La gestion des données de la recherche au laboratoire LIttoral, ENvironnement et Sociétés (LIENSs) : pratiques et besoins.

Elodie Pétorin, Cecilia Pignon-Mussaud

▶ **To cite this version:**

Elodie Pétorin, Cecilia Pignon-Mussaud. La gestion des données de la recherche au laboratoire LIttoral, ENvironnement et Sociétés (LIENSs) : pratiques et besoins.. LIENSs (UMR 7266). 2025. hal-04869558

HAL Id: hal-04869558
https://hal.science/hal-04869558v1

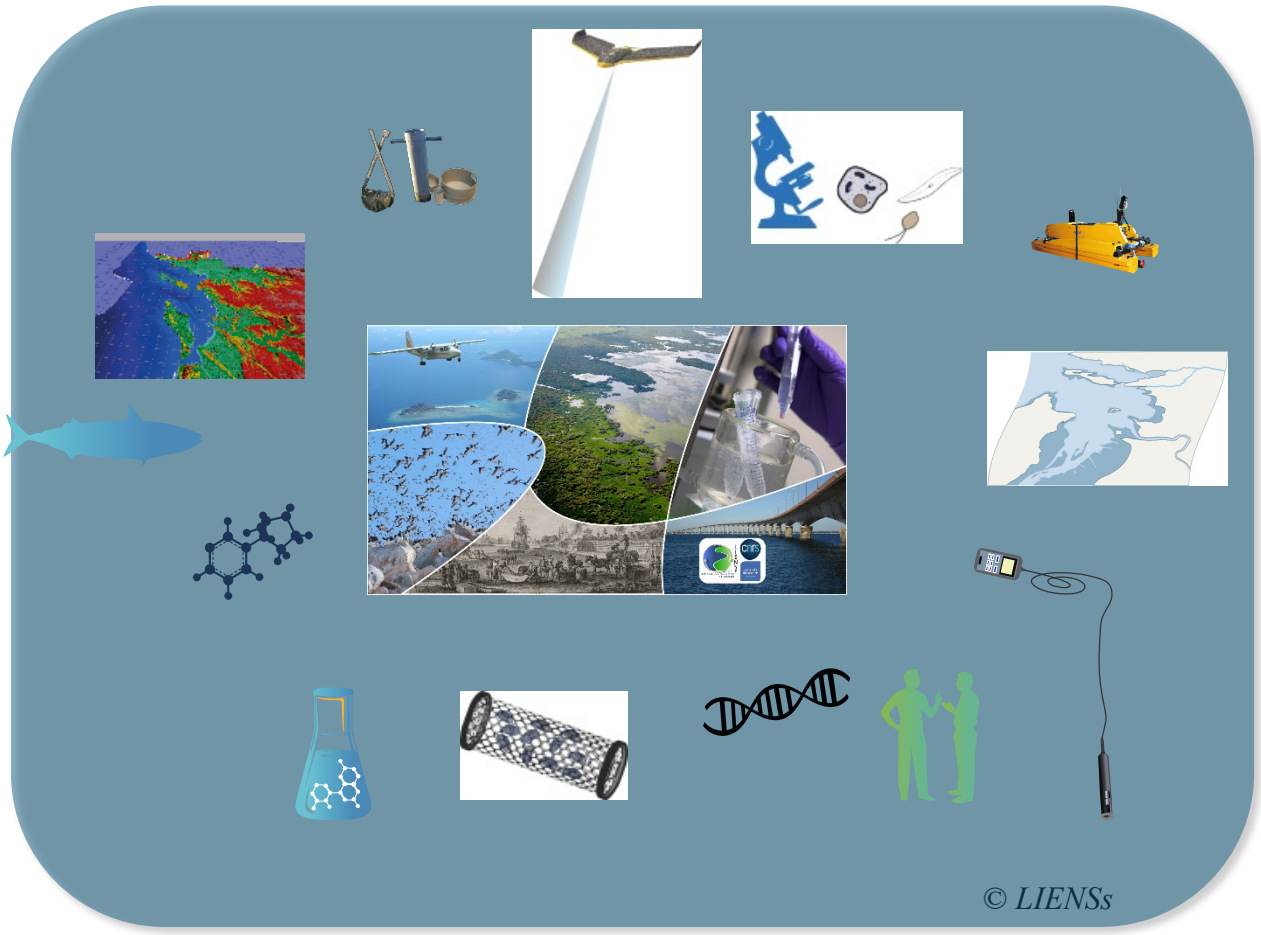
Submitted on 7 Jan 2025

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 - NonCommercial 4.0 International License

La gestion des données de la recherche au laboratoire LIttoral, ENvironnement et Sociétés (LIENSs) : pratiques et besoins.



Rapport de l'enquête réalisée en 2024

Rédigé par le Service d'Aide à la Gestion de la Donnée : Élodie Pétorin et Cécilia Pignon-Mussaud (LIENSs).

Table des matières

Introduction *3*

1. Les données de la recherche au cœur de la science *3*

1.1. Définition et enjeux des données de la recherche **3**

1.2. Le contexte de la Science Ouverte en France : un sujet d'actualité **4**

2. Présentation et spécificités du LIENSs *5*

2.1. Un laboratoire interdisciplinaire pour mieux comprendre le littoral **5**

2.2. Une pluralité et une diversité des données du LIENSs **7**

2.3. Un portail de données pour mieux connaitre les données du LIENSs **8**

3. L'enquête 2024 : connaître les pratiques et les besoins en terme de gestion et de diffusion des données de la recherche *9*

3.1. Objectifs **9**

3.2. Méthodologie et modalités de l'enquête **9**

3.3. Echantillon, nombre et profil des répondants **10**

4. Principaux résultats et synthèse de l'enquête : il reste à FAIRe *13*

4.1 Typologie des données produites **13**

4.2 Pratiques de stockage **15**

4.3 Pratiques de partage **18**

4.4 Les besoins **24**

Conclusion *30*

Liste des figures *31*

Annexe 1 : Questionnaire *32*

Annexe 2 : Autres résultats de l'enquête *41*

Introduction

Cette enquête a été initiée par le Service d'Aide à la GEstion de la DOnnée (SAGEDO) du laboratoire LIENSs. Ce service, créé en 2023 a pour objectif, d'une part, de répondre aux souhaits de la Direction actuelle et d'autre part, de répondre aux engagements pris lors de la dernière évaluation de l'HCERES (2020), en terme de gestion des données du laboratoire. Il a pour mission d'homogénéiser les pratiques et d'organiser au mieux cette gestion, c'est-à-dire de la création à l'archivage, en passant par le stockage et la sauvegarde des données. La mise en place d'une politique de gestion des données permettra, entre autre, d'optimiser la réutilisation des données, mais aussi d'être un des leviers pour renforcer l'interdisciplinarité au sein du laboratoire.

Après un rappel d'une définition des données de la recherche et de ses enjeux en première partie, une présentation du laboratoire et de sa spécificité sera développée dans un deuxième temps. La troisième partie de ce rapport sera consacrée à la génèse, la méthodologie et la construction de cette enquête sur les pratiques de gestion des données au sein du laboratoire. Enfin, les principaux résultats de l'enquête seront présentés, ainsi qu'une conclusion de ce travail de synthèse.

1. Les données de la recherche au cœur de la science

1.1. Définition et enjeux des données de la recherche

Il y a une coexistence de plusieurs définitions des données de la recherche ; la plus communément admise est celle de l'OCDE, révisée en 2021, pour y ajouter la notion de financements publics. C'est cette définition que nous utilisons dans nos présentations et que nous avons utilisé dans cette enquête pour construire le questionnaire.

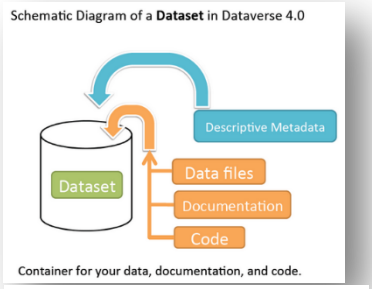
Les **données de la recherche** sont définies « *comme des enregistrements factuels (tels que les chiffres, textes, images et sons) issus de la recherche financée tout ou partie sur fonds publics, qui sont utilisés comme source principale pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche* »[1].

Ces enregistrements factuels peuvent être souvent regroupés pour former un ensemble cohérent de données (brutes, dérivées ou traitées), formant ainsi un jeu de données scientifiques (*dataset*). Un jeu de données est ainsi une agrégation d'enregistrements organisés pour former un ensemble de données homogènes, produits dans le cadre d'un même projet, sur un même objet d'étude et/ou recueillis sur un même lieu. Les jeux de données numériques sont formatés de telle sorte qu'ils soient communicables, interprétables et adaptés à un traitement informatisé. Pour être utilisé et

[1] Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics (https://doi.org/10.1787/9789264034020-en-fr)

cité, un jeu de données doit être accompagné de métadonnées descriptives : titre, description, créateur, date, etc….

Une base de données numérique (*database*) est constituée par un ensemble de jeux de données organisés et structurés pour être accessibles et exploitables au moyen d'un programme informatique.



https://dataverse.org

Par ailleurs, les données de la recherche font ressortir des enjeux quant à leur gestion et leur ouverture qui sont au nombre de cinq[2] :

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

Cette définition et ses enjeux structurent actuellement le paysage de la recherche en France et représentent une volonté forte du ministère et des acteurs de la Recherche.

1.2. Le contexte de la Science Ouverte en France : un sujet d'actualité

La Loi pour une République numérique,[3] datant de 2016, a pour ambition d'encourager l'innovation et le développement de l'économie numérique, de promouvoir une société numérique ouverte, fiable et protectrice des droits des citoyens. Elle vise également à garantir l'accès de tous, dans tous les territoires, aux opportunités liées au numérique. Par la suite, le premier plan national pour la Science Ouverte pour la période 2018-2021 a été mis en place et renouvelé pour la période 2021-2024[4]. Celui-ci prolonge les axes sur lesquels le premier plan avait commencé à travailler :

[2] Ouvrir la Science. *Les enjeux de la science ouverte*. 2023. Consulté sur : https://www.ouvrirlascience.fr/wp-content/uploads/2023/10/Livret_pour_impression.pdf
[3] Dossier de presse. *Loi pour une république numérique*. 2016. Consulté sur : https://www.economie.gouv.fr/files/files/PDF/DP_LoiNumerique.pdf
[4] MESR. *Deuxième Plan national pour la science ouverte*. 2021. Consulté sur : https://www.ouvrirlascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte_2021-2024.pdf

- Généraliser l'accès ouvert aux publications : *« La recherche scientifique est un bien commun que nous devons partager avec tous. »*
- Structurer, partager et ouvrir les données de la recherche : *« Les données de la recherche sont la matière première de la connaissance. Les partager, c'est ouvrir de nouvelles perspectives scientifiques. »*
- Ouvrir et promouvoir les codes sources produits par la recherche : « *L'ouverture des codes sources des logiciels est un enjeu majeur de reproductibilité des résultats scientifiques. »*
- Transformer les pratiques pour faire de la Science Ouverte le principe par défaut : « *L'ouverture des codes sources des logiciels est un enjeu majeur de reproductibilité des résultats scientifiques. »*

Le Comité pour la Science Ouverte (CoSo) est un acteur majeur de la Science Ouverte. Il accompagne les acteurs de l'enseignement supérieur et de la recherche dans la mise en œuvre de la politique nationale de Science Ouverte. Il se compose de plusieurs instances qui proposent des orientations, instruisent les dossiers, effectuent des arbitrages, impulsent et accompagnent les actions associées. Il constitue une structure fluide, facilitant l'expression et la remontée des idées, les engagements et les contributions. Il rédige, publie et diffuse en direction de tous les acteurs de la science, des guides pour promouvoir la Science Ouverte ; pour n'en citer que quelques uns : le guide pédagogique[6] sur les Données de la Recherche, la note sur le choix d'un entrepôt thématique[7] et le passeport pour la Science Ouverte à destination des doctorants.[8]

2. Présentation et spécificités du LIENSs

2.1. Un laboratoire interdisciplinaire pour mieux comprendre le littoral

Le laboratoire LIENSs est une unité mixte de recherche interdisciplinaire (UMR 7266) avec une double tutelle La Rochelle Université – CNRS), et rattaché à 3 instituts du CNRS : CNRS Écologie & Environnement - INEE, CNRS Sciences humaines & sociales – InSHS et CNRS Terre & Univers - INSU. Il met l'interdisciplinarité au service de la compréhension du fonctionnement de la zone littorale et son évolution dans un contexte de changements globaux, de la connaissance des usages du littoral par les sociétés qui l'habitent, de sa valorisation, ainsi que des enjeux du développement durable en zone littorale.

[6] CoSo. *Données de la recherche*. 2024. Consulté sur : https://www.ouvrirlascience.fr/wp-content/uploads/2024/03/24-02-22-Donnees-FR-WEB.pdf
[7] Frédéric de Lamotte, Véronique Stoll, Cécile Arènes, Marie-Emilia Herbet, Stéphane Debard, et al.. Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante. Comité pour la Science Ouverte. 2024. ⟨hal-04534321v1⟩
[8] MESR. *Passeport pour la Science Ouverte*. 2024. Consulté sur : https://www.ouvrirlascience.fr/passeport-pour-la-science-ouverte-guide-pratique-a-lusage-des-doctorants/

Il intègre les compétences de nombreuses disciplines, des sciences de l'environnement aux sciences humaines en passant par la chimie et les biotechnologies.

Le LIENSs est composé de 6 équipes à dominante thématique :

- Équipe AGÎLE (**A**pproche **G**éographique : **I**les, **L**ittoraux, **E**nvironnement) : équipe interdisciplinaire mobilisant des compétences en Sciences Humaines et Sociales (SHS) dans trois domaines qui s'inscrivent tous dans une démarche territorialisée, interdisciplinaire, systémique et multiscalaire. Si l'approche géographique reste centrale au sein de l'équipe, elle est également articulée à des approches juridiques et sciences politiques.
- Équipe AMARE (Réponses des Animaux MARins à la variabilité Environnementale) : L'équipe vise à apporter, à l'aide d'organismes modèles, une vision intégrative des processus adaptatifs en milieu littoral à la variabilité environnementale.
- Équipe BCBS (Biotechnologies et Chimie des Bioressources pour la Santé) : L'équipe regroupe plusieurs thématiques de recherche à l'interface des biotechnologies, de la chimie et de la biologie, sur les ressources marines.
- Équipe BIOFEEL (BIOdiversité et Fonctionnement des EcosystEmes Littoraux) : L'équipe étudie la structure et le fonctionnement des écosystèmes littoraux essentiellement par une approche en écologie trophique. Elle mobilise des approches de suivis et d'expérimentations de terrain, de laboratoire, d'analyses multivariées et de modélisation.
- Équipe DPL (Dynamique Physique du Littoral) : L'équipe DPL vise à mieux comprendre la dynamique physique des littoraux, du point de vue des processus physiques (contrôlant la dynamique hydro-sédimentaire et leurs évolutions morphologiques) et des changements environnementaux récents qui modifient ces systèmes.
- Équipe ESTRAN (Espaces, Sociétés, Territoires des Rivages Anciens et Nouveaux) : L'équipe ESTRAN est une équipe pluridisplinaire composée d'historiens et de géophysiciens travaillant sur les sociétés littorales et leurs rapports à la mer.

