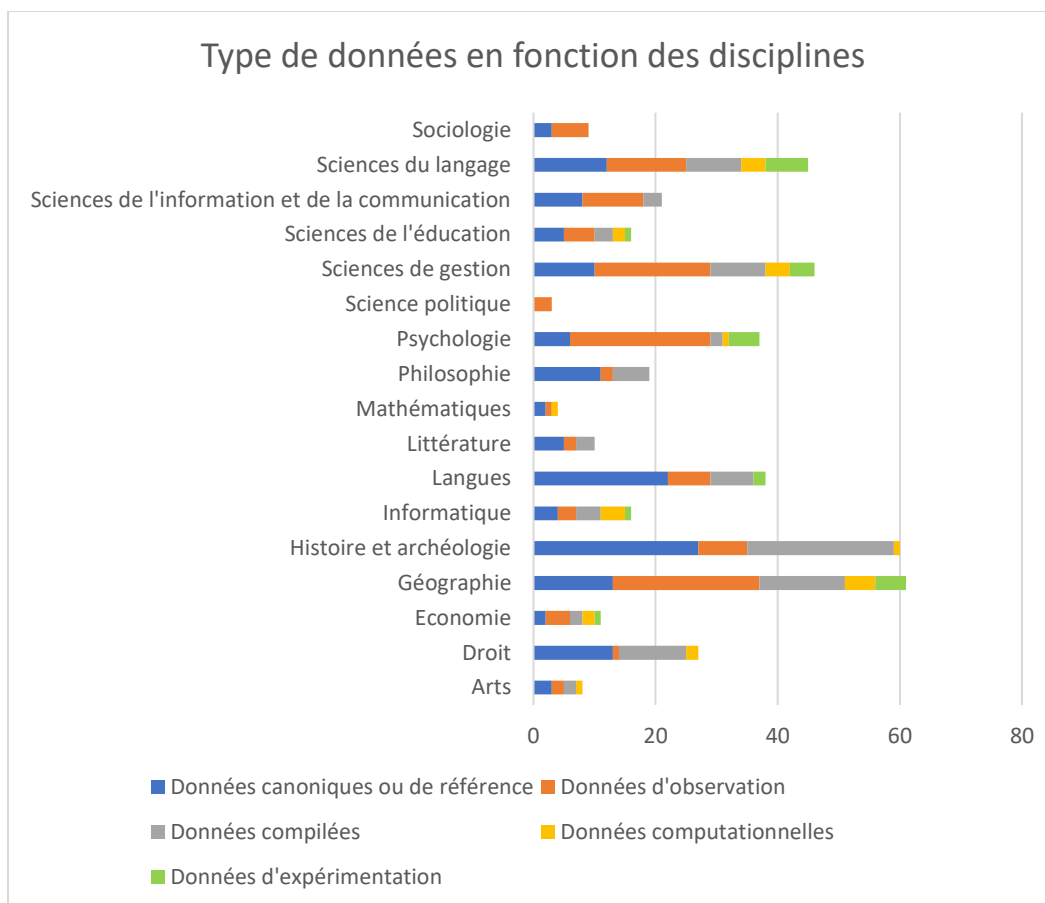


Fig 5.1

Formats informatiques

Les formats les plus fréquemment cités pour les données produites correspondent à des documents textes (doc, txt, pdf, etc.), des tableurs et présentations (**fig. 6**) qui cumulent à eux seuls 60 % des réponses. Le format image arrive en cinquième position. Certains formats sont caractéristiques d'une discipline : les deux tiers des réponses relatives aux formats de dessins vectoriels proviennent de répondants affiliés à des laboratoires en géographie et archéologie. Pour les données audio et vidéo, les chercheurs utilisent majoritairement des formats fermés (MP3, MPEG, AVI). En matière de dessin vectoriel, le résultat est plus contrasté, le format propriétaire AI étant cité en deuxième position derrière le format ouvert SVG.

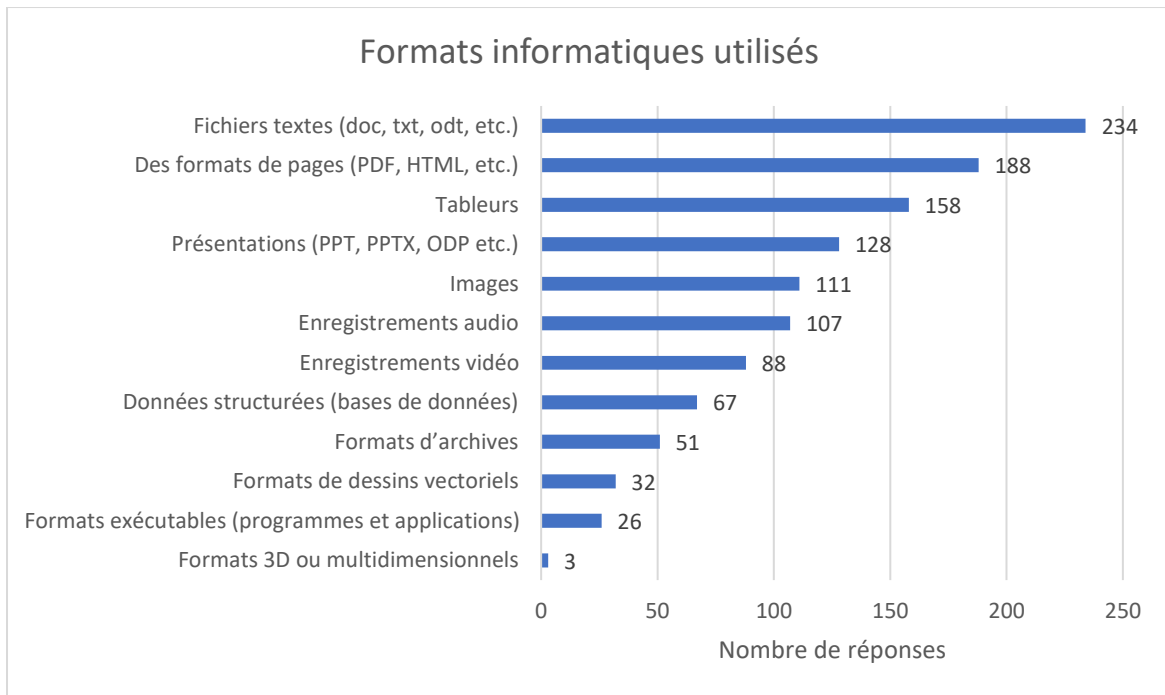


Fig. 6

Volumétrie et stockage

La plupart des répondants produisent moins de 100 Go de données, ce qui représente un volume soutenable (**fig. 7**). Les volumes les plus importants de données, supérieurs à 1 To, concernent 31 répondants sur les 253 ayant répondu à la question, soit 12 %. Un quart environ se situent dans une zone mitoyenne supérieure à 100 Go mais inférieure à 1 To. Les besoins les plus importants de stockage, lorsqu'ils existent, sont surtout exprimés en histoire ou archéologie et en géographie et dans une moindre mesure en langues, psychologie, sciences du langage et littérature (**fig. 7.1**).

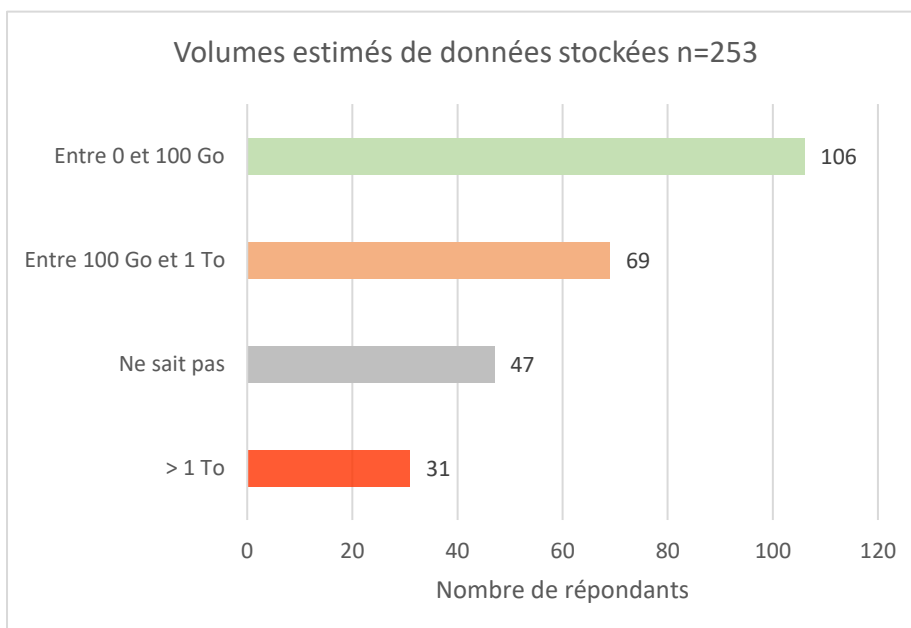


Fig.7

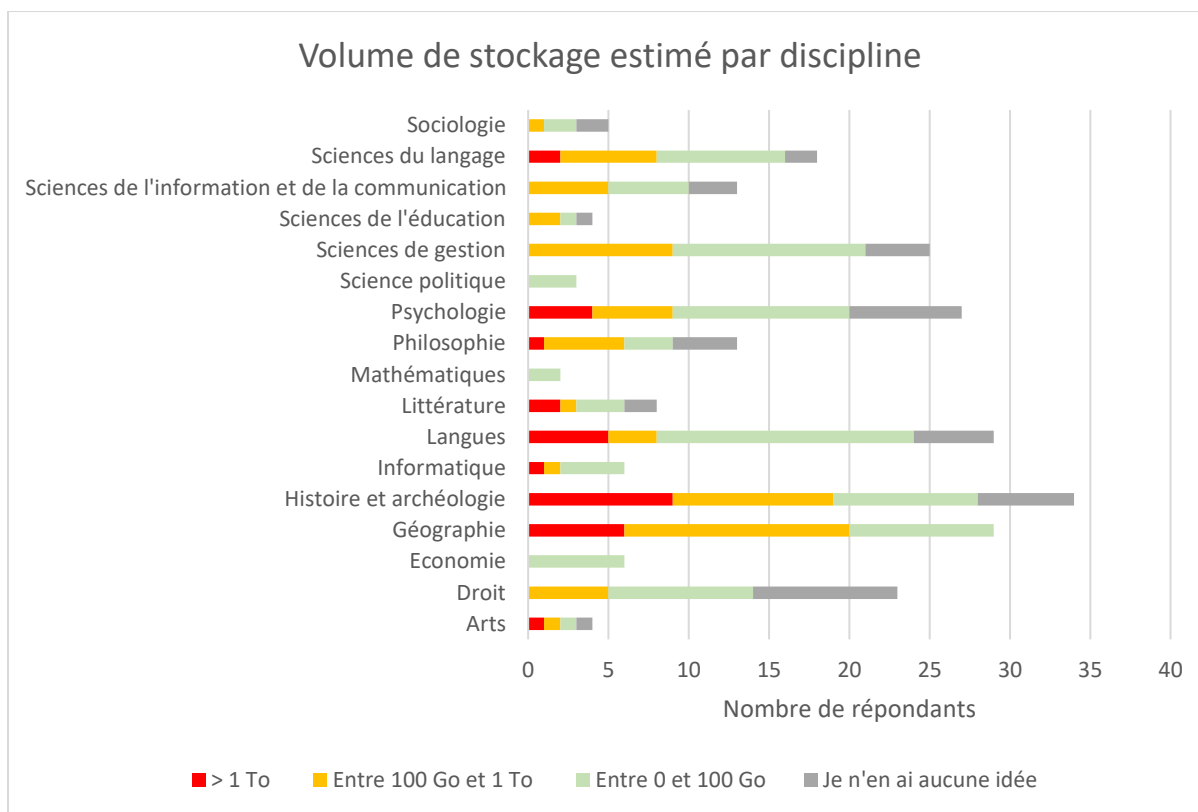


Fig. 7.1

Parmi les différents supports de stockage, le disque dur externe personnel et l'ordinateur personnel (qui devance le recours à l'ordinateur professionnel) sont les plus plébiscités (**fig. 8**). Ce sont des solutions plutôt fragiles qui exposent le chercheur à des pertes de données en cas de défaillance du matériel. Les infrastructures publiques sont quant à elles délaissées : sur les 792 réponses fournies, le stockage sur le serveur du laboratoire revient seulement 32 fois, soit 8 fois moins souvent que les deux solutions précitées. L'usage de clouds institutionnels n'est pas non plus intégré dans les pratiques courantes des chercheurs, cette solution réunissant seulement 6 % de réponses. On observe un décalage entre l'offre institutionnelle de stockage⁸ et les pratiques des chercheurs, ainsi qu'une quasi-absence de recours parallèle aux grandes infrastructures de recherche telle qu'Huma-Num*.

⁸ Plus d'informations sur ce point : voir annexe n° 4 p. 63

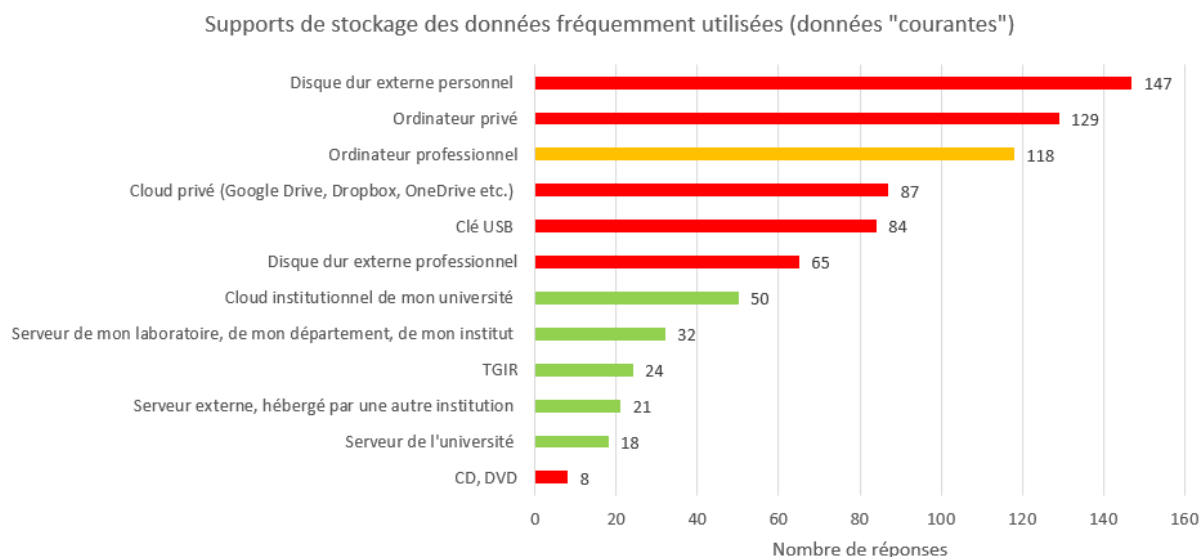


Fig.8

De manière logique, l'absence de recours aux solutions institutionnelles se reflète dans le choix des outils retenus pour partager les données avec les collaborateurs extérieurs dans le cadre d'un projet de recherche. La transmission par mail, solution largement privilégiée récolte 168 réponses (**fig. 9**). En deuxième et troisième position sont cités des solutions privées de transferts adaptés aux fichiers volumineux et les espaces de stockage privés comme Google Drive ou DropBox. Les outils institutionnels nationaux et l'offre propre à chaque établissement ferment la marche, même si des nuances sont à noter. La solution libre OwnCloud proposée par Lyon 3 récolte plus de 50 % du total des réponses relatives à l'offre interne d'espaces de stockage des différents établissements représentés. 17 des répondants de Lyon 3 disent l'utiliser à des fins de partage avec des collaborateurs externes, contre 9 des répondants de Lyon 2, où un outil similaire, baptisé BUL, a été mis à disposition. Inversement, 15 répondants issus du CNRS sur 20 disent utiliser les outils d'envergure nationale comme Renater ou Share-Doc d'Huma-Num, ce qui représente un ratio bien supérieur en comparaison avec les autres établissements.

