



**HAL**  
open science

## Enquête sur les Données de la Recherche au sein de deux universités françaises

Sandrine Wolff, Gabriel Gallezot, Nicolas Jalabert

► **To cite this version:**

Sandrine Wolff, Gabriel Gallezot, Nicolas Jalabert. Enquête sur les Données de la Recherche au sein de deux universités françaises. Université de Strasbourg; Université Côte d'Azur; Université Rennes 2; INRAe. 2022, 82 p. hal-03997711

**HAL Id: hal-03997711**

**<https://hal.science/hal-03997711v1>**

Submitted on 20 Feb 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

# Enquête sur les Données de la Recherche au sein de deux universités françaises

## Rapport rédigé par :

Wolff Sandrine (Université de Strasbourg, BETA)  
Gallezot Gabriel (Université Côte d'Azur, TransitionS)  
Jalabert Nicolas (Université de Strasbourg, BETA)

Date : décembre 2022

## Membres de l'équipe projet :

Gallezot Gabriel, Université Côte d'Azur  
Jalabert Nicolas, Université de Strasbourg  
Jeannaud Viet, Université Côte d'Azur  
Malingre Marie-Laure, Université de Rennes II  
Rennes Stéphanie (INRAe)  
Serres Alexandre, Université de Rennes II  
Wolff Sandrine, Université de Strasbourg

Projet financé par le GIS « Réseau Urfist »



# Table des matières

<b>Introduction</b> .....	<b>4</b>
Contexte de l'étude .....	4
Le questionnaire et la collecte des données .....	5
<b>Partie 1 : Caractéristiques des répondants</b> .....	<b>6</b>
Statut des répondants .....	6
Champ disciplinaire et genre.....	7
Classes d'âge.....	7
Responsabilités .....	8
<b>Partie 2: Cycle des DR, quelles DR ?</b> .....	<b>10</b>
Acception de la notion de données de la recherche .....	10
Méthode d'obtention des données .....	11
Production des données .....	13
Type de données résultats .....	14
<b>Partie 3 : Pratiques de gestion des données : stockage, archivage, description</b> .....	<b>16</b>
Stockage des données chaudes .....	16
Sauvegarde des données chaudes.....	19
Archivage des données froides .....	19
Plan de gestion de données .....	20
<b>Partie 4 : Utilisation des DR : diffusion, partage et réutilisation</b> .....	<b>23</b>
Accès aux données froides des répondants .....	23
Usage de données d'autrui .....	24
Diffusion de ses données sur le Web .....	25
Motifs de non-diffusion des données .....	27
Volonté de partager les données suivant les principes de la science ouverte .....	27
Incitations des chercheurs à partager leurs données en science ouverte .....	29
Freins au partage des données de recherche .....	30
Connaissance des data papers .....	33
<b>Partie 5 : Recensement des besoins pour les DR</b> .....	<b>34</b>
<b>Conclusion</b> .....	<b>37</b>
<b>Les traits les plus saillants de l'enquête</b> .....	<b>37</b>
Des pratiques de production et de stockage différentes selon le champ disciplinaire. ....	37
Une faible connaissance des données de recherche en science ouverte .....	37
Des pratiques de partage encore peu développées, par manque de ressources ou simple ignorance .....	37
Les intentions de partage des données : freins et motivations.....	38
Besoins et attentes .....	38
Le poids du statut et de la discipline .....	38

<b>Points forts et limites méthodologiques.....</b>	<b>38</b>
<b>Recommandations .....</b>	<b>39</b>
<b><i>Annexes</i>.....</b>	<b>42</b>
<b>Annexe 1 - Le questionnaire de l'enquête « amélioré » .....</b>	<b>43</b>
<b>Annexe 2 - Questionnaire Flash .....</b>	<b>74</b>
<b>Annexe 3 – Qualification des données de la recherche .....</b>	<b>82</b>

## Introduction

Dans le contexte actuel d'ouverture des données de la recherche (DR) et afin de contribuer à éclairer les décisions et politiques de science ouverte au niveau des établissements, l'étude présentée ici vise à mieux connaître les perceptions et les pratiques de gestion des données de recherche, mais aussi et surtout les incitations des chercheurs à partager leurs données ainsi que leurs besoins et attentes en la matière. Outre un objectif de pure recherche, il s'agit donc aussi de contribuer à fournir une aide au pilotage des établissements via : un état des lieux, une analyse des besoins, des recommandations et notamment des propositions d'indicateurs.

Nous avons élaboré une enquête et l'avons adressée en ligne à l'ensemble des doctorants et des personnels des universités de Nice (UCA) et Strasbourg (Unistra) ayant des activités de recherche, entre novembre 2019 et mars 2021. Les 496 réponses complètes que nous avons obtenues permettent de dresser un premier état des pratiques et perceptions liées aux données de recherche, un an après le premier Plan National pour la Science Ouverte (PNSO). La synthèse fournit en quelque sorte une photographie des pratiques des DR à un stade très amont, juste avant que les établissements ne mettent en œuvre des politiques de sensibilisation fortes en la matière.

Le présent rapport est consacré à une présentation chiffrée détaillée de ces quelques 500 réponses.

## Contexte de l'étude

Financée par le GIS Réseau Urfist<sup>1</sup> dans le cadre d'un projet multi-établissements intitulé « Enquête Données de la Recherche », cette étude s'appuie sur une expérience antérieure d'enquête conduite à l'Université Rennes 2<sup>2</sup> par deux participants au projet GIS. Le groupe de travail était composé de : Alexandre Serres (Université Rennes 2) , Marie-Laure Malingre (Université Rennes 2), Sandrine Wolf (Université de Strasbourg), Nicolas Jalabert (Université de Strasbourg), Stéphanie Rennes (Inrae) , Viet Jeannaud (Université Côte d'Azur) et Gabriel Gallezot (Université Côte d'Azur).

Le but du projet du GIS était de conduire une extension de l'enquête rennaise dans deux directions : application à d'autres grandes universités française et/ou EPST d'une part, et prise en compte de l'ensemble des champs disciplinaires, au-delà de l'orientation SHS majoritairement représentée à Rennes 2, d'autre part. Le nouveau questionnaire est donc transdisciplinaire. Il concerne tous les chercheurs et tous les types de données (textes, chiffres, images,...)

Les objectifs attendus du projet GIS étaient les suivants :

- Enquête sur les pratiques de DR des chercheurs au sein de plusieurs universités ou organismes
- Méthodologie générale d'enquête sur les données de la recherche, référentiel de questions sur les pratiques informationnelles en DR, et mise à disposition du questionnaire

---

<sup>1</sup> <http://gis-reseau-urfist.fr/enquete-donnees-de-la-recherche/>

<sup>2</sup> Serres A. et alii (2017) : "Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs: une enquête à l'Université Rennes 2." Rapport, novembre 2017. Disponible sur HAL : <https://hal.archives-ouvertes.fr/hal-01635186/>

- Mise à disposition des données anonymisées
- Grille d'indicateurs pour mesurer l'évolution des pratiques des chercheurs.

L'originalité de l'approche, comparée à d'autres enquêtes sur des thèmes similaires, est d'être déployée au niveau local et avec le soutien de l'université, tout en permettant une collecte d'informations et une analyse comparative sur plusieurs établissements, riche d'enseignements. Le questionnaire est librement accessible et peut être étendu aisément à d'autres organismes de recherche. L'enquête dont nous présentons les résultats ici a certes été conduite au sein de deux universités, mais on notera qu'elle a été administrée plus récemment dans les universités de Nantes et du Mans, ainsi qu'à INRAE et au synchrotron SOLEIL.

## Le questionnaire et la collecte des données

Les données d'enquête présentées ci-après ont été obtenues à partir d'un questionnaire (cf Annexe 1) sur LimeSurvey<sup>3</sup>, dont le lien a été envoyé à l'ensemble des doctorants et des personnels des deux universités en ciblant explicitement ceux ayant des activités de recherche. La période d'administration du questionnaire a duré 2 mois, de début novembre 2019 à début février 2020, avec une relance en décembre.

Les canaux de diffusion ont été différents dans les deux universités. À l'UCA, le mail a été adressé à tous les responsables d'Unités de recherche, avec demande de redirection auprès des membres du labo. A Strasbourg, le mail a été envoyé à l'ensemble des personnels (y compris vacataires et contractuels) ainsi qu'à l'ensemble des doctorants. Au final et dans les deux cas, l'ensemble des principaux statuts liés à des activités de recherche a pu être touché: Enseignants-Chercheurs (E-C), Doctorants et Postdocs, bien sûr, mais aussi Chercheurs<sup>4</sup>, Ingénieurs de recherche ou autres personnels administratifs effectuant de la recherche<sup>5</sup>.

Le questionnaire complet comprend 60 questions, dont 20 sont obligatoires, 23 sont facultatives et 17 sont conditionnelles à certaines réponses. Cela induit un nombre variable de questions selon le répondant, mais on peut estimer que le remplir prend une vingtaine de minutes.

Parmi les 496 réponses complètes et exploitables obtenues, 175 concernent UCA et 321 Unistra. Le taux de réponse est de l'ordre de 8%, ce qui est faible mais habituel pour ce genre d'enquête. On note que le taux de participation est double (respectivement 354 UCA et 692 Unistra ont commencé à remplir le questionnaire sans aller au bout) ce qui souligne les limites d'un questionnaire long et possiblement décourageant.

Le présent rapport détaille les résultats obtenus en suivant les 5 parties du questionnaire :

- 1 - Caractéristiques des répondants
- 2 - Cycle des DR : quelles DR ?
- 3 - Gestion des DR : stockage, sauvegarde, archivage
- 4 - Utilisation des DR : diffusion, partage et réutilisation
- 5 - Recensement des besoins pour les DR

---

<sup>3</sup> <https://github.com/LimeSurvey>

<sup>4</sup> Dans la suite de ce rapport, nous distinguerons les « chercheurs » c'est-à-dire les personnes effectuant de la recherche, et les « Chercheurs » c'est-à-dire les personnels ayant un statut de Chercheur affilié à un EPST.

<sup>5</sup> Les statuts de Chercheurs et Ingénieurs relèvent des EPST. Dans le cas de l'UCA, ont été contactés ceux faisant partie d'une UMR. Dans le cas de l'Unistra, ont été contactés ceux donnant des cours / vacations à l'Unistra.

## Partie 1 : Caractéristiques des répondants

La première partie de l'enquête vise à caractériser l'échantillon des répondants selon différents critères : statut, champ disciplinaire, genre, âge, et responsabilités.

Sur les 496 répondants, environ 65% sont affiliés à l'Université de Strasbourg (Unistra) et 35% à Université Côte d'Azur (UCA), ce qui reflète bien la taille respective des deux universités en termes de nombre d'étudiants.

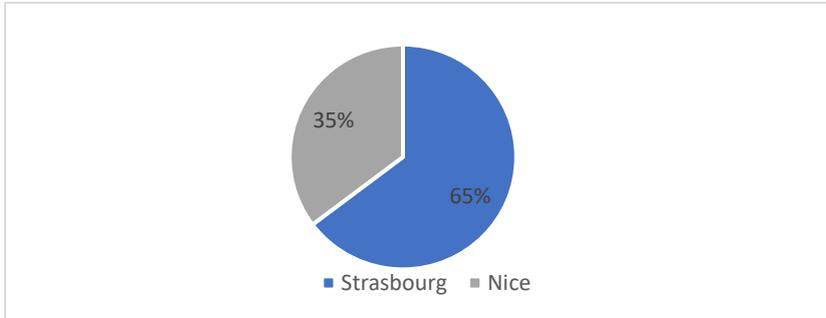


Figure 1 : Répartition des répondants selon l'université d'affiliation

### Statut des répondants

Globalement, tous les statuts de personnels ayant des activités de recherche sont présents dans l'échantillon<sup>6</sup>, avec une légère sur-représentation des Enseignants-Chercheurs et une forte sous-représentation des doctorants, si l'on compare aux données nationales<sup>7</sup>. Ces deux points concernant la représentativité sont plus marqués dans le cas de l'UCA.

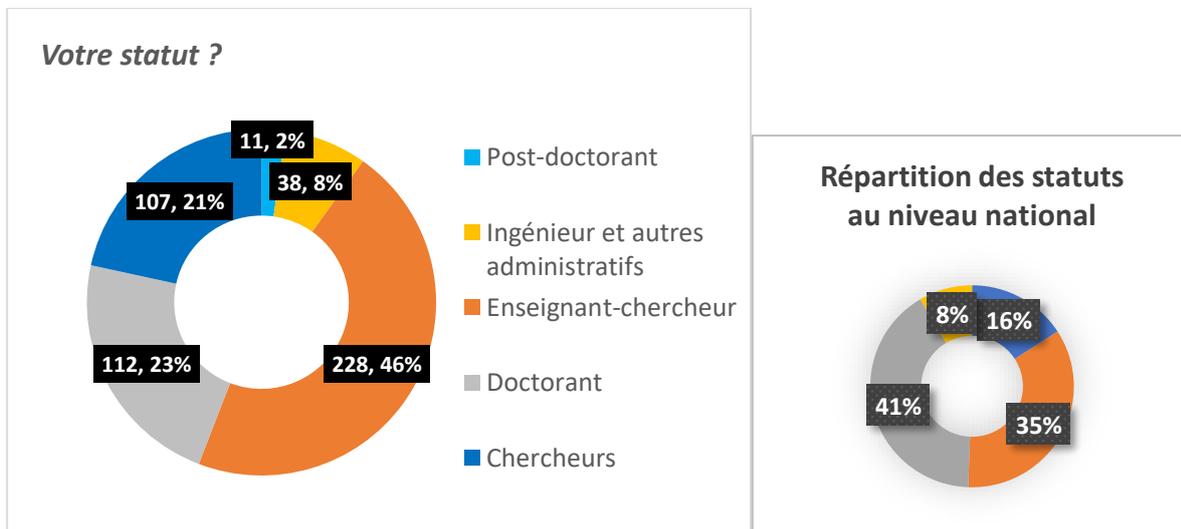


Figure 2 : Répartition des répondants selon leur statut et répartition nationale

<sup>6</sup> Le nombre de Postdocs répondants est négligeable et ne sera pas analysé par la suite.

<sup>7</sup> MESRI (2020) : L'état de l'Emploi scientifique en France. Rapport 2020 (page 92). Disponible à : [https://cache.media.enseignementsuprecherche.gouv.fr/file/2020/87/8/Etat\\_emploi\\_scientifique\\_2020\\_1341878.pdf](https://cache.media.enseignementsuprecherche.gouv.fr/file/2020/87/8/Etat_emploi_scientifique_2020_1341878.pdf)

## Champ disciplinaire et genre

La répartition par discipline offrant une trop grande variété pour être exploitée, nous avons regroupé a posteriori les réponses selon deux grands champs disciplinaires<sup>8</sup> : Sciences Humaines et Sociales (SHS) versus Sciences, Technologies et Médecine (STM), dont on observe qu'ils apparaissent de manière très équilibrée voire paritaire. On note des proportions inversées sur les deux universités à ce niveau, avec 65% de répondants relevant des SHS côté UCA, et 58% de répondants relevant des STM côté Unistra.

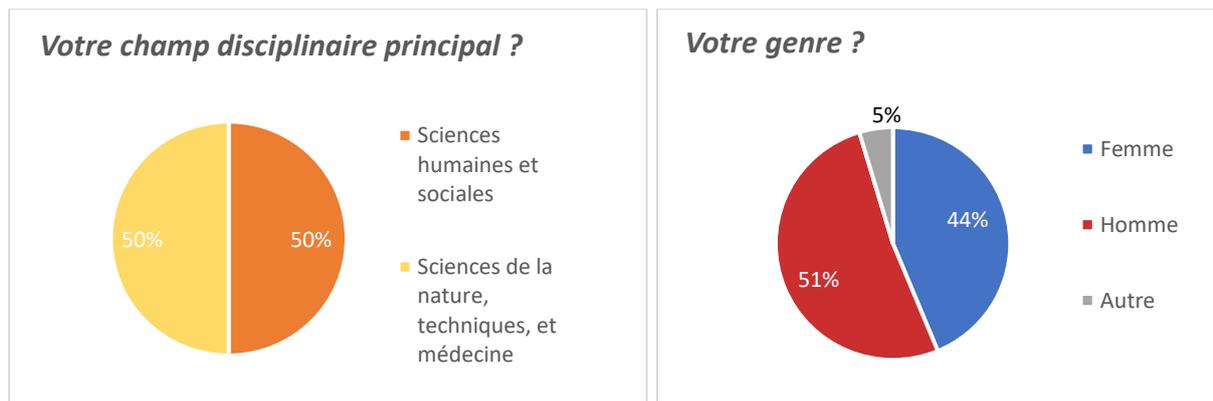


Figure 3 : Répartition des répondants en fonction de la discipline et du sexe

Concernant la représentation des genres, le groupe des répondants se divise en deux sous-ensembles reflétant assez bien les proportions hommes femmes au niveau national. Cette répartition se retrouve au niveau plus fin de chaque université. Un croisement entre le genre et la discipline montre que, conformément à ce qui est attendu, les proportions d'hommes et de femmes se consacrant aux STM et respectivement aux SHS sont inversées : 57,4 % des hommes effectuent leur recherche en STM et 56,7% des femmes font leur recherche en STM.<sup>9</sup> Il est délicat de dissocier l'effet du genre de celui de la discipline sur les pratiques de données de recherche. Il semble bien que le genre ne soit pas discriminant et nous ne le mettrons guère en avant dans la suite de ce rapport.

## Classes d'âge

Dans la Figure 4, les années de naissance ont été regroupées en 6 classes d'âges décennales allant des moins de 30 ans (nées entre 1990 et 1999) aux plus de 60 ans (nés en 1949 ou avant).

La classe d'âge des quadragénaires est la plus représentée, suivie dans l'ordre par celles des quinquagénaires, des moins de trente ans puis des trentenaires. Cela reflète en partie la proportion d'Enseignants-Chercheurs et de doctorants ayant répondu au questionnaire.

<sup>8</sup> Bien qu'imparfaite et critiquable cette distinction en deux masses donne à lire des distinctions.

<sup>9</sup> Autre manière de présenter ce croisement : les répondants de sexe masculin sont majoritaires dans les disciplines STM (ils représentent 60% des répondants STM). Une prépondérance de femmes apparaît dans les profils des répondants relevant des SHS.

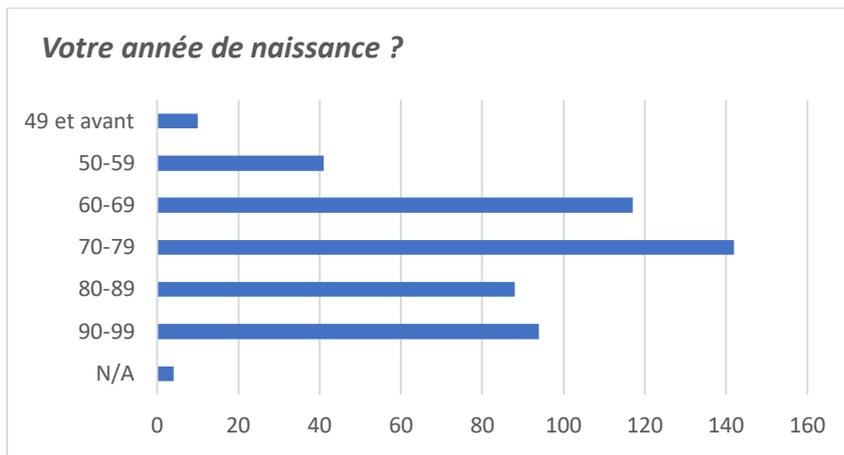


Figure 4 : Répartition des répondants par classe d'âge

## Responsabilités

Plus de la moitié, soit 54%, des répondants, exercent une ou plusieurs responsabilités, ce qui représente 266 personnes. Parmi elles, 103 ont des responsabilités qui vont au-delà de l'encadrement de thèses, de la direction d'une équipe de recherche ou d'une filière de formation, pour nommer celles qui reviennent le plus souvent.