Et de deux pôles de soutien à la recherche :

- Le pôle administratif et financier
- Le **pôle technique de soutien à la recherche** : Pour répondre à la transdisciplinarité des recherches menées, le LIENSs dispose de compétences et de moyens techniques très diversifiés. Les personnels BIATSS[9] (La Rochelle Université) et ITA[10] (CNRS), apportent un appui technique et analytique aux recherches menées par le laboratoire au sein des plateformes analytiques et techniques, des plateaux techniques ou encore en soutien aux équipes de recherche (analyses, missions de terrain, appui logistique ...).

Le Service d'Aide à la GEstion de la DOnnée (SAGEDO) est **un service commun transverse, intégré au pôle technique de soutien à la recherche, et à disposition de tous.** Sa mission principale est une **aide à la mise en œuvre d'une politique de gestion de données** au sein du laboratoire en **conseillant et accompagnant** les membres du laboratoire. Le service accompagne

[9] Bibliothèques, Ingénieurs, Administratifs, Techniciens, Sociaux et de Santé
[10] Ingénieurs, techniciens et personnels administratifs

les membres du LIENSs, tout le long du cycle de vie de la donnée, en aidant à la rédaction de plans de gestion de données (PGD), en conseillant sur des choix de stockage, en accompagant à la rédaction de métadonnées, en aidant à la documentation des données, au choix d'entrepôt de données, etc … Le service souhaite homogénéiser les processus de partage, de diffusion des données de la recherche et répondre ainsi, au plus près aux principes FAIR[11] (Findable, Accessible, Interoperable, Reuseable).

Ce service est composé de trois personnes, deux qui partagent leur temps de travail entre ce service et leur mission première d'ingénieure d'études (rattachées au pôle technique de soutien à la recherche) et une autre personne, directrice de recherche en tant que responsable du service.

2.2. Une pluralité et une diversité des données du LIENSs

La pluridisciplinarité du laboratoire génère, de fait, une pluralité et une diversité des données. Il s'agit essentiellement de :

- **Données d'observation,** qui sont des données capturées en temps réel ; habituellement uniques et donc impossibles à reproduire (exemples: des relevés de température ou de précipitations, données de télédétection, photos d'un événement, données d'enquêtes, etc.). Elles sont essentiellement acquises dans le cadre :
  - des Services Nationaux d'Observation (SNO) de l'IR-ILICO (Infrastructure de Recherche Littorale et COtière) ; le LIENSs est coordinateur de 2 SNO (SONEL et DYNALIT) et contribue également aux SNO SOMLIT, PHYTOBS et BenthOBS ;
  - d'Observatoires régionaux (OBIONE, ECOP), observations réalisées par la plateforme drones aériens et marins ;
  - de nombreux projets de recherche, en métropole, outremer ou à l'international (bouées GPS, prélèvements biologiques, etc…).
- **Données expérimentales :** obtenues à partir d'équipements de laboratoire, souvent reproductibles, mais parfois coûteuses. Par exemple, il peut s'agir de données obtenues suite à l'exposition et à des tests sur des organismes des milieux littoraux, dans des conditions environnementales contrôlées (température, concentrations métalliques, oxygène, dioxyde de carbone, etc…).
- **Données dérivées ou compilées :** issues du traitement ou de la combinaison de données "brutes", souvent reproductibles, mais coûteuses (exemples : fouilles de texte, bases de données compilées)
- **Données d'enquêtes et de sondage, données d'archives et données audiovisuelles :** issues de questionnaires ou entretiens, films, enregistrements audio lors d'enquêtes. Elles sont uniques et donc impossible à reproduire.
- **Les codes sources et modèles** ne doivent pas être considérés comme des données proprement dites, car ils présentent des enjeux, des pratiques et des recommandations de

[11] Urfist Méditerranée. *Les principes FAIR*. 2019. Consulté sur : https://doranum.fr/enjeux-benefices/principes-fair_10_13143_z7s6-ed26/ (DOI : 10.13143/z7s6-ed26)

partage et d'ouverture particuliers. Il existe un guide du CoSo dédié à l'ouverture des codes et logiciels[12] produits dans le cadre de la recherche scientifique.

2.3. Un portail de données pour mieux connaitre les données du LIENSs

Depuis 2011, le LIENSs dispose d'un portail de données reposant sur l'infrastructure de données géographiques marines et littorales, dénommée Sextant et développée par l'IFREMER. L'objectif premier a été d'initier un inventaire des données environnementales du laboratoire, afin d'être en conformité avec la Directive INSPIRE, mais également de favoriser la pluridisciplinarité du laboratoire, de respecter au mieux les recommandations pour l'accès aux données de la recherche (plan national de la science ouverte) et les principes FAIR. Ce portail comprend 1) un catalogue de métadonnées (GeoNetwork) normalisé conformément aux normes ISO 19115 et ISO 19139 et aux spécifications d'INSPIRE, 2) un service de visualisation et de téléchargement (services de l'OGC).

En 2024, plus de 450 fiches de métadonnées y sont disponibles. Mais, comme le présente le graphique ci-dessous, les services de visualisation cartographique (WMS/WFS) et le téléchargement des données sont peu nombreux. Sur la diffusion des données, les réponses obtenues lors de l'enquête (partie 4.3) explicitent les freins et réticences des membres du laboratoire.

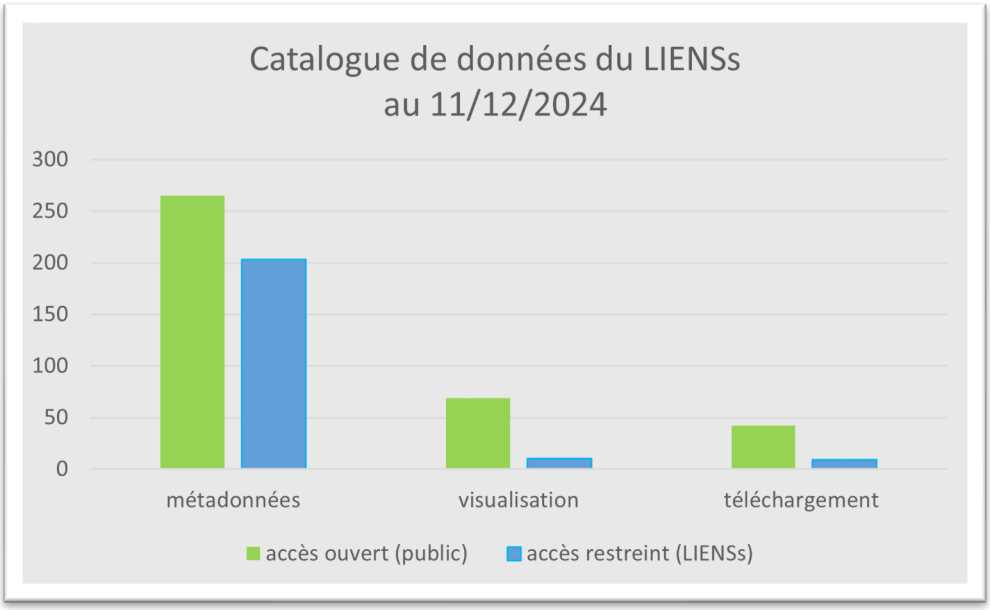


Figure 1 - Statistiques du catalogue du LIENSs

[12] CoSo, *Codes et logiciels*. 2022. Consulté sur : https://www.ouvrirlascience.fr/wp-content/uploads/2022/10/Passeport_Codes-et-logiciels_WEB.pdf

A noter, que pour des raisons internes au laboratoire, pendant longtemps la nécessité d'attribuer un DOI aux données n'était pas jugée prioritaire. D'ici la fin de l'année 2024, l'attribution de DOI pourra être demandée directement depuis le portail auprès du pôle ODATIS pour les données d'observation sur l'océan et le littoral, qui est un des pôles thématiques de l'infrastructure de recherche Système Terre Data Terra.

A noter que le catalogue du LIENSs est actuellement moissonné par plusieurs catalogues, tels que celui du géocatalogue national, de l'Ifremer (Sextant), du pôle ODATIS, de PIGMA (Plateforme d'échange de données de Nouvelle-Aquitaine) et du SNO DYNALIT (réseau d'observation sur la dynamique du littoral).

3. L'enquête 2024 : connaître les pratiques et les besoins en terme de gestion et de diffusion des données de la recherche

3.1. Objectifs

Les objectifs principaux de cette enquête sont :

- d'établir un état des lieux de la gestion actuelle des données de recherche au LIENSs ;
- de proposer des réponses aux besoins exprimés dans l'enquête, en termes d'accompagnement, de formation et de ressources ;
- d'alimenter la recherche et la réflexion sur les pratiques ;

Les résultats permettront de définir et de hiérarchiser les priorités, en terme de gestion des données au sein du laboratoire et de les planifier au travers d'un plan de gestion de données d'entité (PGDE). Ce PGDE sera réalisé à partir de l'outil DMP OPIDoR, un outil en ligne pour rédiger des PGD. Il définira comment sont gérées, au sein du LIENSs en tant qu'entité de recherche[13], les données générées ou provenant de tiers. Ce plan vise à faire adopter les bonnes pratiques de gestion/ouverture des données, à augmenter la qualité des services rendus par le laboratoire, et ainsi à contribuer à la « fairisation » des données et services.

3.2. Méthodologie et modalités de l'enquête

L'enquête s'est déroulée du 14 mars au 2 mai 2024 avec 3 mails de relance sur la période. Le lien du questionnaire a été envoyé sur l'adresse générique du laboratoire comprenant tous les membres

[13] *Une entité de recherche (ou structure) peut correspondre à une plateforme technique, infrastructure ou observatoire de recherche, unité de recherche, laboratoire, centre de ressources biologiques ou génétique.* (source : OPIDOR)

à la date d'envoi (y compris les stagiaires, invités et contractuels). Ainsi, tous les membres du laboratoire ont été invités à le remplir, soit 189 personnes en avril 2024 (à noter que le pôle administratif et financier n'était pas directement concerné par cette enquête, 6 personnes).

Le questionnaire de l'enquête a été construit par l'équipe du SAGEDO, s'inspirant d'une enquête similaire menée par l'Université de Rennes 2[14]. Il a été réalisé sur LimeSurvey et contient 17 questions, toutes obligatoires. 11 questions sont à choix unique et 6 à choix multiples. A noter que 10 questions offrent la possibilité de choisir la modalité « autre » et d'y apporter une précision en champ libre. L'ensemble des questions de l'enquête est disponible en annexe 2.

Après les éléments relatifs au profil des répondants, quatre axes thématiques principaux structurent l'ensemble de l'enquête :

- la typologie des données collectées et produites ;
- les pratiques de stockage et d'archivage des données ;
- les pratiques de partage et de diffusion des données ;
- les besoins et attentes, liés à la gestion des données de recherche ;

Nous avons fait le choix de présenter les principaux résultats ci-dessous. Dans le but de ne pas surcharger ce document, les autres résultats que nous pensons moins significatifs sont disponibles en annexe 2.

3.3. Echantillon, nombre et profil des répondants

56 réponses complètes ont été enregistrées et sont exploitables. 27 sont incomplètes et donc non exploitables. Le taux de participation de 28 % est jugé satisfaisant[15], au regard du taux de réponse des enquêtes en ligne jugées comme telles pour des taux de réponse compris entre 20 et 30 %. Nous considérons ainsi que les résultats de cette enquête sont suffisament représentatifs et peuvent donner une idée et un panorama des pratiques, besoins et attentes des répondants. Les résultats présentés ci-dessous reprennent la numérotation des questions de l'enquête (annexe 1).

A1) Vous êtes ?

Sur 56 répondants, ce sont essentiellement les enseignants-chercheurs (34%) et les chercheurs (18%) qui ont répondu à l'enquête, soit 52% des répondants. Le personnel technique en appui à la recherche représentant un peu moins de 30%.

[14] Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet. Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2. [Rapport de recherche] Université Rennes 2. 2017, 159 p., 47 p., 114 p., 26 p., 23 p. ⟨hal-01635186v2⟩

[15] Aragon Yves, Bertrand Sandrine, Cabanel Magali, Legrand Hervé. *Méthode d'enquête par internet : Leçons de quelques expériences*. Décisions Marketing, 2000, n° 19, p. 29-37.

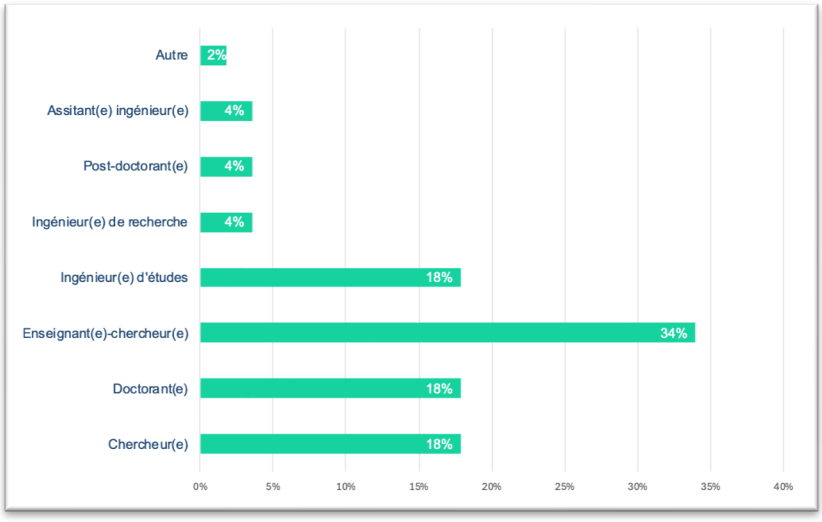
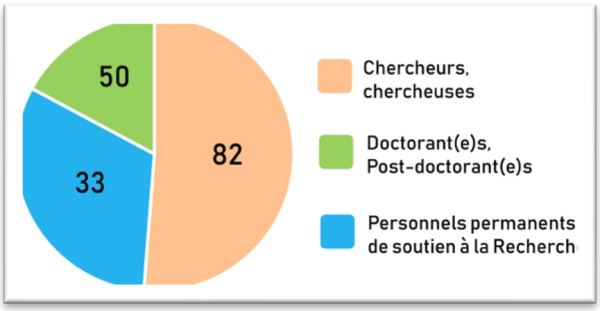


Figure 2 - Répartition par statut des répondants au questionnaire

Sur le site du laboratoire, la répartition au 1er janvier 2024 était la suivante :



Nous avons donc une représentation assez homogène entre ces statuts, et qui reflète assez bien la représentativité du personnel du laboratoire, comme présenté ci-dessous.