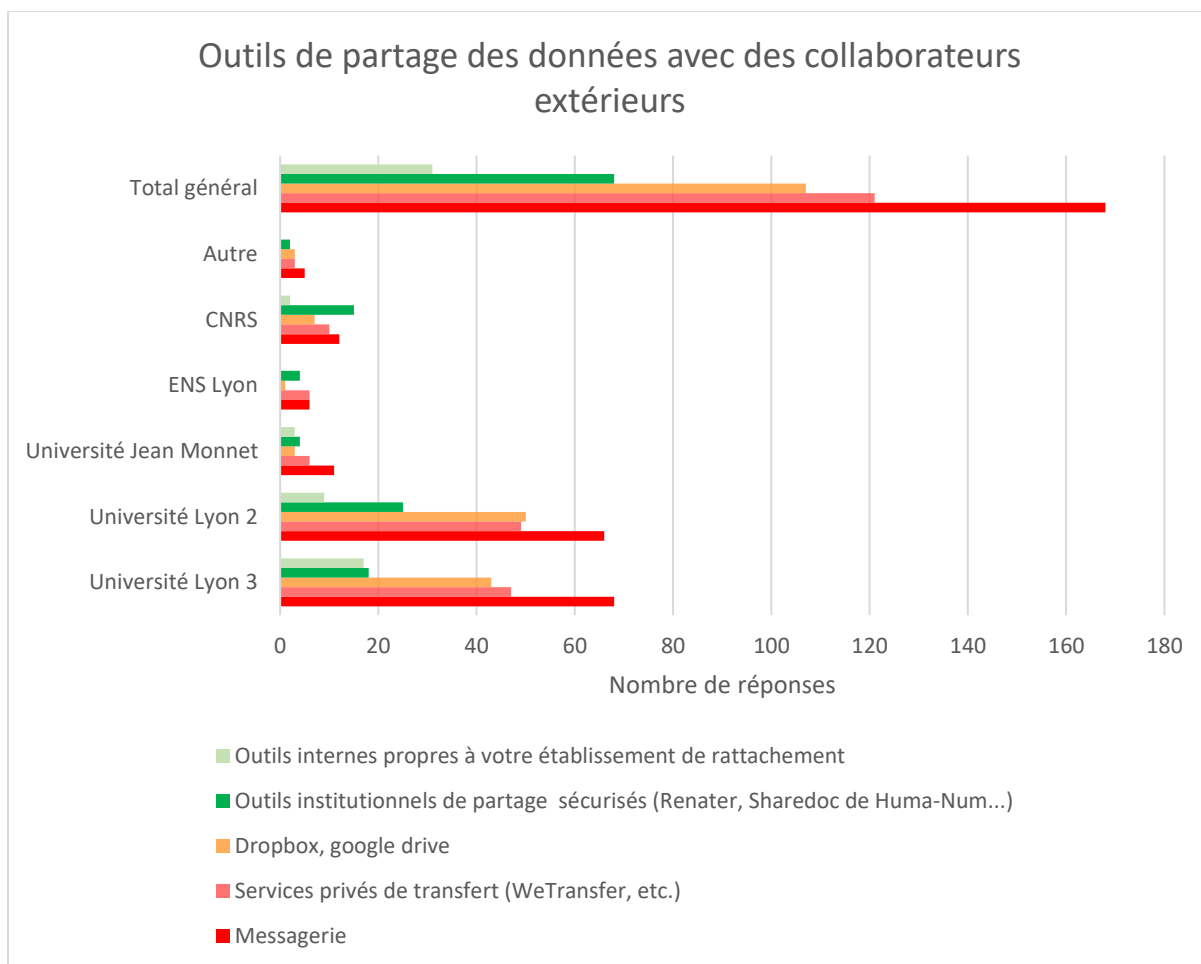


Fig. 9.

Pratiques en matière de gestion des données

Le plan de gestion de données : un exercice encore peu répandu

Seuls 40 répondants, soit 15 % de l'effectif total, se sont déjà livrés à la rédaction d'un plan de gestion de données (**fig. 10**). Un chiffre assez faible qui correspond néanmoins aux proportions affichées dans des enquêtes comparables⁹.

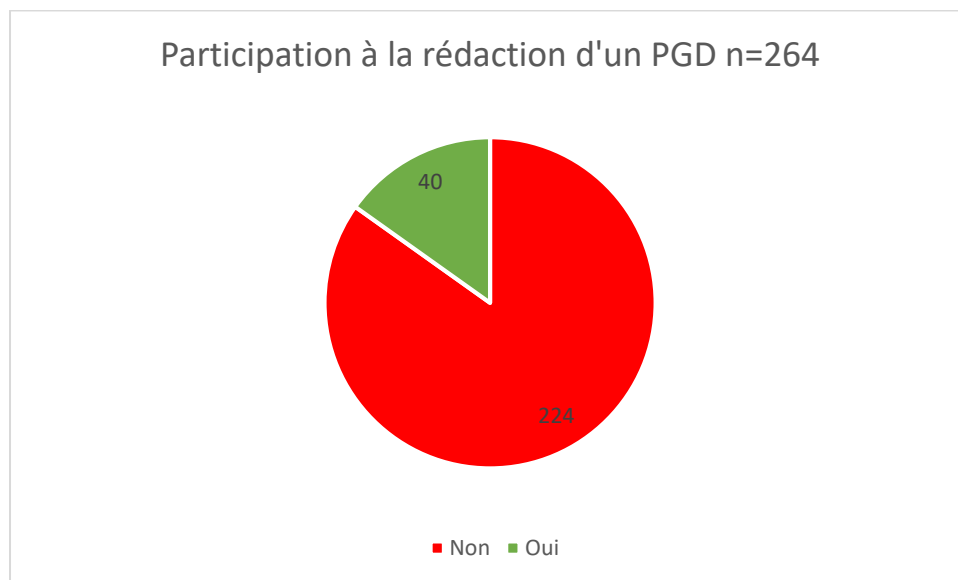


Fig.10

24 des 60 répondants concernés (soit 60 %) ont pu bénéficier d'un accompagnement lors de la préparation du document. L'appui provient principalement des SCD, à égalité avec l'entourage des chercheurs (recours aux collègues ou équipe projet), mais aussi des personnels de laboratoire dédiés, cités en troisième lieu.

Dans la majorité des cas, la rédaction d'un PGD découle d'un projet financé par l'ANR. Néanmoins, de nombreux autres cas de figure sont cités par les répondants (dépôt d'un projet auprès du fonds national suisse pour la recherche, projets européens H2020 et ERC, convention de recherche avec une entreprise privée, etc.).

Seule une infime minorité des répondants (20 sur 261 soit 7 %) affirment disposer de personnels dédiés à la gestion des données au sein du laboratoire (**fig. 11**). Le reste se répartit à parts égales entre les indécis et ceux estimant qu'ils n'en ont pas. Les participants faisant mention d'un appui en interne sont issus de 10 structures différentes, dont 8 sont des UMR (**fig. 11.1**) parmi elles, EVS, la MOM, CIHAM ou encore ICAR, au sein duquel il existe une « cellule qui s'occupe de la conservation centralisée, de l'anonymisation (en cas de données sensibles) et de la pérennisation des données », précise l'un des répondants.

⁹ Cellule Data. *Rapport sur l'enquête sur les usages et les besoins pour la gestion des données de la recherche sur le site de l'université Grenoble Alpes*, 2022, p. 50. Disponible sur : https://scienceouverte.univ-grenoble-alpes.fr/wp-content/uploads/2022/05/enquete_donnees_UGA.pdf

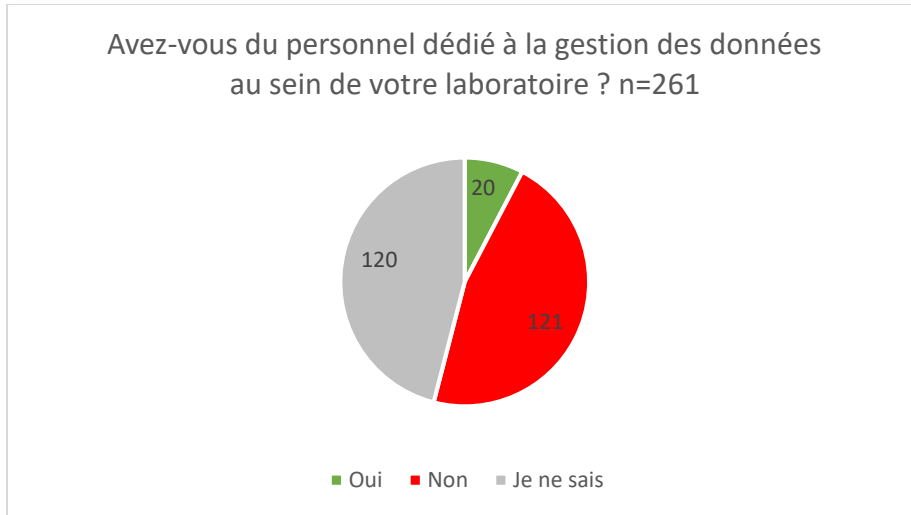


Fig. 11

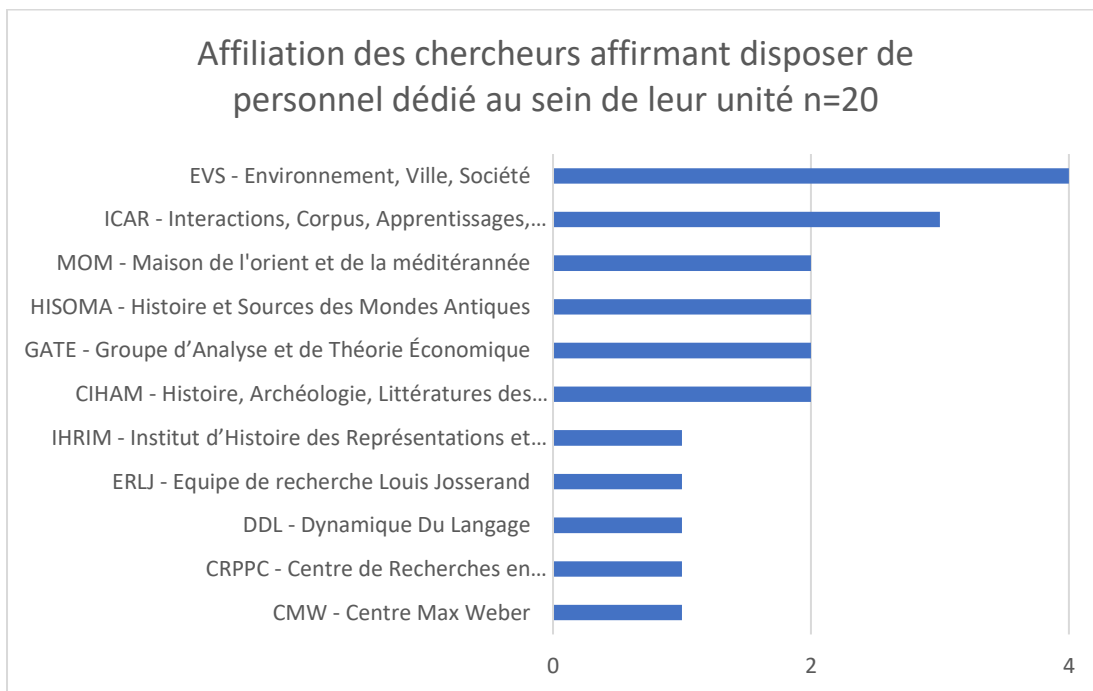


Fig. 11.1

Données sensibles ou personnelles : des doutes sur la conduite à mener

Les réponses obtenues indiquent une méconnaissance globale du délégué à la protection des données (DPD), dont les fonctions sont ignorées par 157 répondants, soit près de 60 % (**fig. 12**).

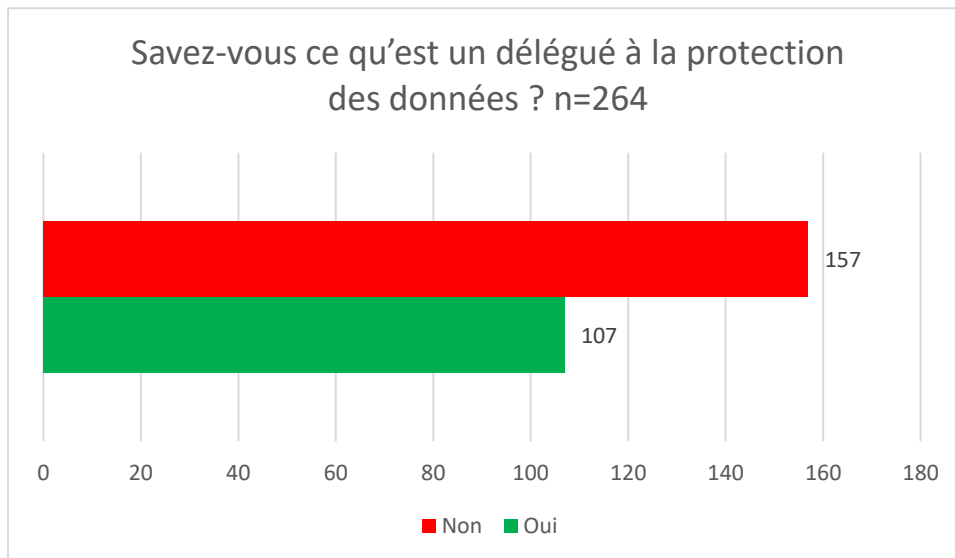


Fig. 12

Cette méconnaissance peut surprendre dans la mesure où la production d'enquêtes et la collecte de données personnelles font partie des activités courantes des chercheurs en SHS, comme en témoignent les chiffres précités, plaçant les données d'observation en deuxième position des données les plus fréquemment produites ou collectées. Pourtant, seuls 79 répondants déclarent être amenés à traiter des données à caractère personnel ou dites sensibles (**fig. 13**). Au sein de ce panel, la moitié indique ne pas avoir fait de déclaration auprès du DPD de leur établissement (**fig. 14**).

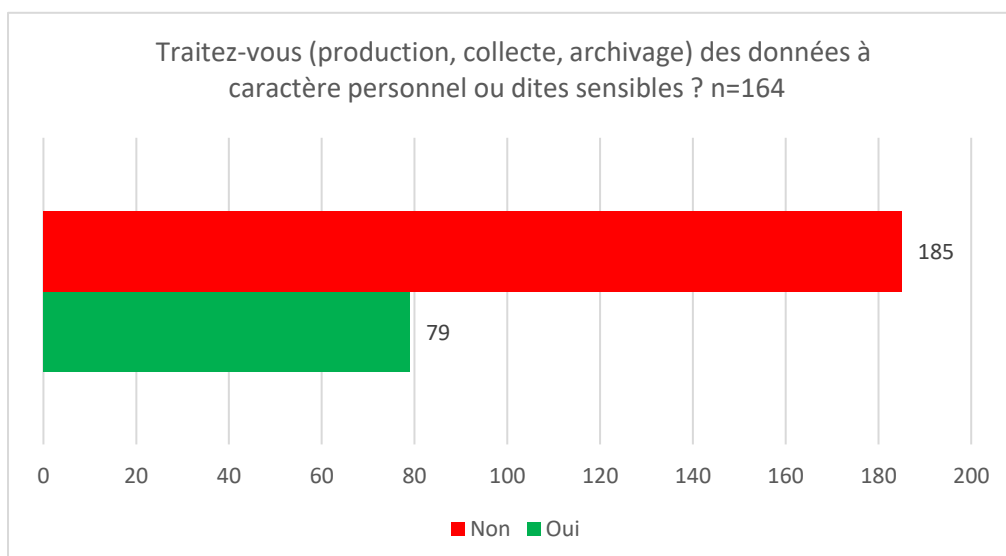


Fig. 13

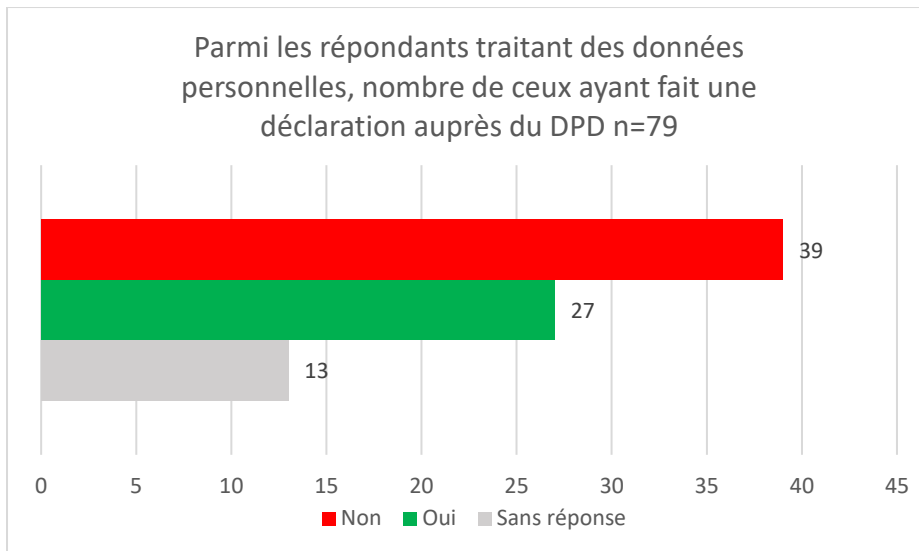


Fig. 14

Sur l'ensemble des répondants, 55 considèrent qu'ils ne sont pas assez protégés contre les failles d'accès sur les données sensibles ou personnelles, 132 expriment leur incertitude sur la protection de leurs données (**fig. 15**), ce qui porte à 187 le nombre de répondants jugeant le niveau de protection des données insuffisant, soit 70 % du total des répondants.