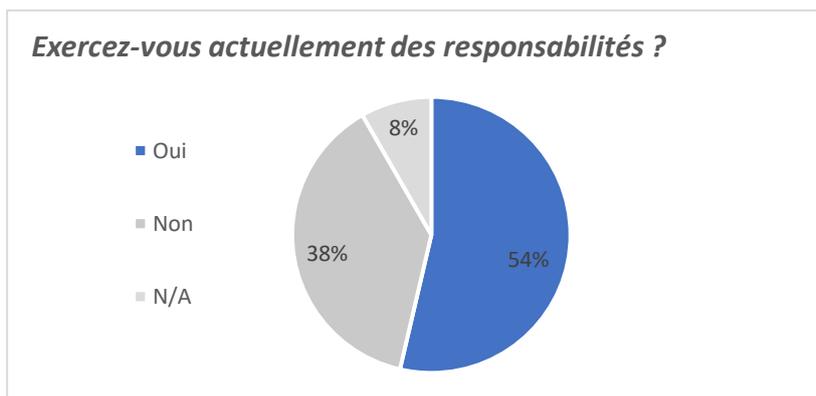
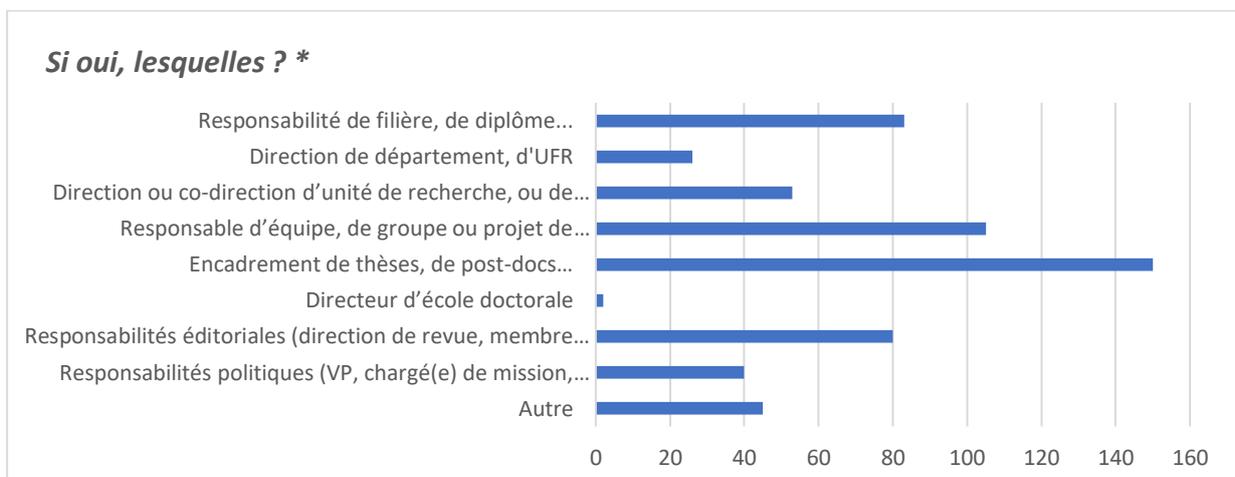


Figure 5 : Répartition des répondants avec ou sans responsabilité(s)



\* Réponses multiples possibles

Figure 6 : Répartition des répondants selon leur(s) responsabilité(s) - Détail

On peut ajouter que les histogrammes des responsabilités détaillées des deux universités montrent des structures très similaires, à ceci près qu'Unistra compte davantage de responsables de filières et d'équipe de recherche.

On notera pour finir que seulement 6% des réponses concernent la version anglaise du questionnaire, ce qui ne reflète pas le poids des chercheurs étrangers dans la recherche publique françaises<sup>10</sup>.

---

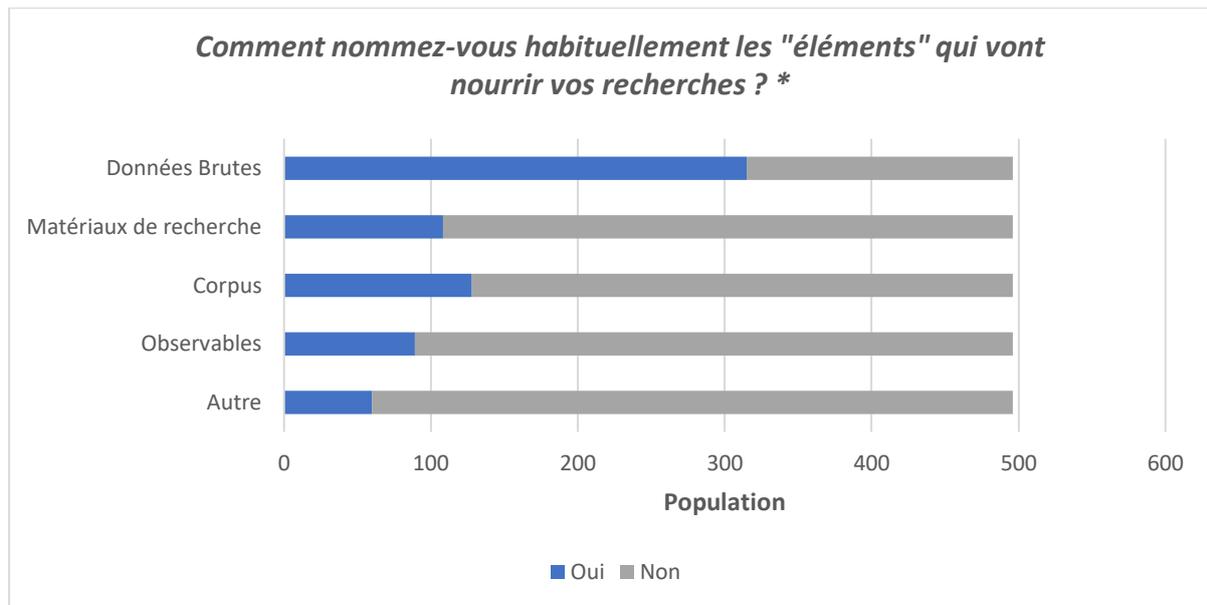
<sup>10</sup> 13% au niveau national en 2015, selon une note du MESRI de 2019, consultée le 01/09/2022 : [https://www.enseignementsup-recherche.gouv.fr/sites/default/files/imported\\_files/documents/NI\\_2019\\_5\\_chercheurs\\_etrangers\\_1099203.pdf](https://www.enseignementsup-recherche.gouv.fr/sites/default/files/imported_files/documents/NI_2019_5_chercheurs_etrangers_1099203.pdf)

## Partie 2: Cycle des DR, quelles DR ?

Cette partie vise à connaître l'acceptation des données par les répondants, qu'elles soient « brutes », « produites », « collectées », « analysées », « publiées » ou autre.

Par la suite nous tentons de caractériser le (ou les) type(s) de données des répondants en distinguant les données qui nourrissent un travail de recherche de celles qui résultent de ce travail, c'est-à-dire en portant une attention particulière au cycle de la donnée.

### Acceptation de la notion de données de la recherche



\* Réponses multiples possibles

Figure 7: Les diverses acceptations de la notion de données de recherche pour les répondants

Si les « données brutes » sont plus citées que les autres propositions on peut noter une différence importante entre STM et SHS (les chercheurs en STM adoptent plus abondamment cette terminologie), ceux en SHS montrent une dispersion sur les différentes acceptations avec une préférence pour le terme « corpus ». On note également de légères variations en fonction des classes d'âges (par décennie), laissant entrevoir des marqueurs d'usage de la langue (« données brutes » pour les plus jeunes, « matériaux de recherche » pour les plus âgés et le terme « observables » très peu usité). On notera également que dans la catégorie « autres », les termes « sources » et « données » (sans « brutes ») sont souvent nommés. Un travail sur « signifiant / signifié » au sujet des DR au sein des disciplines est peut-être à opérer, mais au-delà de l'acculturation langagière c'est surtout l'appréhension du cycle de la donnée qui se joue ici. Faire saisir ou bien comprendre les notions de « *raw data* », « *underlying data* », « *routine gathering data* », « *embeded data* »<sup>11</sup> semble essentiel pour mieux se comprendre, notamment dans le cadre de projets de recherche transdisciplinaire.

Comment appréhender par exemple, des répondants qui désignent les références bibliographiques comme des données, quand d'autres considèrent qu'ils n'ont pas de données parce qu'ils font de la construction théorique ? Qu'est-ce qui fait « donnée » ? Il ne s'agit pas de répondre ici à la question, mais de percevoir que pour communiquer une

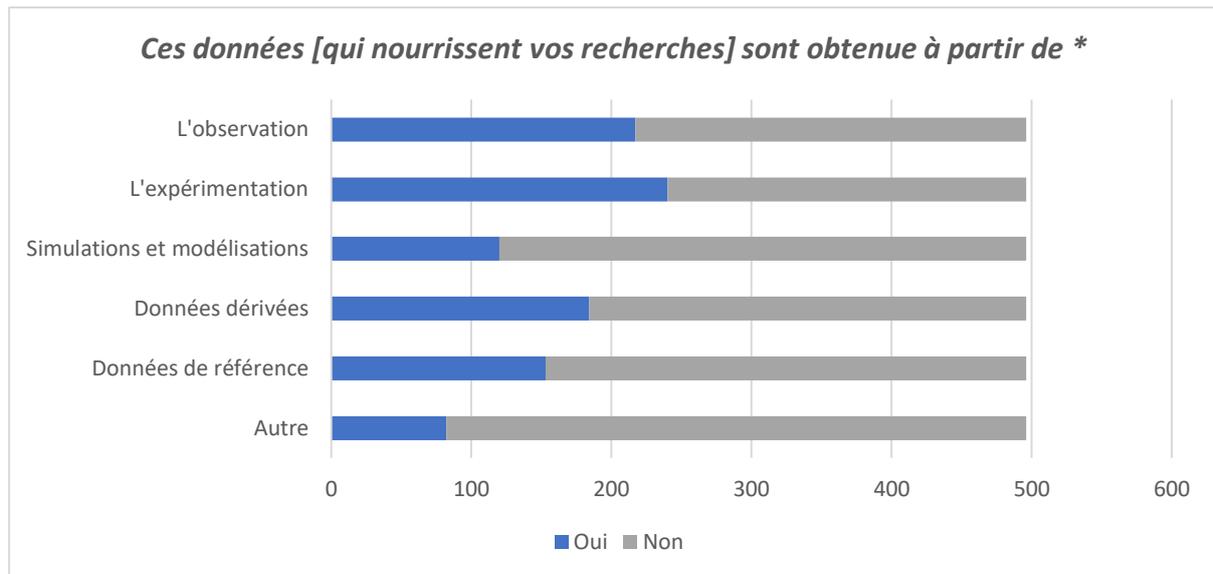
<sup>11</sup> « Données brutes », « Données sous-jacentes », « collecte systématisée de données », « données embarquées », voir à ce sujet le schéma de Francis André en annexe 3.

politique sur les DR ou mettre en œuvre un entrepôt, ou tout autre outil, il conviendra de prendre toute la mesure des acceptions au sein de chaque communauté de recherche.

Si la terminologie autour de la notion de « données de la recherche » est à travailler dans de futures enquêtes, ce peut être également un indicateur en terme d'acculturation au domaine. L'entrepôt national de données (Recherche Data Gouv)<sup>12</sup>, avec les Ateliers de la donnée<sup>13</sup> et les formations associées, façonneront et fixeront certainement la terminologie.

## Méthode d'obtention des données

Afin de mettre en évidence le cycle de la donnée, nous avons d'abord cherché à caractériser les données qui alimentent les processus de recherche des répondants.



\* Réponses multiples possibles

Figure 8 : Les différentes modalités d'obtention des données pour les répondants

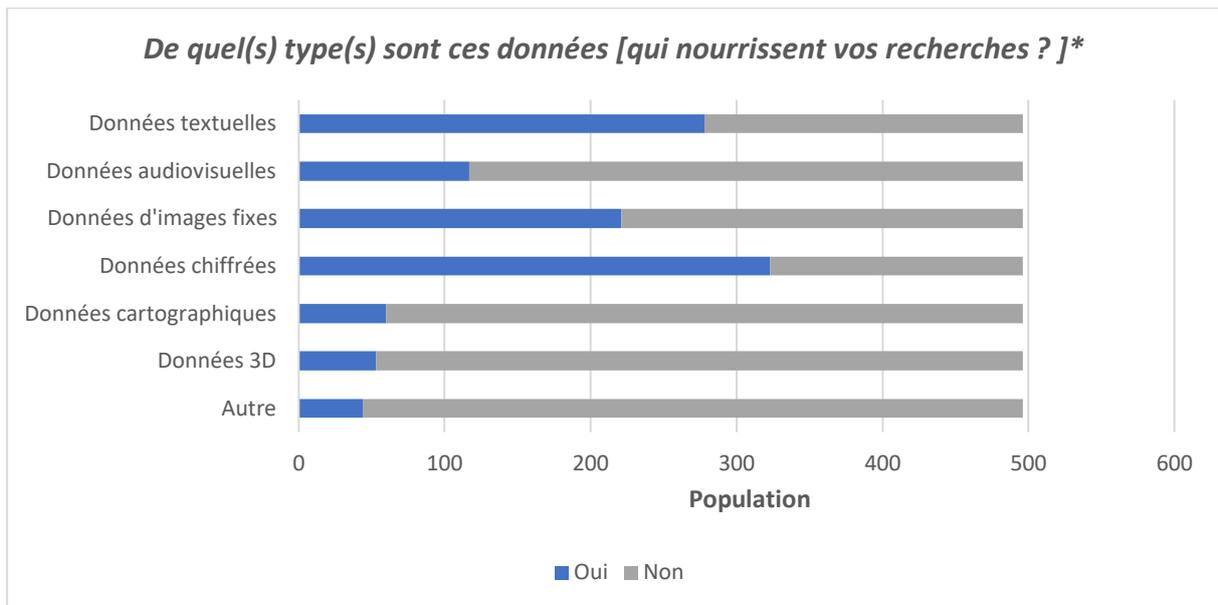
Avant de détailler les différentes modalités d'obtention des données, il convient, comme sur d'autres questions, de noter la forte similitude entre les deux universités interrogées.

Bien que les réponses soient relativement réparties, « l'expérimentation », suivie de près par « l'observation », et les « données dérivées » sont préférées par les répondants pour obtenir leurs données. Avec pour cette question un poids disciplinaire marqué des pratiques « d'observation » (principalement les SHS) et « d'expérimentation » (principalement les STM).

On notera également, pour les précisions de la proposition « autre », les termes « archives », « texte », « entretien », soulignant pour les répondants en SHS le fait de ne pas se reconnaître dans les propositions. Une fois encore l'acception des termes en lien avec le cycle de la donnée, et même, au-delà avec le cycle de la recherche, pointe l'enjeu d'une compréhension et d'échanges transdisciplinaires.

<sup>12</sup> <https://recherche.data.gouv.fr/fr>

<sup>13</sup> <https://recherche.data.gouv.fr/fr/page/ateliers-de-la-donnee-des-services-generalistes-sur-tout-le-territoire>

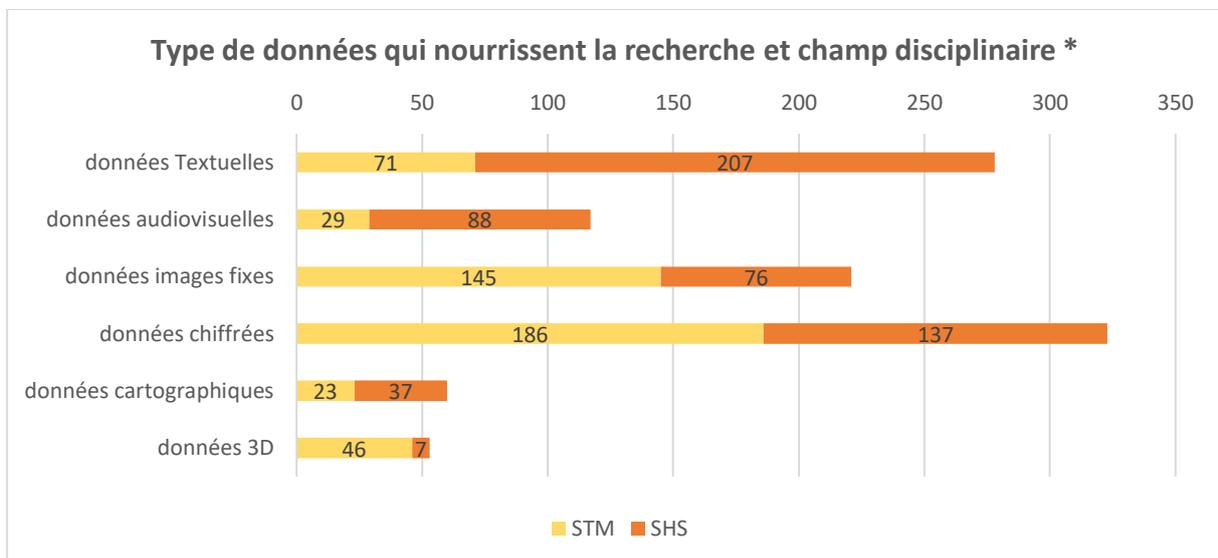


\* Réponses multiples possibles

Figure 9: Les différents types de données nourrissant la recherche des répondants

Les « données chiffrées », « données textuelles » et « données d’images fixes » constituent le trio le plus cité par les répondants.

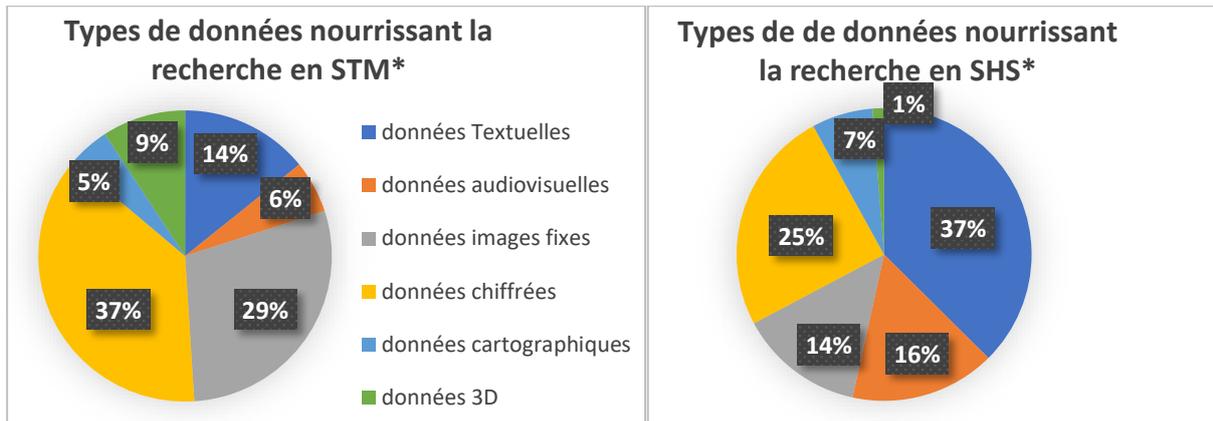
La Figure 10 ci-dessous permet de préciser pour chaque type de données la part des répondants en STM et en SHS. Ainsi les données chiffrées sont utilisées majoritairement par des chercheurs en STM, alors que les données textuelles sont surtout utilisées par des chercheurs en SHS, ce qui est conforme à l’intuition.