Statuts	Répartition du personnel (au 01/01/2024)	Répondants à l'enquête 2024
Chercheurs, Enseignants-chercheurs	50 %	52 %
Doctorants, Post-doctorants	30 %	22 %
Personnels techniques	20%	26%

A2) Vous faites partie de quelle équipe au sein du LIENSs ?

Le LIENSs est composé de 6 équipes thématiques, d'un pôle administratif et financier ainsi que d'un pôle technique de soutien à la recherche qui regroupe le personnel technique en appui à la recherche (dont les membres du SAGEDO). Ce graphique comparatif entre les données de l'enquête et les données administratives du LIENSs peut nous permettre de voir que l'ensemble des équipes est assez bien représenté, sauf pour l'équipe BCBS qui est sous représentée et l'équipe AMARE qui est sur-représentée.

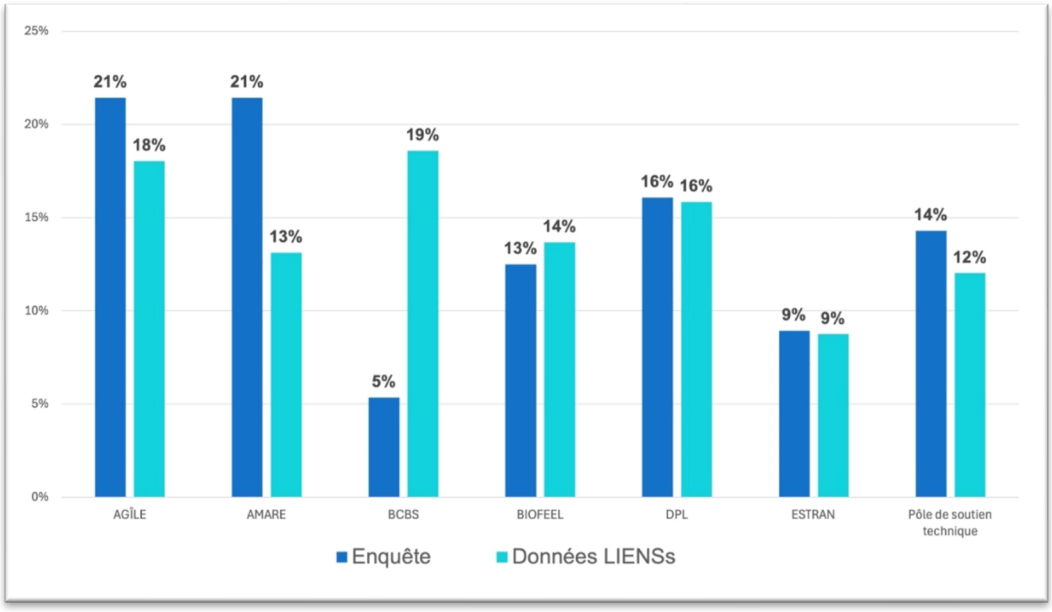


Figure 3 - Répartition par équipes entre les répondants à l'enquête et les données du LIENSs

Les réponses à ces deux premières questions étant très proches de la composition actuelle du LIENSs à la diffusion du questionnaire, nous pouvons donc signifier que les résultats de l'enquête sont significativement représentatifs. Ainsi, les résultats qui suivent peuvent être vus comme une capture en temps réel des pratiques actuelles des membres du laboratoire.

4. Principaux résultats et synthèse de l'enquête : il reste à FAIRe

4.1 Typologie des données produites

B1) Dans quelle(s) catégorie(s) suivantes classeriez-vous vos données produites ?

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

Les données d'observation et expérimentales représentent les principales données créées au laboratoire pour près de la moitié des répondants. Mais, à noter que tous les types de données sont aussi produits.

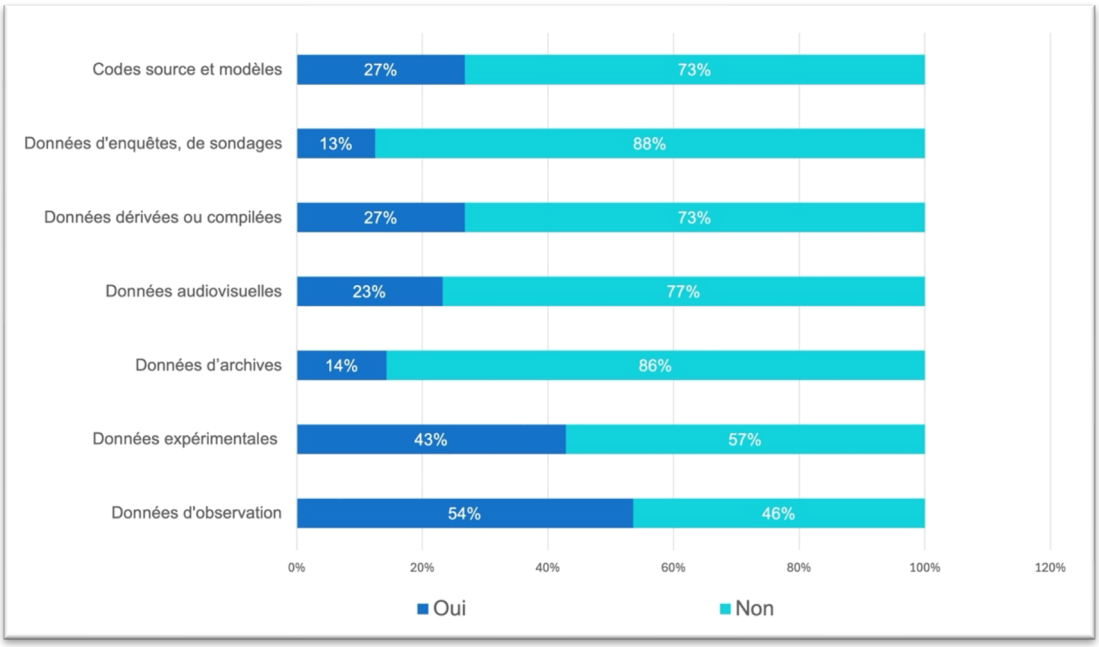


Figure 4 - Types de données produites

Cette diversité montre toute la richesse des données créées au LIENSs, mais elle complexifie aussi leur gestion. Ces données répondent à des normes et standards différents pour leur description, leur documentation, leur structuration et leur diffusion. Il ne peut y avoir de réponse unique et unifiée.

B3) Concernant leur support matériel, vos données produites sont-elles majoritairement ?

Cette question cherche à savoir comment la donnée est produite, et plus précisement, s'il s'agit de données numériques. Près de la moitié (45%) des données sont produites de manière numérique. Il est assez surprenant de voir que 41% des données produites sont numérisées par la suite, cela augmente sensiblement le risque de perte de données, tels que des risques de retranscription de données relevées sur le terrain, par exemple. Les données sont pour 86% d'entre elles, sur des supports numériques. Cela pose donc la question du stockage et de la sauvegarde de ces données numériques, que nous allons voir dans les parties suivantes.

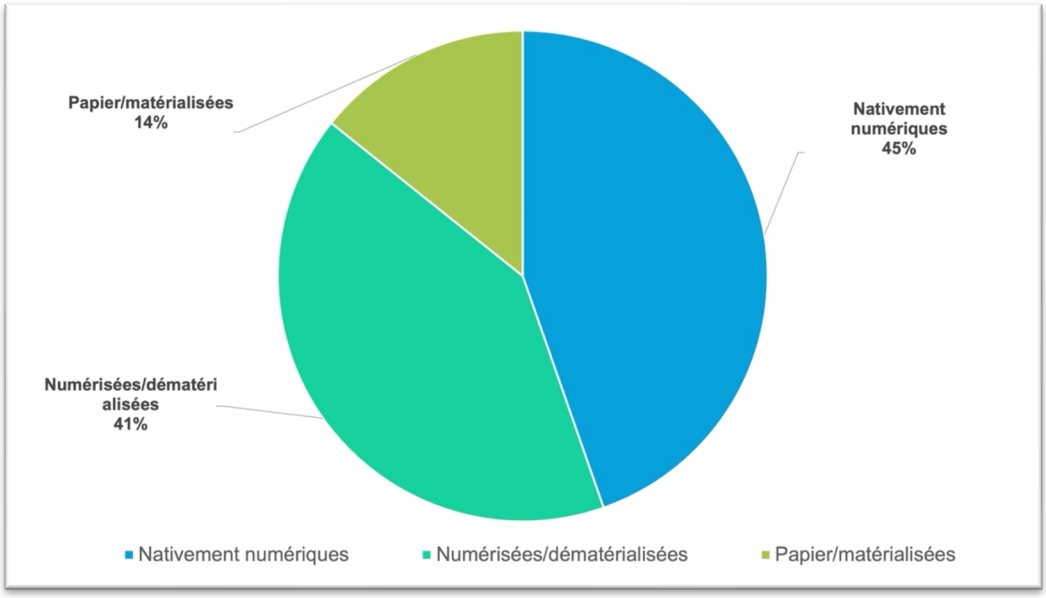


Figure 5 - Type de support des données produites

B4) Si vos données sources sont papier/matérialisées, est-ce que vous les numérisez en général ?

Cette question filtre a pour objectif de préciser le traitement des données papier/matérialisées qui sont produites. La production de données matérielles et/ou papier est tout de même infime (14%) au sein du laboratoire et elles sont quasiment toutes numérisées (88% de Oui).

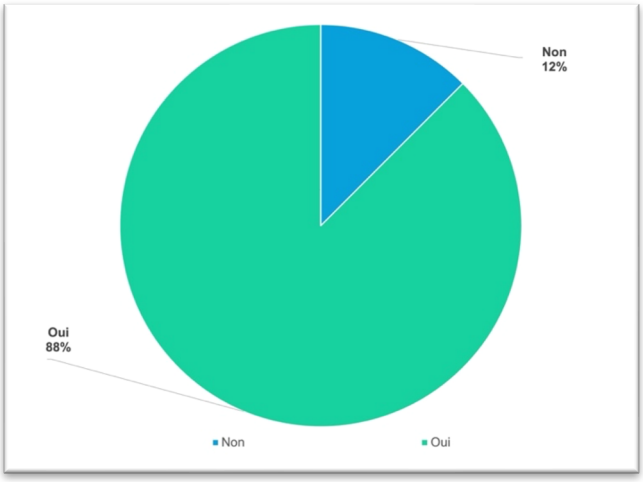


Figure 6 - Numérisation des données papier et/ou matérialisées.

Dans le futur, il serait intéressant de connaître quels types de données non numériques sont produites. Nous pouvons supposer qu'il s'agit des carnets de laboratoire, qui sont incontournables dans la traçabilité des données de recherche. Des solutions numériques existent avec les carnets de laboratoire électronique mis en place par le CNRS (guide de bonnes pratiques du réseau Qualité en Recherche) et utilisable dans tous les domaines de recherche.

4.2 Pratiques de stockage

C1) A combien estimez-vous l'espace nécessaire au stockage de vos données ?

Une des questions importantes au laboratoire est celle de l'espace de stockage nécessaire pour conserver et préserver les données du laboratoire. Les résultats de l'enquête montrent ici des besoins de stockage assez hétérogènes, avec des espaces de stockage relativement grands. Majoritairement (52%), il s'agit de volumes supérieurs à 100 Go, dont la moitié supérieure à 1To par personne.

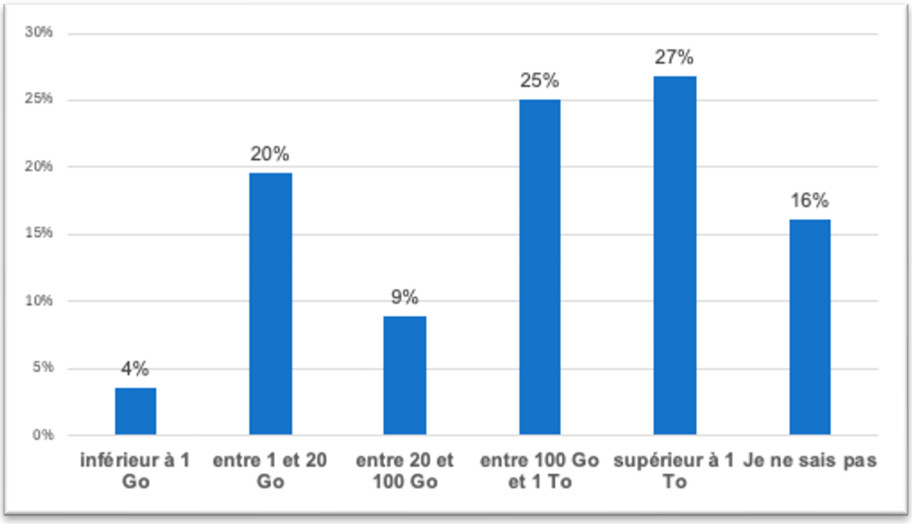


Figure 7 - Estimation du volume de stockage de données nécessaire par personne.

Excepté des espaces de stockage partagés proposés par La Rochelle Université, il n'existe pas, à notre connaissance, de solutions permettant de répondre à des demandes de stockage supérieurs à 1 To avec un espace de travail collaboratif sur des serveurs institutionnels. Il peut exister des solutions de stockage supérieure à 1 To mais elles sont souvent proposées par des serveurs de calculs (par exemple le CC-IN2P3 ou Datarmor).

A noter ici qu'il n'existe pas encore d'espace de stockage partagé à l'échelle du laboratoire.

C2) Où stockez-vous les données de recherche ?

La question du stockage des données est essentielle, pour assurer la conservation des données et leur sécurité dans le cadre des projets de recherche. Cette question se limite donc au stockage sécurisé et non à l'archivage de données, qui est une conservation sur du moyen et long terme (plus de 30 ans)[16].

Une grande majorité des répondants a opté pour un stockage local et personnel : l'ordinateur professionnel et le disque dur externe arrivent en tête ex-aequo avec 82%. Les serveurs ou plateformes institutionnelles proposés par le CNRS ou La Rochelle Université arrivent en troisième position avec 39% des personnels qui l'utilisent. Près de 30% utilisent des serveurs d'une autre institution ou d'une Infrastructure de Recherche (exemple : Huma-Num IR pour les SHS, mais aussi de plus en plus pour les projets pluridisciplinaires). A noter que certains programmes de recherche bénéficient d'espaces de stockage partagés, financés par le projet, mais bien souvent non pérennes, faute de financements continus.

Enfin, presque 20% utilisent leur ordinateur personnel, nous pouvons supposer ici, qu'ils dupliquent leurs données de recherche sur un autre espace de stockage (un disque dur externe ou un autre serveur, par exemple).