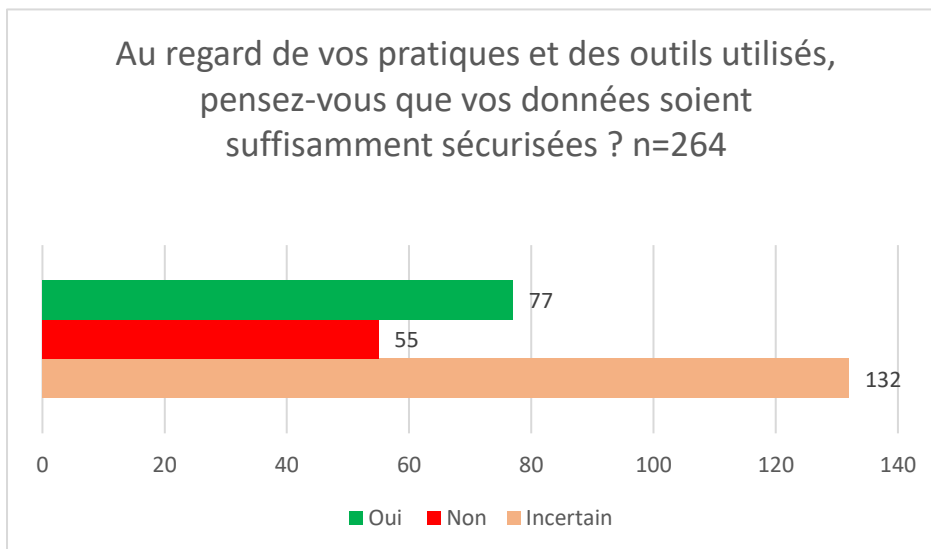


Fig. 15

Recours aux métadonnées : un usage équivoque

Seuls 51 chercheurs sur les 213 ayant répondu à cette question (soit moins d'un quart) ont connaissance de standards de métadonnées couramment utilisés dans leur domaine de recherche (fig. 16). En revanche, le fait d'ajouter des mots clés à leurs fichiers pour faciliter le classement, la traçabilité ou la découvrabilité de leurs données (projet de recherche, auteur, thématiques, etc.) est une pratique plus courante. Elle concerne 98 chercheurs sur les 224 ayant répondu, soit 43 % (fig. 17). Ce résultat révèle que les chercheurs saisissent individuellement l'intérêt des métadonnées sans que cette pratique ne soit nécessairement structurée à l'échelle de leur discipline.

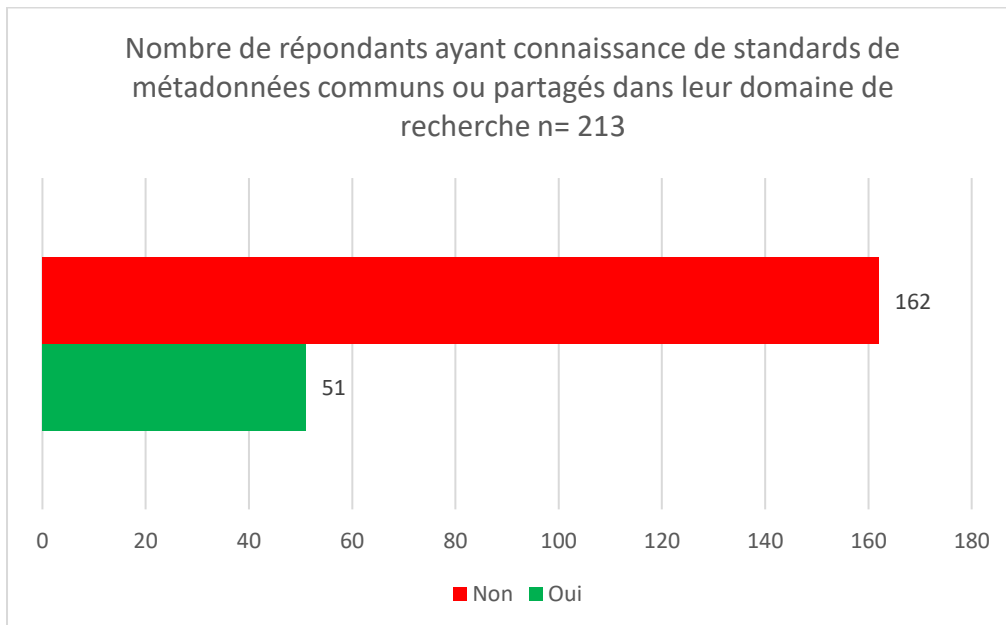


Fig. 16

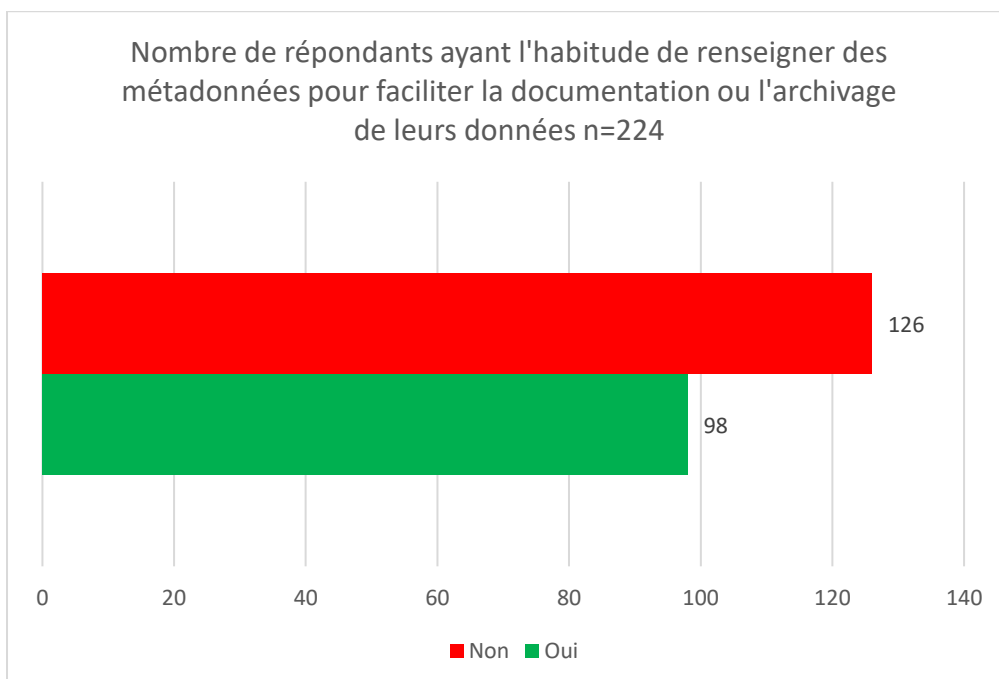


Fig. 17

Diffusion des données : un intérêt non identifié

Le partage en ligne de jeux de données n'est pas du tout intégré dans la routine du chercheur. Les trois quarts des répondants ne s'y sont jamais livrés et seuls 6,5 % disent publier leurs données en accès ouvert régulièrement (fig. 18). Les chercheurs en informatique et en sciences du langage semblent être un peu plus enclins à ouvrir leurs données, ces disciplines affichant un peu plus de répondants dans ce cas (fig. 18.1). Dans les autres disciplines, le non-partage est la règle dominante, voire exclusive.

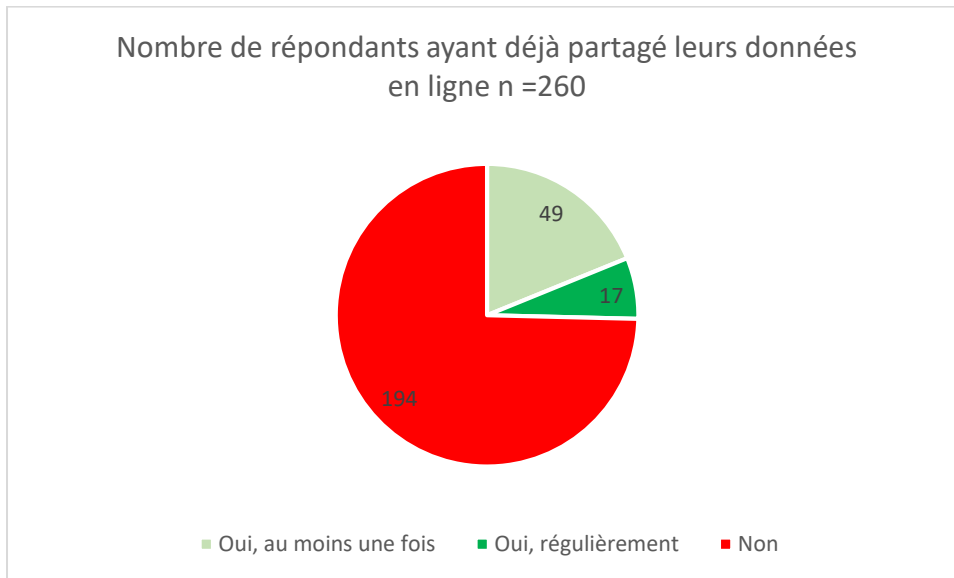


Fig. 18

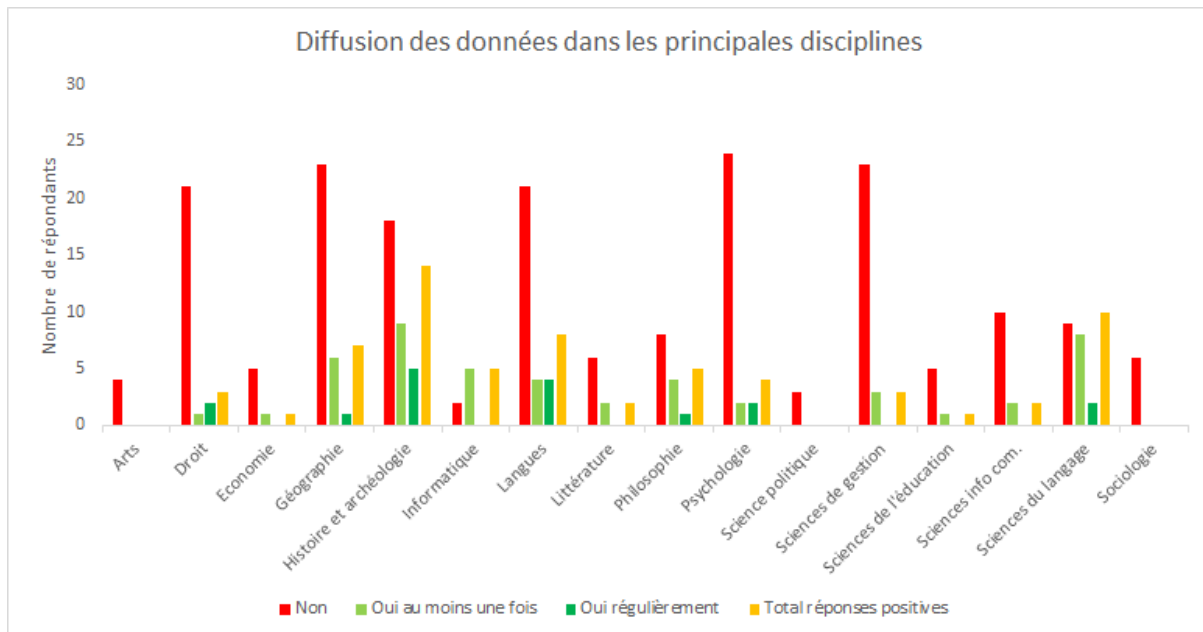


Fig. 18.1

Lorsque le partage a lieu, le motif premier invoqué est la valorisation des résultats, loin devant les obligations qui seraient formulées par un éditeur ou un financeur. Il semble donc que les chercheurs identifient d’eux-mêmes la valeur scientifique et l’utilité associées à certains jeux de données, qu’ils déposent la plupart du temps, indiquent-ils, dans une archive ouverte de type HAL (31 réponses) ou sur le site du projet de recherche (23 réponses). Si une plateforme comme HAL n’a pas été conçue à l’origine comme réceptacle des données de recherche, l’outil prévoit en revanche une garantie de préservation que n’offre pas le site web créé pour un projet ad hoc. Le recours aux entrepôts de données reste plus minoritaire (11 réponses), à parts égales avec les réseaux sociaux de chercheurs comme ResearchGate. Enfin, le chercheur a tendance à partager ses données sans se soucier des conditions de réutilisation de ces dernières. Sur les 68 répondants ayant déjà publié leurs données, 46 ignorent la nature de la licence régissant ce partage.

L’effort de valorisation émanant de certains chercheurs est par ailleurs éclipsé par les nombreuses raisons justifiant la non-communication des données. D’une part, ce partage aurait peu d’intérêt, à en croire les répondants, qui citent ce motif en tout premier lieu (**fig. 19**). Mais la rétention des données est aussi le fruit de doutes ou d’une méconnaissance de l’écosystème. Parmi les motifs invoqués, 107 réponses proviennent de chercheurs ne sachant pas s’ils ont le droit de déposer, ignorant que cela était possible ou n’y ayant pas pensé (**fig. 19**). Le manque de temps, cité en cinquième position, ne s’impose pas comme le frein le plus important à l’ouverture des données pour les répondants, ce qui fait contraste avec d’autres études menées sur le sujet¹⁰. Au final, loin d’être considérées comme des données publiques, les données de recherche restent des matériaux relativement confidentiels qui, s’ils sont partagés, le sont au sein du cercle restreint des membres de l’équipe projet ou collaborateurs proches (**fig. 20**).

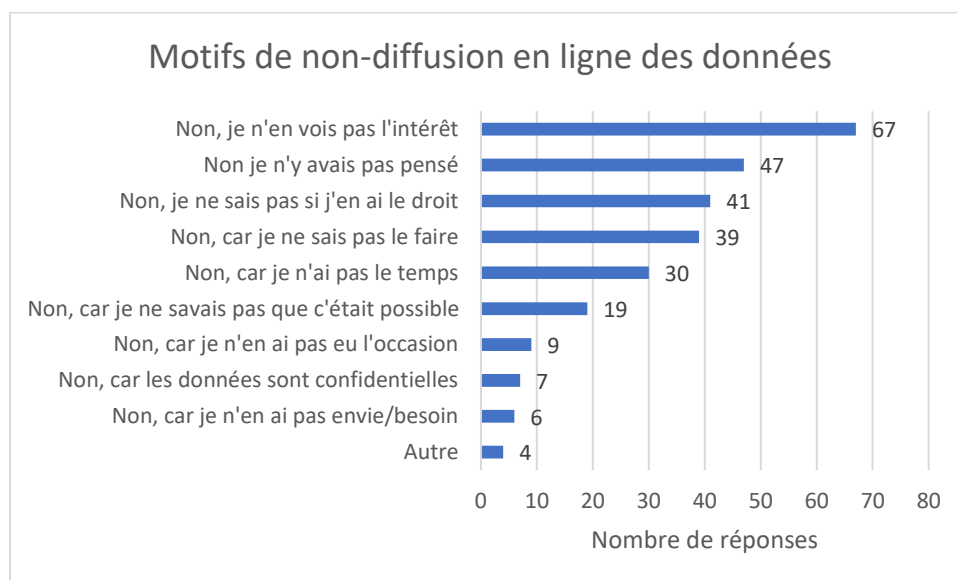


Fig. 19

¹⁰ Brenel, Mireille, et al. *Rapport d'analyse – Enquête : Les données de la recherche à l'université ParisSaclay, panorama et perspectives*. Université Paris-Saclay, 2022, p.48. Disponible sur : <https://hal-universite-paris-saclay.archives-ouvertes.fr/hal-03857804> et Dillaerts, Hans, et al. *Usages et pratiques en lien avec les données de recherche. Une enquête menée auprès des chercheurs de l'université Paul-Valéry Montpellier 3*. Université Paul-Valéry Montpellier 3, 2020, p.7. Disponible sur : <https://hal.archives-ouvertes.fr/halshs-02902710>

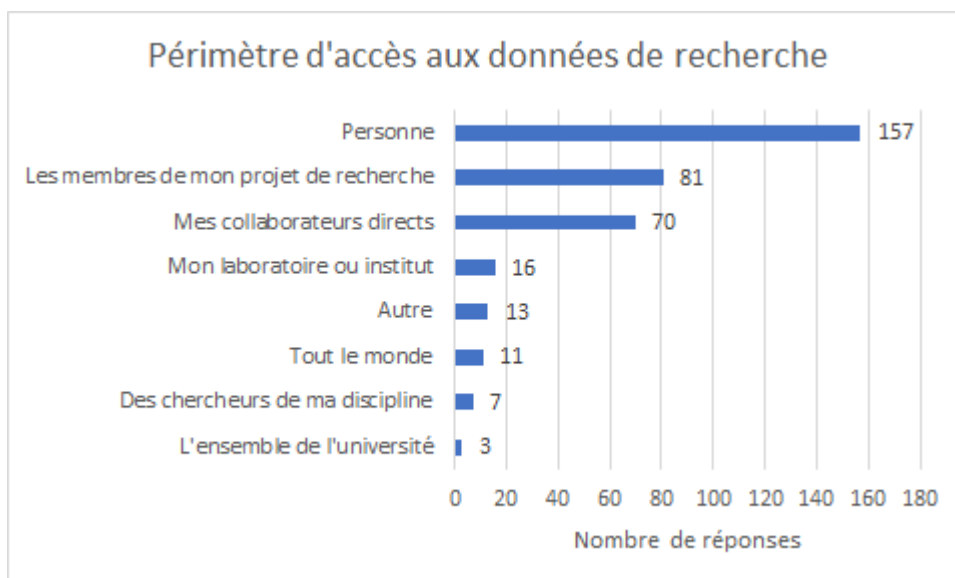


Fig. 20

Des pratiques d'archivage quasi inexistantes

L'enquête révèle la quasi-inexistence des pratiques d'archivage des données de la recherche parmi les répondants quel que soit leur statut, tout comme la méconnaissance des services d'archivage. En effet, 49 répondants sur 250, soit 76 %, ignorent l'existence d'un pôle archives au sein de leur établissement de rattachement (**fig. 21**) et 70 % indiquent n'avoir aucune politique en la matière au regard de leurs données de recherche dites « froides », qui ne sont pas consultées au quotidien mais peuvent faire l'objet de réutilisations ultérieures (**fig. 22**). Pour les 30 % restants, l'archivage ne comprend pas forcément de méthode de tri particulière puisque la conservation de l'ensemble des données est citée comme la pratique la plus courante. Les données susceptibles d'être réutilisées dans des travaux ultérieurs ou celles qui revêtent un intérêt historique ou patrimonial interviennent respectivement en deuxième et troisième position.