\* Réponses multiples possibles

Figure 10: Nombre de répondants par discipline pour chaque type de données nourrissant la recherche

Le graphique suivant (Figure 11) montre des priorités similaires sous un angle différent, en examinant, parmi la totalité des modalités de réponses cochées (et non sur l’ensemble des répondants), le pourcentage de chaque type de données mentionné par les chercheurs en STM d’une part, et par les chercheurs en SHS d’autre part. Les données chiffrées dominent largement en STM, suivies par les données d’imagerie. Les données textuelles dominent et représentent plus du tiers des données utilisées en SHS.



\* Réponses multiples possibles, % établi à partir de la totalité des modalités de réponse cochées  
 Figure 11 : Part des différents types de données nourrissant la recherche en STM et en SHS

On peut noter une petite différence (non représentée ici) entre UCA et Unistra pour les données chiffrées en STM versus SHS. A UCA il y a une quasi équivalence entre STM et SHS (près de 30% des réponses cochées portent sur les données chiffrées pour les deux domaines disciplinaires), alors qu'à l'Unistra les données chiffrées nettement plus citées en STM (38% des modalités cochées) qu'en SHS (23% des modalités cochées).

Il ressort dans tous les cas que les « données textuelles » sont majoritaires en SHS. Là encore il conviendrait de lever des ambiguïtés terminologiques en lien avec le cycle de la donnée. S'agit-il « d'inputs » textuels dédiés à des processus de traitement informatique (on pensera au *Text Data Mining* par exemple), ou la lecture d'une publication qui donne lieu à une prise de note (auquel cas ce serait plutôt les notes qui seraient les données textuelles) ?

### Production des données

La question suivante porte sur la contribution personnelle du chercheur à la production des données nécessaires à sa recherche.

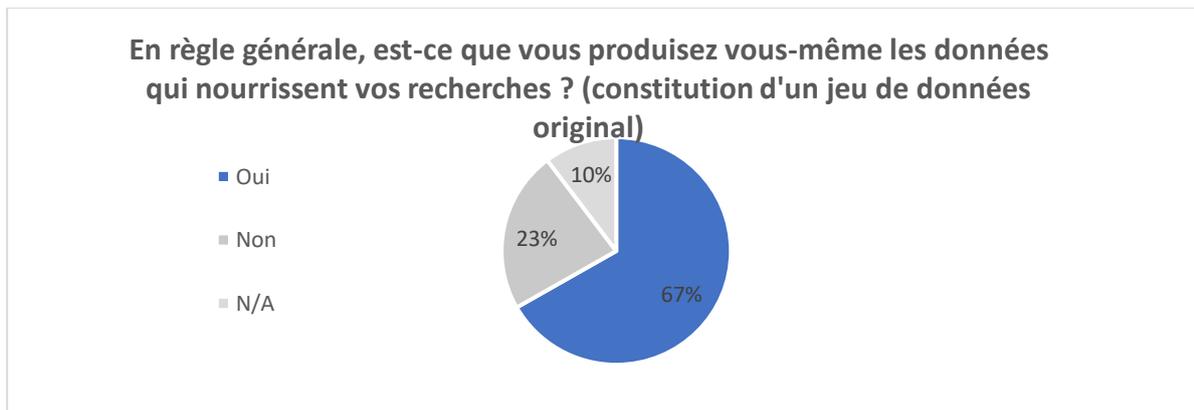


Figure 12: Part des répondants produisant eux-mêmes leurs données

Cette question, associée à celles sur le partage en partie 4, pourrait permettre de mieux cerner l'hypothèse selon laquelle les chercheurs qui ont passé beaucoup de temps à produire des données sont soit plus enclins à les voir réutilisées (flux d'échanges entre chercheurs), soit au

contraire réticents à les partager (compétition entre équipes)<sup>14</sup>. La distinction « données froides » / « données chaudes » est utile pour mieux cerner les comportements.

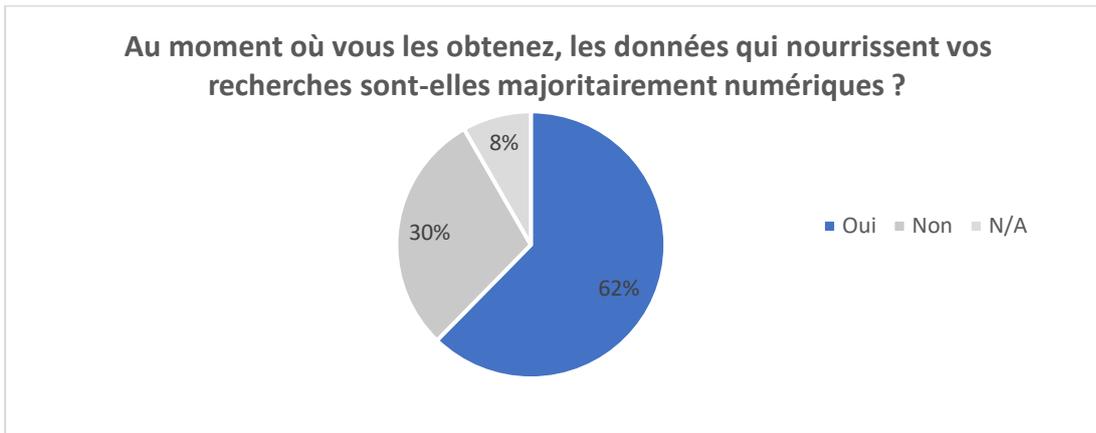


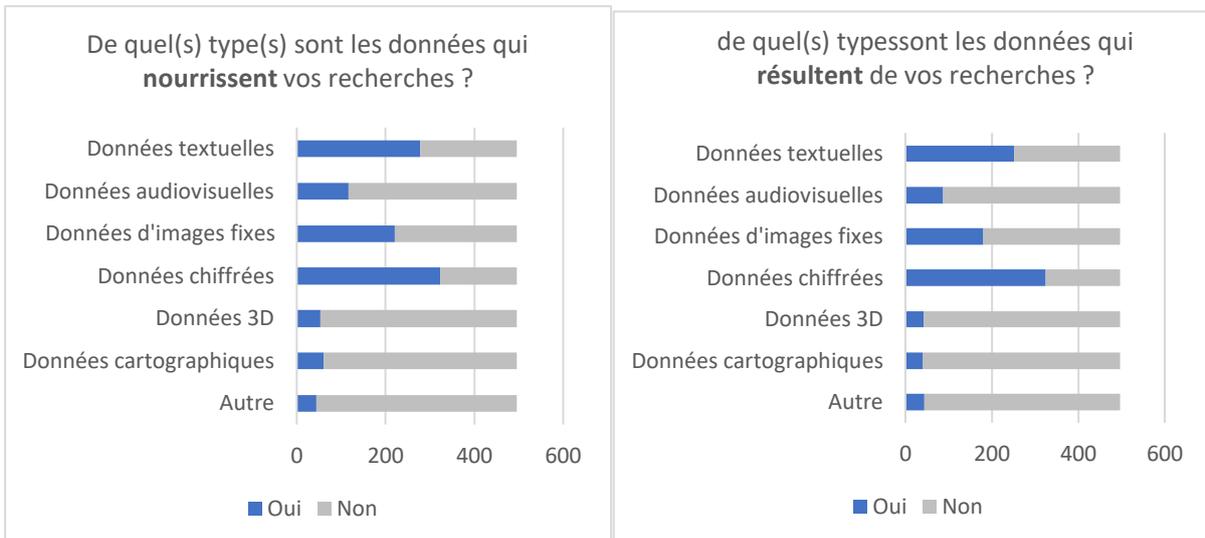
Figure 13: Part des répondants travaillant directement à partir de données numériques

On pourra s'étonner des 30% de « non », alors que tout semble « nativement numérique », du moins « numérisé » à la source. S'agit-il réellement d'archives et de manuscrits ? Ou bien de photos et autres enregistrements non numériques (par ex argentique ou super8) ?

Dans tous les cas il sera difficile ou plus complexe d'évoquer le partage de données via un entrepôt si ces dernières ne sont pas numériques.

### Type de données résultats

On s'intéresse ensuite aux types de données qui résultent de la recherche.



\* Réponses multiples possibles

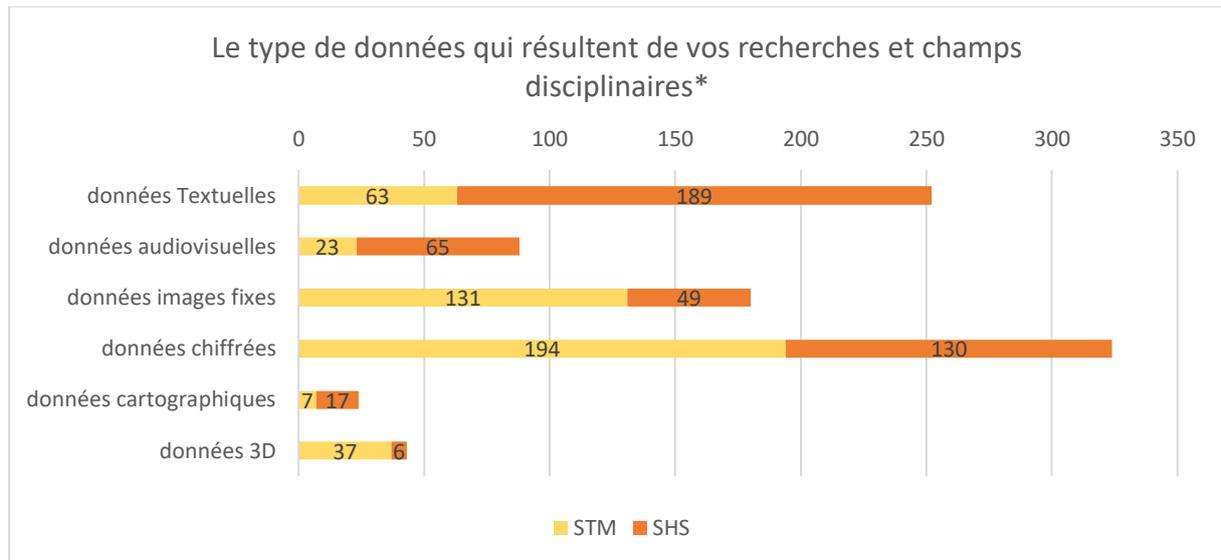
Figure 14 : Comparaison entre les données qui « nourrissent » et « résultent » de la recherche

On observe une similitude entre les données qui « nourrissent » les recherches et celles qui en « résultent » : les données chiffrées, textuelles, et d'image fixes sont les plus fréquemment citées et dans le même ordre.

<sup>14</sup> De fait, les 67% de répondants indiquant produire leur données ne se comportent pas de manière très différente des autres, en matière de diffusion et d'intention de partage.

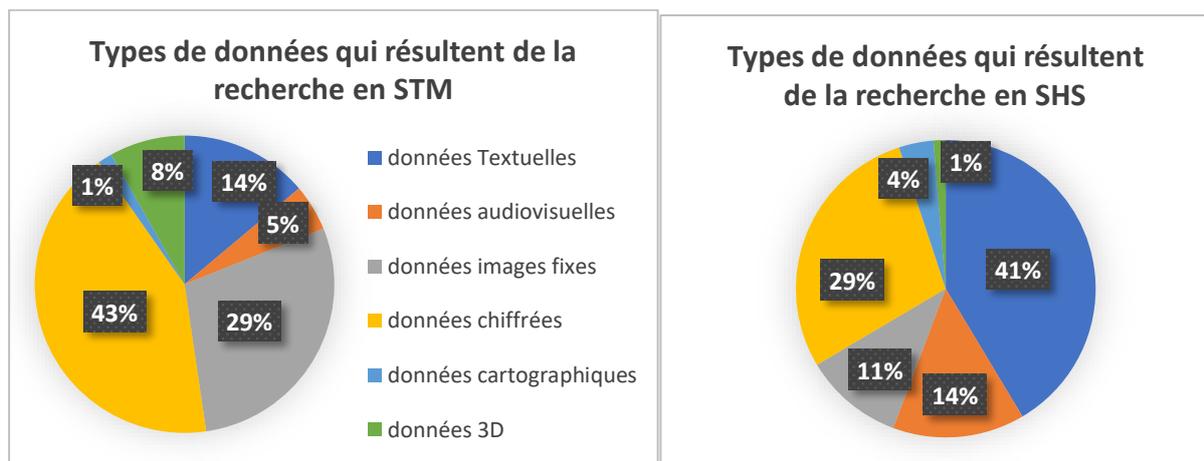
La distinction disciplinaire montre aussi des similitudes entre les *inputs* (données qui nourrissent) et les *ouputs* (données qui résultent) si l'on pense en termes de cycle de la donnée. Ainsi, les « données chiffrées » sont à nouveau plus souvent citées par les STM, et les données textuelles le sont davantage par les SHS (cf. Figures 15 et 16).

L' *Output* « données textuelles » interroge encore une fois : doit-on considérer qu'il s'agit simplement de publications (articles) ou bien doit-on également considérer l'enregistrement de notes ou annotations diverses et variées (par exemple, des verbatims, des notices bibliographiques, des relevés d'observation, des termes issus de *topic modeling* ou de TDM,...). Qu'il s'agisse d'*inputs* ou d'*ouputs*, l'ambiguïté serait à lever par des entretiens.



\* Réponses multiples possibles

Figure 15 : Nombre de répondants par discipline pour chaque type de données résultant de la recherche



\* Réponses multiples possibles, % établi à partir de la totalité des modalités de réponse cochées

Figure 16 : Part des différents types de données résultant de la recherche en STM et en SHS

## Partie 3 : Pratiques de gestion des données : stockage, archivage, description

La troisième partie du questionnaire traite des pratiques de gestion de données des chercheurs, en commençant par les données chaudes, définies comme étant « liées au projet de recherche, dont le contenu peut évoluer au cours de la recherche, faisant par définition l'objet d'un stockage ». Il s'intéresse ensuite aux données froides, c'est-à-dire celles qui sont « produites à l'issue du projet de recherche, dont le contenu est stabilisé, faisant l'objet d'un archivage ». Les pratiques individuelles et les perceptions sur les politiques de données sont questionnées. La fin de cette troisième partie est dédiée à la description des données.

### Stockage des données chaudes

Nous questionnons successivement le volume et les lieux de stockage des données chaudes. Les besoins de stockage sont évalués sur l'ensemble des projets de l'individu, suivant 5 classes allant de très faiblement volumineux (« moins de 50 Go ») à extrêmement volumineux (« plus d'un Po »). On peut noter des besoins de stockage assez importants, un peu plus de la moitié des répondants ayant besoin de plus de 500 Go, le reste situant ses besoins à moins de 500 Go. Si on extrapole aux non-répondants, des capacités de de stockage importantes semblent donc nécessaires, même si ce constat doit être nuancé par la difficulté pour tout un chacun de se représenter des unités de mesure de l'ordre du To ou du Po, et par la difficulté d'estimer le volume de stockage dont il aura besoin dans un avenir proche, compte tenu de l'évolution rapides des technologies et des services numériques ainsi que de l'incertitude intrinsèque aux activités de recherche.

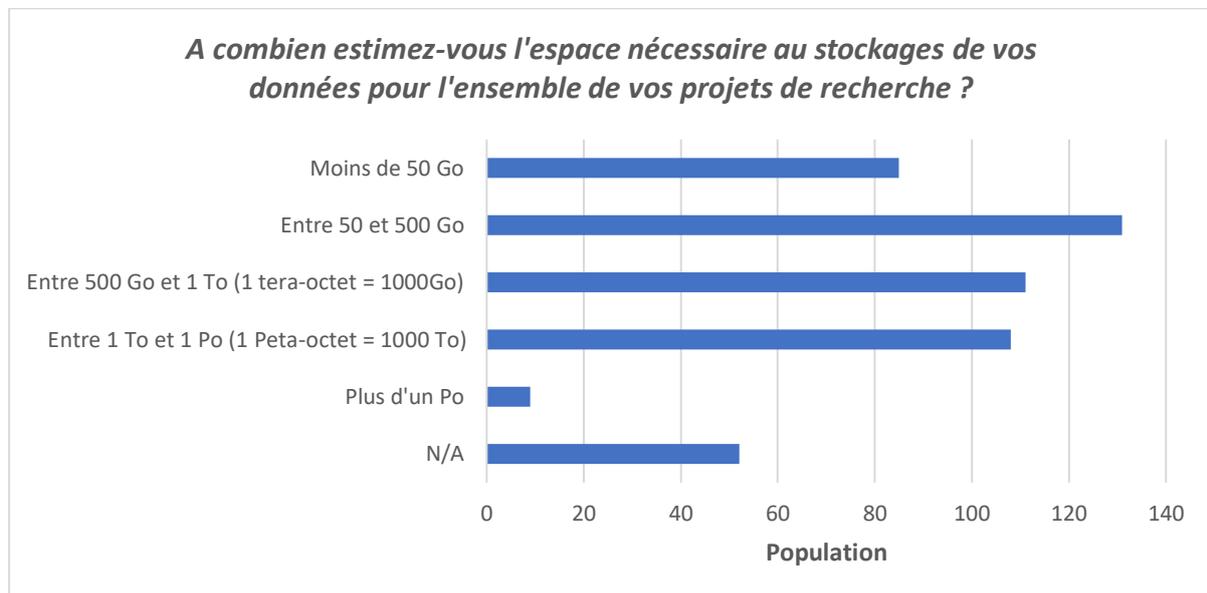


Figure 17: Estimation du volume de stockage par les répondants

Lorsque l'on affine l'analyse en croisant les besoins de stockage avec le champ disciplinaire, on observe que les SHS sont globalement moins gourmandes que les STM