[16] Inist-CNRS. *Stockage et archivage : fiche synthétique*. 2017. Consulté sur : https://doranum.fr/stockage-archivage/stockage-et-archivage-fiche-synthetique_10_13143_0c4b-2743/ (DOI : 10.13143/0c4b-2743)

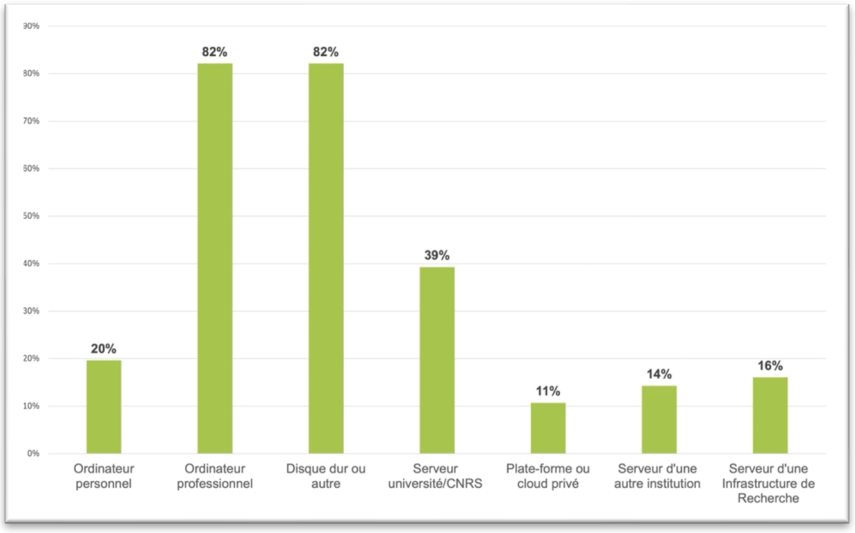


Figure 8 - Lieu de stockage des données

A ce jour, selon la tutelle, les membres du laboratoire peuvent disposer individuellement de diverses solutions de stockage, 300 Go avec NextCloud côté La Rochelle Université et 100 Go OwnCloud/SDrive, côté CNRS.

C4) Faites-vous des sauvegardes de vos données ?

La sauvegarde est un processus fondamental et incontournable dans la conservation des données. Elle est gérée individuellement, par chaque membre du laboratoire. Il est donc indispensable de connaitre les pratiques de sauvegarde.
Un peu plus de la moitié (61%) des répondants réalise une sauvegarde manuelle de leurs données et 11% seulement le font automatiquement. 23% utilisent les deux méthodes de sauvegarde. Par ailleurs, 5% des répondants ne font aucune sauvegarde de leurs données.

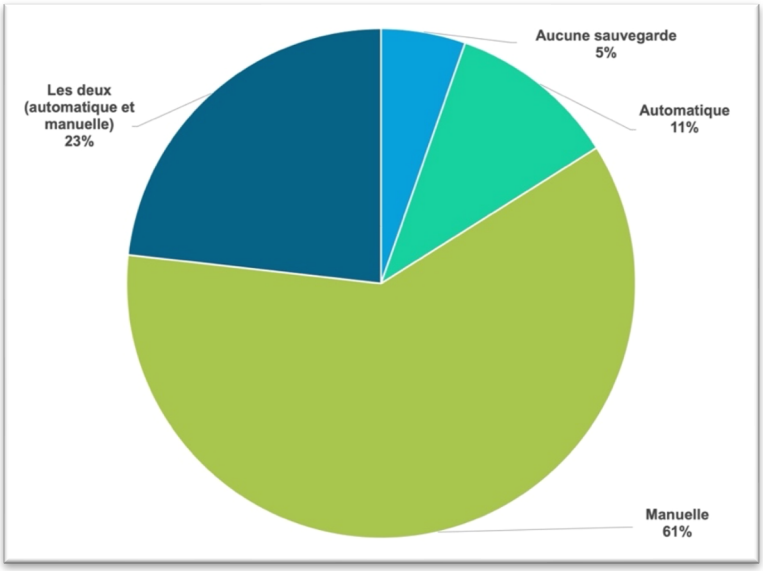


Figure 9 - Types de sauvegarde

C5) A quelle fréquence les données sont-elles sauvegardées ?

Nous souhaitions poser cette question de la fréquence des sauvegardes des données, suite à la volonté de la Direction de mettre en place une solution de sauvegarde automatique et systématique des postes de travail (solution non disponible pour le moment à LRU). Les résultats de la question précédente peuvent paraitre satisfaisants. Mais, lorsque la fréquence de ces sauvegardes est demandée, quelques inquiétudes légitimes peuvent être exprimées : plus d'1/3 font des sauvegardes tous les mois voir plus d'un mois, 1/3 en fonction des besoins, ce qui est relativement flou. Enfin, près de 28% réalisent des sauvegardes a *minima* toutes les semaines voir tous les jours. Tandis que la très grande majorité des répondants réalisent des sauvegardes de manière irrégulière.

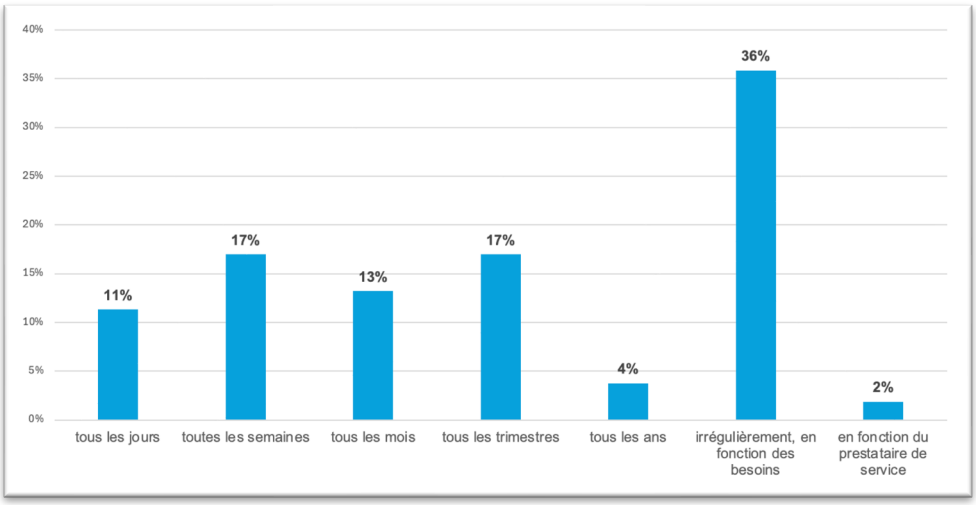


Figure 10 - Type de fréquence des sauvegardes

Plusieurs réflexions ont été lancées pour trouver un système de sauvegarde adapté et automatique pour les membres du laboratoire. La solution Lina Atempo a été testée et présentée aux membres du laboratoire. Elle a été ensuite validée par le Conseil de laboratoire, le Conseil Scientifique et la Direction début 2024. Cette solution a été déployée fin 2024, après acquisition des licences Lina Atempo, des serveurs et disques de stockage. Elle assure la sauvegarde continue, automatisée et incrémentale de 1 To pour chaque membre du laboratoire.

4.3 Pratiques de partage

C7) Attribuez-vous habituellement des informations à vos données de recherche ?

Avant de pouvoir partager les données du laboratoire, il est essentiel de comprendre leur contexte de production et d'utilisation, pour faciliter leur compréhension et leur potentielle réutilisation, par les membres du laboratoire ou toute autre futur utilisateur.

Nous le voyons ici, 45% des répondants n'ajoutent pas d'informations supplémentaires à leurs données de recherche. Ces informations supplémentaires peuvent regrouper les métadonnées, un dictionnaire de données ou encore un fichier de type *readme*, afin de donner des éléments

supplémentaires sur le contexte d'acquisition des données, des variables utilisées, des commentaires qui pourraient être utiles à leur utilisation (matériel ou problème de calibrage d'un instrument, etc…). Ce sont seulement 21% des répondants qui ajoutent des informations scientifiques, tels que l'origine des données, des mots-clés, le projet de recherche, etc … et 13% des répondants ajoutent les trois types d'informations, que l'on retrouve essentiellement dans les métadonnées.

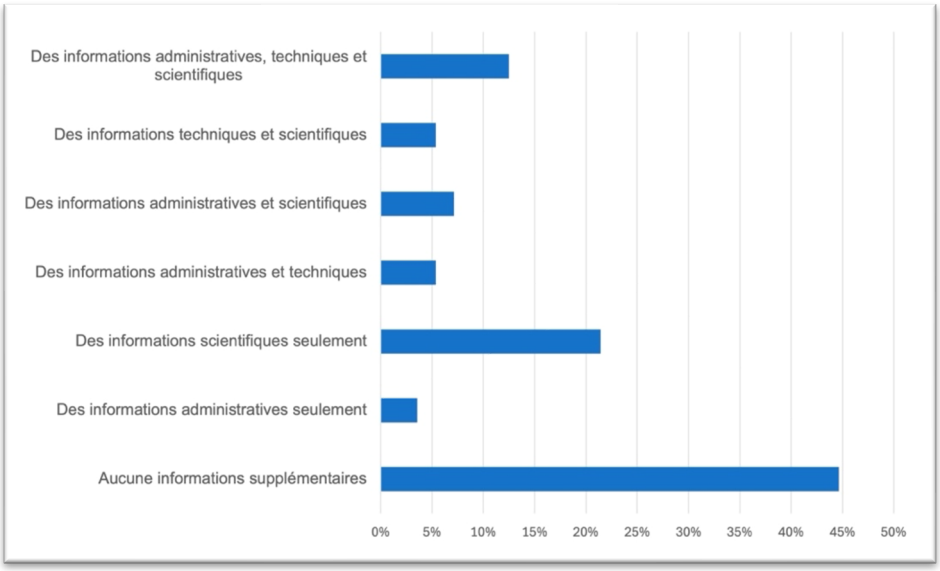


Figure 11 - Informations sur les données de recherche produites.

Ces résultats démontrent que le travail de sensibilisation sur la description des données produites doit continuer et se renforcer, afin de mesurer un réel changement dans les pratiques pour répondre aux principes FAIR.

C8) Si vous attribuez des informations supplémentaires, comment procédez-vous ?

Cette question permet de comprendre les habitudes actuelles pour documenter les données produites. Plus de la moitié des répondants rédige une documentation qui lui est propre. L'utilisation de standard de métadonnées est très faible. Mais, nous pensons que le terme « standard de métadonnées » a pu gêner les répondants et biaiser les résultats.

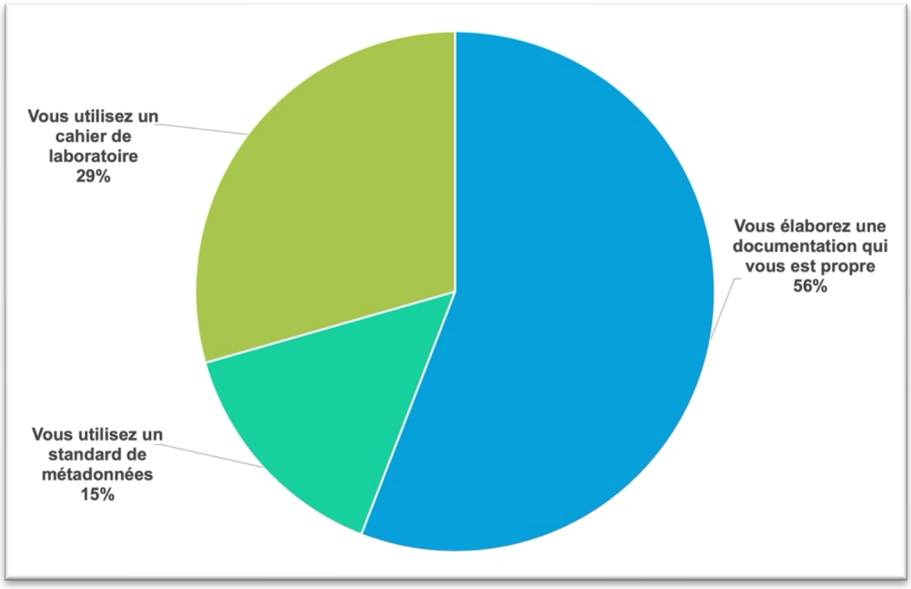


Figure 12 - Type de métadonnées ajoutées

Dans le catalogue de données du LIENSs, l'authentification prévue en début d'année 2025 via la fédération eduGAIN devrait simplifier l'accès au catalogue de données du laboratoire (aujourd'hui accessible via un compte sextant) et ainsi faciliter la création de fiche de métadonnées. L'organisation de futurs ateliers sur les métadonnées, avec les diverses options pour créer une fiche dans le catalogue seront présentées et explicitées.

D1) Seriez-vous d'accord pour diffuser vos données de recherche en libre accès ?

La très grande majorité des répondants sont d'accord pour diffuser en libre accès les données de recherche qu'ils produisent même si 31% d'entre eux sont d'accord pour que cette diffusion soit réalisée après un délai. Dans la pratique, l'embargo sur les données est effectivement possible et toléré avant diffusion. Mais, plusieurs expériences nous ont montré que cet engagement de diffusion au-delà des 2 ans est parfois oublié. Néanmoins, ces résultats témoignent d'une réelle volonté de diffuser les données en libre accès (72% des répondants) permettant de répondre aux incitations de la Science Ouverte.

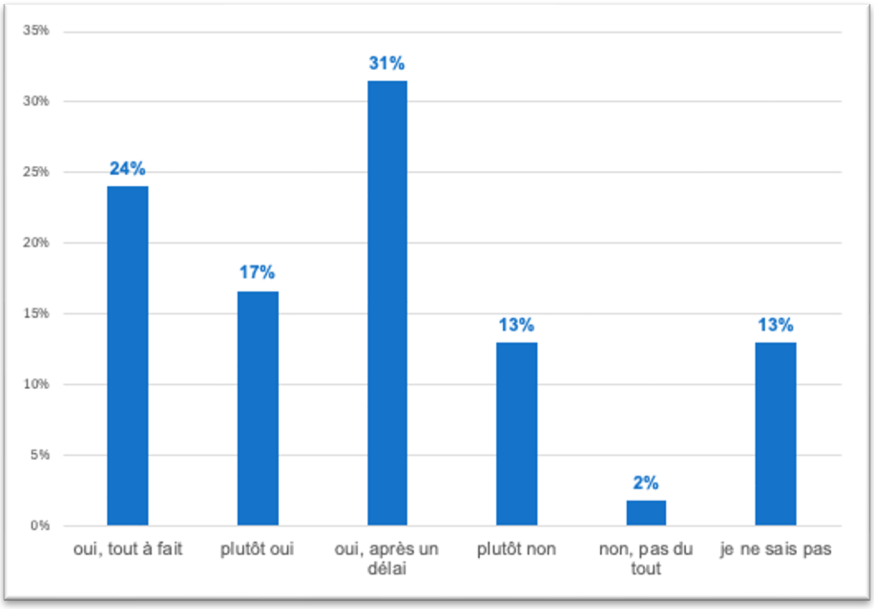


Figure 13 - Accord ou pas pour diffuser les données en libre accès

La communication renforcée du SAGEDO, mais aussi la multiplication des divers vecteurs de communications (Doranum, Inist…), n'ont pu qu'encourager ces intentions de diffusion en libre accès. Il reste à définir une feuille de route pour organiser sa mise en pratique pour les membres qui ne sont pas encore dans cette démarche de diffusion de leurs données.