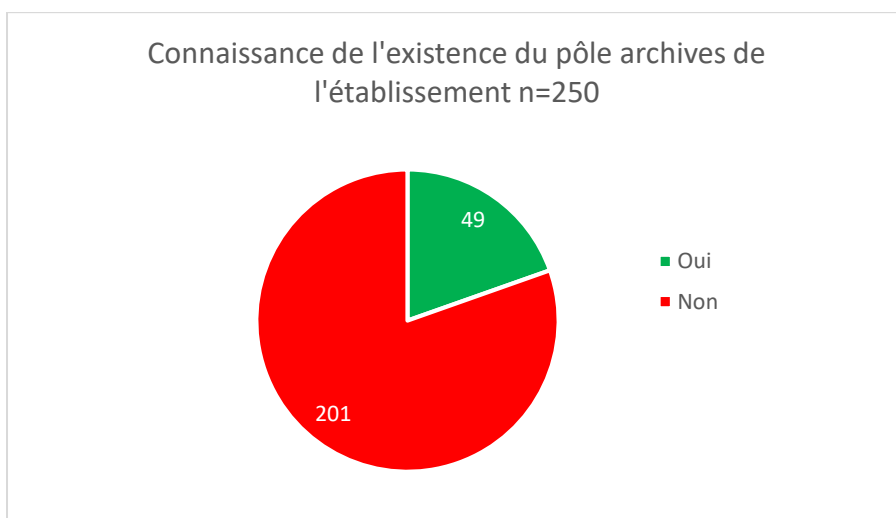


Fig. 21

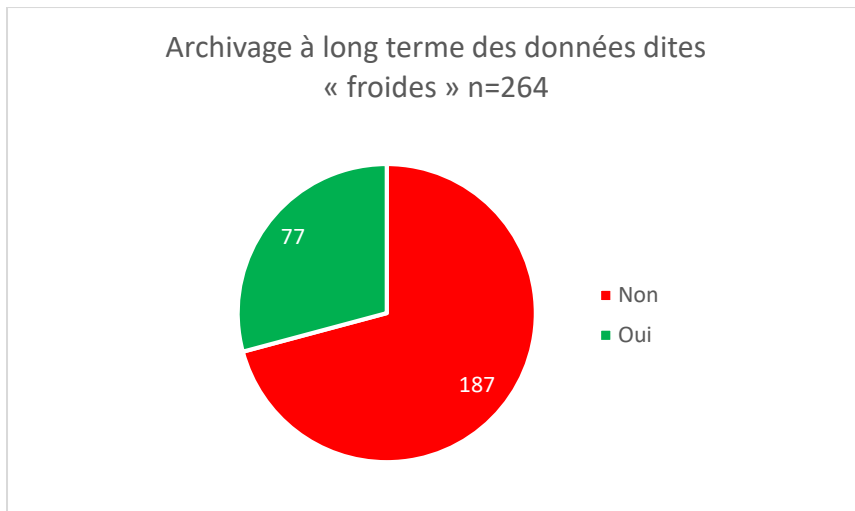


Fig. 22

Des besoins traduisant le positionnement inconfortable des SHS

Cette partie se propose de mettre en avant les besoins exprimés afin d'émettre une série de propositions pour accompagner au mieux ces mutations. Néanmoins, les attentes en matière de gestion des données sont encore floues pour certains.

La nécessité de recevoir des conseils d'ordre général ou de l'aide à la gestion informatique des données se hisse en tête des demandes formulées, suivies de l'aide à la pérennisation des données (**fig. 23**). La diffusion des données, encore peu répandue comme nous avons pu le voir, a néanmoins suscité 100 réponses de chercheurs souhaitant être accompagnés en la matière. L'aide juridique et la rédaction de PGD occupent respectivement les 5^e et 6^e rangs. L'aide à la citation des données produites par d'autres chercheurs ferme la marche.

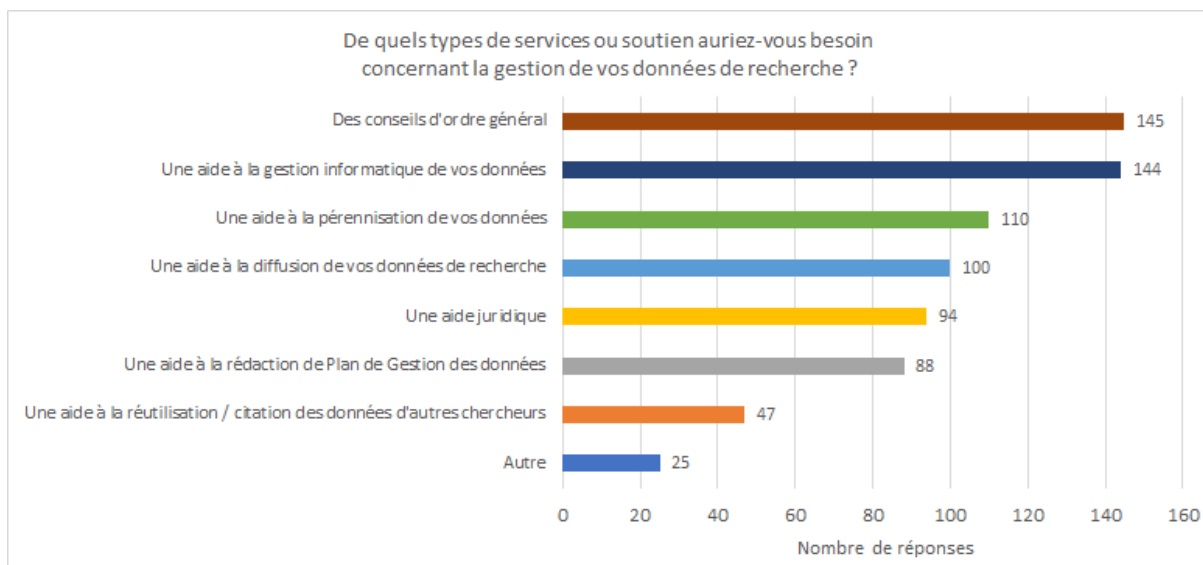


Fig. 23

Parmi les modalités d'accompagnement souhaitées, les ateliers de groupe représentent le format le moins plébiscité, puisqu'il est cité près de deux fois moins que le conseil ponctuel ou l'accès à une documentation en ligne (**fig. 24**). Selon les besoins exprimés, les chercheurs peuvent privilégier un canal plutôt qu'un autre : le recours à une assistance ponctuelle en cas de problème d'ordre juridique arrive par exemple largement en tête par rapport aux autres modalités proposées (documentation en ligne, formation, ateliers, etc.) Les réponses libres révèlent également que 15 répondants sur 25 estiment ne pas savoir de quelle aide ils auraient besoin ou ne pas avoir du tout besoin d'accompagnement.

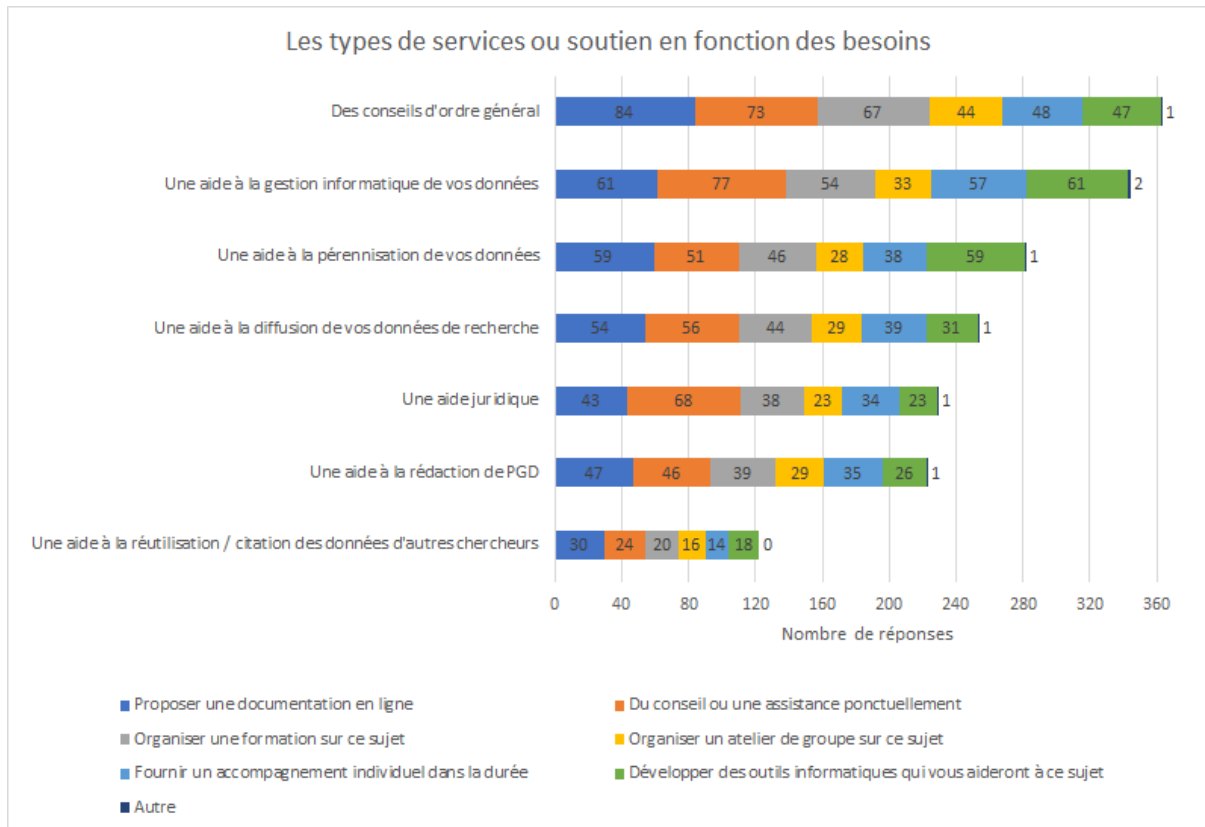


Fig. 24

Enquête qualitative : des résultats clivants

Les commentaires libres exprimés par les répondants au questionnaire, enrichis par près de 30 entretiens menés par les trois établissements, permettent de tirer plusieurs enseignements :

- Une partie des SHS reste étrangère aux politiques actuelles de gestion et d'ouverture des données de recherche, dont la raison d'être et l'application sont davantage perçues comme compatibles avec les sciences dures ;
- Une partie des chercheurs, confrontés à des enjeux éthiques ou de droit d'auteur, ne trouvent pas de solutions ou réponses à leurs problématiques, malgré la sollicitation des services compétents ;
- L'enjeu des données comme source de mutations voire d'opportunités dans la façon de conduire un projet de recherche émerge ponctuellement, mais de manière minoritaire. L'enjeu de la pérennité des données fait en revanche davantage consensus.

Des politiques de gestion et d'ouverture des données mal articulées avec certaines disciplines et pratiques en SHS

Un cadre normatif hérité des sciences exactes

Certains chercheurs pointent le décalage entre la nature de leurs pratiques et les attendus croissants sur la gestion des données issues de la recherche. « Les chercheurs en SHS ont parfois l'impression qu'ils n'ont pas de données », pointe une chercheuse en philosophie. La finalité des politiques actuelles, axées sur la documentation de données susceptibles d'être reproduites en vue d'une réutilisation future, repose sur des méthodes étrangères aux disciplines interrogées. Comme le souligne une chercheuse en droit :

« Il ne s'agit pas de reprendre, comme en sciences dures, les résultats ou travaux qui appartiennent au laboratoire. Ce n'est pas du tout le cas en droit. »

Un point de vue également exprimé par un chercheur en linguistique :

« Je pense que la question se pose différemment pour un linguiste comme moi qu'elle ne se poserait pour des biologistes, des physiciens, des personnes qui sont dans des projets très collectifs et où il y a une continuité. »

Un chercheur en sociologie/anthropologie refuse quant à lui d'utiliser le terme « résultats », lui préférant celui « d'analyse ». Il s'interroge également sur le sens même de déposer des données dans sa discipline et souligne qu'il n'a pas connaissance de revues formulant cette demande dans son domaine.

Interrogé sur ses pratiques, un chercheur en sciences économiques corrobore l'idée selon laquelle les politiques actuelles seraient davantage en cohérence avec les sciences dures :

« Je constate que mes collègues en sciences expérimentales sont de plus en plus dans une approche open science et doivent rendre les données plus ou moins disponibles pour pouvoir publier. Honnêtement, je ne sais pas trop comment ça marche. Si moi, je me propose d'aller récupérer des données ailleurs, je ne vois pas comment on y arrive. »

Une doctorante en littérature confirme le fossé entre les règles édictées et la réalité de certaines disciplines.

« Les données suscitent un non-intérêt au sein du labo. Je comprends que cela ne soit pas passionnant, c'est pas très sexy ! »

Dans cette discipline, l'approche est encore « artisanale », observe un chercheur en littérature, comme « déconnectée de ce qui se passe en humanités numériques ». Le potentiel issu de la réutilisation de données produites par des tiers ne serait pas non plus un argument suffisant, y compris en économie :

« Vous avez rarement l'opportunité de retourner dans vos anciennes données. Moi, je ne le vois pas. Peut-être pour faire des méta-analyses... En économie, je vois très exceptionnellement des gens qui vont rechercher des vieilles données et qui refont des analyses dessus. Donc l'intérêt, il n'est pas évident ».

Gestion et documentation des données : de nouvelles obligations jugées lourdes ou inadéquates

Alors que l'ouverture des données tend à devenir un principe par défaut, plusieurs personnes interrogées rappellent que le partage des données ne peut être systématique tant il dépend étroitement du projet lui-même et de ses finalités. Conséquence des évolutions du cadre normatif, la multiplication des services d'accompagnement à la rédaction de plans de gestion de données conduit à développer des services d'appui potentiellement déconnectés de la réalité des pratiques de recherche en SHS, alerte un chercheur en sociologie, avec le risque de services inadaptés ou construits en silos. Une solution, pour éviter cet écueil, consisterait selon lui à organiser des séminaires sur les données propres à sa discipline afin de favoriser le débat.