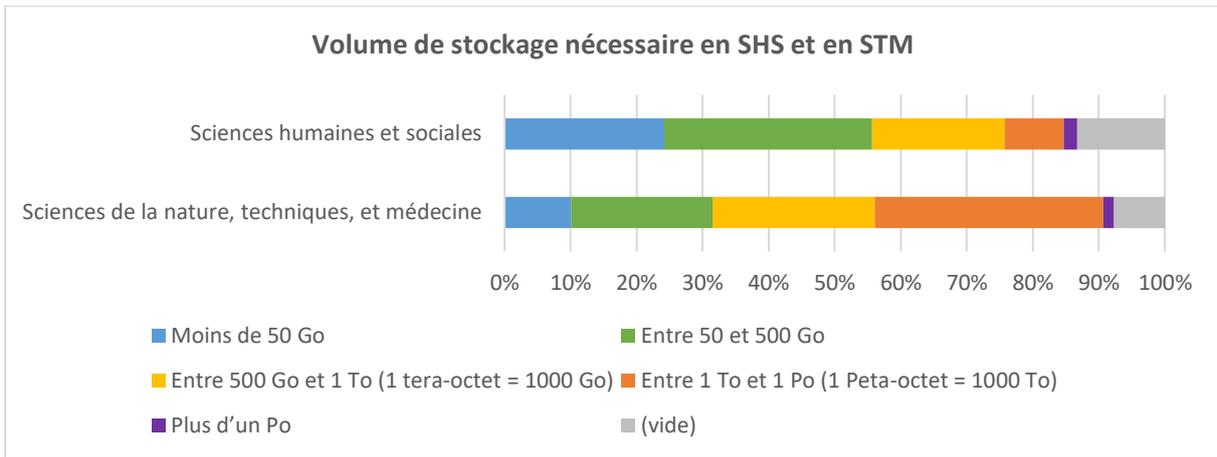
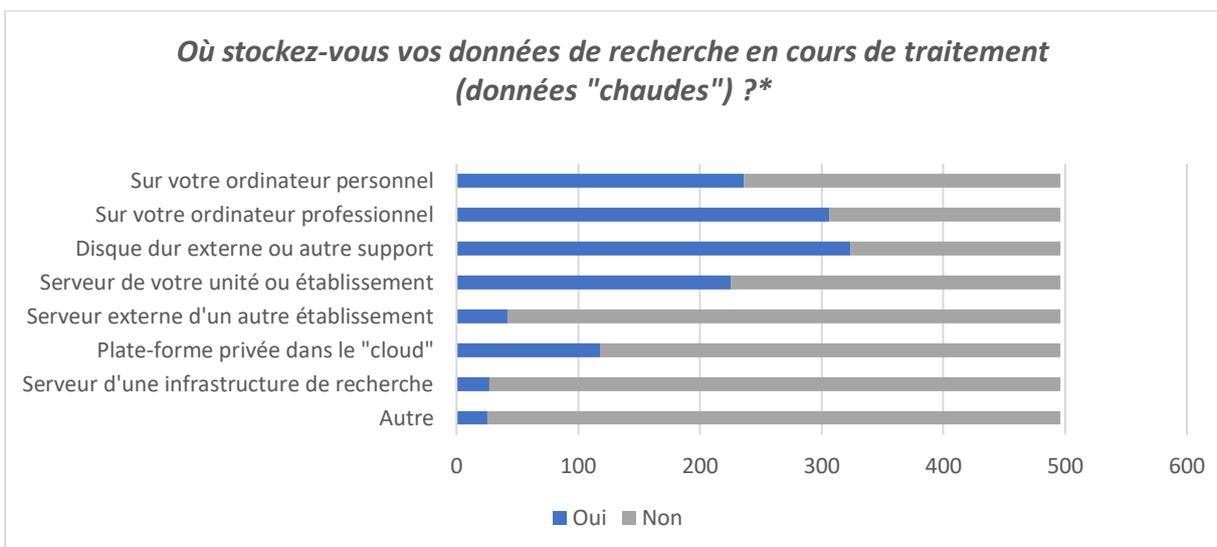


Figure 18 : Volume de stockage selon le champ disciplinaire

Plus précisément, il y a peu de différences majeures entre SHS et STM sur les besoins de stockage de niveau intermédiaire (entre 50 Go et 1 To) qui concernent environ la moitié des répondants dans les deux cas ; idem pour les besoins extrêmement volumineux (« plus d'1 Po ») qui concernent de toutes les façons peu de monde. En revanche, les deux grands champs disciplinaires s'opposent sur les classes périphériques : les SHS comprennent davantage de chercheurs ayant de faibles besoins (« moins de 50 Go ») alors que les STM ont davantage de chercheurs ayant de gros besoins de stockage (c'est-à-dire « entre 1 To et 1 Po »).

Le schéma suivant montre les lieux de stockage de données chaudes les plus utilisés. Les répondants pouvaient cocher plusieurs réponses et beaucoup de répondants stockent sur plusieurs supports : ordinateur personnel, professionnel, disques externes, serveurs de labo ou d'établissement, ou encore *cloud* privé. Le « disque dur externe » est utilisé en priorité, c'est-à-dire par près des 2/3 des répondants des 2 universités, suivi par « l'ordinateur professionnel » (plus de 60% des répondants) et « l'ordinateur personnel » (la moitié des répondants). Le stockage sur un « serveur de l'unité ou de l'établissement » concerne un peu moins de la moitié des répondants. Les autres modes de stockage sont nettement moins fréquents, cependant l'usage de « plateformes privées dans le cloud » est fait par un quart des chercheurs.



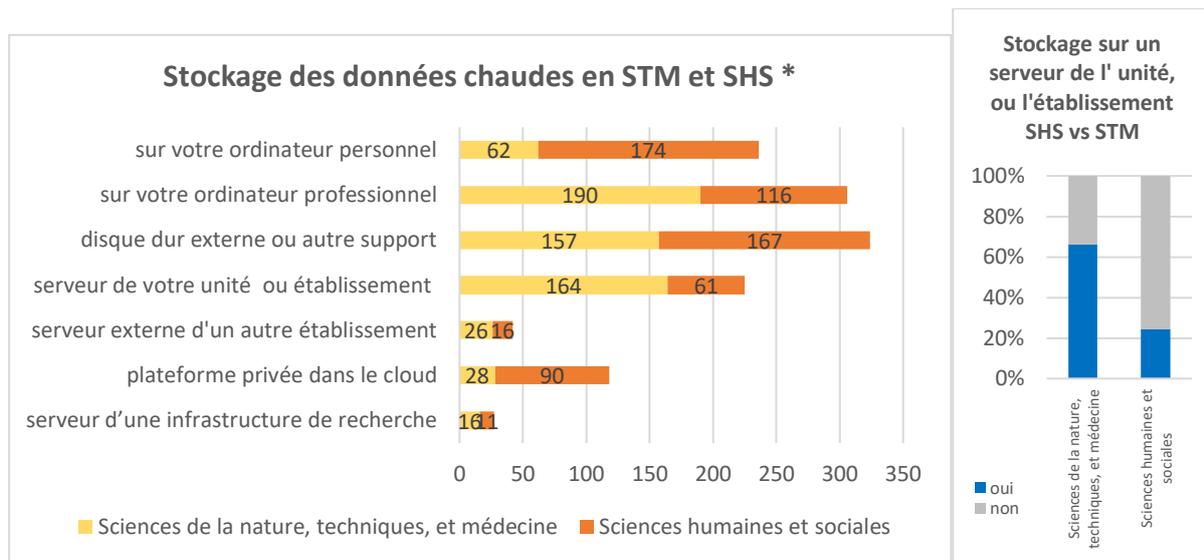
\* Réponses multiples possibles

Figure 19 : Lieux de stockage des données chaudes choisis par les répondants

En combinant les modalités de réponses on observe que le stockage reste largement une démarche individuelle : 239 personnes soit presque la moitié des répondants ne stocke jamais

ses données dans un entrepôt public (serveur de labo/établissement ou de TGIR), un tiers stocke ses données uniquement sur son ordinateur personnel/professionnel et/ou sur disque externe. Il est intéressant à ce stade d'examiner de plus près l'influence des pratiques disciplinaires.

La Figure 20 décompose les réponses positives selon les deux grands champs disciplinaires. STM/SHS. Sachant qu'il y a exactement autant de répondants en SHS qu'en STM, les croisements ci-dessous montrent des pratiques de stockage assez distinctes. Ainsi les SHS privilégient encore beaucoup l'ordinateur personnel<sup>15</sup> et utilisent davantage les plateformes privées, alors que les STM ont une propension nettement plus forte à utiliser des serveurs « internes » d'établissement. Cela pourrait s'expliquer par une différence de moyens informatiques selon les disciplines, ou peut-être par la prépondérance du travail individuel en SHS par opposition avec le poids du travail en équipe en STM.



\* Réponses multiples possibles

Figure 20: Nombre de répondants par discipline pour chaque lieu de stockage des données chaudes

Le lieu de stockage dépend aussi de l'âge du répondant. Ainsi les plus jeunes comme les plus âgés tendent à utiliser davantage leur ordinateur personnel. Les décennies intermédiaires (notamment sexa-, quinqu- et quadragénaires) utilisent plus largement les entrepôts de leur laboratoire ou de l'établissement (ce qui n'exclut pas l'usage d'ordinateur individuel).

Ceci est plus ou moins corroboré par l'analyse des mêmes pratiques de stockage en fonction du statut (cf. Figures 21 ci-après) : près de 70% des doctorants utilisent leur ordinateur personnel comme lieu de stockage. Ils mobilisent plus volontiers les plateformes Cloud que les autres catégories, et relativement moins les entrepôts de labos ou d'établissement. Cela est vrai dans une moindre mesure pour les Enseignants-Chercheurs, et cela présente une opposition nette avec le comportement des statuts de Chercheurs, et davantage encore avec les pratiques des Ingénieurs/ITA.

<sup>15</sup> Reste à préciser toutefois ce que l'acceptation ordinateur personnel recouvre exactement.

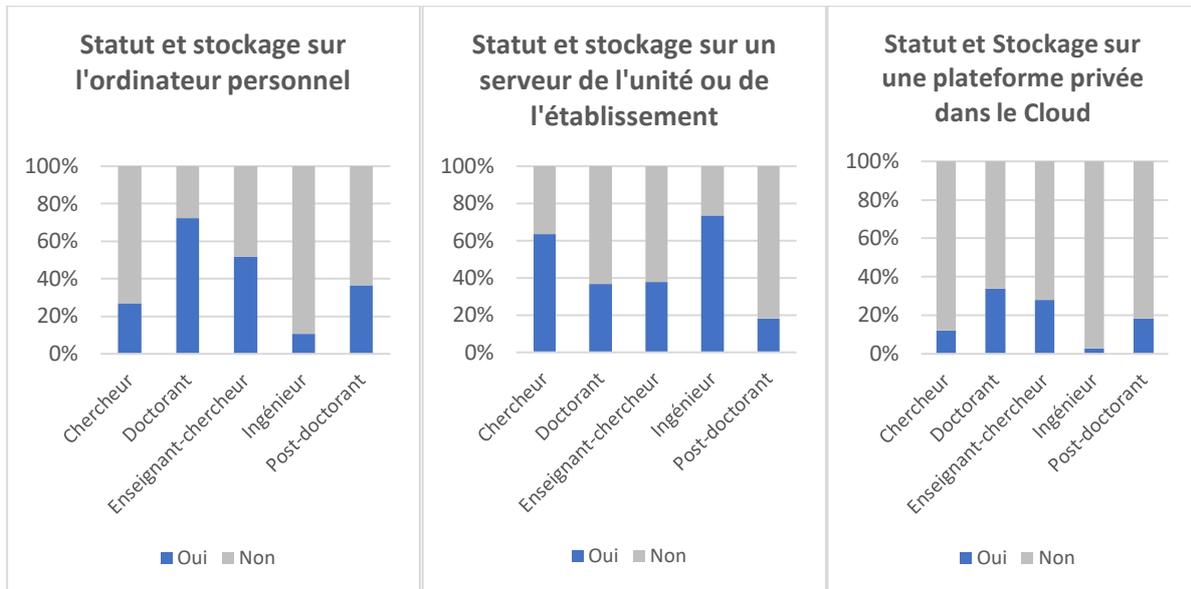


Figure 21 : Choix de différents lieux de stockage selon le statut du répondant

### Sauvegarde des données chaudes

Concernant par ailleurs les pratiques de sauvegarde des données, on observe que des copies sont faites massivement (par 83 % des répondants), mais qu'elles restent relativement peu automatisées (46 %).

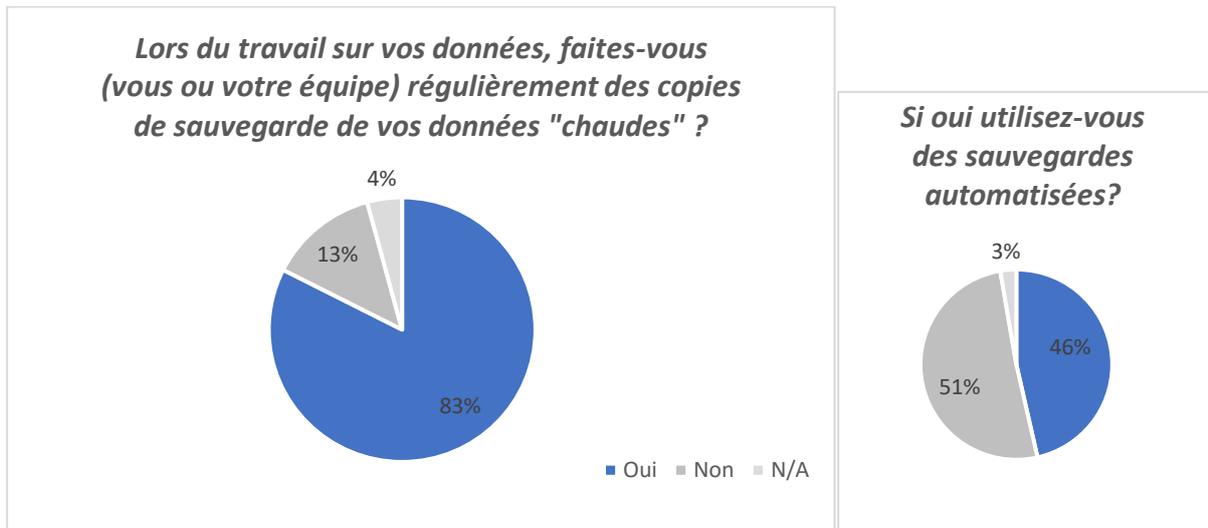
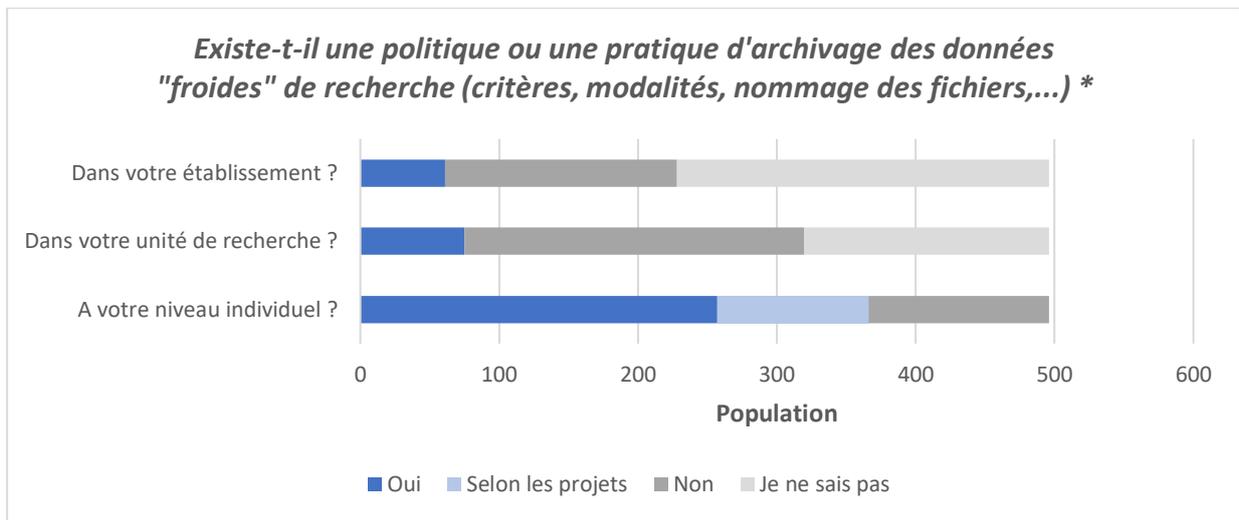


Figure 22 : Part des répondants effectuant des sauvegardes régulières et/ou automatisées de leur données chaudes

### Archivage des données froides

Les questions suivantes portent sur les pratiques individuelles d'archivage de données froides et les politiques des établissements en la matière (telles que perçues par les répondants). Le schéma ci-après montre que l'essentiel de l'archivage se fait avant tout « au niveau individuel ». Quand bien même les politiques de labo ou d'établissement existent, elle sont encore peu ou mal connues.



\* Réponses multiples possibles

Figure 23 : Perceptions et pratiques d'archivage des répondants à différents niveaux

Des croisements supplémentaires (non représentés ici) suggèrent que le niveau de connaissance s'améliore sans surprise avec le niveau de responsabilité.

Un aspect central des politiques de données concerne la conservation /pérennisation des données des chercheurs non permanents comme les doctorants et post doctorants. Moins d'un quart des répondants considère qu'une telle politique existe au niveau du labo. Près de la moitié (45 %) ne sait pas répondre. Cette lacune politique et le manque de connaissance sont plus marquées dans le cas des SHS.

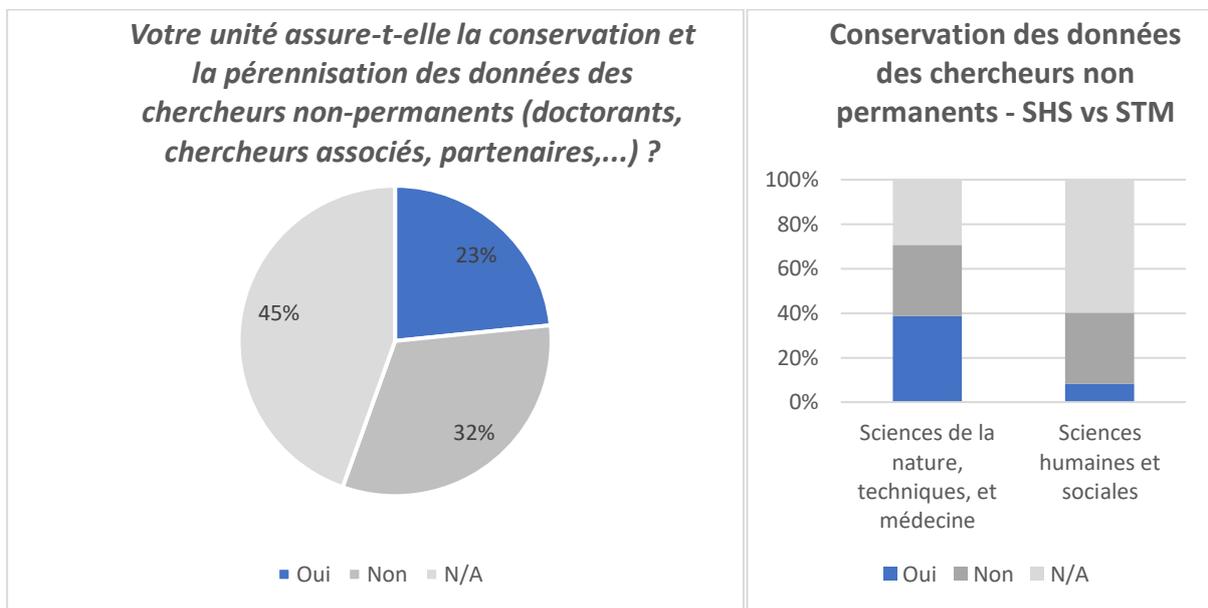


Figure 24 : Perceptions des répondants concernant la sauvegarde des données des chercheurs non permanents de leur laboratoire

## Plan de gestion de données

Un marqueur assez intéressant du degré de maturité des pratiques de gestion des données des chercheurs est la connaissance ou non des plans de gestion des données (PGD, ou DMP en anglais).