D2) Quelles seraient les principales raisons qui vous empêcheraient de partager vos données de recherche ?

Actuellement, le catalogue du laboratoire recense 450 fiches de métadonnées, mais toutes n'ont pas de données associées. Ainsi, nous souhaitions appréhender quels sont les potentiels freins à leur diffusion. A noter que les répondants avaient la possibilité de choisir plusieurs modalités de réponses et ajouter d'éventuels commentaires.

La principale raison (52%) qui ressort de l'enquête est le manque de temps. En effet, celle du partage des données est bien définie comme une mission à réaliser, et semble être importante, mais le temps de travail à y consacrer est difficile. La deuxième raison principale au non-partage des données est le manque de connaissance pour le faire (39%).
La crainte du plagiat (36%) arrive en troisième position, puis des raisons scientifiques (32%) et des freins juridiques liés à la nature des données (30%).
Dans un champ libre de réponse, voici une citation que nous pensons intéressante à mettre en avant :

« Aucune idée... je crois que c'est un manque de temps. Avant de gérer mes données pour que les autres puissent les utiliser je dois déjà trouver le temps de les exploiter moi-même. Ici ce n'est pas une question de vouloir mais de pouvoir (avoir le temps de le faire) ».

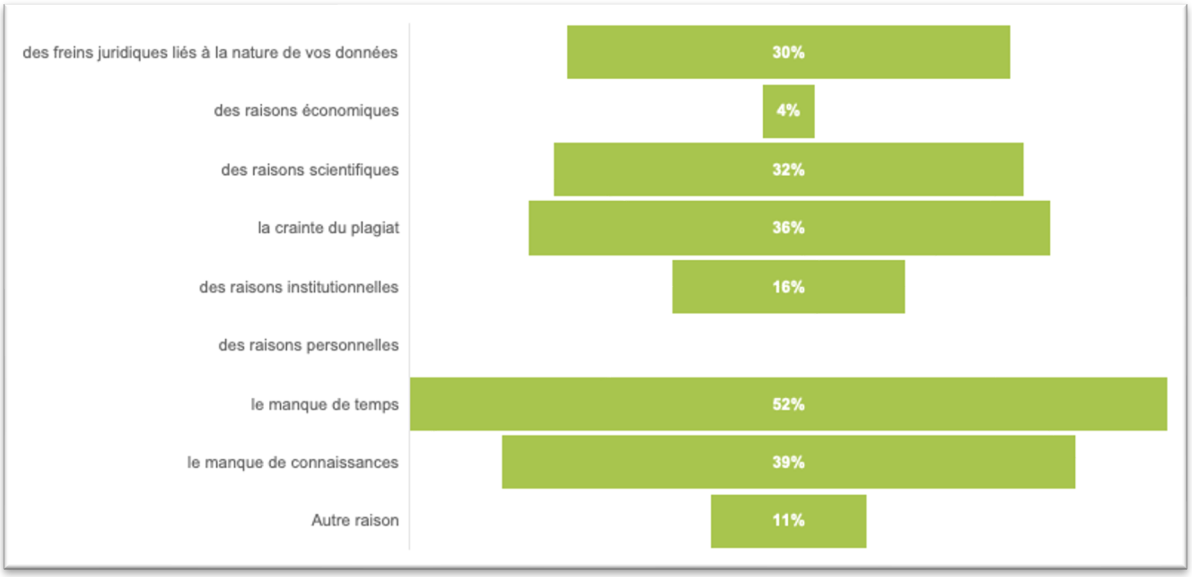


Figure 14 – Raisons du non-partage

Le côté positif est que le SAGEDO peut influer sur certaines de ces raisons, en apportant des outils, des conseils, des guides, des contacts permettant de faire gagner du temps sur le dépôt mais aussi d'améliorer la connaissance et favoriser le partage.

D4) En ce qui concerne le partage de vos données, quel est (ou quel serait) votre support principal ?

Les choix proposés dans cette question sont au nombre de 7, comprenant : le catalogue du laboratoire, une plateforme thématique, une plateforme généraliste, une plateforme imposée par une revue scientifique, un réseau social académique, un blog ou site personnel, puis une catégorie « Autre » avec un champ libre laissant la possibilité d'ajouter un autre support de partage. Le terme plateforme a été préféré à celui d'entrepôt, jugé par expérience plus compréhensible.

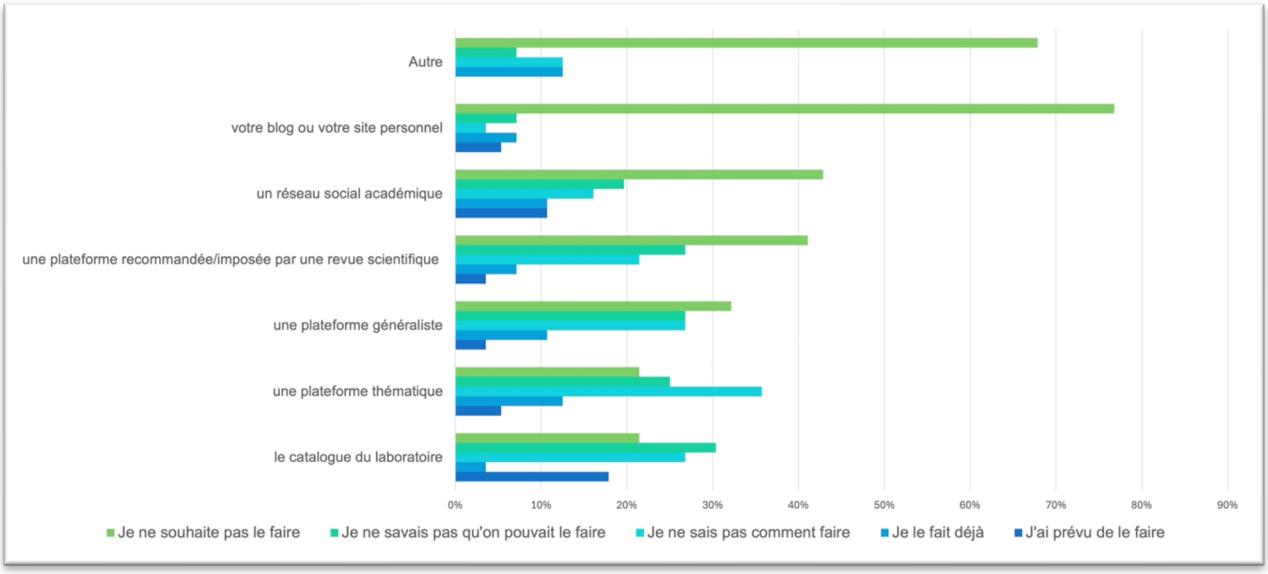


Figure 15 - Support principal pour le partage des données

Seul 22% des répondants, qui ont choisi le catalogue du laboratoire, sont des personnes qui le font déjà ou ont prévu de le faire, mais rappelons que ces données ne représentent que 10% des jeux de données en libre téléchargement et qu'ils sont actuellement sans DOI.

D4a : Le catalogue du LIENSs

Pour compléter, le graphe ci-dessous apporte des informations complémentaires sur trois de ces supports : le catalogue du LIENSs, une plateforme thématique et une plateforme généralise.

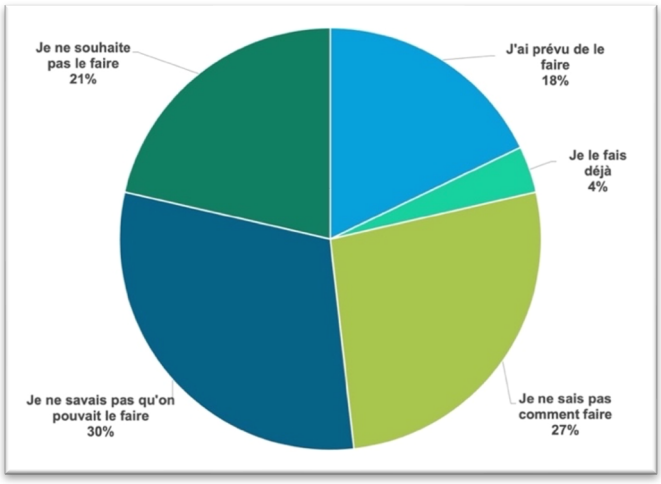


Figure 16 - Partage via le catalogue du laboratoire

Rappelons que le partage des données issu des activités de recherche nécessite de les mettre à disposition dans des entrepôts, tout en respectant le principe « aussi ouvert que possible, aussi fermé que nécessaire ». La préconisation de la plateforme nationale des données de recherche

Recherche Data Gouv est de privilégier les dépôts dans des entrepôts thématiques/disciplinaires et de confiance.[17]

Une certaine méconnaissance de cette pratique est révélée par l'enquête. Seulement 13% des répondants partagent les résultats de leurs travaux de recherche sur un entrepôt thématique. La grande majorité ne savent pas qu'on pouvait le faire (25%) ou ne savent pas comment faire (36%).

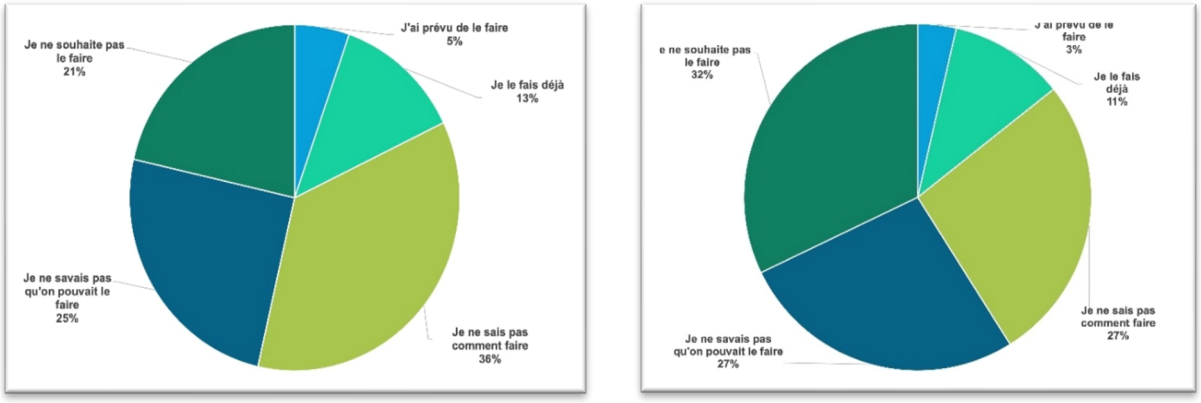


Figure 17 - a) Partage via une plateforme thématique ; b) Partage via une plateforme généraliste

Le SAGEDO, par sa veille scientifique et ses expertises, a pour mission de sensibiliser, mais également de conseiller et d'aider aux dépôts. Ces résultats vont ainsi nous permettre de mettre en place du soutien ou des formations dans ce domaine.

A ce jour, les principaux entrepôts thématiques identifiés sont, ODATIS pour les données relatives à l'Océan et au littoral, Data.Indores pour les données d'écologie et sur la biodiversité, Nakala pour les données en SHS. Bien évidemment, d'autres entrepôts thématiques peuvent être conseillés, mais sont plus à la marge en ce qui concerne les données produites et identifiées actuellement au laboratoire.

4.4Les besoins

E1) Avez-vous déjà rédigé un Plan de gestion de données ?

La question du Plan de Gestion de Données (PGD) nous semblait incontournable. En effet, il permet de prévoir tout le cycle de vie de la donnée d'un projet, mais aussi d'une plateforme d'analyse d'un observatoire, etc. On y décrit : la création, le stockage, les analyses, le partage, la conservation, l'archivage des données et métadonnées.
Au LIENSs, la rédaction de PGD n'est pas encore devenue une habitude : 77% des répondants n'ont jamais rédigé de PGD et 7% indique que cela ne s'applique pas à leurs recherches.

[17] Frédéric de Lamotte, Véronique Stoll, Cécile Arènes, Marie-Emilia Herbet, Stéphane Debard, et al.. *Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante*. Comité pour la Science Ouverte. 2024. ⟨hal-04534321v1⟩

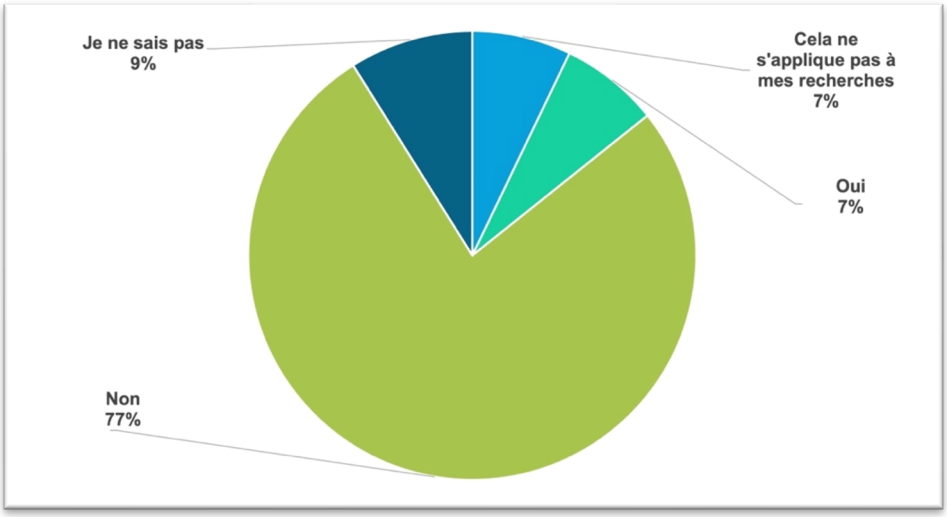


Figure 18 - Rédaction d'un PGD

Une des missions du SAGEDO est l'appui à la rédaction de PGD. Depuis un an, 4 demandes ont été faites au service, mais plusieurs demandes de conseils ou de questions rapides ont également été formulées. Au vu de ces résultats, nous souhaitons proposer des ateliers réguliers, pour sensibiliser à la gestion de la donnée et particulièrement à la rédaction de PGD, voir d'organiser d'éventuelles formations sur l'outil DMP Opidor.

E2) Dans quelle mesure, êtes-vous d'accord ou pas d'accord avec ce qui suit ?

Cette question nous est apparue essentielle pour comprendre les attentes et proposer de futures actions du SAGEDO.