À l'heure actuelle, un sentiment de rejet naît parfois des nouvelles injonctions adressées aux chercheurs :

« Il y a un problème d'inadéquation entre le plan de gestion de données et la réalité concrète des disciplines, réagit un chercheur en philosophie, ce qui pose la question de l'utilité de ce travail et d'y consacrer autant de temps. Le PGD donne une impression de complication extrême pour dire des choses banales et communes qui ne méritent pas tant de réflexion que cela. »

À l'issue de l'enquête à laquelle il a participé, un chercheur en gestion tranche : « Je ne vois pas trop l'utilité de ce questionnaire. D'autres actions sont prioritaires pour nous aider. »

Même dans l'hypothèse favorable où l'intérêt est identifié et qu'il n'y a pas d'obstacle d'ordre juridique ni de désaccords sur l'ouverture entre les parties prenantes du projet, la diffusion des données est compromise par l'effort colossal de documentation. C'est le constat dressé par une chercheuse en histoire, qui se heurte à cette difficulté, en plus d'être isolée au sein de son laboratoire, où la documentation en vue du partage n'est pas entrée dans les pratiques :

« Pour mes travaux, depuis la thèse j'ai beaucoup de fichiers et de tableurs sur mon ordi. Ce ne sont pas des informations sensibles, mais des personnes décédées depuis longtemps. Ces tableurs ne sont jamais parfaits et je n'ai pas envie de les mettre comme ça associés à son nom sur internet en disant : « Allez-y servez-vous ! » Par contre je trouve ça très frustrant d'avoir tous ces fichiers qui ne peuvent pas être réutilisés par d'autres. Il faudrait les associer à chaque fois à un guide pour expliquer comment ça a été conçu - avec les faiblesses à chaque fois - pour qu'un autre chercheur ou chercheuse puisse repartir de cette base, plutôt que de refaire ce qui a déjà été fait. [...] Mais il y a beaucoup de collègues qui estiment que cela relève un peu de leur cuisine interne et qu'ils ont pas à rendre compte - c'est un peu fort - et qu'ils n'ont pas à montrer comment ils font ou à discuter comment ils font. »

L'effort de reconstitution de la méthode, des différentes étapes ou corrections et du cheminement d'ensemble en vue d'une réutilisation par des tiers place le chercheur dans une position presque intenable estime un chercheur en économie, car disproportionnée au regard du temps et des moyens à disposition :

« Ça, c'est en fait un gros boulot. Documenter en vue d'un usage par d'autres, c'est se retrouver dans la position de l'INSEE, ou des autres fournisseurs de données généralistes. C'est très difficile et du coup, personne ne veut le faire. »

Des chercheurs démunis face aux règles encadrant le traitement des données

Les données sensibles

Adopté en 2018 dans un contexte de préoccupation croissante de l'exploitation des données personnelles par les GAFAM, le règlement général pour la protection des données (RGPD) a un impact considérable sur l'activité des chercheurs en SHS amenés à traiter de sujets sensibles (groupes sociaux marginalisés, trafics, opinions politiques ou religieuses, prostitution, etc.) Dans le cadre de ses recherches sur des personnes atteintes d'une maladie sexuellement transmissible, une chercheuse en psychologie s'est par exemple heurtée au délégué à la protection des données (DPD), qui refusait la collecte de l'orientation sexuelle des patients sondés, alors que cette information représentait une variable indispensable du questionnaire. Entravée dans son projet, la chercheuse estime que les DPD ont également à apprendre des pratiques de recherche : « Il faut que ce soit une formation mutuelle. »

Autre corollaire de l'application du RGPD, le recueil du consentement de la personne interviewée, via la signature d'un formulaire qui décline de manière administrative les visées de l'étude. Une approche difficilement compatible avec la recherche consignée dans des carnets ou collectée sur le terrain. Pour un chercheur en sociologie :

« Nous assistons à une bureaucratisation et un contrôle du travail d'enquête. Exit le travail sur des activités illégales ou officieuses comme je le fais. J'ai fait une demande de suivi et de conseil via mon labo auprès de l'institution tutelle mais elle a été refusée. Dont acte. »

Un doctorant en gestion fait également le constat d'un accompagnement défaillant et d'un déficit criant de connaissance de l'impact de ces nouvelles règles :

« Avant de commencer à collecter mes données via entretiens qualitatifs, j'ai sollicité différents services de mon université : responsable de groupe au labo, directeur de labo, référent intégrité scientifique de l'université, service juridique de l'université. Je me suis fait balader de service en service pour connaître la marche à suivre pour bien respecter la loi sur la collecte et le stockage de données collectées dans le cadre d'une recherche. Je n'ai eu aucune réponse satisfaisante et pragmatique (et encore moins coordonnée) à ce sujet. Le monde universitaire est complètement largué en la matière. »

Au-delà de ce constat, c'est aussi le manque d'infrastructures idoines qui peut aboutir à des pratiques de stockage inadaptées. Comme en témoigne une chercheuse en psychologie :

« Très clairement, on est hors les clous de tout. J'ai des bases de données sur mon ordinateur que je ne devrais pas avoir. On est sur une hétérogénéité des pratiques. »

En cause notamment, l'absence de serveur sécurisé institutionnel agréé pour accueillir des données sensibles relevant du domaine de la santé. Une autre chercheuse de la même discipline réagit :

« Si les données appartiennent à l'État, l'État doit nous donner les moyens de stockage. »

Dans ce paysage, seul un laboratoire en économie semble faire exception, l'un de ses membres précisant que des procédures standardisées sont en place pour les données sensibles.

Propriété intellectuelle

Lorsque l'intérêt de publier ses données existe et est identifié, des craintes émergent parfois sur les enjeux de propriété intellectuelle. Qu'en est-il de l'exploitation ou de la diffusion d'images satellites obtenues par un Equipex¹¹ ? s'interroge un ingénieur en sciences de l'information géographique. Quid de la réutilisation des cartes anciennes numérisées ? Ai-je le droit d'utiliser des matériaux issus d'archives en vue de la publication d'un ouvrage ? se demande une chercheuse en sciences de l'information. L'incertitude gouverne également le statut des données produites par les chercheurs eux-mêmes. Un chercheur en économie l'admet :

¹¹ Equipex : est ici question ici d'une infrastructure nationale de données satellitaires. Equipex, pour « équipement d'excellence », désigne des projets financés par le grand emprunt et destinés à améliorer les équipements des laboratoires de recherche scientifique français.

« Je ne sais pas trop ce que ce droit de propriété intellectuelle veut dire. Je comprends bien que si c'est moi qui ai collecté la donnée, l'établissement a en quelque sorte quelque chose à dire aussi. Mais honnêtement, je n'ai jamais bien compris jusqu'où ça allait. Moi, en tous cas, je suis souvent dans des situations où les jeux de données qu'on utilise sont des trucs assez hybrides, et il n'y a pas qu'un seul détenteur de la donnée. »

Un autre chercheur, en linguistique, assimile également la donnée de recherche à un bien public, tout en portant une inflexion sur le rôle joué par l'établissement employeur.

« Je ne perçois pas personnellement cela comme une propriété mais comme quelque chose qui appartient à tout le monde. C'est un bien commun quoi, ce n'est pas comme un brevet. Je n'ai peut-être pas vraiment le sentiment que les données elles appartiennent à l'institution, parce que je n'ai pas vraiment le sentiment qu'elle me permette d'en produire. J'ai tellement l'impression de faire beaucoup ma recherche sur mon temps libre que je n'ai pas l'impression de devoir à mon établissement. D'ailleurs, c'est l'établissement qui a la responsabilité de vous aider aussi en tant que chercheur pour bien stocker et puis gérer vos données aussi. »

Un chercheur en aménagement du territoire résume la situation en indiquant être soumis à des injonctions contradictoires : ouverture des données versus propriété intellectuelle, protection des données versus valorisation économique, peinant à trouver le bon équilibre.

Les chercheurs se trouvent confrontés à de multiples cas, freinant parfois leur volonté d'ouverture ou les amenant à prendre des « chemins de traverse », concède l'un d'entre eux.

Vers une meilleure utilisation des données, une opportunité en SHS ?

Le défi de la conservation et de l'archivage

L'enjeu de la pérennité d'accès aux données produites semble constituer un sujet fédérateur pour une majorité des chercheurs interrogés, dont la carrière a pu être marquée par des incidents plus ou moins douloureux.

« J'ai déjà perdu des documents qui m'auraient bien été utiles plus tard, admet un chercheur en sciences de gestion. Je suis sur des schémas de sauvegarde qui ne sont pas satisfaisants et je me rends compte que des projets que j'avais menés il y a une trentaine d'années m'auraient été bien utiles. La recherche est un horizon long et il faudrait de vrais plans de pérennisation. »

D'ici là, ce même interlocuteur souhaiterait a minima une solution permettant de « créer, projet par projet, un Intranet sécurisé pour mes collègues externes à Lyon 3 et moi-même ». Une chercheuse en informatique a quant à elle mis en place son propre protocole : « Pour pérenniser

mes données, je n'ai pas trouvé d'autre solution que d'utiliser 2 disques externes de sauvegarde disposés à des endroits différents (un au bureau et un chez moi). »

Lorsqu'aucune politique de conservation n'a été pensée dès le début du projet, la pertinence de certaines données est parfois réduite à néant. « J'avais une base de données assez unique, explique un chercheur en histoire. Elle est désormais complètement éclatée. » Deux ans après avoir été muté dans un autre établissement, le chercheur n'a pu que déplorer la cessation de l'hébergement et la disparition de cet outil.

Un risque également identifié par un chercheur en littérature :

« On monte des sites web, sur la littérature francophone par exemple. Ce n'est pas très difficile en prenant un prestataire. Le problème, c'est la maintenance. Les URL sont fragiles, les noms de domaine peuvent ne pas être maintenus, etc. Nous cherchons une solution pérenne et nous souhaiterions nous tourner vers les services d'accompagnement sur les données, car là, il y a un manque. »

Dans certaines disciplines, la conservation de certaines images revêt une importance cruciale, voire « patrimoniale », insiste un chercheur en archéologie. C'est le cas de « photographies constituant les seules traces de monuments détruits depuis ».

La conservation impose des choix clairs, souligne une ingénieure de recherche. Elle signale d'ailleurs que l'enjeu de la conservation et de l'archivage devrait être pensées de manière rationnelle : tout ne pourra pas être conservé au vu de la quantité que permet de produire le numérique ou du travail que représentent la documentation et contextualisation des données en vue de leur archivage.

Un gisement scientifique à l'épreuve de compétences en devenir

Quelques chercheurs tendent à nourrir beaucoup d'espoirs, percevant l'accès aux données massives comme un levier scientifique ouvrant la voie à de nouvelles méthodes voire, de nouvelles hypothèses. « L'apparition des données est assez révolutionnaire, elle ouvre énormément de perspectives en recherche », estime un chercheur en littérature, pour qui les outils de traitement associés au big data ouvrent un vaste champ d'exploration :

« Il ne s'agira plus de travailler sur les nouvelles de Maupassant mais sur l'ensemble des nouvelles publiées pendant la deuxième moitié du XIXe siècle. Passer de corpus réduits à des données énormes, c'est une rupture qui n'est pas forcément facile. »

Un chercheur en économie souligne également l'intérêt actuel du contexte d'ouverture des données, propice aux traitements et croisements dans sa discipline en citant l'exemple de données d'enquêtes issues du ministère américain du développement économique.

Par ailleurs, la formalisation du processus de recherche par le biais du plan de gestion de données rencontre parfois un écho positif, lorsqu'il est porteur de sens dans la discipline en question. Comme le souligne une chercheuse en sciences de l'éducation :

« J'estime qu'on ne va pas juste dialoguer sur les résultats mais que l'on va aussi travailler à plusieurs sur comment on arrive aux résultats. »

« Le PGD permet de planifier, complète une chercheuse en didactique de l'anglais, de ne pas se retrouver ensuite avec des données qui manquent, tel qu'un corpus incomplet. Pour les partenaires également, c'est un gage éthique qu'il y ait un plan de gestion des données. »

Si certains soulignent le caractère déclaratif de cet exercice, « j'explique comment j'ouvre et où sans que cette diffusion ne soit une réalité », cette obligation participe d'un « système vertueux de la culture de la donnée », précise un chercheur en archéologie.

Selon plusieurs interlocuteurs, l'impulsion donnée par un directeur ou une directrice de laboratoire sensible à l'ouverture des données peut avoir un impact concret sur les pratiques. Comme en témoigne une chercheuse en didactique des langues :

« Je pense que la culture du labo fait que l'on est sensibilisé à la fois aux données numériques et à la science ouverte [...] C'est vraiment la directrice du laboratoire qui a initié un gros projet, avec dès le départ, l'idée de partager le corpus. »

Dans ce laboratoire, les chercheurs peuvent également demander de l'aide s'ils veulent mettre en place un plan de gestion des données et ce, même en dehors d'un financement ANR.

Néanmoins les pratiques sont en général encore balbutiantes, d'où l'importance de sensibiliser les doctorants à ces enjeux : « Ce sont eux qui permettront de faire évoluer les pratiques au sein des laboratoires », complète une chercheuse en sciences de l'information.

À terme, le développement de connaissances liées à la gestion des données pourra cependant représenter un atout dans une carrière académique. « Me construire une compétence là-dessus sera bientôt vendable », prédit un chercheur en sciences de gestion.

Mais, pour l'heure, le constat est plus mesuré : « La recherche numérique n'est pas encore valorisée », objecte une chercheuse en études lusophones. Un chercheur en aménagement du territoire, conscient de la charge supplémentaire que représente le travail sur des données qualitatives (parfois fruit de plusieurs années de recherche), s'interroge quant à lui sur la reconnaissance d'un tel investissement : « Quelle incitation positive avons-nous à ouvrir nos données ? » La publication d'un article demeure la préoccupation majeure des chercheurs, bien avant la diffusion ou mise à disposition des données.

Enfin, plusieurs des personnes interrogées soulignent également le besoin et la nécessité de compétences spécifiques ou pointues, qu'elles ne possèdent pas forcément, pour assurer les phases de collecte et de traitement de données plus ou moins massives. Ce qui pose un défi en termes d'accompagnement des chercheurs. Interrogé à ce sujet, un directeur adjoint de laboratoire estime que les enjeux du big data et de l'intelligence artificielle relèvent des compétences d'ingénieurs spécialisés, en lien étroit avec les disciplines, et non des professionnels de la documentation ou des bibliothèques.

Conclusion et perspectives

Les résultats de cette enquête reflètent le décalage parfois béant entre les injonctions formulées dans les politiques nationales ou européennes et les pratiques constatées sur le terrain. L'incitation à une bonne gestion (incluant une logique de conservation) et à une ouverture élargie des données de recherche se heurte en SHS à un triple obstacle :

- L'incapacité ou la difficulté à cerner ce qu'est une donnée de recherche, dont le caractère reproductible n'a pas d'application concrète dans toutes les disciplines relevant des SHS ;
- la pratique très minoritaire de la diffusion des données de recherche ;
- l'utilisation d'infrastructures informatiques de stockage décorrélées de l'offre institutionnelle.

À ces observations s'ajoute une préoccupation liée au traitement des données personnelles et au formalisme induit par l'application du règlement européen sur la protection des données. Si ce cadre éthique s'impose dans le cadre d'études cliniques et de données de patients, certains chercheurs s'interrogent néanmoins sur son incidence dans d'autres domaines de recherche, lorsque cette réglementation entrave la conduite de certains projets de recherche ayant recours aux méthodes d'enquêtes, à l'instar de la sociologie ou de la psychologie sociale, par exemple.

L'analyse des besoins exprimés révèle un besoin d'accompagnement en matière de gestion informatique des données, confirmé par une préoccupation assez générale sur l'accès pérenne aux données les plus pertinentes en vue de réutilisations ultérieures. Ces besoins interrogent donc la capacité des établissements à fournir toute une série de services et d'infrastructures : serveurs pour les chercheurs et doctorants, outils collaboratifs, protocole en cas de stockage de données confidentielles ou sensibles, maintenance de bases de données, pérennité des sites web adossés aux projets, etc.

Soulignons sur ce point la création récente des pôles d'archives dans les différents établissements du site lyonnais, avec une organisation fonctionnelle et un travail en réseau incluant les archives départementales. La présence d'archivistes dans les universités étant recommandée, mais pas obligatoire, c'est un atout fort du site lyonnais. Une suggestion pour les porteurs de projets serait de se rapprocher des services d'archives de leurs établissements de rattachement. Mobiliser les archives départementales pourrait par ailleurs être pertinent pour archiver et valoriser les données issues de projets de recherche en lien avec le territoire.

Par ailleurs, le partage et la diffusion des données sont régis par un cadre juridique lié aux licences et au droit d'auteur, pour lesquels une sensibilisation, une formation voire un accompagnement au cas par cas serait à envisager. Il est aussi essentiel que les répondants puissent identifier les personnes-ressources pour les questions juridiques, en s'appuyant sur des interlocuteurs pérennes pour les accompagner au cours d'un projet de recherche. En ce sens, depuis plusieurs années, les SCD créent de nouveaux postes de personnels dédiés à la gestion des données de recherche au sein de services spécifiques de soutien à la recherche. Ces personnels accompagnent par exemple les chercheurs dans la rédaction de PGD et proposent des formations.

La faible pénétration des pratiques d'ouverture des données et la part importante de chercheurs n'identifiant pas l'intérêt de celle-ci ou ignorant qu'elle était envisageable invitent par ailleurs à

repenser le type d'accompagnement proposé. L'organisation de cycles de rencontres favorisant le dialogue entre chercheurs et le retour d'expériences sur le partage des données peut faire partie des hypothèses à envisager. Cette méthode permettrait de réintroduire du sens à des incitations politiques calquées sur les pratiques des sciences dures (à l'instar du principe de reproductibilité des données) et difficilement applicables à l'ensemble des SHS. En outre, cette diversification d'approches prendrait en compte le fait que l'accompagnement à la rédaction des plans de gestion de données, sur lequel les établissements ont eu tendance à se concentrer, n'est pas la demande première exprimée par les chercheurs. Enfin, la mise en place du module de formation Isidoc't, à destination des écoles doctorales et comprenant un volet sur la gestion des données, gagnerait à toucher un nombre plus important de doctorants.

L'ensemble de ces conclusions pourra alimenter la feuille de route d'un futur atelier de la donnée en cours de construction au sein du pôle Lyon/St-Étienne en 2023-2024.