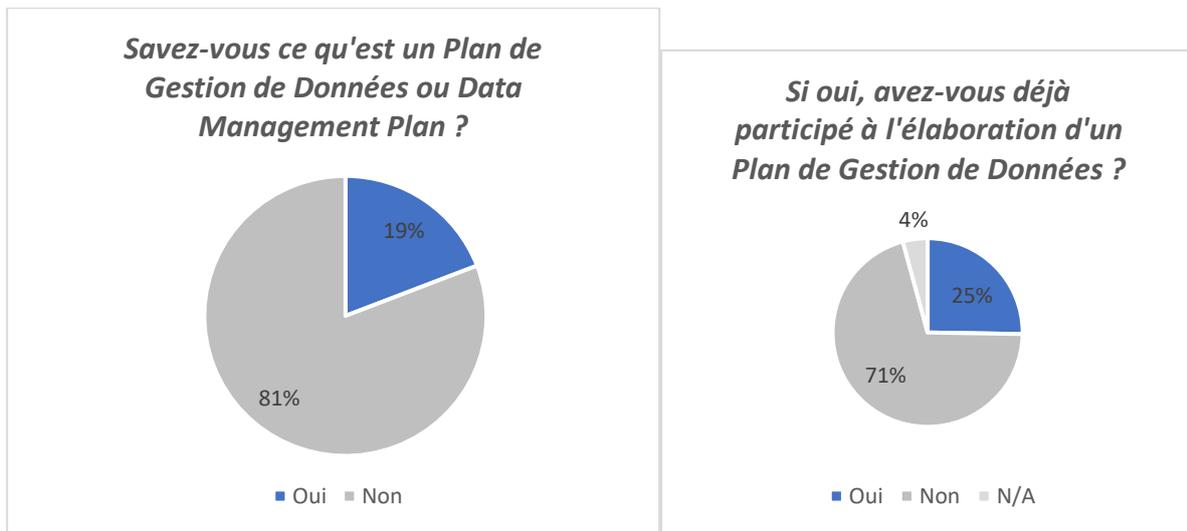


Figure 25: Pourcentage de répondants connaissant le PGD - en théorie et en pratique

En janvier 2020, moins d'un cinquième des répondants sait ce qu'est un PGD, et seulement un quart d'entre eux (24 personnes) a déjà participé à l'élaboration de ce type de document. Cette question extrêmement simple et rapide pourrait être utilisée dans un questionnaire Flash pour créer un indicateur du degré de connaissance des chercheurs en matière de données de recherche. Le même type de question pourrait être posé concernant les principes FAIR pour construire un indicateur complémentaire.

Bien que le niveau de connaissance des PGD ne soit pas très élevé, 28% des répondants ajoutent des informations décrivant leur jeu de données. Presque tous ajoutent des informations scientifiques, la plupart ajoutent des informations administratives et la moitié aussi des informations techniques. Cela se fait majoritairement sous forme de spécifications *ad hoc*, mais aussi pour 40% des cas dans le cadre de cahiers de laboratoires.

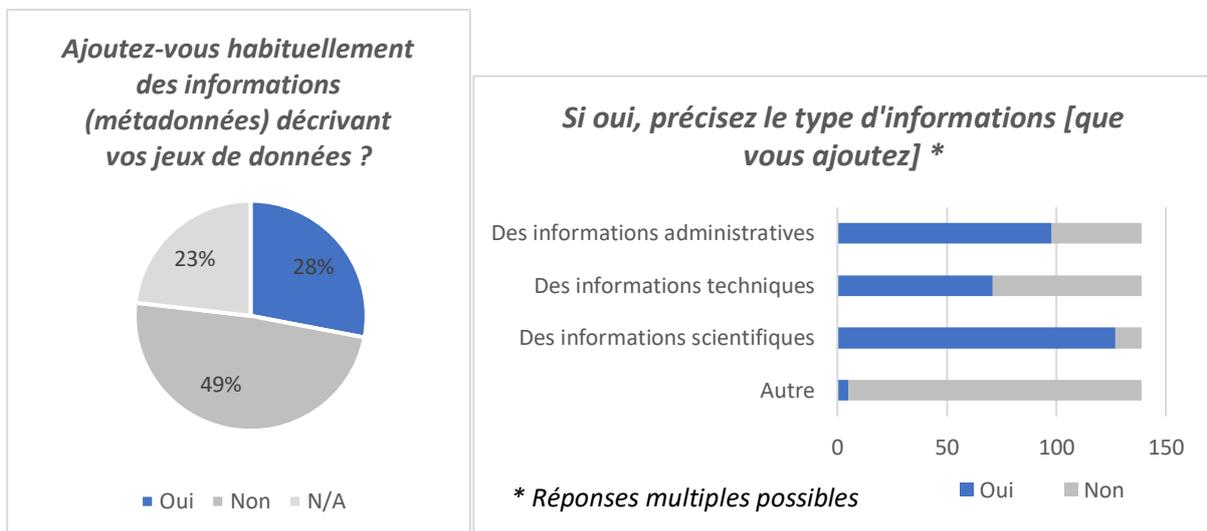


Figure 25 : Ajout de métadonnées par les répondants et types de métadonnées

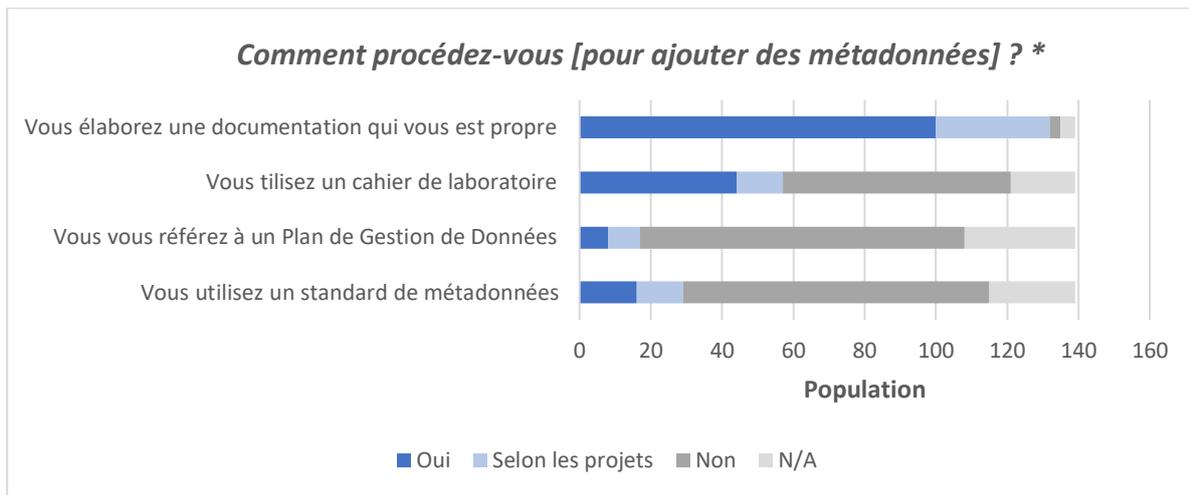


Figure 26 : Méthodes d'ajout de métadonnées par les répondants

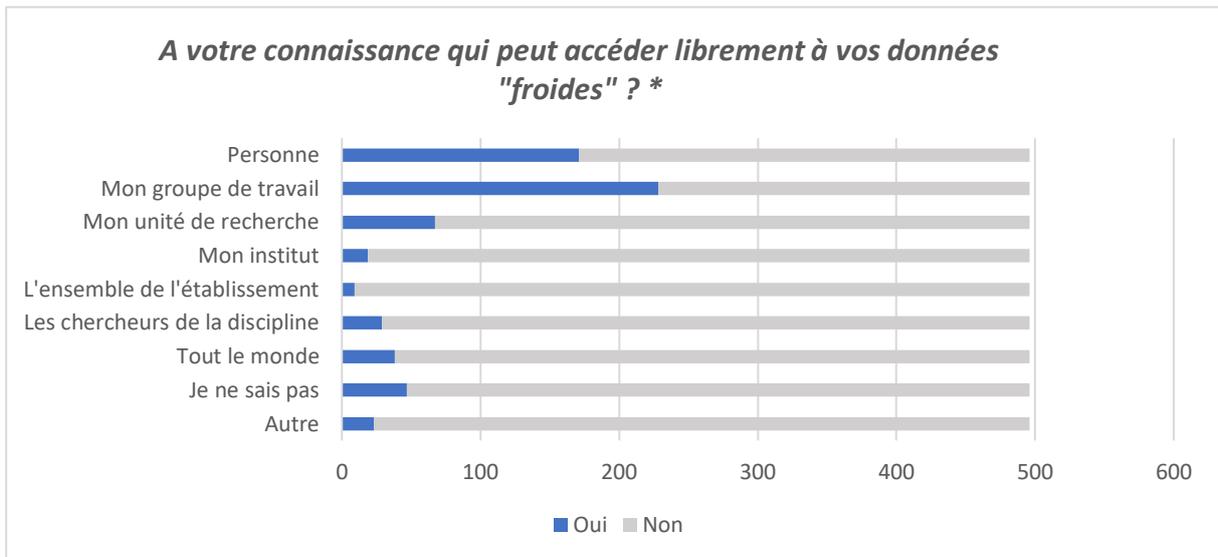
Dans l'ensemble, on retiendra de cette partie que les perceptions et les pratiques de gestion des données sont très marquées par le champ disciplinaire et par le statut du répondant. De plus, il semble que le niveau de connaissance des politiques en cours, des bonnes pratiques et des outils correspondants soient encore à un stade embryonnaire au moment de l'administration du questionnaire, c'est-à-dire au début de l'année 2020.

## Partie 4 : Utilisation des DR : diffusion, partage et réutilisation

Cette partie porte sur l'accès et les pratiques de partage des données de recherche, mais aussi sur l'usage de données d'autrui, ainsi que les incitations et les freins du chercheur à ouvrir ses données suivant les principes de la science ouverte.

### Accès aux données froides des répondants

La première question, à choix multiples, cherche à identifier les différents cercles qui ont accès aux données des chercheurs interrogés.

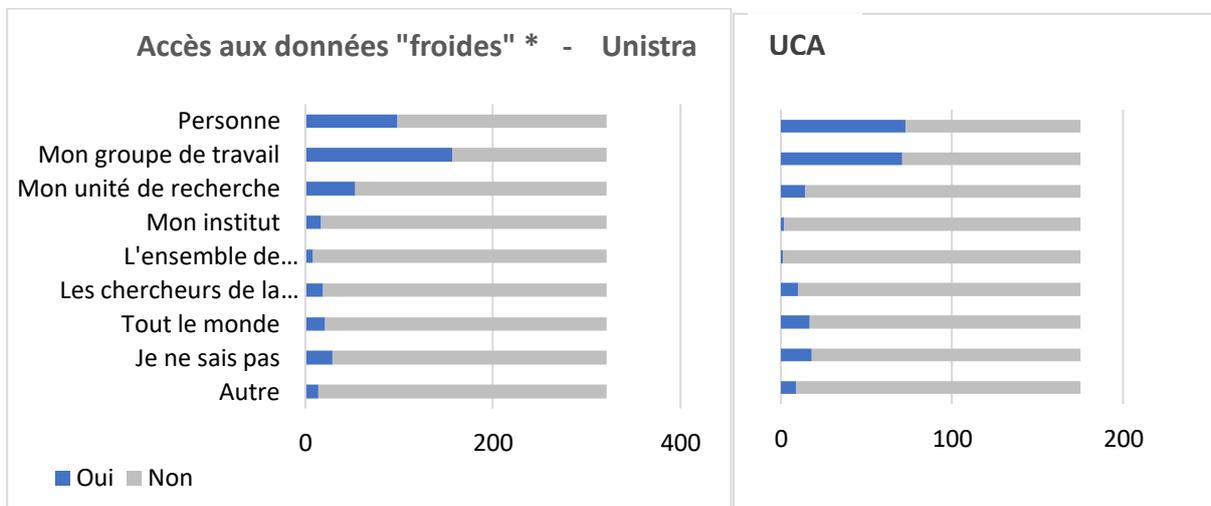


\* Réponses multiples possibles

Figure 27 : Degré d'accès aux données froides des répondants

Le constat est clair : de nombreux répondants considèrent que l'accès à leur données est très limité. Ainsi, un tiers pensent que « personne » n'y a accès, et 46% que l'accès est restreint à leur « groupe de travail ».

Les réponses sont proches dans les deux universités (Figure 28), avec cependant un accès plus restreint dans le cas d'UCA (cf. réponse « Personne »).



\* Réponses multiples possibles

Figure 28 : Degré d'accès aux données froides par établissement

Ceci peut être lié à la part dominante des SHS parmi les répondants d’UCA, dans la mesure où le travail individuel est plus fréquent en sciences humaines, et le travail en équipe plus habituel en sciences dites « dures » (cf. Figure 29).

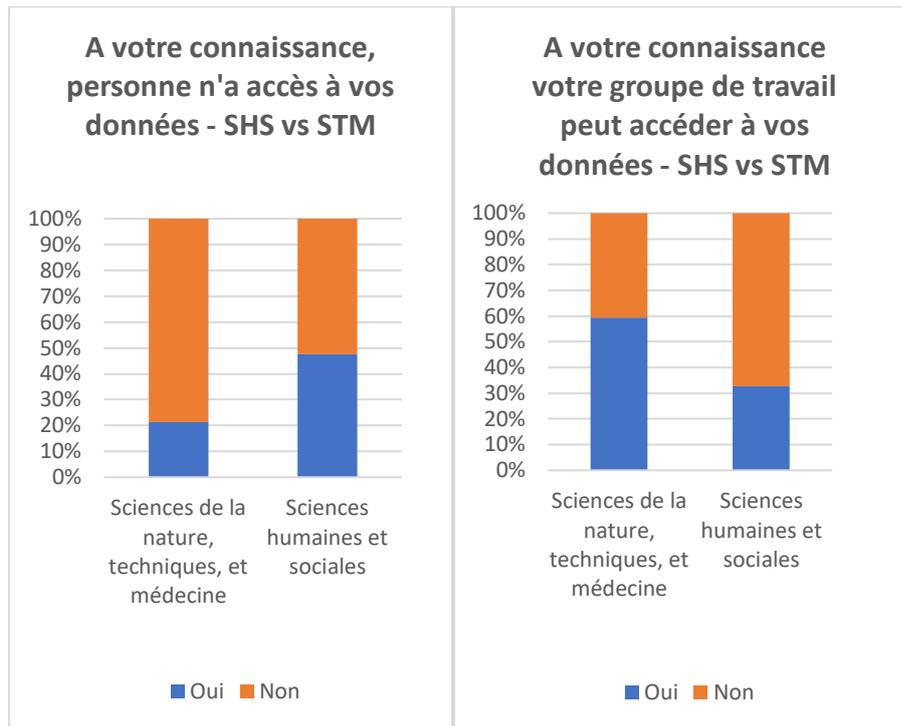


Figure 29 : Différences d'accès aux données en SHS et en STM

### Usage de données d'autrui

Si une large majorité de chercheurs ne donne pas accès à ses données, plus de la moitié, soient 263, ont déjà téléchargé les données d'autres chercheurs disponibles sur le web. A nouveau la configuration des réponses est similaire dans les deux universités, à ceci près qu'ils sont légèrement plus nombreux à avoir téléchargé les données d'autrui « au moins une fois » à UCA.

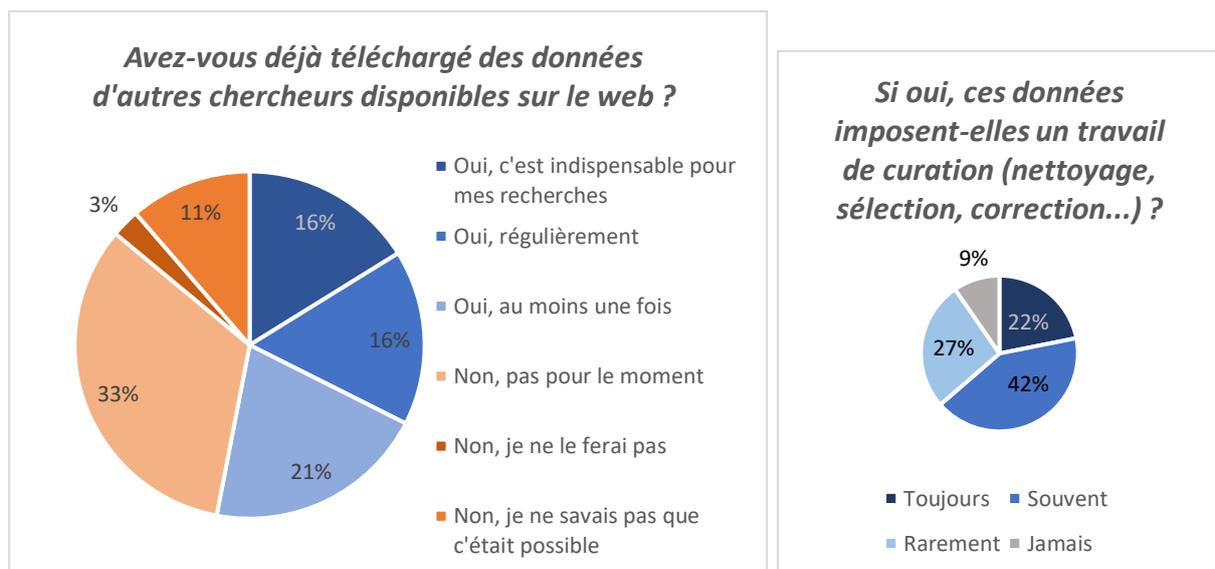


Figure 30 : Fréquence de téléchargement de DR et besoin de curation

Ces résultats doivent toutefois être utilisés avec précaution, car les termes « téléchargement de données sur le web » restent assez ambigus (par exemple : s’agit-il véritablement de jeux de données ou bien plus simplement de publications d’autrui ? S’agit-il vraiment de données produites par d’autres chercheurs et mises à disposition par ces mêmes chercheurs, ou bien plus largement de données de référence mises à disposition dans des bases dédiées et ouvertes à tous ?).

Pour ceux qui ont déjà téléchargé les données des autres, celles-ci font assez logiquement l’objet d’un travail de curation (au moins souvent dans 64 % des cas). Cependant ces données sont considérées par tous, du moins ceux qui se prononcent, comme étant plutôt de qualité (« Bonne » à 38% et « Moyenne » à 23%). La qualité se mesure selon la notoriété des auteurs et non de l’entrepôt qui les héberge (cf Figure 31 ci-dessous). Il conviendra de s’interroger ultérieurement sur les répondants qui n’ont pas répondu à ces deux questions (lassitude du long questionnaire ou réelles interrogations sur la qualité des données d’autrui ?).

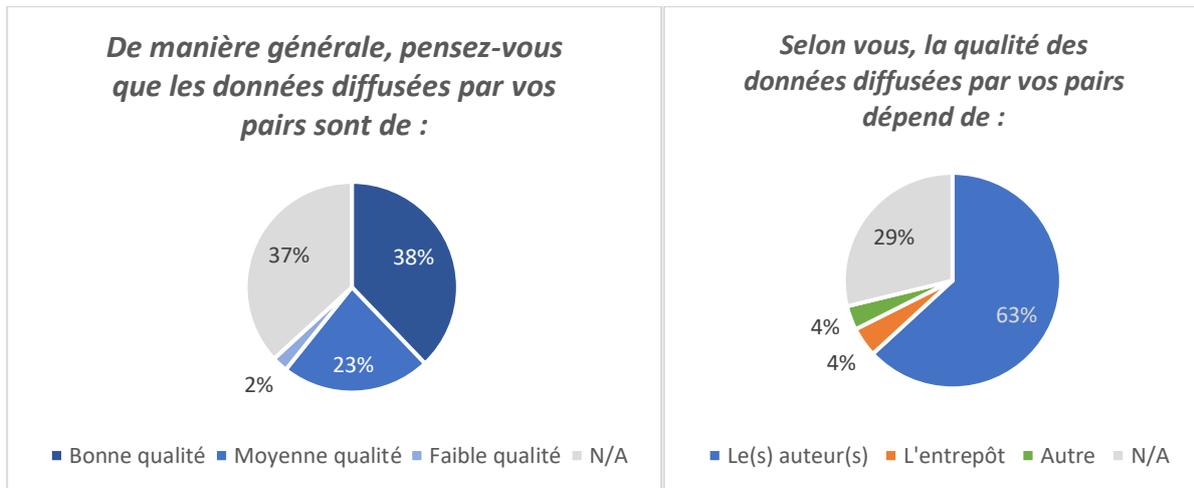


Figure 31 : Appréciation de la qualité des données téléchargées

### Diffusion de ses données sur le Web

La suite du questionnaire examine la diffusion des données « froides » sur le web au sens large, c’est à dire aussi bien leur mise à disposition sur des plateformes privées que leur partage via des plateformes respectant les principes de la Science Ouverte. Début 2020, assez rares sont les chercheurs qui ont déjà diffusé leurs données froide sur le web : 26% des personnes interrogées seulement.