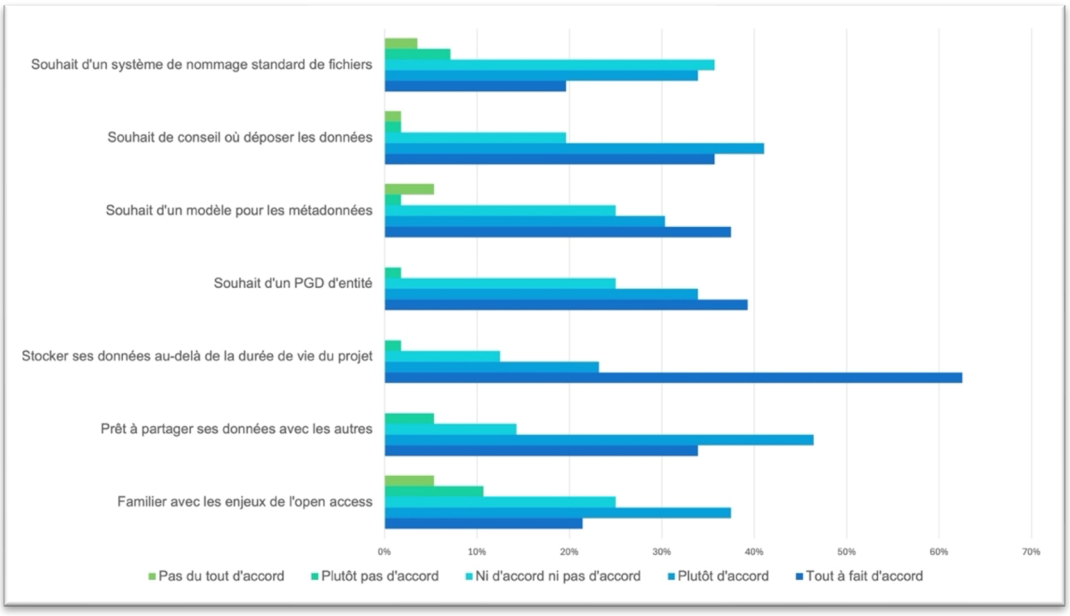


Figure 19 - Répartition des réponses à la question E2

Ce sont de nombreuses et fortes attentes sur tout le cycle de vie de la donnée, en particulier sur le partage et la diffusion, mais aussi sur les questions de stockage et d'archivage des données à plus de 60%, pour les projets de recherche finalisés.

Ci-dessous, nous présentons un focus sur deux modalités de réponse. Les autres modalités de réponse sont à retrouver en annexe 2.

E2d : Souhaitez-vous que le labo dispose d'un PGD d'entité ?

Le plan de gestion des données (PGD) est bien connu des chercheurs, car il est souvent demandé depuis 2019, par les financeurs des projets de recherche (ANR, Horizon Europe, etc…). Mais, il est également possible de mettre en place un PGD pour son entité de recherche.

Nous avons proposé la rédaction de ce PGDE lors de différentes réunions avec les membres du laboratoire, dans le but de disposer d'un ensemble de bonnes pratiques dans la gestion de leur donnée, mais aussi et surtout de pouvoir avoir des réponses simples, claires et unifiées au sein du laboratoire. Ce sont ainsi plus de 70% des répondants qui sont favorables pour qu'un plan de gestion de données d'entité soit rédigé et accessible à tous.

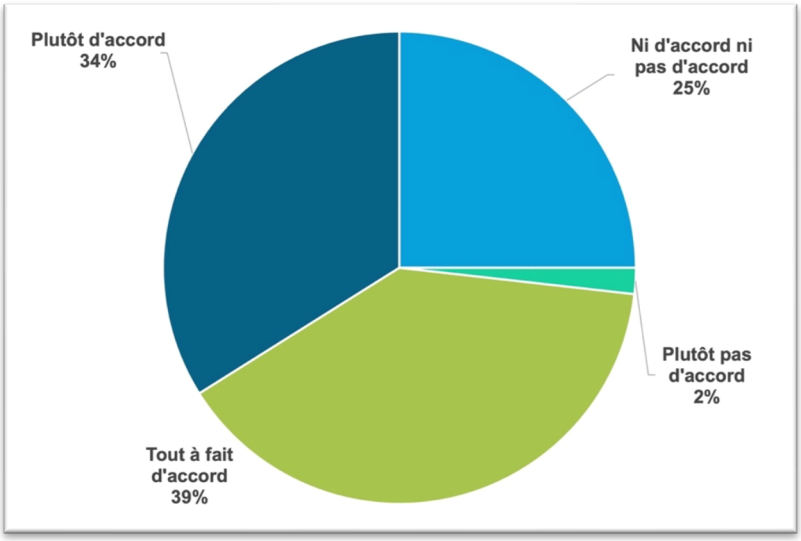


Figure 20 - Volonté de disposer d'un PGD entité

La rédaction de ce PGDE est un processus long et collaboratif. Le SAGEDO propose de le rédiger, mais l'investissement des responsables d'équipes, du conseil de laboratoire et de la Direction sont incontournables pour atteindre cet objectif. Enfin, il est à souligner qu'actuellement, le laboratoire ne dispose que de peu de personnel avec des compétences et des connaissances approfondies en informatique qui sont indispensables, pour du conseil et la mise en place de solutions. A clarifier dans quelles mesures, le Centre de Données pour l'Observation en Nouvelle-Aquitaine (CeDONA de l'OASU) dont le LIENSs fait partie, pourrait être, ici, en appui.

E2f : Souhait de conseil pour savoir où déposer les données pour les partager

Une des missions du SAGEDO est de promouvoir le partage et la diffusion des données, mais également d'orienter les déposants. Cette question permet de mieux définir nos futures propositions d'actions.

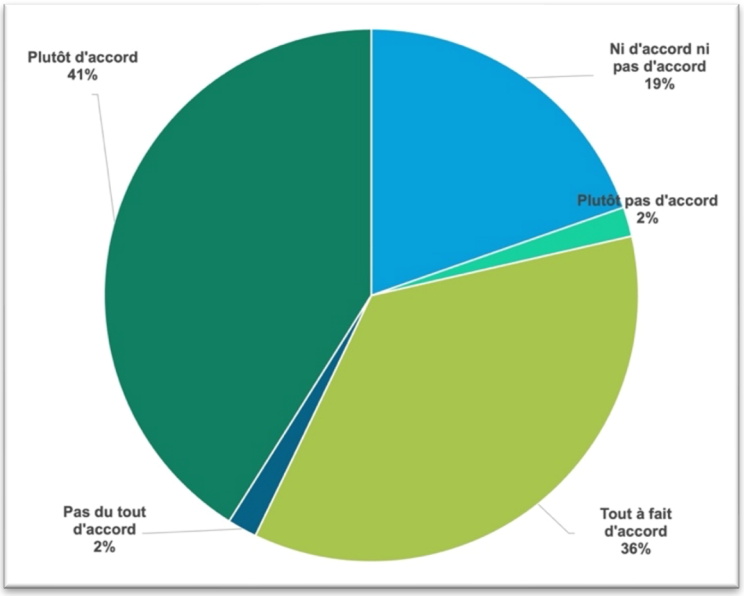


Figure 21 - Volonté d'être conseillé sur le dépôt de données

Ce sont 77% des répondants qui souhaitent des conseils pour déposer leurs données.

Ces résultats confortent les choix de mettre en place des temps dédiés aux conseils de dépôt par le SAGEDO, afin d'orienter les déposants vers l'entrepôt ou le catalogue le plus approprié et tendre ainsi vers un lieu de partage offrant une visibilité et une valorisation maximale de leurs données.

E3) Sur quels points souhaiteriez-vous bénéficier de conseils, d'une aide et/ou d'une formation ?

Cette question avait pour but de connaître les réels besoins en termes de conseils, d'appui et/ou de formation, afin d'organiser et d'optimiser les actions futures à mettre en place. En effet, effectuer une formation demande plus de temps que de prodiguer de « simples » conseils. Ainsi, et nous le voyons ici, la majorité des répondants souhaitent des conseils (entre 39 à 52%) ou une aide (entre 20 à 32%) plus que des formations.

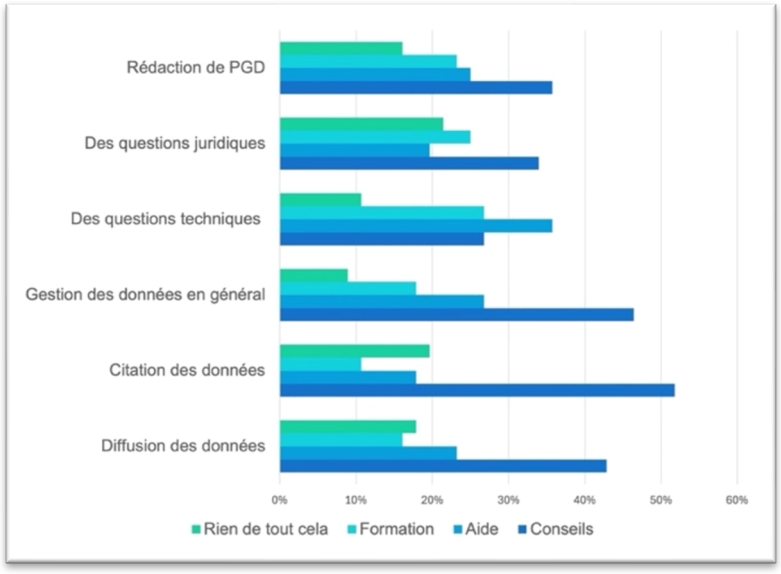
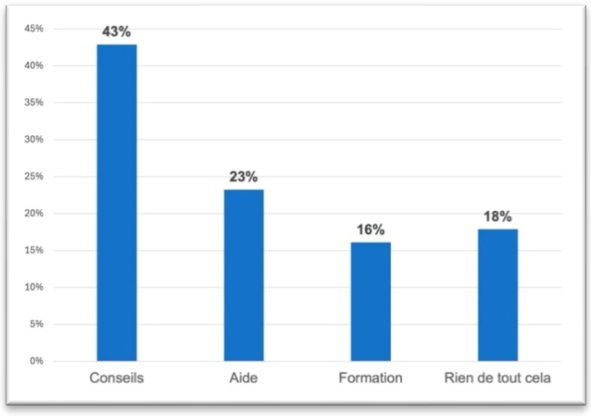


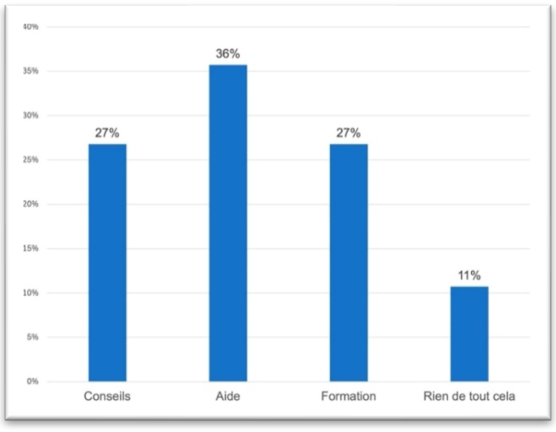
Figure 22 - Type de besoins (conseils, aide et/ou formation)

Ci-dessous, nous avons réalisé un focus sur 4 modalités de réponse, les autres sont à retrouver en annexe 2. Il est intéressant de constater que selon les modalités, le besoin change. En effet, la demande de conseil semble être la plus souhaitée, cela peut démontrer une connaissance (et donc une volonté de faire) dans le domaine de la rédaction de PGD ou de diffusion des données. Mais, concernant les questions techniques, l'aide est la plus demandée. Enfin, les besoins sur les questions techniques resteront à clarifier.

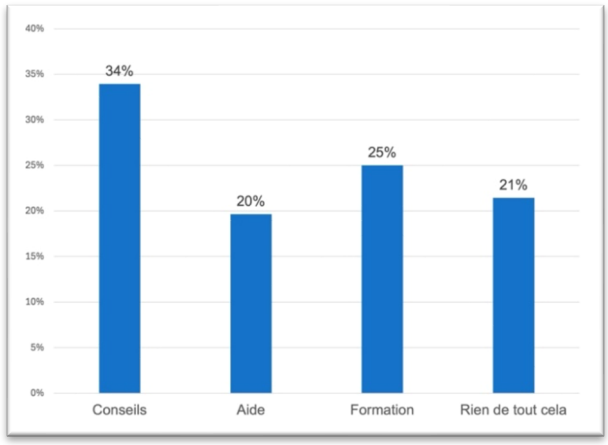
a)



b)



c)



d)

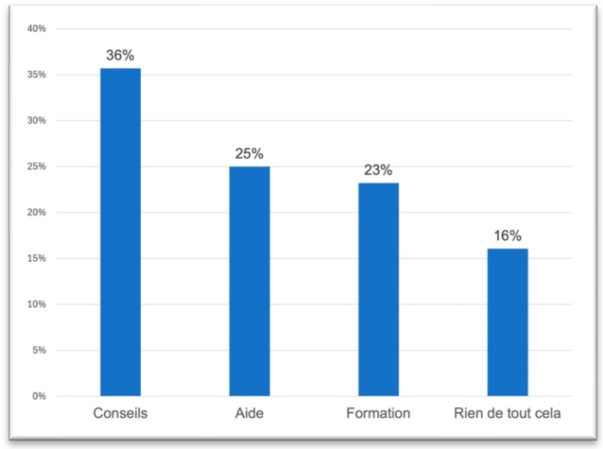


Figure 23 - a) Diffusion des données ; b) Questions techniques ; c) Questions juridiques ; d) Rédaction de PGD

Cette question conforte notre choix de proposer régulièrement des temps d'échanges et d'accompagnement, de type cafés-discussions. L'objectif est de prévoir sur un temps court (30-45 mn) et ouvert à tous, où chacun pourrait venir discuter (sans jugement et sans complexe), poser des questions et boire un café. Cela permettra d'apporter des conseils dans un premier temps, puis éventuellement selon les demandes, proposer un atelier aux personnes ayant plutôt besoin d'une aide.

Par ailleurs, le SAGEDO souhaite également proposer des formations. Les modalités restent à définir. Pour des formations plus longues, spécifiques ou nécessitant des compétences que le SAGEDO n'a pas forcément, elles pourraient s'organiser en partenariat avec des personnes extérieures au laboratoire. Par exemple, les services de la Bibliothèque Universitaire proposent des formations sur le règlement général de protection des données (RGPD) ou encore sur les données sensibles avec le délégué à la protection des données de LRU.

Conclusion

Les résultats de cette enquête ont permis de dresser un bilan en 2024 de la gestion des données de recherche du laboratoire LIENSs. Ils mettent en lumière les besoins, tant en termes de méthodes, d'appui, d'accompagnement, d'outils et de matériels pour répondre au mieux à la gestion et au partage de données, définis au travers de la politique d'Open Data et plus largement de l'Open Science. Ils présentent clairement la nécessité de définir conjointement une politique du laboratoire en termes de gestion, de partage et de diffusion des données.

En s'appuyant sur les résultats de cette enquête, le SAGEDO propose que cette politique soit décrite dans un plan de gestion d'entité, qui sera la feuille de route du laboratoire. Ce plan présentera ce qui est déjà opérationnel et perfectible, comme la gestion des métadonnées, et d'autre part, ce qui devra être mis en œuvre pour tendre vers une gestion efficace et optimale des données du laboratoire. Ce travail ne pourra se faire sans un engagement fort de la Direction, des responsables d'équipe, des chercheurs, enseignants-chercheurs et du personnel technique. Il éclairera et facilitera, sans aucun doute, la mise en place d'une politique de gestion de la donnée de recherche, répondant aux principes FAIR.