TABLE DES FIGURES

Fig. 1. Affiliation des répondants

Fig. 2. Profil des répondants

Fig. 3. Principales disciplines des répondants

Fig. 4. Affiliation des répondants par unité de recherche

Fig. 5. Type de données produites

Fig. 5.1. Type de données en fonction des disciplines

Fig. 6. Formats informatiques utilisés

Fig. 7. Volumes estimés de données stockées

Fig. 7.1. Volume de stockage de données nécessaire par discipline

Fig. 8. Supports de stockage pour les données fréquemment utilisées

Fig. 9. Outils de partage des données avec des collaborateurs extérieurs

Fig. 10. Participation à la rédaction d'un PGD

Fig. 11. Présence de personnels dédiés à la gestion des données au sein du laboratoire

Fig. 11.1. Affiliation des chercheurs affirmant disposer de personnel dédié au sein de leur unité

Fig. 12. Connaissance de la fonction de délégué à la protection des données

Fig. 13. Traitement des données à caractère personnel ou dites sensibles

Fig. 14. Déclaration du traitement des données personnelles auprès du DPD

Fig. 15. Sécurisation des données

Fig. 16. Connaissance des standards de métadonnées communs ou partagés ? Domaine de recherche

Fig. 17. Renseignement des métadonnées pour faciliter la documentation ou l'archivage des données

Fig. 18. Partage des données en ligne

Fig. 18.1. Diffusion des données dans les principales disciplines

Fig. 19. Motifs de non-diffusion en ligne des données

Fig. 20. Périmètre d'accès aux données de recherche

Fig. 21. Connaissance de l'existence du pôle archives de l'établissement

Fig. 22. Archivage à long terme des données dites « froides »

Fig. 23. Besoins de services ou soutien concernant la gestion des données de recherche

Fig. 24. Types de services ou soutien en fonction des besoins

GLOSSAIRE

ANR : L'Agence nationale de la recherche (ANR) est un établissement public à caractère administratif, placé sous la tutelle du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. L'Agence met en œuvre le financement de la recherche sur projets, pour les opérateurs publics en coopération entre eux ou avec des entreprises.

AMI - appel à manifestation d'intérêt : appel à projets pour inviter les acteurs de l'enseignement supérieur et de la recherche à se positionner sur des projets.

DPD - délégué à la protection des données : une fonction située au cœur de la conformité au règlement européen sur la protection des données (RGPD), le délégué à la protection des données (DPO) conseille et accompagne les organismes qui le désignent. La désignation d'un délégué à la Protection des données est obligatoire pour certaines entités privées, ainsi que pour l'ensemble des autorités publiques ou organismes publics (à l'exception des juridictions agissant dans l'exercice de leur fonction juridictionnelle).

Huma-Num : infrastructure de recherche dédiée aux lettres, sciences humaines et sociales et aux humanités numériques mise en œuvre par le ministère de l'Enseignement supérieur et de la Recherche et portée par le Centre national de la recherche scientifique, Aix-Marseille université et le campus Condorcet.

RGPD : signifie « règlement général sur la protection des données ». Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne. Le contexte juridique s'adapte pour suivre les évolutions des technologies et de nos sociétés (usages accrus du numérique, développement du commerce en ligne...). Depuis 2016 ce règlement européen s'inscrit dans la continuité de la loi française Informatique et Libertés de 1978 et renforce le contrôle par les citoyens de l'utilisation qui peut être faite des données les concernant.

IR* (ex TGIR – très grande infrastructure de recherche) : infrastructures qui, bien qu'étant sous la responsabilité scientifique des opérateurs de recherche, relèvent d'une politique nationale et font l'objet d'un fléchage budgétaire du MESRI. Par exemple : Huma-Num ou Progedo sont deux IR* dans le domaine des SHS.

ANNEXES

Annexe 1 : questionnaire en ligne

Enquête sur la gestion des données de la recherche en SHS

OBJECTIF ET CONTEXTE

Dans une optique de développement des services et des infrastructures, ce questionnaire a pour but de recenser les pratiques, les besoins et les outils utilisés par les chercheurs et personnels de la recherche de Lyon 3, de Lyon 2 et de la MSH Lyon Saint-Étienne quant à la gestion de leurs données de recherche.

Dans le droit fil du deuxième plan national pour la science ouverte, de la charte de la science ouverte de Lyon 3, de la politique de science ouverte de Lyon 2 et du plan « Données de la recherche du CNRS », les universités Lyon 2 et Lyon 3 et la MSH Lyon Saint-Étienne souhaitent renforcer leur dispositif de soutien et d'accompagnement à la gestion des données de recherche. Le lancement de cette enquête commune participe de cette démarche et de leur volonté d'apporter une réponse consolidée à cette problématique.

Les données de la recherche sont devenues un enjeu majeur pour la gestion et la diffusion des connaissances scientifiques. Sous forme d'enregistrements factuels (chiffres, textes, images et sons), elles constituent des sources principales pour la recherche et sont généralement reconnues par la communauté scientifique comme nécessaires à la validation des résultats issus de la recherche.

La rédaction d'un plan de gestion de ces données est en phase de devenir la norme pour obtenir des financements dans un cadre européen (AAP Horizon Europe) et français (ANR depuis 2019).

DESTINATAIRES ET DIFFUSION

Cette enquête s'adresse donc à l'ensemble des chercheurs, enseignants-chercheurs, doctorants et personnels de la recherche de ces établissements, producteurs ou gestionnaires de données de recherche afin de mieux les aider et les orienter dans leurs démarches. Elle sera mise en ligne jusqu'au 10 décembre 2021 à 23h00.

Les résultats seront publiés et diffusés sur les plateformes HAL des deux universités. Répondre à ces questions prendra environ quinze à vingt minutes. Les participations pourront rester strictement anonymes mais des entretiens avec des personnes volontaires pourront également être proposés.

Il y a 75 questions dans ce questionnaire.

* = question obligatoire

IDENTITÉ

* Quel est votre statut ?

Veillez sélectionner une seule des propositions suivantes :

- Chercheur-se ou enseignant-e-chercheur-se
- Doctorant-e
- Personnel de soutien à la recherche ITA ou ITRF
- Maître de conférences
- Post-doc
- Professeur-e d'université
- Autre

* Quelle est votre université ou institution de rattachement ?

Veillez sélectionner une seule des propositions suivantes :

- Université Lyon 2
- Université Lyon 3
- Université Jean Monnet
- CNRS
- ENS Lyon
- Autre

Votre laboratoire est-il associé à la MSH Lyon Saint-Étienne ?

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

La réponse était « Autre » à la question 2 [Q02] (Quelle est votre université ou institution de rattachement ?)

- Oui
- Non

Pouvez-vous préciser de quelle unité de recherche il s'agit ?

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

La réponse était « Oui » à la question [Q02bis] (Votre laboratoire est-il associé à la MSH Lyon Saint-Étienne ?)

Veillez choisir toutes les réponses qui conviennent :

- Centre Gabriel Naudé
- ECCLA - Études du contemporain en littératures, langues et arts
- LVIS - Laboratoire sur les vulnérabilités et l'innovation dans le sport
- MAELYSE - Management économie Lyon St-Étienne
- RESHAPE - Recherche en santé publique SAF - Sciences actuarielle et financière
- S2HEP - Sciences, société, historicité, éducation, pratiques

* À quelle unité ou à quels groupements de recherche de Lyon 3 êtes- vous rattaché.e ?

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

La réponse était « Université Lyon 3 » à la question 2 [Q02] (Quelle est votre université ou institution de rattachement ?)

Veillez choisir toutes les réponses qui conviennent :

- CEL - Centre d'études linguistiques
- CIHAM - Histoire, archéologie, littérature des mondes chrétiens et musulmans médiévaux

- CLHDPP - Centre lyonnais d'histoire du droit et de la pensée politique
- CRDMS - Centre de recherche en droit et management des services de santé
- EDIEC - Équipe de droit international et comparé
- EDPL - Équipe de droit public de Lyon
- ELICO - Équipe lyonnaise en information et communication
- ERLJ - Équipe de recherche Louis Josserand
- EVS - Environnement ville société
- Francophonie - Francophonie, mondialisation et relations internationales
- HISOMA - Histoire et sources des mondes antiques
- IETT - Institut d'études transtextuelles et transculturelles IRPHIL - Institut de recherches philosophiques de Lyon
- LARHRA - Laboratoire de recherche historique Rhône-Alpes
- LIRIS - Laboratoire d'informatique en image et systèmes d'information
- Magellan - iaelyon School of Management
- MARGE
- GIS Asie
- GIS IDA - Institut des Amériques GIS IDG - Institut du genre.
- GIS IXXI - Institut rhônalpin des systèmes complexes
- GIS Moyen Orient et mondes musulmans
- GIS Religions, pratiques, textes, pouvoirs
- GIS Z ABR - Zone atelier bassin de Rhône
- Autre

*** À quelle unité ou à quels groupements de recherche de Lyon 2 êtes -vous rattaché.e ?**

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

La réponse était « Université Lyon 2 » à la question 2 [Q02] (Quelle est votre université ou institution de rattachement ?)

Veillez choisir toutes les réponses qui conviennent :

- Archéorient – Environnements et sociétés de l’Orient ancien - UMR 5133
- Arar - Archéologie et archéométrie - UMR 5138
- CeRCriD - Centre de recherches critiques sur le droit - UMR 5137
- CeRLA - Centre de recherche en linguistique appliquée - EA 4162
- CIHAM - Histoire, archéologie, littérature des mondes chrétiens et musulmans médiévaux – UMR 5648
- CMW - Centre Max Weber - UMR 5283
- COACTIS – Conception de l’action en situation - EA 4161
- CRPPC - Centre de recherche en psychopathologie et psychologie clinique - EA 653
- DCT - Droit, contrats, territoires - EA4573
- DDL - Dynamique du langage - UMR 5596
- DIPHE - Développement, individu, processus, handicap, éducation - EA 4129
- DISP - Décision et information pour les systèmes de production - EA 4570
- ECP - Éducation, cultures, politiques - EA 4571
- ELICO - Équipe lyonnaise en information et communication - EA 4147
- EMC - Laboratoire d’étude des mécanismes cognitifs - EA 3082
- ERIC – Entrepôts, représentation et ingénierie des connaissances - EA 3083
- EVS - Environnement, ville, société - UMR 5600

- GATE LSE - Groupe d'analyse et de théorie économique Lyon Saint-Étienne - UMR 5824
- GRePS - Groupe de recherche en psychologie sociale - EA 4163
- HISOMA - Histoire et sources des mondes antiques - UMR 5189
- IAO - Institut d'Asie orientale - UMR 5062
- ICAR - Interactions, corpus, apprentissages, représentations - UMR 5191
- IHRIM - Institut d'histoire des représentations et des idées dans les modernités – UMR 5317
- IRAA - Institut de recherche sur l'architecture de l'antiquité - USR 3155
- LADEC - Laboratoire d'anthropologie des enjeux contemporains - FRE 2002 - UR
- LAET - Laboratoire aménagement économie transports - UMR 5593
- LARHRA - Laboratoire de recherche historique Rhône-Alpes - UMR 5190 LCE - Lettres et civilisations étrangères - UR 1853
- LER - Laboratoire d'études rurales - EA 3728
- LIRIS - Laboratoire d'informatique en image et systèmes d'information - UMR 5205
- Passages arts & littératures (XX-XXI) - EA 4160
- Triangle - Action, discours, pensée politique et économique - UMR 5206
- Autre

*** À quelle unité ou à quels groupements de recherche du CNRS êtes-vous rattaché.e ?**

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

La réponse était « CNRS » à la question 2 [Q02] (Quelle est votre université ou institution de rattachement ?)

Veillez choisir toutes les réponses qui conviennent :

- AGORANTIC - FR 3621
- ARAR - UMR 5138
- Archéorient - UMR 5133
- CERCOR - UMR 8584 - Centre européen de recherche sur les communautés, congrégations et ordres religieux.
- CERCRID - UMR 5137 - Centre de recherches critiques sur le droit
- CIHAM - Histoire, archéologie, littérature des mondes chrétiens et musulmans médiévaux- UMR 5648
- CMW - UMR 5283 - Centre Max Weber
- DDL - UMR 5596 - Dynamique du langage
- EVS - UMR 5600 - Environnement ville société
- GATE LSE- UMR 5824 - Groupe d'analyse et de théorie économique
- HISOMA - UMR 5189 - Histoire et sources des mondes antiques
- IAO - UMR 5062 - Institut d'Asie orientale
- ICAR - UMR 5191
- IHRIM - UMR 5317 - Institut d'histoire des représentations et des idées dans les modernités
- IRAA - USR 3155 - Institut de recherche sur l'architecture antique
- ISERL - FR Institut supérieur d'étude des religions et de la laïcité
- LAET-UMR 5593 - Laboratoire aménagement économie transports
- LARHRA - UMR 5190 - Laboratoire de recherche historique Rhône-Alpes

- LIRIS - UMR 5205 - Laboratoire d'informatique en image et systèmes d'information
- MOM - FR 3747 - Maison de l'Orient et de la Méditerranée
- SFR-RELYS - Recherche en éducation Lyon Saint-Étienne.
- Triangle - UMR 5206
- UAR 2000 - MSH Lyon Saint-Étienne
- Autre

À quelle unité ou à quels groupements de recherche de l'ENS Lyon êtes-vous rattaché.e ?

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

La réponse était « ENS Lyon » à la question 2 [Q02] (Quelle est votre université ou institution de rattachement ?)

Veillez choisir toutes les réponses qui conviennent :

- CERCC - Centres d'études et de recherche comparées sur la création
- CIHAM - Histoire, archéologie, littérature des mondes chrétiens et musulmans médiévaux
- CMW - Centre Max Weber
- EVS - Laboratoire environnement ville société
- GATE - Groupe d'analyse et de théorie économique
- HISOMA - Laboratoire histoire et sources des mondes antiques
- IAO - Institut d'Asie orientale
- ICAR - Interactions, corpus, apprentissages, représentations
- IHRIM - Institut d'histoire des représentations et des idées dans les modernités
- IXXI - Institut rhônalpin des systèmes complexes
- LARHA - Laboratoire de recherche historique Rhône-Alpes
- LIP - Laboratoire de l'informatique du parallélisme
- LLE - Laboratoire de l'éducation Persée
- Triangle - Triangle. Action, discours, pensée politique et économique

*** À quelle unité ou à quels groupements de recherche de l'Université Jean Monnet êtes-vous rattaché.e ?**

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

La réponse était « Université Jean Monnet » à la question 2 [Q02] (Quelle est votre université ou institution de rattachement ?)

Veillez choisir toutes les réponses qui conviennent

- CERCRID - Centre de recherches critiques sur le droit
- CMW - Centre Max Weber
- COACTIS - Conception de l'action en situation
- ECLLA - Études du contemporain en littératures, langues, arts
- ECP - Éducation, culture, politiques
- EVS - Environnement ville société
- GATE - Groupe d'analyse et de théorie économique
- HISOMA - Histoire et sources des mondes antiques
- IHRIM - Institut d'histoire des représentations et des idées dans les modernités
- LEM - Laboratoire d'études sur les monothéismes
- Triangle : action, discours, pensée politique et économique

- Autre

À quelle(s) section(s) du CNU êtes-vous affiliée ?

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

La réponse était « Université Lyon 2 » ou « Autre » ou « Université Lyon 3 » à la question 2 [Q02]

(Quelle est votre université ou institution de rattachement ?)

Veillez choisir toutes les réponses qui conviennent :

1. Droit privé et sciences criminelles
 2. Droit public
 3. Histoire du droit et des institutions
 4. Science politique
 5. Sciences économiques
 6. Sciences de gestion
 7. Sciences du langage : linguistique et phonétique générale
 8. Langues et littératures anciennes
 9. Langue et littérature françaises
 10. Littératures comparées
 11. Langues et littératures anglaises et anglo-saxonnes
 12. Langues et littératures germaniques et scandinaves
 13. Langues et littératures slaves
 14. Langues et littératures romanes : espagnol, italien, portugais, autres langues romanes
 15. Langues et littératures arabes, chinoises, japonaises, hébraïques, d'autres domaines linguistiques
 16. Psychologie, psychologie clinique, psychologie sociale
 17. Philosophie
 18. Architecture (ses théories et ses pratiques), arts appliqués, arts plastiques, arts du spectacle, épistémologie des enseignements artistiques, esthétique, musicologie, musique, sciences de l'art
 19. Sociologie, démographie
 20. Ethnologie, préhistoire, anthropologie biologique
 21. Histoire, civilisations, archéologie et art des mondes anciens et médiévaux
 22. Histoire et civilisations : histoire des mondes modernes, histoire du monde contemporain ; de l'art ; de la musique
 23. Géographie physique, humaine, économique et régionale
 24. Aménagement de l'espace, urbanisme
 26. Mathématiques appliquées et applications des mathématiques
 27. Informatique
 70. Sciences de l'éducation
 71. Sciences de l'information et de la communication
 00. Aucun
- Autre

*** Quelle est ou quelles sont vos/votre section(s) de rattachement disciplinaire au CNRS ?**

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

La réponse était « CNRS » à la question 2 [Q02] (Quelle est votre université ou institution de rattachement ?)