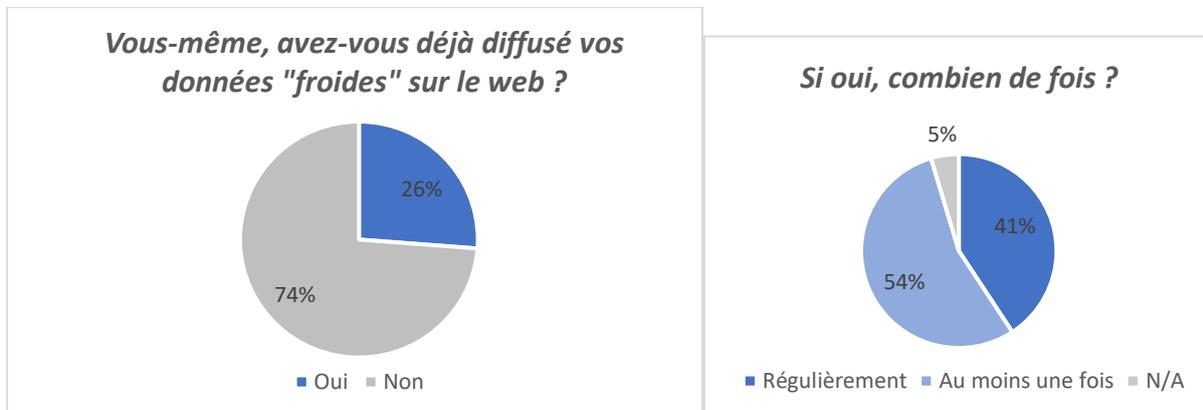
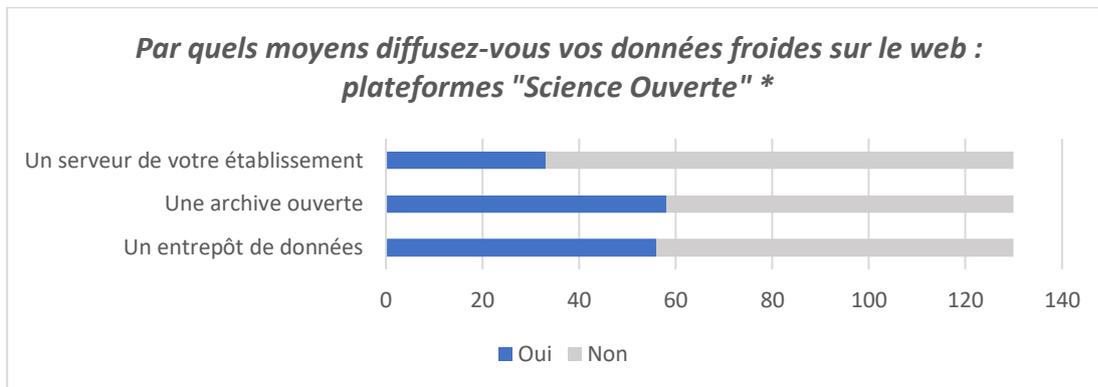


Figure 32 : Diffusion des données froides par les répondants

Une analyse plus fine par champ disciplinaire ne montre pas de différence significative à ce niveau (la diffusion semble un peu plus répandue en STM qu'en SHS, mais il n'y a guère que 5 points d'écart). En revanche, le fait d'exercer des responsabilités influence positivement la diffusion des données sur le web. C'est aussi vrai pour le statut : ainsi il y a proportionnellement plus de Chercheurs que d'E-C qui ont déjà effectué une diffusion des données au moment de l'enquête (un tiers versus un quart environ), et les doctorants diffusent nettement moins de données que les autres (12 à 13% seulement).

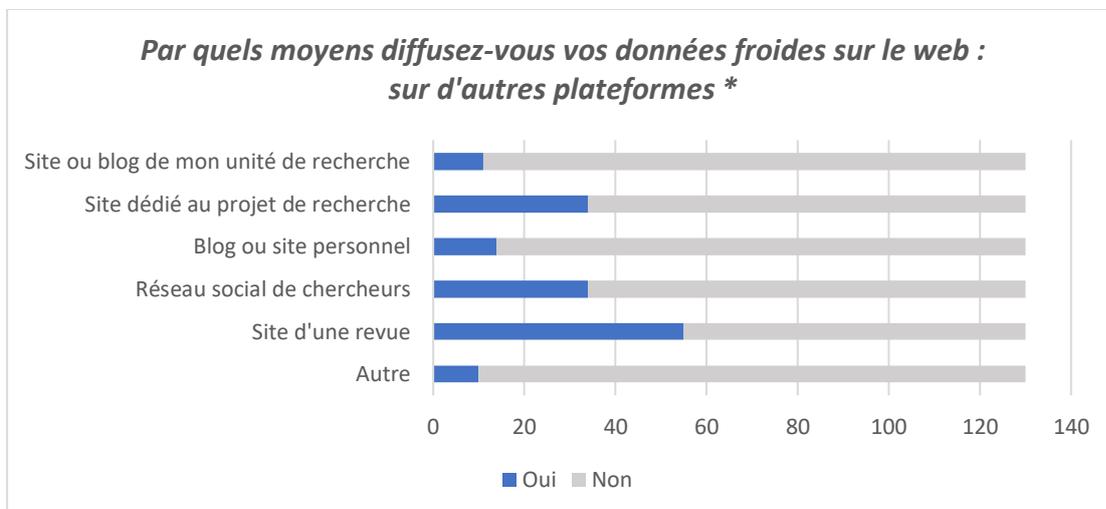
Pour ceux qui ont répondu oui, à la question de la diffusion de leurs données, les moyens utilisés méritent examen. Ceux qui partagent leurs données via des plateformes labellisées « Science Ouverte » semblent privilégier les Archives Ouvertes et les entrepôts de données « extérieurs » à leur établissement. Au moment de l'enquête, *Recherche Data Gouv* n'existait pas. Les établissements n'avaient pas tous des entrepôts dédiés, même s'il existait quelques entrepôts disciplinaires de longue date. Il était possible de déposer un jeu de données comme fichier attaché à un papier dans une archive ouverte telle que HAL.



\* Réponses multiples possibles

Figure 33 : Les différents types de plateformes SO utilisés par les répondants pour diffuser leurs données froides

Enfin, on peut noter que la terminologie « plateforme Science Ouverte » est vague et peut porter à confusion, puisqu'elle englobe aussi bien des archives ouvertes que des entrepôts de données ou encore des plateformes dédiées aux algorithmes (telle que *Software Heritage*).



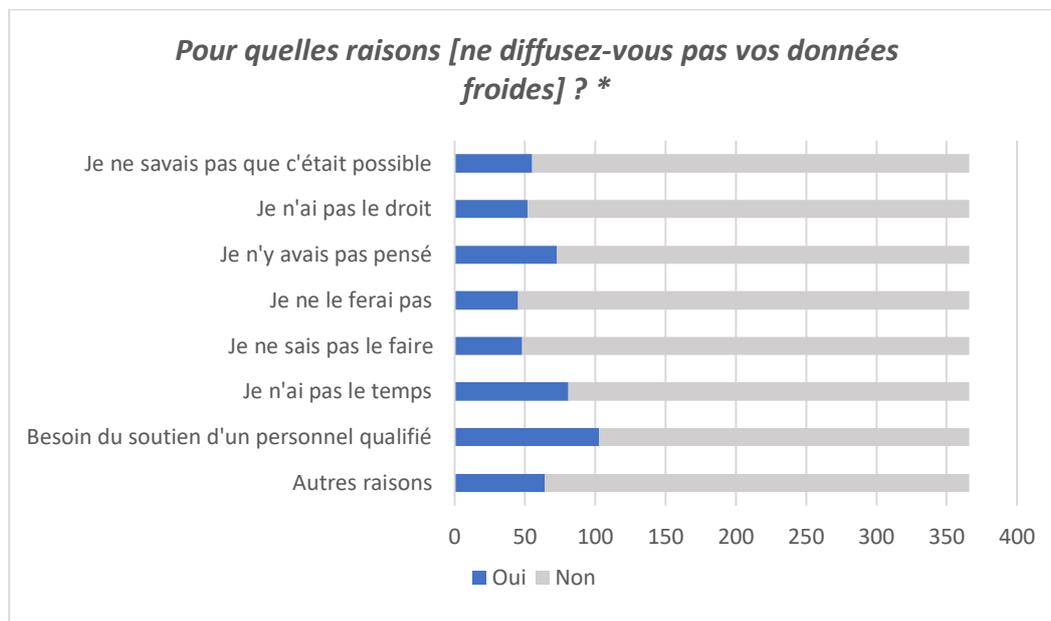
\* Réponses multiples possibles

Figure 34 : Autres types de plateforme utilisées pour diffuser les données froides

Quant aux autres plateformes de partage (non labellisées SO); « le site de la revue », un « réseau social de chercheurs » et le « site du projet dédié » constituent le trio de tête. On s'étonnera de la place accordée au « site de la revue ». Il serait intéressant, par exemple en entretien, de questionner la notion d'*embedded data*<sup>16</sup>. Des croisements avec les champs disciplinaires suggèrent que le partage via « réseau social » pourrait être marqué SHS, alors que l'entrepôt SO est plus fréquent en STM (mais cela reste à confirmer à une plus large échelle).

## Motifs de non-diffusion des données

Pour les 366 personnes qui ne partagent pas leurs données, une question additionnelle concernent les raisons de ce non partage.



\* Réponses multiples possibles

Figure 35 : Raisons de la non diffusion des données froides

Les différentes modalités de réponses possibles sont toutes assez bien représentées dans le schéma ci-dessus, cependant la raison la plus citée semble être le besoin de personnel qualifié. Ce manque de ressources humaines (RH) compétentes concerne plus du quart de ceux qui ne diffusent pas leurs données. En seconde position se trouve le manque de temps, puis le fait de « ne pas y avoir pensé ».

## Volonté de partager les données suivant les principes de la science ouverte

Les questions suivantes cherchent à déterminer si les répondants sont *a priori* favorables à la diffusion de leurs données froides, quelles pourraient être leurs incitations à le faire, et quels pourraient être selon eux les facteurs de blocage.

De très nombreux chercheurs (78% des répondants) sont au moins plutôt favorables au partage de leurs données, et 56% pensent que leurs données seront réutilisables. On pourrait y voir un certain paradoxe dans la mesure où « accéder », « télécharger », « diffuser » se fait

<sup>16</sup> « données embarquées », avec l'idée d'interroger la place des données embarquées (encapsulées) dans les pratiques des chercheurs.

de manière restreinte (1/3 des répondants) alors qu’être prêt à le faire dans le cadre de la SO est largement accepté ( ¾ des répondants).

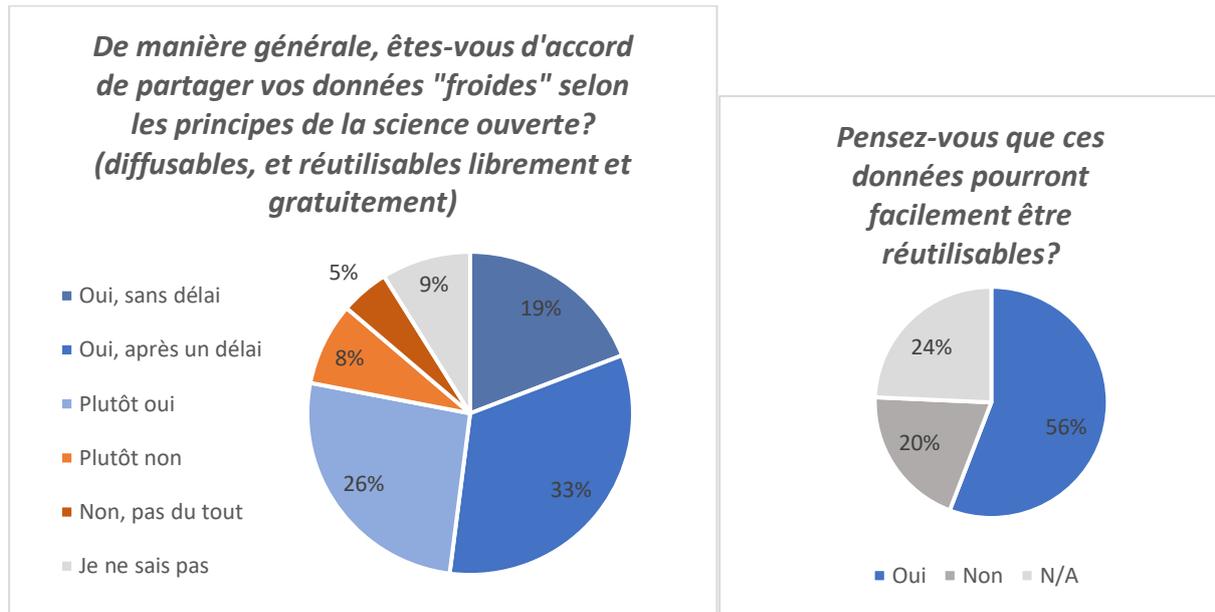


Figure 36 : Degré d’adhésion au partage des données froides selon les principes de la SO et caractère réutilisable de DR

On observe en outre que, comme pour de nombreuses autres questions, les résultats sont assez similaires dans les deux universités, comme le montrent les deux graphiques ci-dessous :

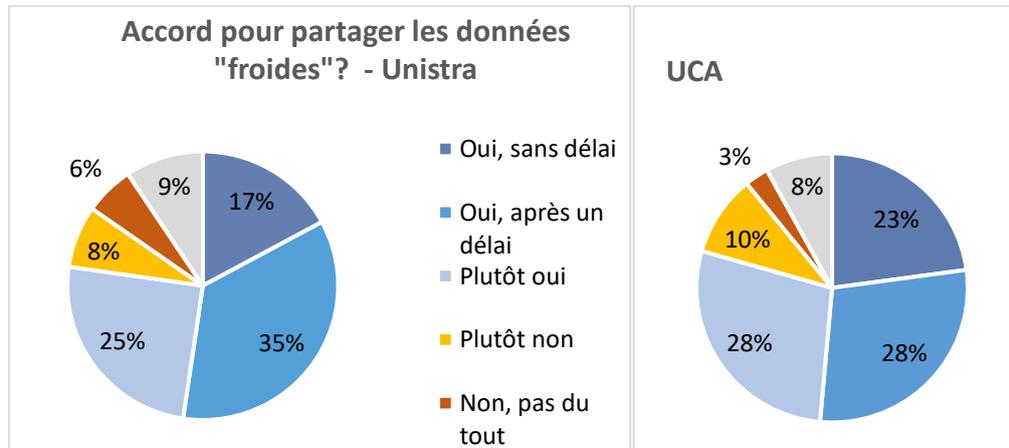


Figure 37 : Degré d’adhésion au partage de DR en SO pour chaque université

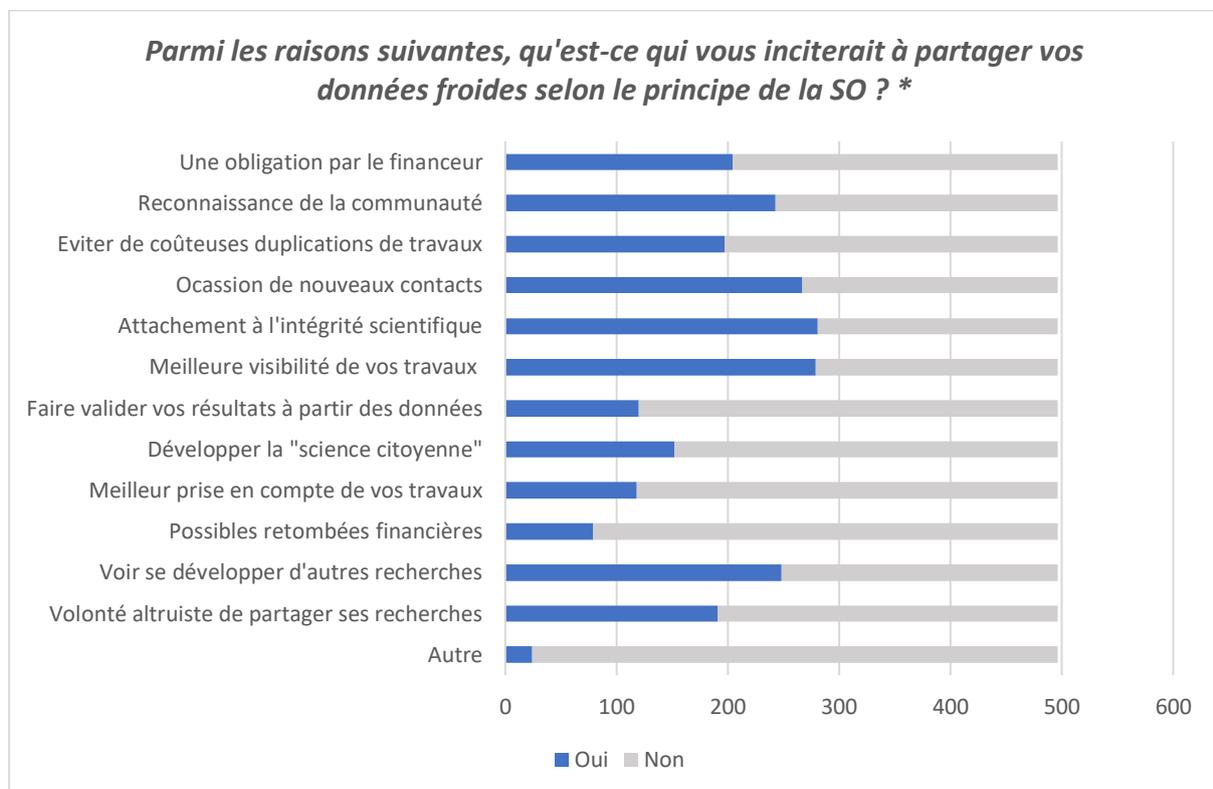
Lorsque l’on creuse un peu les résultats en fonction du profil des répondants, les chercheurs en STM semblent un peu plus favorables à la diffusion de leurs données que ceux en SHS<sup>17</sup>, en particulier après un délai d’embargo (40% des chercheurs en STM y sont favorables en ce cas, et 10% seulement s’opposent au partage, contre respectivement 25% et 15% en SHS). Considérant les statuts il semble que les Chercheurs et les EC comprennent davantage d’opposants à la diffusion, même si l’opposition reste peu importante dans l’ensemble au sein de ces catégories (5% de « non pas du tout » et 10% de « plutôt non »). De même avoir une « responsabilité » semble freiner le souhait de partager (plus de 15% d’opposants contre 8%

<sup>17</sup> Les femmes comptent aussi un pourcentage un peu moins élevé d’avis négatifs à la diffusion que les hommes (moins de 10% contre 15% pour ces derniers) mais ceci demande confirmation du fait du petit nombre d’opposants et de la probable interférence disciplinaire (plus de femmes en SHS).

d'opposants au partage pour les « sans responsabilité »), et la réticence au partage semble augmenter avec l'âge, à l'exception notable des deux dernières décennies (les plus de 60 ans). Ces derniers croisements laissent percevoir des différences inter-disciplinaires, des doutes plus importants pour ceux qui ont des responsabilités et une moindre concurrence en fin de carrière.