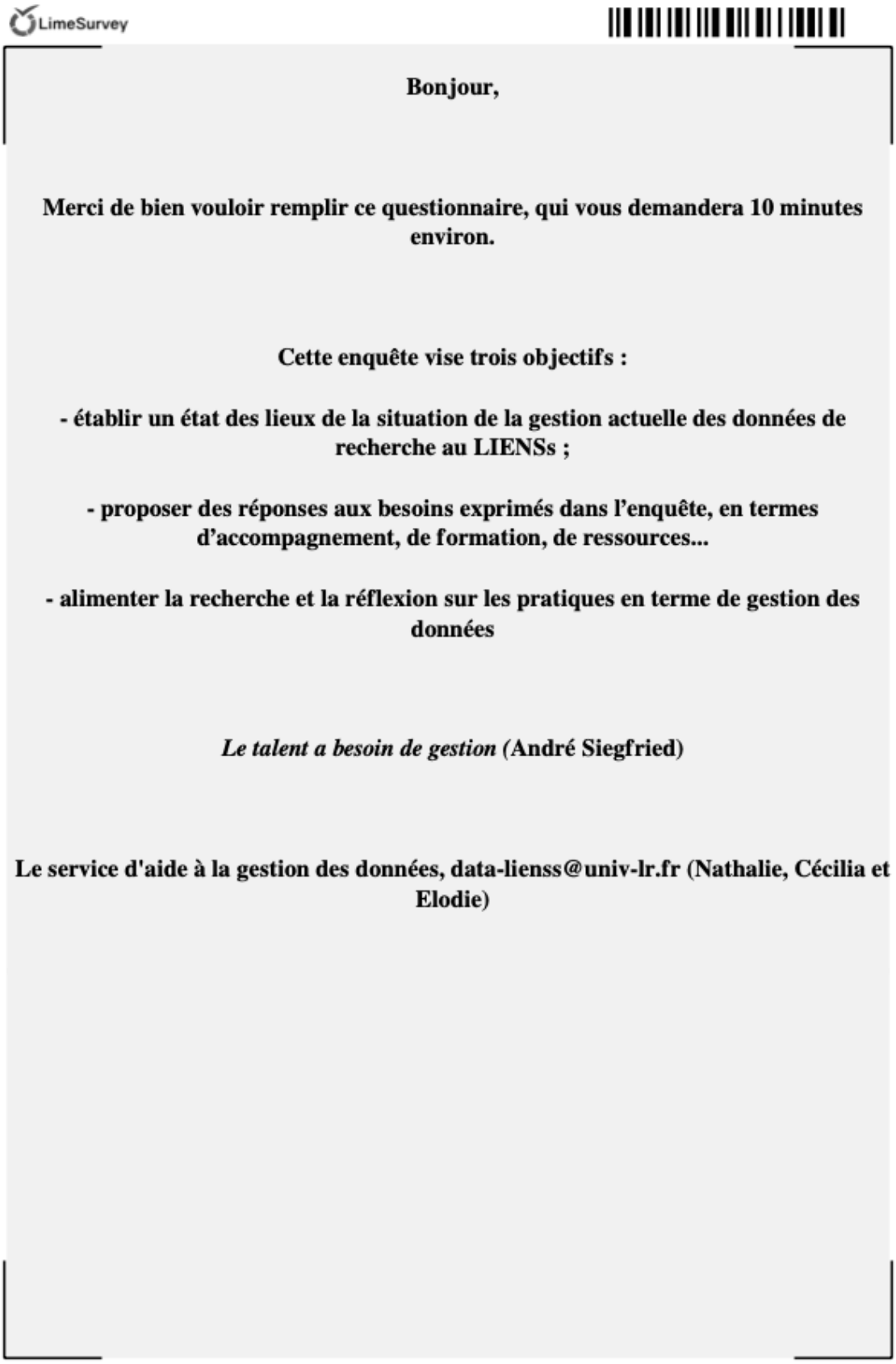
Au sein de LRU, le LIENSs est un des laboratoires actuellement pris en référence pour sa gestion des métadonnées, l'accompagnement dans le choix des entrepôts ou encore la mise en place de DOI via le pôle ODATIS dans son catalogue. Ces efforts réalisés s'inscrivent complétement dans le projet UCLR qui vise à mettre en place une plateforme numérique de données de LRU, mais aussi de partage de modèles, de codes informatiques et d'outils numériques.

Nous tenons à remercier tous les répondants de l'enquête, ainsi que le soutien de la Direction du laboratoire pour la réalisation de cette enquête.

Liste des figures

Figure 1 - Statistiques du catalogue du LIENSs 8
Figure 2 - Répartition par statut des répondants au questionnaire 11
Figure 3 - Répartition par équipes entre les répondants à l'enquête et les données du LIENSs.. 12
Figure 4 - Types de données produites 13
Figure 5 - Type de support des données produites 14
Figure 6 - Numérisation des données papier et/ou matérialisées. 15
Figure 7 - Estimation du volume de stockage de données nécessaire par personne. 16
Figure 8 - Lieu de stockage des données 17
Figure 9 - Types de sauvegarde 17
Figure 10 - Type de fréquence des sauvegardes 18
Figure 11 - Informations sur les données de recherche produites. 19
Figure 12 - Type de métadonnées ajoutées 20
Figure 13 - Accord ou pas pour diffuser les données en libre accès 21
Figure 14 – Raisons du non-partage 22
Figure 15 - Support principal pour le partage des données 23
Figure 16 - Partage via le catalogue du laboratoire 23
Figure 17 - a) Partage via une plateforme thématique ; b) Partage via une plateforme généraliste 24
Figure 18 - Rédaction d'un PGD 25
Figure 19 - Répartition des réponses à la question E2 25
Figure 20 - Volonté de disposer d'un PGD entité 26
Figure 21 - Volonté d'être conseillé sur le dépôt de données 27
Figure 22 - Type de besoins (conseils, aide et/ou formation) 28
Figure 23 - a) Diffusion des données ; b) Questions techniques ; c) Questions juridiques ; d) Rédaction de PGD 29
Figure 24 - Partage via une plateforme recommandée par une revue scientifique 41
Figure 25 - Partage via un réseau social académique 41
Figure 26 - Partage via un blog personnel et autre moyen. 42
Figure 27 – Familier avec les enjeux de l'open access 42
Figure 28 - D'accord avec le partage des données de recherche qu'il produit 43
Figure 29 - Partage après la fin du projet 43
Figure 30 - Souhait d'un modèle de métadonnées 44
Figure 31 - Souhait d'un nommage standard de fichiers 44
Figure 32 - Souhait sur la citation des données 45
Figure 33 - Souhait sur la gestion des données en général 45

Annexe 1 : Questionnaire



LimeSurvey

Bonjour,

Merci de bien vouloir remplir ce questionnaire, qui vous demandera 10 minutes environ.

Cette enquête vise trois objectifs :

- établir un état des lieux de la situation de la gestion actuelle des données de recherche au LIENSs ;

- proposer des réponses aux besoins exprimés dans l'enquête, en termes d'accompagnement, de formation, de ressources...

- alimenter la recherche et la réflexion sur les pratiques en terme de gestion des données

***Le talent a besoin de gestion* (André Siegfried)**

Le service d'aide à la gestion des données, data-lienss@univ-lr.fr (Nathalie, Cécilia et Elodie)

LimeSurvey

Partie A: Vous-même

A1. Vous êtes ?

- [ ] Enseignant(e)-chercheur(e)
- [ ] Chercheur(e)
- [ ] Post-doctorant(e)
- [ ] Doctorant(e)
- [ ] Ingénieur(e) de recherche
- [ ] Ingénieur(e) d'études
- [ ] Assitant(e) ingénieur(e)
- [ ] Technicien(ne) de recherche
- [ ] Autre

Autre

A2. Vous faites partie de quelle équipe au sein du LIENSs ?

- [ ] AGÎLE : Approche Géographique : Iles, Littoraux, Environnement
- [ ] AMARE : Réponses des Animaux MARins à la variabilité Environnementale
- [ ] BCBC : Biotechnologies et Chimie des Bioressources pour la Santé
- [ ] BIOFEEL : BIOdiversité et Fonctionnement des EcosystEmes Littoraux
- [ ] DPL : Dynamique Physique du Littoral
- [ ] ESTRAN : Espaces, Sociétés, Territoires des Rivages Anciens et Nouveaux
- [ ] Pôle de soutien technique
- [ ] Autre

Autre

LimeSurvey

Partie B: Typologie des données produites

B1. Dans quelles(s) catégorie(s) suivantes classeriez-vous vos données produites ?

- [ ] des données d'observation (levés de terrain, mesures de capteurs, drône...)
- [ ] des données expérimentales (depuis un équipement du labo biologie, biochimie...)
- [ ] des données d'archives (iconographie, documents historiques, etc...)
- [ ] des données audiovisuelles (enregistrements audio, séquences filmiques...)
- [ ] des données dérivées ou compilées (textes, données d'une base, web-scraping)
- [ ] des données d'enquêtes, de sondages (questionnaires, entretiens)
- [ ] codes source et modèles
- [ ] Autre

B2. Autre, merci de préciser :

B3. Concernant leur support matériel, vos données produites sont- elles majoritairement ?

- [ ] papier/matérialisées (par exemple : carnet de laboratoire, des entretiens...)
- [ ] numérisées/dématérialisées (par exemple : textes numérisés, images scannées...)
- [ ] nativement numériques (par exemple : capteurs, etc ...)

B4. Si vos données sources sont papiers/matérialisées, est-ce que vous les numérisez en général (saisie d'entretiens ou de carnet de laboratoire...) ?

- [ ] Oui
- [ ] Non

LimeSurvey

Partie C: Pratiques de stockage et d'archivage des données (collectées et produites)

C1. À combien estimez-vous l'espace nécessaire au stockage de vos données ?

- ☐ inférieur à 1 Go
- ☐ entre 1 et 20 Go
- ☐ entre 20 et 100 Go
- ☐ entre 100 Go et 1 To (1 tera-octet = 1000 Go ou 1000 milliards d'octets)
- ☐ supérieur à 1 To
- ☐ Je ne sais pas

C2. Où stockez-vous les données de recherche ?

- ☐ sur votre ordinateur personnel
- ☐ sur votre ordinateur professionnel
- ☐ sur un disque dur externe ou autre support (clé USB)
- ☐ sur un serveur ou une plateforme de l'université ou du CNRS (NextCloud, OwnCloud)
- ☐ sur une plate-forme privée dans le "cloud" (Google Drive, Dropbox, etc.)
- ☐ sur un serveur ou une plateforme externe d'une autre institution
- ☐ sur le serveur d'une Infrastructure de Recherche (par ex : ShareDoc Huma-Num)
- ☐ Autre

C3. Autre, merci de préciser :

C4. Faites-vous des sauvegardes de vos données ?

	Oui	Non
manuellement	☐	☐
automatiquement	☐	☐

C5. À quelle fréquence les données sont-elles sauvegardées ?

- [ ] 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

C6. Autre, merci de préciser

C7. Attribuez-vous habituellement des informations supplémentaires (métadonnées, Readme, autre documentation) à vos données de recherches ?

- [ ] des informations administratives (auteur, date de création, etc.)
- [ ] des informations techniques (format des fichiers, taille, logiciel, variables utilisées, etc.)
- [ ] des informations scientifiques (origine des données, projet de recherche, mots-clés, etc.)
- [ ] je n'attribue pas d'informations supplémentaires

C8. Si vous attribuez des informations supplémentaires, comment procédez-vous ?

- [ ] vous élaborez une documentation qui vous est propre
- [ ] vous utilisez un standard de métadonnées (Iso 19 115, Dublin Core)
- [ ] vous utilisez un cahier de laboratoire (ou de recherche)
- [ ] autre

C9. **Autre, merci de préciser**

Partie D: Pratiques de partage et de diffusion des données

D1. **Seriez-vous d'accord pour diffuser vos données de recherche en libre accès (à condition qu'elles soient diffusables et protégées juridiquement) ?**

- ☐ oui, tout à fait
- ☐ oui, après un délai
- ☐ plutôt oui
- ☐ plutôt non
- ☐ non, pas du tout
- ☐ je ne sais pas

D2. **Quelles seraient les principales raisons qui vous empêcheraient de partager vos données de recherche ?**

- ☐ des freins juridiques liés à la nature de vos données (Exemple : données à caractère personnel, données dites sensibles (espèces protégées, etc.))
- ☐ des raisons économiques (Exemple : brevet)
- ☐ des raisons scientifiques (Exemple : embargo)
- ☐ la crainte du plagiat
- ☐ des raisons institutionnelles (Exemple : mon labo ne veut pas diffuser les données…)
- ☐ des raisons personnelles
- ☐ le manque de temps
- ☐ le manque de connaissances/compétences
- ☐ Autre

D3. **Autre, merci de préciser :**

D4. En ce qui concerne le partage de vos données, quel est (ou serait) votre support principal ?

	Je le fait déjà	J'ai prévu de le faire	Je ne sais pas comment faire	Je ne savais pas qu'on pouvait le faire	Je ne souhaite pas le faire
le catalogue du laboratoire (Catalogue LIENSs)	☐	☐	☐	☐	☐
une plateforme thématique (Form@ter, PNDB, ODATIS)	☐	☐	☐	☐	☐
une plateforme généraliste (Zenodo, Recherche Data Gouv)	☐	☐	☐	☐	☐
une plateforme recommandée/imposée par une revue scientifique	☐	☐	☐	☐	☐
un réseau social académique (ResearchGate, Academia…)	☐	☐	☐	☐	☐
votre blog ou votre site personnel	☐	☐	☐	☐	☐
Autre	☐	☐	☐	☐	☐

D5. Autre, merci de préciser :

D6. Pouvez-vous préciser quelle(s) revue(s) et quelle(s) plateforme(s) ?

D7. Sur quelle plateforme thématique dans l'archivage et la diffusion des données de recherche avez-vous partagé vos données ?