Veillez choisir toutes les réponses qui conviennent :

- 6. Sciences de l'information : fondements de l'informatique, calculs, algorithmes, représentations, exploitations
- 7. Sciences de l'information : signaux, images, langues, automatique, robotique, interactions, systèmes intégrés matériel-logiciel
- 31. Hommes et milieux : évolution, interactions
- 32. Mondes anciens et médiévaux
- 33. Mondes modernes et contemporains
- 34. Sciences du langage
- 35. Philosophie, littératures, arts
- 36. Sociologie et sciences du droit
- 37. Économie et gestion
- 38. Anthropologie et étude comparative des sociétés contemporaines
- 39. Espaces, territoires, sociétés
- 40. Politique, pouvoir, organisation
- 41. Mathématiques et interactions des mathématiques

Je ne sais pas

Autre

Caractérisation des données

*** Quels grands types de données produisez-vous ou collectez-vous ?**

Veillez choisir toutes les réponses qui conviennent :

- Des données d'observation [données capturées en temps réel, habituellement uniques et donc impossibles à reproduire. Sondages, enquêtes, captation d'images...]
- Des données d'expérimentation [données obtenues à partir d'équipements de laboratoire, souvent reproductibles mais parfois coûteuses, tests, modélisations...]
- Des données computationnelles ou de simulation [données générées par des modèles informatiques ou de simulation, souvent reproductibles si le modèle est correctement documenté]
- Des données compilées ou dérivées [données obtenues par de la fouille de textes ou de données d'une base ou par la combinaison de données « brutes », souvent reproductibles mais coûteuses]
- Des données canoniques ou de référence [corpus de textes, de statistiques, de cartes et plans...collection ou accumulation de petits corpus et jeux de données revus publiés ou par les pairs, annotés et mis à disposition]
- Autre

De quel(s) type(s) sont vos données d'observation ?

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

La réponse était à la question 12 [Q05] (Quels grands types de données produisez-vous ou collectez-vous ?)

Veillez choisir toutes les réponses qui conviennent :

- Remontées de terrain
- Données de capteurs
- Entretiens questionnaires séquences filmées photographies images satellites
- Données de géolocalisation (SIG) cartographie de sites web réseaux sociaux
- Autre

De quel(s) type(s) sont vos données d'expérimentation ?

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

La réponse était à la question 12 [Q05] (Quels grands types de données produisez-vous ou collectez-vous ?)

Veillez choisir toutes les réponses qui conviennent :

- Mises en situation
- Tests
- Panels
- Modélisations
- Autre

De quel(s) type(s) sont vos données de simulation ?

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

La réponse était à la question 12 [Q05] (Quels grands types de données produisez-vous ou collectez-vous ?)

Veillez choisir toutes les réponses qui conviennent :

- Modélisation
- Calcul
- Autre

De quel(s) type(s) sont vos données compilées ou dérivées ?

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

La réponse était à la question 12 [Q05] (Quels grands types de données produisez-vous ou collectez-vous ?)

Veillez choisir toutes les réponses qui conviennent

- Statistiques
- Tableaux de données
- Bases de données
- Autre

De quel(s) type(s) sont vos données canoniques ?

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

La réponse était à la question 12 [Q05] (Quels grands types de données produisez-vous ou collectez-vous ?)

Veillez choisir toutes les réponses qui conviennent :

- Corpus de textes
- Catalogue
- Recueil statistique
- Recueils d'informations géographiques
- Bases de données publiées
- Cartes et plans
- Autre

Pouvez-vous préciser les formats de fichiers que vous utilisez ?

Veillez choisir toutes les réponses qui conviennent :

- Des formats de fichiers textes (ODT, TXT, DOC, RTF, etc.) des formats de tableurs (XLS, XLSX, ODS, CSV, etc.)
- Des formats de présentations (PPT, PPTX, PPTM, ODP, etc.)
- Des formats de pages (PDF, HTML, etc.)
- Des formats 3D ou multidimensionnels
- Des formats de dessins vectoriels des formats d'enregistrements audio des formats d'enregistrements vidéo des formats d'images
- Des formats d'archives
- Des formats de données structurées (bases de données) des formats exécutables (programmes et applications) des formats de spectrométries
- Autre

Pouvez-vous préciser les formats de fichiers 3D ou multidimensionnels que vous utilisez ?

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

La réponse était à la question 18 [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- .XCF
- .BLEND
- .SKP
- .SKB
- DXF
- .3DS Max
- .C4D
- .VRML
- .X3D
- .IFC
- .DWG
- Autre

Pouvez-vous préciser les formats de fichiers d'enregistrements audio ?

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

La réponse était à la question [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- OGG
- FLAC
- MP3
- WAV
- WMA
- AAC
- Autre

Pouvez-vous préciser les formats de fichiers de dessin vectoriel ?

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

La réponse était à la question [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- VML
- SVG
- Silverlight
- SWF
- AI
- EPS
- DXF
- Autre

Pouvez-vous préciser les formats d'enregistrements vidéo que vous utilisez ?

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

La réponse était à la question [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- MPEG
- OGM
- DVD
- Xvid
- DivX
- AVI
- Theora
- FLV
- Autre

Pouvez-vous préciser les formats de fichiers d'images que vous utilisez ?

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

La réponse était à la question [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- PNG
- MNG
- TIFF
- JPEG
- GIF
- TGA
- OpenEXR
- BMP
- FITS
- Autre

Pouvez-vous préciser les formats de fichiers de données structurées (bases de données) que vous utilisez ?

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

La réponse était à la question [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- CSV
- XML
- JSON
- Autre

Pouvez-vous préciser les formats exécutables (programmes et applications) que vous utilisez ?

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

La réponse était à la question [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- BIN
- ELF
- EXE
- SDC
- BAT
- Autre

Pouvez-vous préciser les formats de fichiers de spectrométrie que vous utilisez ?

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

La réponse était à la question [Q10bis] (Pouvez-vous préciser les formats de fichiers que vous utilisez ?)

Veillez choisir toutes les réponses qui conviennent :

- JCAMP
- mzML
- MOL
- SDF
- RXN
- RDF
- CML
- SMILES
- INCHI
- XYZ
- FID
- CIF
- Autre

Avez-vous pour habitude d'ajouter des informations supplémentaires (métadonnées) à ces fichiers ? (Pour documenter la collecte et la production de vos données ou jeux de données, notamment en vue de leur diffusion et archivage)

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui, avec quels types d'informations ?

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

La réponse était « Oui » à la question [Q24] (Avez-vous pour habitude d'ajouter des informations supplémentaires (métadonnées) à ces fichiers ? (Pour documenter la collecte et la production de vos données ou jeux de données, notamment en vue de leur diffusion et archivage)

Veillez choisir toutes les réponses qui conviennent :

- Des métadonnées administratives (auteur, date de création, etc.)
- Des métadonnées techniques (format des fichiers, taille, logiciel, etc.)
- Des métadonnées scientifiques en lien avec la discipline (origine des données, projet de recherche, mots clefs, etc.)
- Autre

Avez-vous connaissance de standards ou formats de métadonnées communs ou partagés au sein de votre domaine de recherche ?

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui, pouvez-vous les lister ?

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

La réponse était « Oui » à la question [Q26] (Avez-vous connaissance de standards ou formats de métadonnées communs ou partagés au sein de votre domaine de recherche ?)

Veillez choisir toutes les réponses qui conviennent :

- XML
- TEI
- DDI
- Dublin Core
- Autre

*** Savez-vous ce qu'est un délégué à la protection des données (DPD ou DPO- Data Protection Officer) ?**

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

La réponse était « Non » à la question [Q28a] (Savez-vous ce qu'est un délégué à la protection des données (DPD ou DPO- Data Protection Officer) ?) :

Qualifié de « chef d'orchestre » par la CNIL, le délégué à la protection des données, aussi appelé DPO pour « Data Protection Officer », est la personne en charge de la protection des données à caractère personnel au sein des organismes publics ou privés. Le délégué veille à mettre en œuvre la conformité au règlement européen sur la protection des données au sein de l'organisme qui l'a désigné s'agissant de l'ensemble des traitements mis en œuvre par cet organisme.

Connaissez-vous l'identité du délégué à la protection des données (DPD ou DPO- Data Protection Officer) de votre établissement ?

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

La réponse était « Oui » à la question [Q28a] (Savez-vous ce qu'est un délégué à la protection des données (DPD ou DPO- Data Protection Officer) ?)

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

*** Traitez-vous (production, collecte, archivage) des données à caractère personnel ou dites sensibles ?**

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui, avez-vous déjà fait une déclaration de traitement personnel auprès de votre DPD/DPO ?

Cette déclaration permettra d'alimenter le registre de traitement prévu par [l'article 30 du RGPD](#). Il participe à la documentation de la conformité. Il s'agit d'un document de recensement et d'analyse qui doit refléter la réalité de vos traitements de données personnelles et vous permet d'identifier notamment :

- Les parties prenantes (représentant, sous-traitants, co-responsables, etc.) qui interviennent dans le traitement des données,
- Les catégories de données traitées,
- À quoi servent ces données (ce que vous en faites), qui accède aux données et à qui elles sont communiquées,
- Combien de temps vous les conservez,
- Comment elles sont sécurisées.

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

La réponse était « Oui » à la question [Q30] (Traitez-vous (production, collecte, archivage) des données à caractère personnel ou dites sensibles ?)

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Gestion des données

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

Veillez sélectionner une seule des propositions suivantes :

- Entre 0 et 100 Go (pour les corpus rassemblant textes, fichiers audios et photos en basse ou moyenne qualité. On stocke 4 000 photos de 5 Mo dans 20 Go...)
- Entre 100 Go et un To (un téra-octet = 1 000 Go ou 1 000 milliards d'octets) (pour les corpus réunissant photographies haute définition et vidéos en grand nombre)
- 1 To (pour un gros projet utilisant une grosse base de données)
- Je n'en ai aucune idée

Où stockez-vous vos données courantes ?

Les données courantes ou « données chaudes » sont des données fréquemment utilisées, actualisées ou consultées.

Veillez choisir toutes les réponses qui conviennent :

- Sur une clé USB
- Sur des CD, DVD
- Sur un disque dur externe personnel
- Sur un disque dur externe professionnel
- En local, sur mon ordinateur privé
- En local, sur mon ordinateur professionnel
- En réseau dans mon espace personnel sur le cloud institutionnel de mon université
- En réseau, sur le serveur de mon laboratoire, de mon département, de mon institut
- En réseau, sur le serveur de l'université
- Sur un serveur externe, hébergé par une autre institution : une unité ou un labo partenaire
- Sur le cloud d'un prestataire privé (Google Drive, Dropbox, OneDrive etc.)
- Sur une grande infrastructure publique (grid, plateforme d'instrumentation ou de données, TGIR, ShareDoc-HumaNum...)
- Autre

Effectuez-vous des opérations de sauvegardes régulières sur vos données ?

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui, ces sauvegardes sont-elles automatisées ?

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

La réponse était « Oui » à la question [Q34] (Effectuez-vous des opérations de sauvegardes régulières sur vos données ?)

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui à quels intervalles de temps ?

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

La réponse était « Oui » à la question [Q34] (Effectuez-vous des opérations de sauvegardes régulières sur vos données ?)

Veillez sélectionner une seule des propositions suivantes :

- Tous les jours
- Toutes les semaines
- Tous les mois
- Tous les trimestres
- Tous les ans
- Irrégulièrement, en fonction des besoins
- En fonction du prestataire de service ou du gestionnaire de site
- Autre

Si oui, sur quels supports ?

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

La réponse était « Oui » à la question [Q34] (Effectuez-vous des opérations de sauvegardes régulières sur vos données ?)

Veillez choisir toutes les réponses qui conviennent :

- Sur une clé USB
- Sur des CD, DVD
- Sur un disque dur externe personnel
- Sur un disque dur externe professionnel
- En local, sur mon ordinateur privé
- En local, sur mon ordinateur professionnel
- En réseau dans mon espace personnel sur le cloud institutionnel de mon université
- En réseau, sur le serveur de mon laboratoire, de mon département, de mon institut
- En réseau, sur le serveur de l'université
- Sur un serveur externe, hébergé par une autre institution : une unité ou un labo partenaire
- Sur le cloud d'un prestataire privé (Google Drive, Dropbox, OneDrive etc.)
- Sur une grande infrastructure publique (grid, plateforme d'instrumentation ou de données, TGIR, ShareDoc-HumaNum...)
- Autre

*** Une fois vos recherches finalisées, avez-vous une politique de préservation ou d'archivage sur le long terme de vos données « froides » ?**

Les « données froides » sont les données de recherche rarement actualisées ou consultées.

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui, quelles données sont concernées ?

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

La réponse était « Oui » à la question [Q37] (Une fois vos recherches finalisées, avez-vous une politique de préservation ou d'archivage sur le long terme de vos données « froides » ? Les « données froides » sont les données de recherche rarement actualisées ou consultées.)

Veillez choisir toutes les réponses qui conviennent :

- Toutes mes données sans distinction
- Uniquement celles ayant une valeur probante ou venant en appui de mes publications
- Celles réutilisables dans le cadre de recherches futures
- Celles pouvant présenter un intérêt historique ou patrimonial
- Celles présentant un caractère unique ou non reproductible
- Autre

Selon quels critères déterminez-vous la durée de conservation de ces données ?

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

La réponse était « Oui » à la question [Q37] » (Une fois vos recherches finalisées, avez-vous une politique de préservation ou d'archivage sur le long terme de vos données « froides » ? Les « données froides » sont les données de recherche rarement actualisées ou consultées.)

Veillez écrire votre réponse ici :

Sur quel support, stockez-vous vos « données froides » ?

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

La réponse était « Oui » à la question [Q37] » (Une fois vos recherches finalisées, avez-vous une politique de préservation ou d'archivage sur le long terme de vos données « froides » ? Les « données froides » sont les données de recherche rarement actualisées ou consultées.)

Veillez choisir toutes les réponses qui conviennent :

- Sur une clé USB
- Sur des CD, DVD
- Sur un disque dur externe personnel
- Sur un disque dur externe professionnel
- En local, sur mon ordinateur privé
- En local, sur mon ordinateur professionnel
- En réseau dans mon espace personnel sur le cloud institutionnel de mon université
- En réseau, sur le serveur de mon laboratoire, de mon département, de mon institut
- En réseau, sur le serveur de l'université
- Sur un serveur externe, hébergé par une autre institution : une unité ou un labo partenaire
- Sur le cloud d'un prestataire privé (Google Drive, Dropbox, OneDrive etc.)
- Sur une grande infrastructure publique (grid, plateforme d'instrumentation ou de données, TGIR, ShareDoc-HumaNum...)
- Autre :

*** Au regard de vos pratiques et des outils utilisés, pensez-vous que vos données soient suffisamment sécurisées contre les risques suivants :**

Choisissez la réponse appropriée pour chaque élément :

	Oui	Incertain	Non
perte de données	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
corruption de données	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
faible d'accès sur données sensibles ou personnelles	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Concernant les données produites par des chercheurs non permanents de votre laboratoire (doctorants, chercheurs associés, partenaires...), comment votre laboratoire assure-t-il la conservation et la pérennisation de ces données ?

Veillez écrire votre réponse ici :

Partage et diffusion de vos données

En dehors de vous, qui peut accéder directement (en lecture et/ou écriture) à vos données de recherche ?

Veillez choisir toutes les réponses qui conviennent :

- Personne
- Mes collaborateurs directs
- Les membres de mon projet de recherche
- Mon laboratoire, mon institut
- L'ensemble de l'université
- Des chercheurs de ma discipline
- Tout le monde
- Autre

Dans le cadre de vos travaux ou projets de recherche, comment partagez-vous vos données chaudes avec vos collaborateurs (partenaires externes ou internes à votre laboratoire) ?

Veillez choisir toutes les réponses qui conviennent :

- Outils internes de l'université (Owncloud Lyon3)
- Messagerie
- Dropbox, google drive
- Services de transferts de fichiers (We transfer, etc.)
- Plateformes collaboratives propriétaires (microsoft teams, etc.)
- Outils de partages institutionnels sécurisés (Renater, Sharedoc de Huma-Num...)
- Autre

*** Avez-vous déjà publié ou diffusé des données ou jeux de données en libre accès sur le web, via une archive ouverte ou un entrepôt de données ?**

Veillez sélectionner une seule des propositions suivantes :

- Oui, au moins une fois
- Oui, régulièrement
- Non
- Autre

Si oui, dans quel contexte ?

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

La réponse était « Oui, au moins une fois » ou « Oui, régulièrement » à la question [Q45] (Avez-vous déjà publié ou diffusé des données ou jeux de données en libre accès sur le web, via une archive ouverte ou un entrepôt de données ?)

Veillez choisir toutes les réponses qui conviennent

- Pour répondre aux exigences d'un financeur
- Pour valoriser les résultats d'une recherche
- Pour publier un article à la demande d'un éditeur
- Dans le cadre de la rédaction d'un data-paper (Le data paper ou data article est une publication scientifique dont le but principal est de décrire un ou plusieurs jeux de données, plutôt que des résultats d'analyse)
- Autre

Si oui, sous quelle licence avez-vous placé ces données ?

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

La réponse était « Oui, régulièrement » ou « Oui, au moins une fois » à la question [Q45] (Avez-vous déjà publié ou diffusé des données ou jeux de données en libre accès sur le web, via une archive ouverte ou un entrepôt de données ?)

Veillez choisir toutes les réponses qui conviennent :

- La licence ouverte version 2.0 : etalab-2.0 (licence établie par le Gouvernement)
- L'ODC Open Database License (ODbL) version 1.0 : ODbL-1.0 (licence avec obligation de partage à l'identique)
- Je ne sais pas
- Autre

Par quel(s) moyen(s) diffusez-vous vos données « froides » sur le web ?