## Incidations des chercheurs à partager leurs données en science ouverte

Derrière les velléités de partage se trouve la question des motivations et incitations à partager (que les répondants aient déjà partagé leurs données ou non). Les réponses de la Figure 37 sont assez diffuses, mais parmi les motifs les plus cités (c'est-à-dire par plus de 200 répondants) on trouve : « l'attachement à l'intégrité scientifique », « la meilleure visibilité des travaux », « la reconnaissance scientifique », « l'occasion de nouer de nouveaux contacts », « la volonté de voir se développer d'autres recherches à partir du jeu de données », et puis « l'obligation par le financeur ». A l'inverse, les « possibles retombées financières », « la meilleure prise en compte des travaux pour l'évaluation », le fait de « faire valider les résultats » ou de « développer la science citoyenne » sont relativement peu mentionnés.



\* Réponses multiples possibles

Figure 37 : Incitations et motivations à partager les données froides selon le principe de la SO

On peut affiner ces résultats à l'aide de quelques croisements.

« L'attachement à l'intégrité scientifique » est le 1<sup>er</sup> motif et il bien partagée par tous, quel que soit le profil (aucune caractéristique personnelle ne semble discriminante). Il en va de même pour la « reconnaissance de la communauté » et la « visibilité des travaux », à ceci près que ce motif semble un peu plus important pour les personnes en SHS (environ 10 points d'écart avec les STM) et les doctorants (concernés par la visibilité pour plus de 65%). Ces deux derniers profils mentionnent aussi plus souvent « l'occasion de nouer de nouveaux contacts », dans des écarts et proportions similaires. L'aspect « nouveaux contacts » est nettement moins

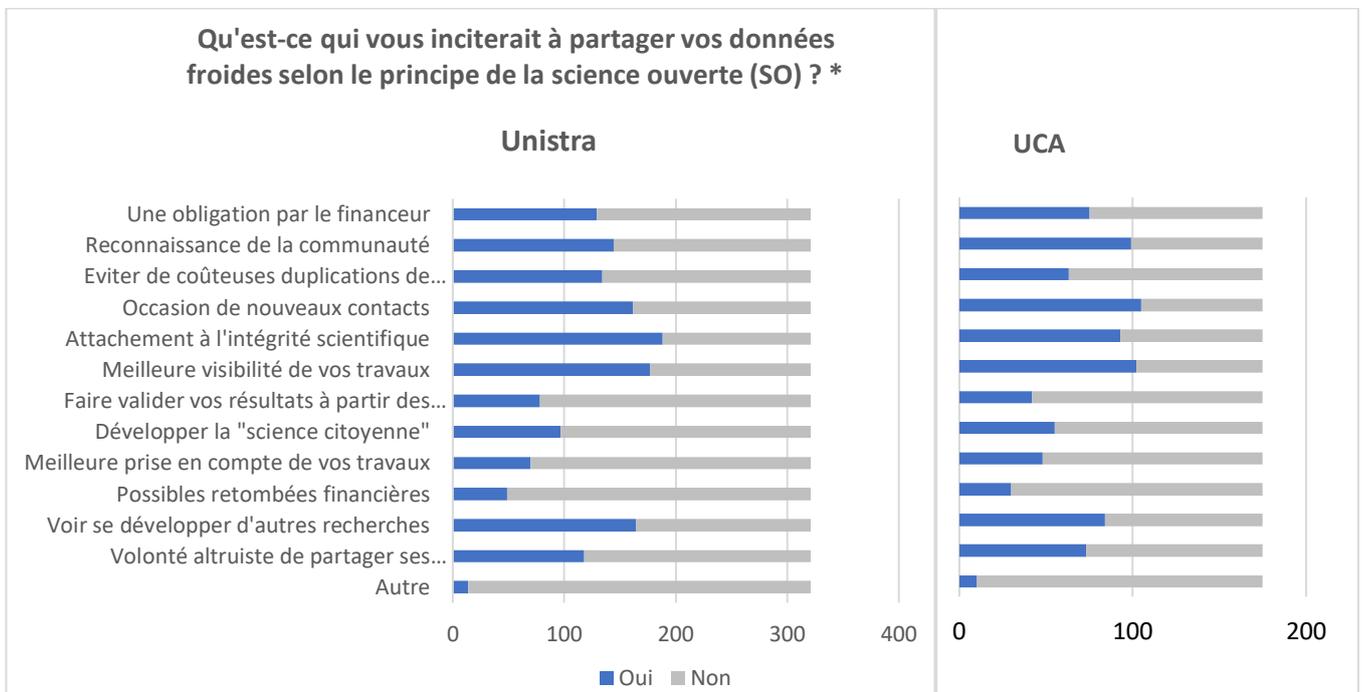
affirmé pour les ingénieurs. En outre ce motif lié au besoin de se créer un réseau diminue sans surprise avec l'âge et la prise de responsabilités.

Inversement, la sensibilité à une « obligation faite par le financeur » d'ouvrir ses données est nettement plus faible pour ceux qui relèvent des STM (20 points d'écart avec les SHS), ou qui ont un statut de permanent (c'est-à-dire des non doctorants), ainsi que pour les classes d'âge des plus âgés, ou pour ceux qui ont des responsabilités.

La sensibilité au non gaspillage et/ou aux économies de coût (cf. le motif « éviter de coûteuses duplications des travaux ») semble plus importante chez les ingénieurs et dans le domaine des STM. Elle l'est nettement moins pour les EC et les détenteurs de responsabilités.

Considérant enfin les motifs minoritaires, peu de caractères ressortent de manière significative, si ce n'est l'attention portée par les doctorants à la « volonté altruiste de partager ses données ».

Un dernier résultat, assez frappant, concerne la similitude des réponses dans les deux universités. Les profils de réponses sont en effet très semblables dans les schémas ci-dessous.



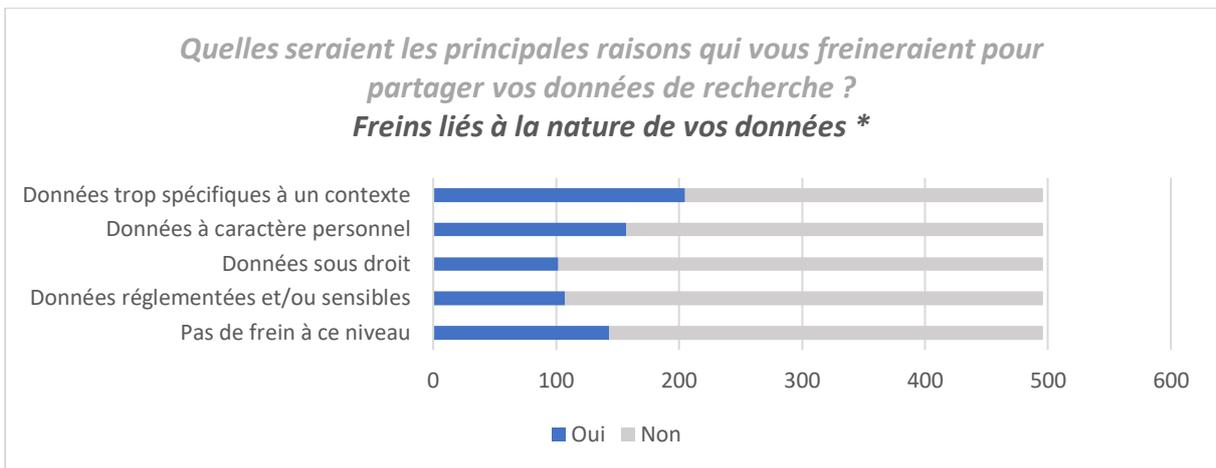
\* Réponses multiples possibles

Figure 38 : Incitations à partager les données froides selon les principes de la SO pour chaque université

### Freins au partage des données de recherche

Nous nous intéressons dans la suite aux freins susceptibles de bloquer le partage des données, d'après les répondants. Il s'agit à nouveau d'un ressenti, que les enquêtés aient déjà de l'expérience en la matière ou non. Nous distinguons plusieurs types de freins : ceux liés à la « nature même des données », ceux liés à des contraintes « économiques ou contractuelles », ceux relevant de « raisons d'ordre stratégique », et enfin les « raisons liées à un manque de ressources ».

Le 1<sup>er</sup> type de frein porte sur la nature même des données.

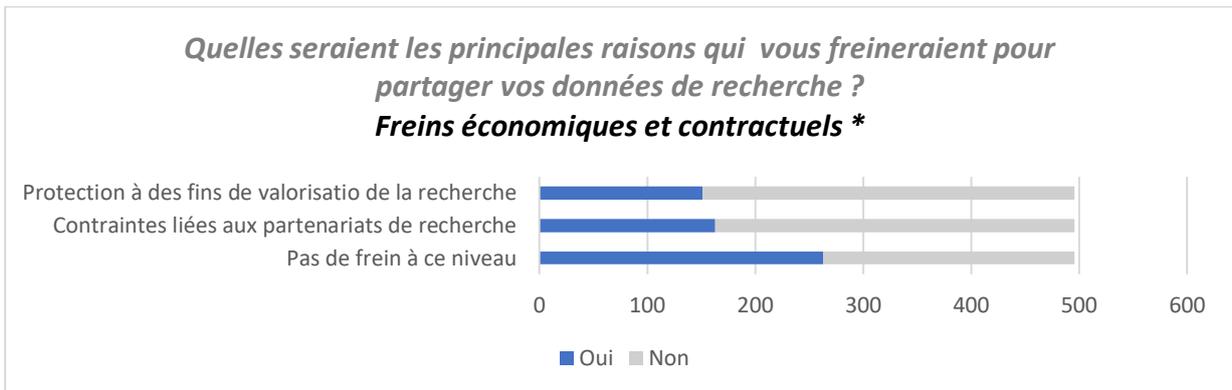


\* Réponses multiples possibles

Figure 39 : Freins au partage liés à la nature des données

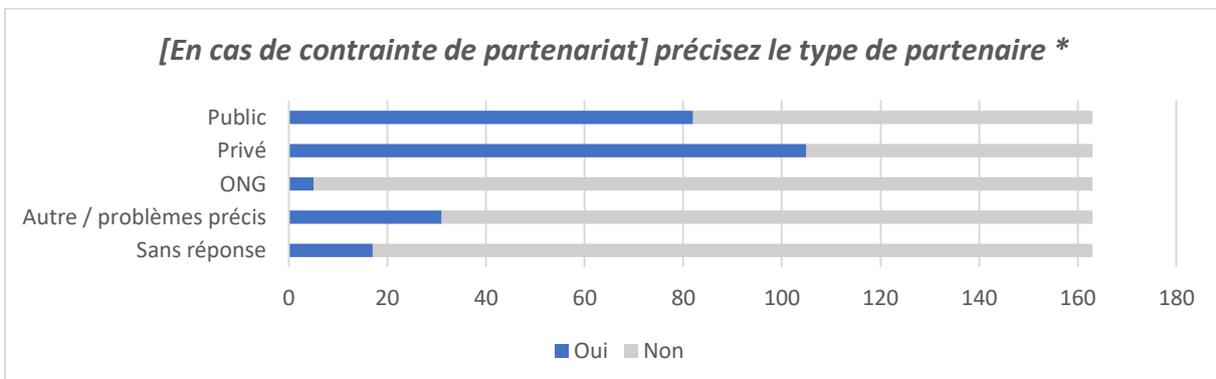
La raison citée le plus souvent à ce niveau est le fait que les données soient « trop spécifiques pour être réutilisées ». Il ne semble pas y avoir de différence marquée entre les SHS et les STM à ce niveau. En revanche les « données à caractère personnel » et les « données sensibles » concernent deux fois plus de répondants en SHS (40% d’entre eux) qu’en STM

Les « freins économiques et contractuels » sont ceux pour lesquels l’absence de freins est la plus prononcée globalement, mais ils restent importants et concernent plus de la moitié des répondants. La différence selon le champ disciplinaire est marquée : près de 60% des répondants en STM évoquent ce problème (contre seulement 35% des SHS).



\* Réponses multiples possibles

Figure 40 : Freins au partage d’ordre économique et contractuel

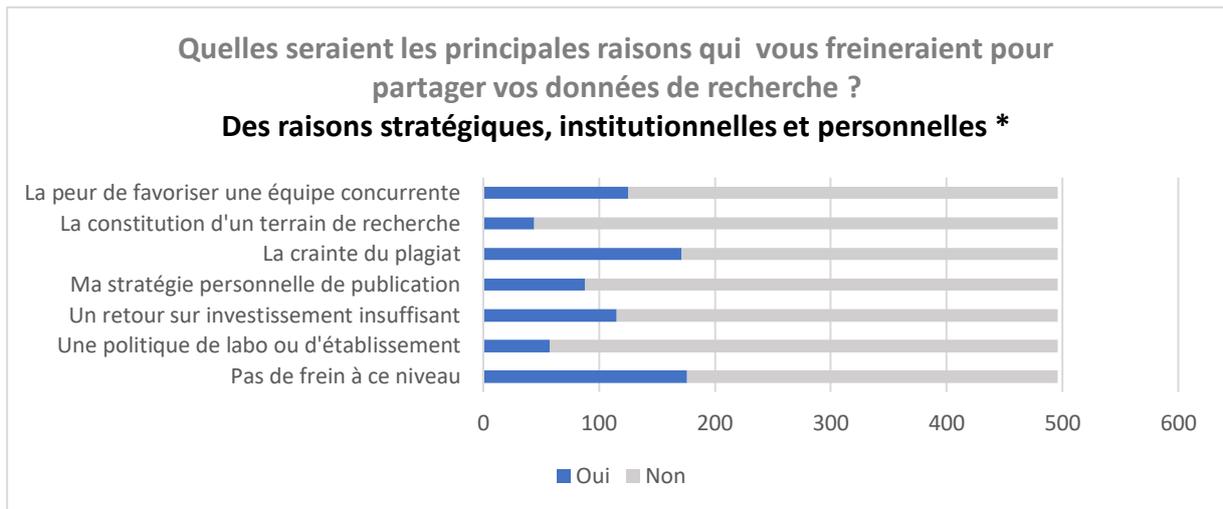


\* Réponses multiples possibles

Figure 41 : Nature du partenaire en cas de frein au partage des DR dû à un partenariat

En outre, on observe dans la Figure 41 que pour les 163 personnes concernées par des problèmes dans le cadre d'un partenariat de recherche, le type du partenaire, public versus privé, ne semble pas très discriminant.

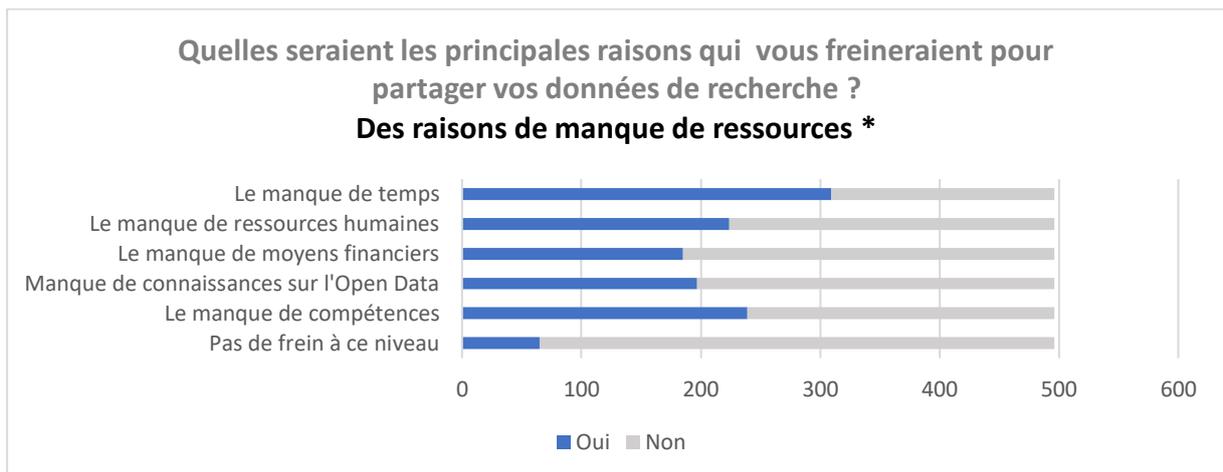
Le troisième type de freins portent sur les « stratégies personnelles ou institutionnelles ». Parmi elles, la « peur de favoriser une équipe concurrente » et la « crainte du plagiat » dominant. La première concerne principalement les STM (40% d'entre eux contre 12% des SHS). Le plagiat est mentionné de manière similaire par les deux champs disciplinaires (avec peut-être une petite prééminence pour les SHS). On observe aussi que la « crainte du plagiat » décroît clairement et régulièrement avec l'âge. De manière générale, ce type de frein concerne davantage les doctorants (70% d'entre eux) que les autres statuts, et parmi ceux-ci, encore moins les ingénieurs (45% environ). Il tend aussi à décroître avec l'âge.



\* Réponses multiples possibles

Figure 42 : Freins au partage d'ordre stratégique (aux niveaux institutionnel et personnel)

Le dernier type de freins est clairement le plus important. Il est partagé par 87 % des répondants, indépendamment de la discipline et du genre. Il s'agit du manque de ressources.



\* Réponses multiples possibles

Figure 43 : Freins au partage liés au manque de ressources

On peut toutefois souligner quelques nuances en entrant dans le détail des ressources en question.

Le fait d'avoir des responsabilités impacte certains types de frein. Ainsi 68% des responsables signalent un « manque de temps » (contre 52% des répondants sans responsabilité). Le statut

est discriminant aussi : 65% des Chercheurs et 68% des E-C sont concernés par le « manque de temps ».

De même les « responsables » pointent davantage le « manque de RH » que les autres répondants (55% contre 33%). C'est aussi le cas des statuts Chercheurs et surtout d'E-C (55% mentionnent ce problème de RH), alors que moins de 20% des doctorants semblent en souffrir.

La sensibilité au « manque de moyens financiers » et au « manque de compétences » est équitablement répartie sur l'ensemble de la population (pas de caractéristique personnelle discriminante).

Dans l'ensemble on peut retenir que les doctorants manquent un peu moins de ressources que les autres statuts. Cependant cela reste un frein au partage des données pour 78% d'entre eux.

### Connaissance des data papers

Les dernières questions de cette partie concernent la diffusion de données dans des « *data journals* » ou des publications similaires.

Seulement 13% des répondants connaissent les « *data journals* » (ou équivalents) et bien moins encore ont déjà publié des « *data papers* » ou les ont utilisés pour faire de la veille. Il s'agit donc d'une pratique encore très marginale en 2020.