D8. **Sur quelle plateforme thématique dans l'archivage et la diffusion des données de recherche avez-vous prévu de partager vos données ?**

Partie E: Vos besoins, vos attentes

E1. **Avez-vous déjà rédigé un Plan de Gestion de Données (PGD ou DMP) ?**

- ☐ Oui
- ☐ Non
- ☐ Je ne sais pas
- ☐ Cela ne s'applique pas à mes recherches

E2. **Dans quelle mesure êtes-vous d'accord ou pas d'accord avec ce qui suit ?**

	tout à fait d'accord	plutôt d'accord	ni d'accord ni pas d'accord	plutôt pas d'accord	pas du tout d'accord
Je suis familier(e) avec les enjeux de l'open access (accès ouvert)	☐	☐	☐	☐	☐
Je suis prêt(e) à partager mes données de recherche avec les autres	☐	☐	☐	☐	☐
Je souhaiterais pouvoir stocker mes données de recherche au-delà de la durée de vie d'un projet	☐	☐	☐	☐	☐
Je souhaiterais que le laboratoire dispose d'un plan de gestion de données	☐	☐	☐	☐	☐
Je souhaiterais que le laboratoire me propose un modèle pour saisir mes métadonnées	☐	☐	☐	☐	☐
Je souhaiterais que le laboratoire me conseille où déposer mes données pour les partager	☐	☐	☐	☐	☐
Je souhaiterais que le laboratoire utilise/recommande un système de nommage standard de fichiers	☐	☐	☐	☐	☐

E3. **Sur quels points souhaiteriez-vous bénéficier de conseils, d'une aide et/ou d'une formation ?**

	Des conseils	Une aide	Des formations	Rien de tout cela
sur la diffusion des données de la recherche	☐	☐	☐	☐
sur la citation des données de la recherche	☐	☐	☐	☐
sur la gestion des données de la recherche en général	☐	☐	☐	☐

	Des conseils	Une aide	Des formations	Rien de tout cela
sur des questions techniques (par exemple, métadonnées, normes, archivage à long terme, etc ...)	☐	☐	☐	☐
sur des questions juridiques (par exemple, les restrictions d'accès,le traitement des données sensibles , l'utilisation de licences)	☐	☐	☐	☐
sur la rédaction d'un plan de gestion de données	☐	☐	☐	☐
autre	☐	☐	☐	☐

E4. Autre, merci de préciser

Merci d'avoir complété ce questionnaire.

Annexe 2 : Autres résultats de l'enquête

D4) En ce qui concerne le partage de vos données, quel est (ou serait) votre support principal ?

D4d : Une plateforme recommandée par une revue scientifique

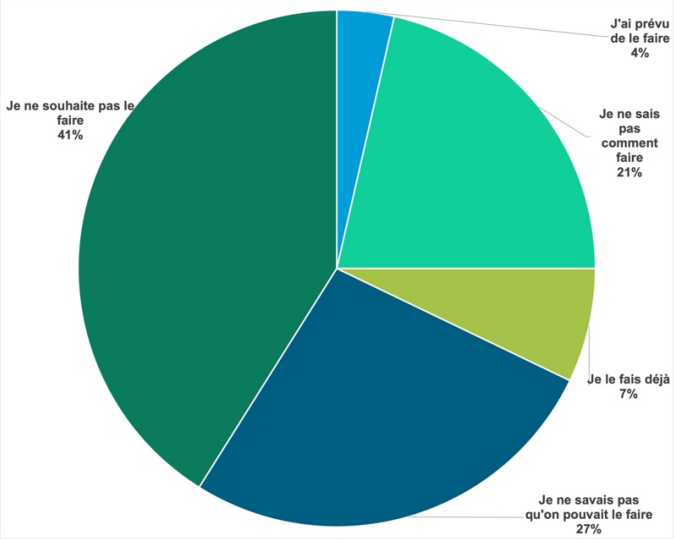


Figure 24 - Partage via une plateforme recommandée par une revue scientifique

D4e : Un réseau social académique (ResearchGate, Academia)

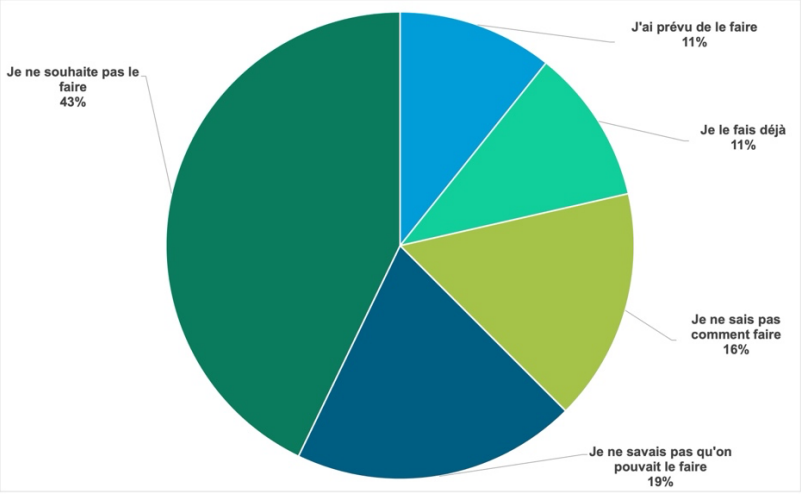


Figure 25 - Partage via un réseau social académique

D4e : Blog personnel et autre

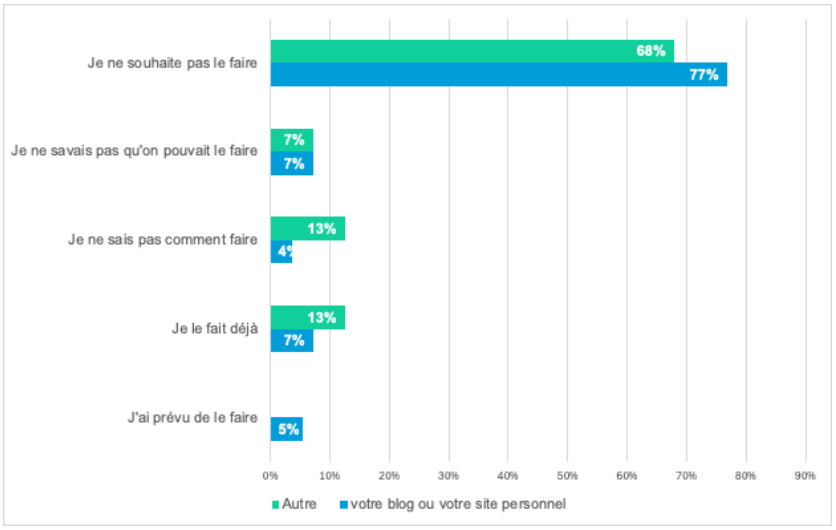


Figure 26 - Partage via un blog personnel et autre moyen.

E2) Dans quelle mesure, êtes-vous d'accord ou pas d'accord avec ce qui suit ?

E2a : Etes-vous familier avec les enjeux de l'open access ?

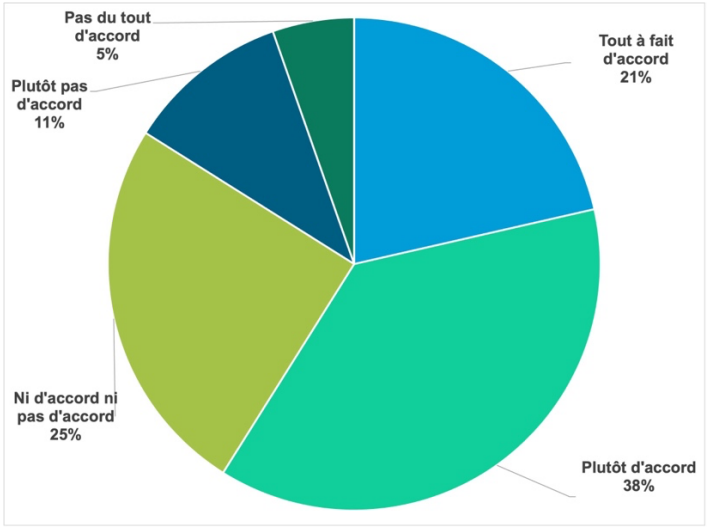


Figure 27 – Familier avec les enjeux de l'open access

E2b : Prêt à partager ses données de recherche

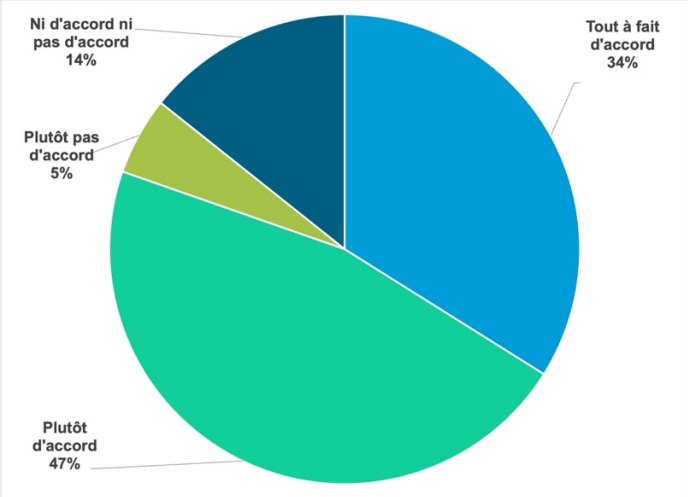


Figure 28 - D'accord avec le partage des données de recherche qu'il produit

E2c : Souhait de pouvoir partager ses données de recherche au-delà de la durée de vie du projet

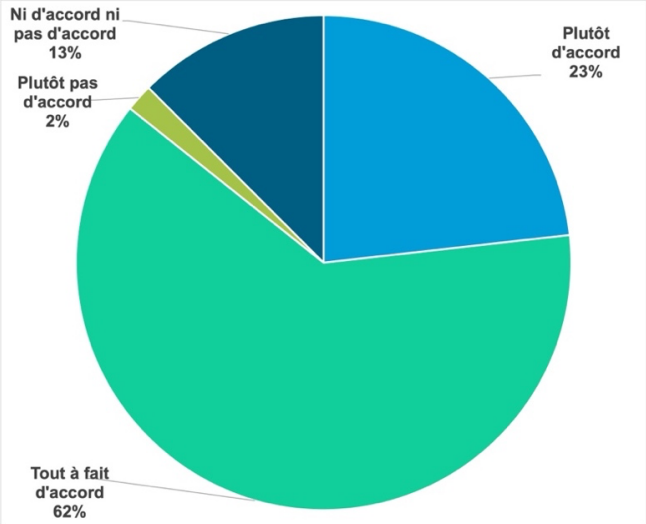


Figure 29 - Partage après la fin du projet

E2e : Souhaitez-vous que le labo propose un modèle pour les métadonnées ?

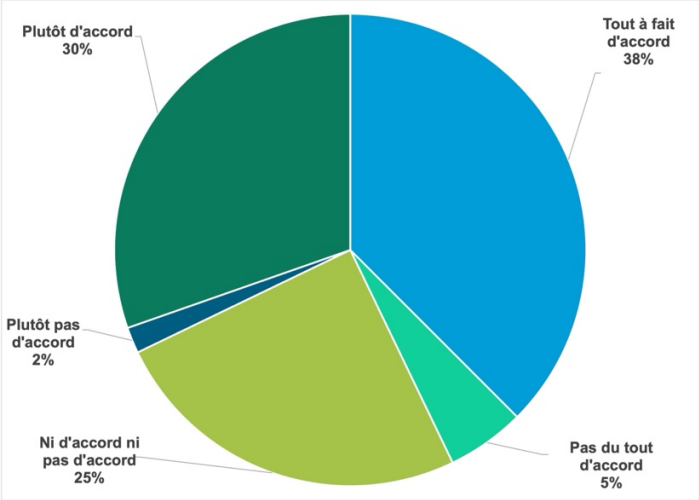


Figure 30 - Souhait d'un modèle de métadonnées

E2g : Souhait d'un système de nommage standard de fichiers

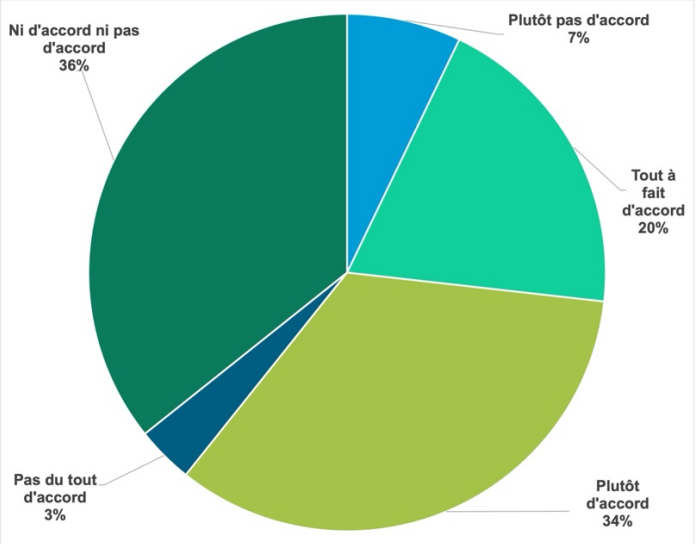


Figure 31 - Souhait d'un nommage standard de fichiers

E3) Sur quels points souhaiteriez-vous bénéficier de conseils, d'une aide et/ou d'une formation ?

E3b : Sur la citation des données de la recherche

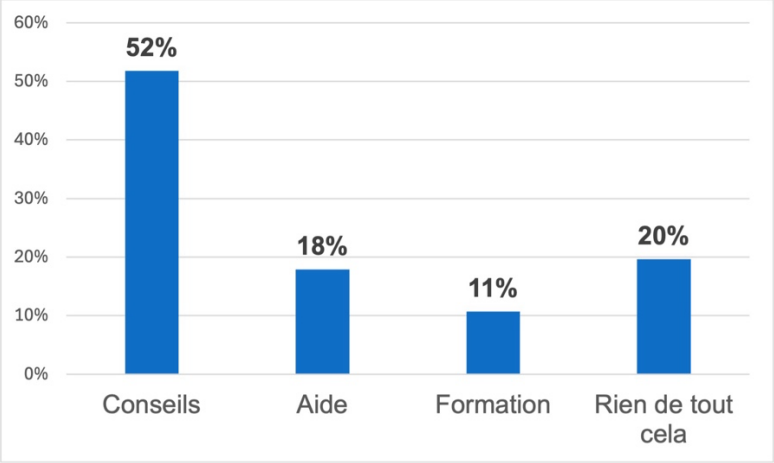


Figure 32 - Souhait sur la citation des données

E3c : Sur la gestion des données de la recherche en général

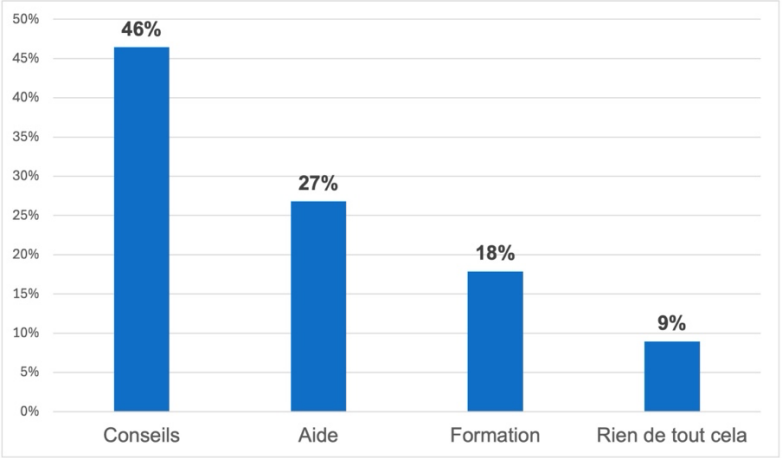


Figure 33 - Souhait sur la gestion des données en général