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

La réponse était « Oui, régulièrement » ou « Oui, au moins une fois » à la question [Q45] (Avez-vous déjà publié ou diffusé des données ou jeux de données en libre accès sur le web, via une archive ouverte ou un entrepôt de données ?)

Veillez choisir toutes les réponses qui conviennent :

- Un entrepôt de données (plateforme spécialisée dans l'archivage et la diffusion des données de recherche ; liste disponible des entrepôts sur re3data.org.)
- Une archive ouverte (de type HAL), associées à une publication
- Un serveur de votre (vos) établissement(s)
- Le site ou le blog de votre unité de recherche
- Le site dédié au projet de recherche
- Votre blog ou votre site personnel
- Un réseau social de chercheurs (ResearchGate, Academia...)
- Le site d'une revue (jeu de données associé à une publication)
- Autre

Si non pourquoi ?

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

La réponse était « Non » à la question [Q45] (Avez-vous déjà publié ou diffusé des données ou jeux de données en libre accès sur le web, via une archive ouverte ou un entrepôt de données ?)

Veillez choisir toutes les réponses qui conviennent :

- Non car je ne sais pas si j'en ai le droit
- Non, car je ne savais pas que c'était possible
- Non, je n'y avais pas pensé
- Non, et je n'en vois pas l'intérêt
- Non, car je ne sais pas le faire
- Non, car je n'ai pas le temps
- Autre

Documentation et description des données.

*** Avez-vous déjà participé à la rédaction d'un plan de gestion des données ?**

Un plan de gestion des données (PGD ou DMP - Data management plan) est un document évolutif qui aide le chercheur ou le chargé de projet de recherche à définir un plan pour gérer les données utilisées et générées dans le cadre de son activité ou de son projet de recherche.

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui, dans quel contexte ?

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

La réponse était « Oui » à la question [Q48] (Avez-vous déjà participé à la rédaction d'un plan de gestion des données ?)

Veillez écrire votre réponse ici :

Si oui, avez-vous été accompagné ?

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

La réponse était « Oui » à la question [Q48] (Avez-vous déjà participé à la rédaction d'un plan de gestion des données ?)

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Par qui ?

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

La réponse était « Oui » à la question [Q50] (Si oui avez-vous été accompagné ?)

Veillez écrire votre réponse ici :

*** Y a-t-il une politique formalisée de gestion et d'archivage des données de recherche au sein de votre laboratoire (charte, procédure, critères, modalités, nommage des fichiers...) ?**

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas
- Autre

Si oui, qui en est le/la responsable ?

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

La réponse était « Oui » à la question [Q51] (Y a-t-il une politique formalisée de gestion et d'archivage des données de recherche au sein de votre laboratoire (charte, procédure, critères, modalités, nommage des fichiers...) ?)

Veillez sélectionner une seule des propositions suivantes

- Une personne de votre équipe de recherche
- Un membre d'un service de soutien à la recherche (BU, DSI...)
- Un prestataire extérieur public

- Un prestataire extérieur privé
- Il n'y a pas de responsable identifié
- Autre

Connaissez-vous l'existence du pôle archives de votre établissement de rattachement ?

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Organisation sur la question des données

Combien de temps passez-vous à gérer vos données ?

Veillez sélectionner une seule des propositions suivantes :

- Pas assez de temps
- Juste ce qu'il faut, au fil de l'eau
- De façon intermittente
- Le temps nécessaire quand ce n'est plus gérable
- Trop de temps, par manque d'organisation personnelle /de procédures adaptées
- Beaucoup trop
- Autre

Avez-vous du personnel dédié à la gestion des données au sein de votre unité de recherche ?

Veillez choisir toutes les réponses qui conviennent :

- Oui
- Non
- Je ne sais pas
- Autre

Si oui, qui ?

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

La réponse était « Oui » à la question [Q55] (Avez-vous du personnel dédié à la gestion des données au sein de votre unité de recherche ?)

Veillez choisir toutes les réponses qui conviennent :

- Doctorant
- Enseignant-chercheur
- Personnel de soutien à la recherche (ITA ou ITRF)
- Post-doc
- Gestionnaire de laboratoire
- Vous
- Autre

*** De quels types de services ou soutiens auriez-vous besoin concernant la gestion de vos données de recherche ?**

Plusieurs réponses sont possibles. Nous vous remercions d'être le plus exhaustif possible afin de nous permettre de couvrir l'ensemble de vos besoins.

- Des conseils d'ordre général

- Une aide juridique
- Une aide à la gestion informatique de vos données
- Une aide à la rédaction de plan de gestion des données
- Une aide à la diffusion de vos données de recherche
- Une aide à la réutilisation / citation des données d'autres chercheurs
- Une aide à la pérennisation de vos données
- Autre

Pour [option de réponse], quels services vous semblent préférables ?

Pour toutes les options de réponses à la question [Q57] (De quels types de services ou soutiens auriez-vous besoin concernant la gestion de vos données de recherche ?), sauf « Autre ».

Veillez choisir toutes les réponses qui conviennent :

- Proposer une documentation en ligne
- Du conseil ou une assistance ponctuellement
- Organiser une formation sur ce sujet
- Organiser un atelier de groupe sur ce sujet
- Vous fournir un accompagnement individuel dans la durée
- Développer des outils informatiques qui vous aideront à ce sujet
- Autre

Acceptez-vous d'être contacté.e pour un entretien individuel, pour que nous en sachions plus sur vos besoins en termes de gestion des données de la recherche ?

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Si oui pouvez-vous nous donner votre adresse mail ?

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

La réponse était « Oui » à la question [Q58] (Acceptez-vous d'être contacté.e pour un entretien individuel, pour que nous en sachions plus sur vos besoins en termes de gestion des données de la recherche ?)

Veillez écrire votre réponse ici :

Avez-vous quelques remarques, suggestions ou questions à formuler ?

Veillez écrire votre réponse ici :

Annexe 2 : grille d'entretien

NB : demander si autorisation d'enregistrer l'entretien

Préciser les objectifs des entretiens qualitatifs

- Préciser les pratiques en matière de gestion de données et de *workflow*
- Vérifier s'il existe des différences entre les pratiques par disciplines
- Recueillir et vérifier les attentes en matière de formations / accompagnements

Partie systématique « socle commun » :

1) Identité de la personne et définition de l'objet du questionnaire

- Proposer à la personne de se présenter
- Lui demander de préciser ce que sont pour elle les données de la recherche
- (Vérifier qu'elle différencie science ouverte et gestion des données de la recherche)

2) Préciser ses pratiques / son workflow

- Demander à la personne comment elle gère ses données
- Existe-t-il des standards/normes propres à sa discipline ?

I. Partie variable selon le profil -> orienter vers les besoins et les attentes

1) Besoin d'un intranet projet par projet pour ses collaborateurs externes à Lyon 3 et lui-même.

- Définir une sorte de cahier des charges des infrastructures nécessaires (Qui sont les partenaires externes, quel volume de données, quels types de fichiers, accès sécurisé hors site ?)
- En l'absence de cet intranet, quelle(s) solution(s) a-t-il trouvée(s) ?
- A-t-il déjà réalisé un DMP au cours de ses projets ?

2) **Questionnaire** : se sent démuné devant cet aspect des choses

- Vous dites-vous sentir démuné > pouvez-vous préciser ? Sentiment personnel ou partagé ?
- Quelles seraient vos attentes / besoins > en termes de formation (sous quelles modalités) ? de ressources accessibles de manière autonome ?

3) **Questionnaire** : Globalement, l'accompagnement pour la capitalisation des données est dans les faits, pour l'instant, inexistant, alors que le potentiel est très important.

- Vous avez proposé le terme « capitalisation des données » plutôt que pérennisation, pouvez-vous le définir ? Expliciter la différence entre les deux ?
- Définir un cahier des charges de l'accompagnement pour la capitalisation des données (stockage, type d'institution, etc.) ?

4) **Questionnaire** : Dans le cadre des projets financés (ANR, ERC...) identifiés par le Service de la recherche, pour lesquels la livraison d'un plan de gestion de données est obligatoire et attendue à des durées fixes, il pourrait être pertinent que le Service de la recherche soit davantage proactif. On pourrait imaginer que, après qu'un projet a été sélectionné par son agence, le SGR envoie au porteur un document ou un mail indiquant en gros : vous allez devoir produire un PGD à tels intervalles, voici vos interlocuteurs, etc.

- Vous avez identifié un besoin d'accompagnement dans le PGD, pouvez-vous préciser ce besoin ? (Aide à la rédaction, rappel des dates, identification des interlocuteurs etc.)
- Vous êtes vous-même porteur d'un projet ANR, quelles sont vos attentes d'accompagnement à la gestion des données sur la durée ? (Incluant stockage, valorisation, archivage etc.)

5) Formations

- D'après vous, quel est le rôle des laboratoires dans la formation à la gestion sur les données de la recherche ?
- Quels seraient vos besoins de formation ?
- (Pour les doctorant.es) Avez-vous suivi la formation IsiDoc't ?
- Si oui et vous avez encore des questions / si non et vous avez des questions, vers qui vous tournez vous ?

6) Questionnaire juristes

- Quel type de données produisez-vous ?
- Comment gérez-vous les données sensibles ?

7) Questionnaire doctorants (si focus group, pas plus de 5 personnes [autour d'un plateau-repas ? suggestion OJC])

- Connaissance de la formation IsiDoc't ?
- Êtes-vous accompagnés dans la gestion de vos données de la recherche ?
- Avez-vous défini un *workflow* pour / avec vos données ?

8) Questionnaire enseignants-chercheurs (possibles HDR)

- Commun HDR et non HDR
- D'après vous, à qui appartiennent les données produites ?
- Dans vos recherches réutilisez-vous des données déjà existantes – comment les avez-vous obtenues ?
- Quand vous partirez (mutation, retraite) qu'advient-il de vos données ?
- Si HDR
- Comment accompagnez-vous les étudiants / doctorants que vous encadrez dans la gestion de leurs données ?
- Connaissez-vous la formation IsiDoc't proposée par la BU à destination des doctorants ?

9) Questionnaire directeur unités de recherche

- Quel est le rôle de l'UR dans la gestion des données ?
- L'UR propose-t-elle des infrastructures pour la gestion des données ? (stockage, valorisation)
- Quelle intégration des personnels de soutien à la recherche > qui sont les producteurs de données dans l'UR ?

Annexe 3 : services institutionnels de stockage en date (février 2023)

Serveurs	Sauvegardes	Accès partagé	Volume maximal	Accès distant
Espace disque de service – Université Lyon 3	Oui	À l'échelle du service	150 Go	Par VPN
Espace disque personnel – Université Lyon 3	Oui	Non	15 Go	Par VPN
Owncloud - Université Lyon 3	Non	Oui	50 Go	Oui
BUL – Boîte Université Lyon 2 (OwnCloud)	Oui	Oui Y compris externe à Lyon 2	10 Go (augmentation à la demande)	Oui
OneDrive - Université Lyon 2	Non	Oui	1 To pour OneDrive	Oui
Espace disque de service/composante (Partage) - Université Lyon 2	Oui	Par groupe, rattachement en fonction de SIHAM	50 Go par groupe (augmentation à la demande)	Par VPN
Stockage MSH (Nuxeo, solution remplacée dans un proche avenir)	Oui	Oui	10 Go (possible de l'augmenter)	Oui
Sharedoc (CNRS / Humanum)	Oui (au centre de calcul IN2P3) – pas de sauvegarde à l'extérieur	Oui Y compris externe	90 Go par espace individuel 300 Go par projet de recherche	Oui

Annexe 4 : liste des unités de recherche des répondants

Sigle	Label	Intitulé complet	Tutelle(s)
CEL	EA 1663	Centre d'études linguistiques	Univ. Lyon 3
CERCOR	UMR 85 84	Centre européen de recherche sur les communautés, congrégations et ordres religieux	CNRS
			Sorbonne Université
			EPHE
			Univ. St-Étienne
CERLA	EA 4162	Centre de recherche en linguistique appliquée	Univ. Lyon 2
CIHAM	UMR 56 48	Histoire archéologie littératures des mondes chrétiens et musulmans médiévaux	CNRS
			Univ. Lyon 2
			EHESS
			ENS de Lyon
			Univ. Lyon 3
			Univ. d'Avignon
CLHDPP	EA 669	Centre lyonnais d'histoire du droit et de la pensée politique	Univ. Lyon 3
CMW	UMR 52 83	Centre Max Weber	CNRS
			ENS de Lyon
			Univ. Lyon 2
			Univ. St-Étienne
COACTIS	EA 4161	Conception de l'action en situation	Univ. St-Étienne
			Univ. Lyon 2

CRDMS	EA 4588	Centre de recherche en droit et management des services de santé	Univ. Lyon 3
CRPPC	EA 653	Centre de recherche en psychopathologie et psychologie clinique	Univ. Lyon 2
DCT	EA 4573	Droits, contrats et territoires	Univ. Lyon 2
DDL	UMR 5596	Dynamique du langage	CNRS
			Univ. Lyon 2
DIPHE	UR	Développement, individu, processus, handicap, éducation	Univ. Lyon 2
ECP	EA 4571	Education, cultures, politiques	Univ. Lyon 2
			ENS de Lyon
			Univ. St-Étienne
EDIEC	EA 4185	Equipe de droit international européen et comparé	Univ. Lyon 3
EDPL	EA 666	Équipe de droit public de Lyon	Univ. Lyon 3
ELICO	EA 4147	Équipe de recherche de Lyon en sciences de l'information et de la communication	ENSSIB
			Sciences Po Lyon
			Univ. Lyon 1
			Univ. Lyon 2
			Univ. Lyon 3
ELJ	EA 3707	Équipe Louis Josserand	Univ. Lyon 3
EMC	UR	Étude des mécanismes cognitifs	Univ. Lyon 2
ERIC	EA 3083	Entrepôts, représentation et ingénierie des connaissances	Univ. Lyon 1
			Univ. Lyon 2

EVS	UMR 56 00	Environnement, ville, société	CNRS
			Univ. Lyon 2
			Univ. Lyon 3
			Univ. St-Étienne
			ENTPE
			ENS de Lyon
			INSA Lyon
			ENSA Lyon
			Mines St-Étienne
GATE	UMR 58 24	Groupe d'analyse et de théorie économique Lyon Saint-Étienne	CNRS
			Univ. St-Étienne
			Univ. Lyon 2
			ENS de Lyon
GREPS	EA 4163	Groupe de recherche en psychologie sociale	Univ. Lyon 2
HISOMA	UMR 51 89	Histoire et sources des mondes antiques	CNRS
			ENS de Lyon
			Univ. Lyon 2
			Univ. Lyon 3
			Univ. St-Étienne
ICAR	UMR 51 91	Interactions, corpus, apprentissage, représentation	CNRS
			Univ. Lyon 2
			ENS de Lyon
IETT	EA 418 6	Institut d'études transtextuelles et transculturelles	Univ. Lyon 3

IHRIM	UMR 53 17	Institut d'histoire des représentations et des idées dans les modernités	CNRS
			Univ. Lyon 2
			Univ. Clermont Auvergne
			Univ. St-Étienne
			Univ. Lyon3
			ENS de Lyon
			CNRS
			Univ. Pays de l'Adour
			Univ. Lyon 2
IRPHIL	EA 418 7	Institut de recherches philosophiques de Lyon	Univ. Lyon 3
LADEC	UR	Laboratoire d'anthropologie des enjeux contemporains	Univ. Lyon 2
			ENS de Lyon
LAET	UMR 55 93	Laboratoire aménagement économie transports	CNRS
			Univ. Lyon 2
			ENTPE Lyon
LARHRA	UMR 51 90	Laboratoire de recherche historique Rhône-Alpes	CNRS
			Univ. Grenoble Alpes
			Univ. Lyon 2
			Univ. Lyon 3
			ENS de Lyon
LCE	EA 185 3	Lettres et civilisations étrangères	Univ. Lyon 2

LER	EA 3728	Laboratoire d'études rurales	Univ. Lyon 2
			INRA
LIRIS	UMR 52 05	Laboratoire d'informatique en image et systèmes d'information	CNRS
			INSA Lyon
			Univ. Lyon 1
			Univ. Lyon 2
			EC Lyon
LVIS	UR	Laboratoire sur les vulnérabilités et l'innovation dans le sport	Univ. Lyon 1
MAGELLAN	EA 371 3	Centre de recherche Magellan	Univ. Lyon 3
MARGE	EA 371 2	Zones marginales de l'écriture et de l'image	Univ. Lyon 3
MOM	FR 3747	Maison de l'Orient et de la Méditerranée - Jean Pouilloux	CNRS
			Univ. Lyon 2
			Univ. Lyon 1
			Univ. Lyon 3
			ENS de Lyon
			Univ. St-Étienne
			Aix-Marseille Univ.
PASSAGES	EA 416 0	Passages, arts et littératures (XX-XXI)	Univ. Lyon 2
UMR1296	U 1296	Radiations : défense, santé, environnement	Univ. Lyon 2
			Inserm
SAF	EA 242 9	Sciences actuarielle et financière	Univ. Lyon 1

TRIANGLE	UMR 52 06	Action, discours, pensée politique et économique	CNRS
			Univ. Lyon 2
			Sciences Po Lyon
			ENS de Lyon
			Univ. St-Étienne