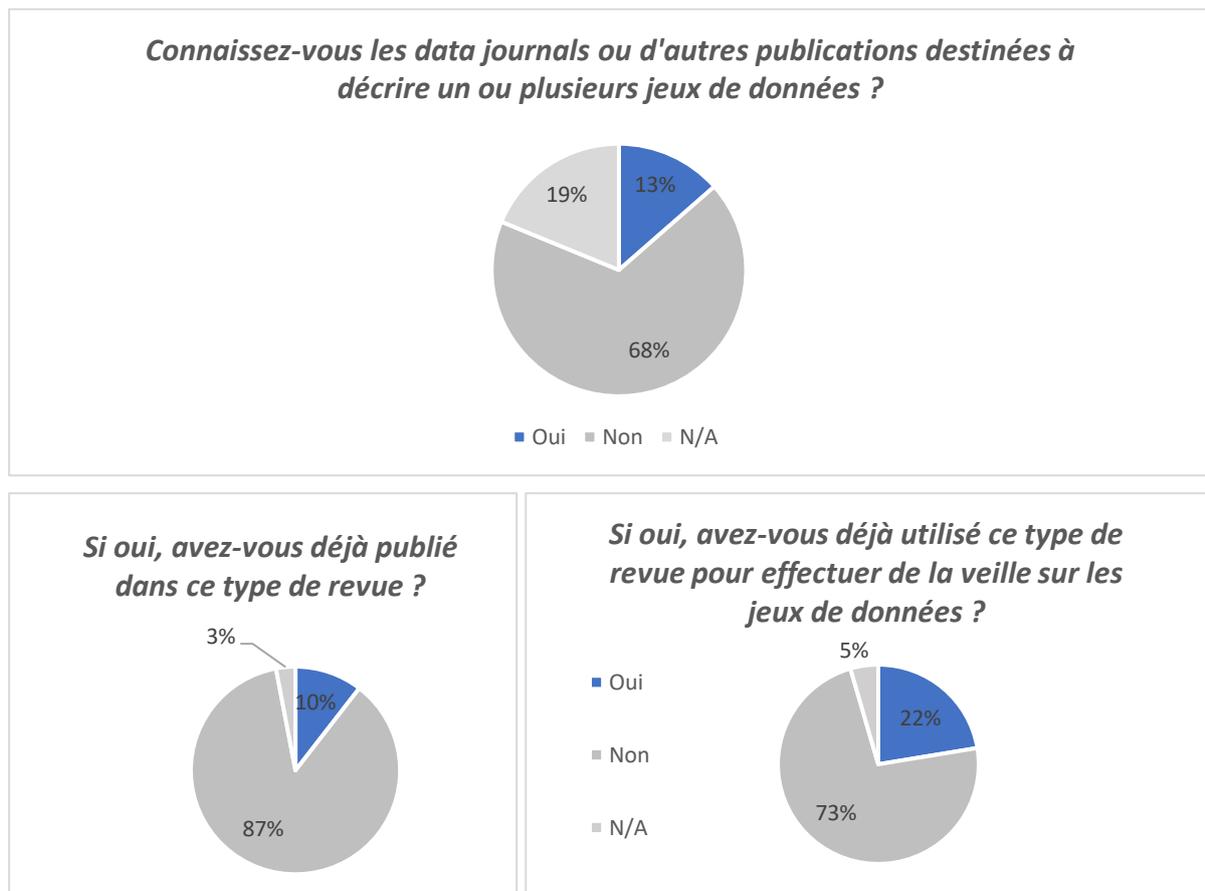
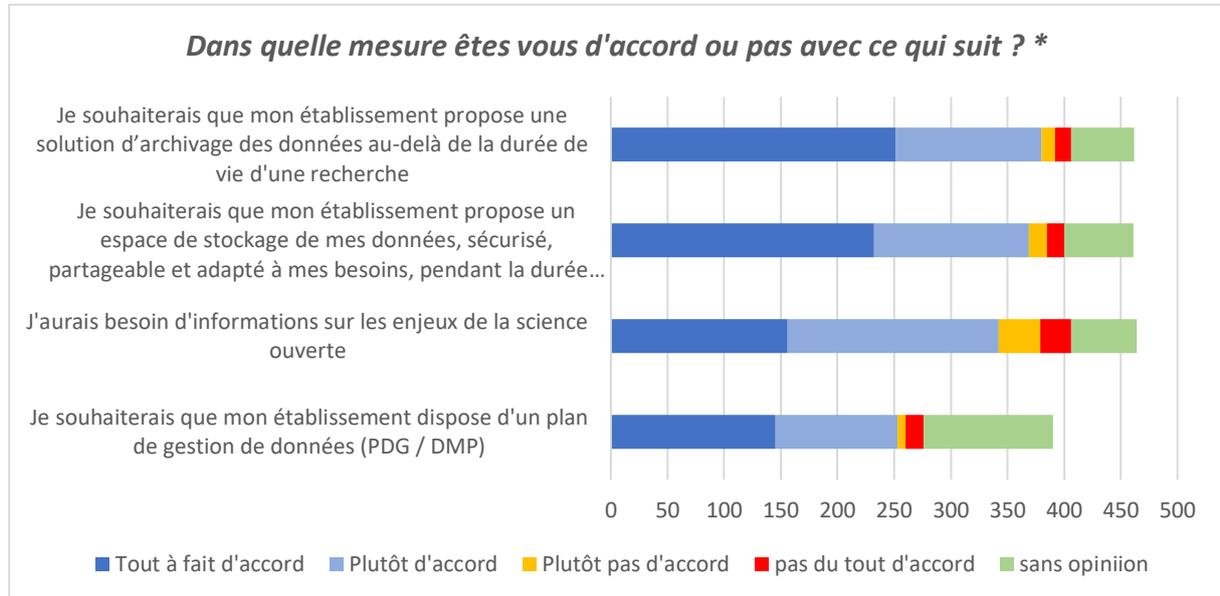


Figure 44 : Connaissance des data journals – théorique et pratique

## Partie 5 : Recensement des besoins pour les DR

La dernière partie du questionnaire vise à recueillir les attentes et besoins des personnels de recherche des universités en termes de ressources, de compétences ou connaissances, de formation, etc. Elle doit permettre d'élaborer un certain nombre de recommandations et d'alimenter utilement la politique de l'université en matière de données de recherche.

Les répondants sont invités à donner leur avis sur différentes propositions.



\* Réponses multiples possibles

Figure 45 : Degré d'adhésion à différentes propositions relevant de la politique d'établissement

Si toutes les propositions recueillent majoritairement un « accord », on peut noter le souhait affirmé que l'établissement dispose d'un espace de stockage (travail) et d'archivage des données et ce, quel que soit le domaine disciplinaire. On pourra penser à des carnets de recherche électronique (ELN)<sup>18</sup> ou des serveurs de stockage (« données chaudes ») et un entrepôt (« données froides ») avec toutefois la question de l'articulation du local et du global selon les disciplines, les types de données et *Recherche Data Gouv* l'entrepôt national pour les « DR de longue traîne ».

Plus des deux tiers sont demandeurs d'information « sur les enjeux de la science ouverte », c'est à dire au moins « plutôt d'accord » sur leur besoin en la matière, soulignant en cela l'intérêt pour la thématique et la volonté de comprendre les recommandations, les incitations, les obligations autour de l'*open access* (peut-être déjà mieux connu) et de l'*Open Research Data*, voire de distinguer les deux et comprendre leur articulation.

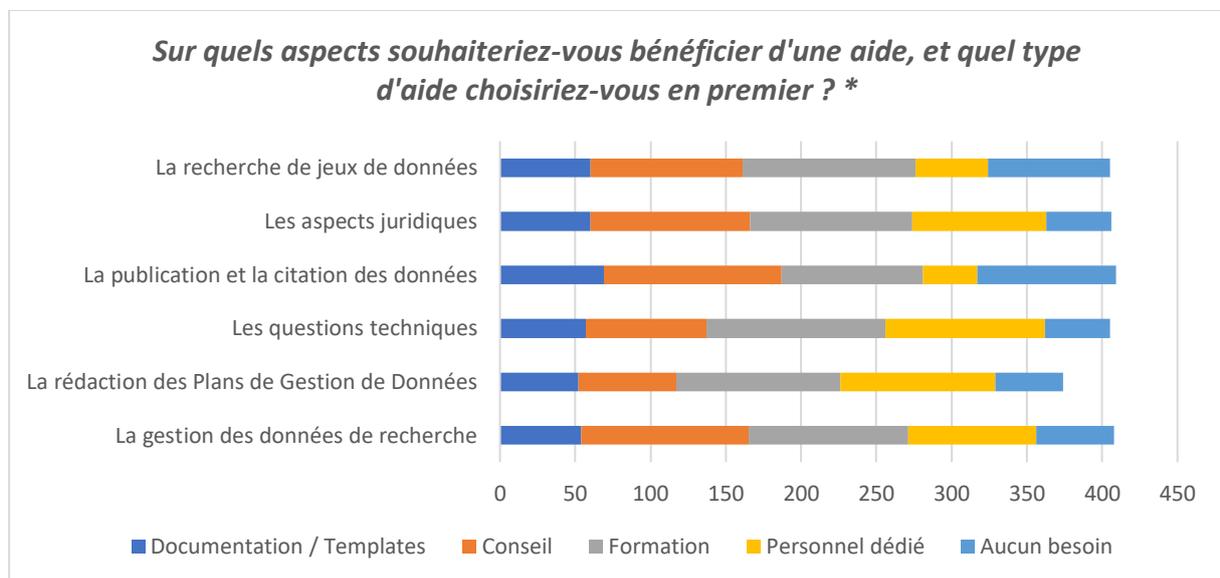
La question du PGD se pose peut-être de manière moins prégnante aujourd'hui, qu'au moment de la rédaction du questionnaire : l'Europe, l'ANR, certaines disciplines proposent des modèles de PGD, et des outils comme *DMP Opidor* facilitent ce travail. Malgré cela un accompagnement / une information pour la rédaction d'un PGD (pour une structure, une thèse, ...) peut s'avérer utile, en particulier si les DMP deviennent « *machine actionable* »<sup>19</sup>.

<sup>18</sup> Aspect corroboré par la réponse positive à près de 58% à la question « Souhaiteriez-vous des outils/espaces de travail collaboratifs pour vos projets de recherche ? »

<sup>19</sup> « maDMP » : rendre lisible et exploitable par les machines pour permettre l'exploitation optimale de son contenu par tous les acteurs du DMP (chercheurs, financeurs, fournisseur d'infrastructures...)

Nous avons procédé à des croisements systématiques avec les caractéristiques de profil des répondants. Les tendances qui ressortent semblent être :

- Les STM tendent à exprimer davantage de besoins que les SHS, à tous les niveaux présentés ci-dessus.
- Parmi les profils les plus demandeurs d'informations générales sur les enjeux de la SO, on trouve les personnes ne détenant pas de responsabilités, les femmes et les classes d'âge les plus jeunes et les plus âgées (c'est-à-dire hors quadra et quinquagénaire) ainsi que les statuts de Chercheurs et de doctorants.
- En matière d'espace de stockage des données chaudes et d'archivage des données froides, le statut d'E-C semble moins demandeur que les autres types de personnels. Les doctorants expriment davantage de besoins en matière d'archivage.
- Il n'y a guère de différences de profils en ce qui concerne les besoins de PGD.



\* Réponses multiples possibles

Figure 46 : Contenu et type des aides souhaités par les répondants à propos des DR

En matière d'accompagnement des chercheurs, bien que les réponses soient très réparties, elles indiquent finalement un besoin d'aide pour l'ensemble des propositions, le souhait du « conseil » et de la « formation » semble toutefois primer alors que la « documentation » emporte peu d'adhésion. Pour la « rédaction de PGD », les « questions techniques » et les « aspects juridiques », le besoin de « personnel dédié » est marqué, soulignant en cela la nécessité d'experts ou d'expertise dans ces domaines.

Les SHS ont tendance à préférer la « formation » quand les STM préfèrent les « conseils ». De même on notera un besoin de « personnel dédié » plus marqué pour les STM en matière de « gestion de DR » et de « rédaction de PGD ».

Assez logiquement pour un public de chercheur, les besoins sont moins importants (davantage de réponses cochées sur « aucun besoin ») à propos de la publication et la recherche de jeux de données.

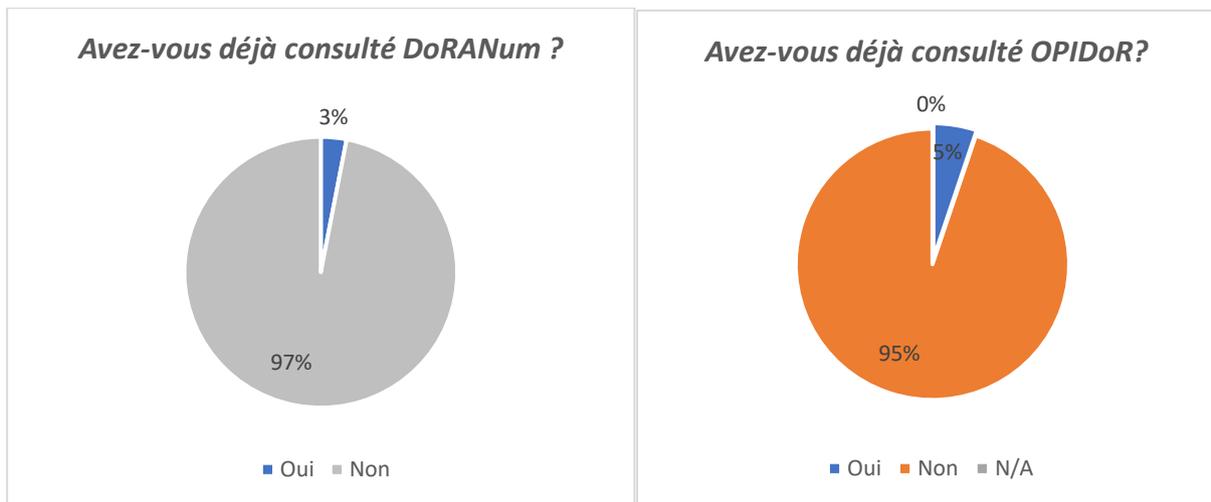


Figure 47 : Connaissance des plateformes d'aide à la gestion des DR DoRANum et OPIDoR

On notera que *Doranum*<sup>20</sup> et *Opidor*<sup>21</sup> ne sont pas (ou très peu) consultés (en 2020), ce qui serait à mettre en regard de la question sur les aides et propositions de « conseils » et de « documentation ». Ces sites d'auto-formation et de conseil sur les DR ne sont pas connus, ou pas suffisamment valorisés auprès des publics académiques, ou bien encore les contenus sont peu ciblés et ne correspondent pas aux attentes et besoins. Il convient de moduler ce propos au regard du temps écoulé entre la passation du questionnaire et la rédaction de ce document. Ces services et la communication sur ces services se sont depuis développés.

<sup>20</sup> <https://doranum.fr/>

<sup>21</sup> <https://opidor.fr/>

## Conclusion

Nous concluons ce rapport en revenant sur les traits les plus saillants de l'enquête, les points forts et les limites méthodologiques et enfin quelques remandations.

### Les traits les plus saillants de l'enquête

Les presque 496 réponses complètes obtenues auprès des « personnels de recherche » de l'Université de Strasbourg (321) et d'Université Côte d'Azur (175) montrent un taux de réponse assez faible (de l'ordre de 7%), une sous-représentation des doctorants et une sur-représentation des Enseignant-Chercheurs par rapport aux chiffres nationaux. La parité homme/femme (51/44) et des disciplines STM/SHS (50/50) est quasiment respectée.

*Des pratiques de production et de stockage différentes selon le champ disciplinaire.*

Les deux tiers des répondants produisent eux même leurs données de recherche, issues principalement de l'expérimentation (surtout pour les STM) et de l'observation (surtout pour les SHS). Elles prennent majoritairement la forme de données chiffrées (pour les 2/3 d'entre eux) et/ou de texte (pour la moitié d'entre eux), et dans une moindre mesure d'images (hors audiovisuel).

La population des répondants peut être séparée en deux parties quasi égales : celle qui a besoin de moins de 500 Go d'espace de stockage pour ses données chaudes et celle qui a besoin de davantage d'espace de stockage. Les différences disciplinaires semblent marquées à ce niveau, les SHS étant nettement moins gourmandes en forte capacité de stockage. Le stockage sur des supports individuels (ordinateur professionnel voire personnel, disque dur) reste dominant, même si la moitié des répondants stockent également leurs données sur un serveur du labo d'appartenance ou de l'établissement. Le lieu de stockage montre clairement des différences disciplinaires, les SHS ayant plus tendance à stocker leurs données sur leur ordinateur personnel et les STM ayant plus tendance à stocker leurs données sur un serveur interne.

*Une faible connaissance des données de recherche en science ouverte*

Un fait particulièrement important est le niveau apparent de méconnaissance (du moins à la date de passation du questionnaire, soit début 2020) de nombreux répondants concernant les politiques de données de recherche dans un contexte de science ouverte : ainsi, moins d'un quart savent ce qu'est un plan de gestion de données, la moitié ne savent pas s'il existe une politique d'archivage des données au niveau de l'établissement, et un tiers ignore s'il existe une quelconque politique d'archivage au niveau de leur unité de recherche.

*Des pratiques de partage encore peu développées, par manque de ressources ou simple ignorance*

En ce qui concerne les pratiques de partage des données au cœur du questionnaire, on observe que pour la grande majorité des personnes interrogées, l'accès à ses données se limite généralement à soi-même ou à son groupe de travail. Seul 13,5% des répondants donnent accès à leur données aux membres de leur laboratoire, et une part marginale (moins de 8%) à l'ensemble de sa discipline ou à tout le monde. La plupart des répondants n'ont jamais partagé de données (74%). Les raisons sont multiples, mais on retrouve au premier plan un manque de temps et de compétence, ainsi qu'une forme d'ignorance sur la possibilité

de partage. Dans l'ensemble, un quart des répondants donne accès librement à ses données froides sur le Web, et en ce cas cela se fait principalement via le site d'une revue, un entrepôt ou une archive ouverte. A contrario, plus de la moitié des personnes interrogées reconnaissent avoir déjà téléchargé au moins une fois les données d'autres chercheurs.

#### *Les intentions de partage des données : freins et motivations*

Si les pratiques de partage sont encore peu développées, les intentions des répondants y sont plutôt favorables : plus des trois quarts sont au moins « plutôt d'accord » avec l'idée de partager leurs données froides selon les principes de la science ouverte (avec ou sans délai d'embargo). Ils sont motivés en priorité par des questions déontologiques comme l'intégrité scientifique et le souhait de voir les données utilisées pour d'autres travaux, mais aussi par la visibilité et la reconnaissance des pairs, l'occasion d'étendre leur réseau ou l'obligation faite par un financeur.

Les principaux freins quant eux sont surtout liés à un manque de temps et de ressources (humaines et/ou compétences). Viennent ensuite des raisons liées à la nature des données et des raisons stratégiques, et en dernier lieu des raisons économiques liées à la valorisation et à la recherche contractuelle.

#### *Besoins et attentes*

Les besoins de stockage et d'archivage sont importants : plus des deux tiers des répondants souhaitent que leur établissement leur fournisse des solutions à ces deux niveaux. Conseil et formation sont attendus sur tous les aspects des données de recherche (recherche de jeux de données, gestion des données, ainsi que leur publication et citation, aspects juridiques et techniques, et dans une moindre mesure rédaction de PGD). Des besoins en personnel dédié sont également exprimés de manière forte, sauf pour la recherche et la publication/citation des données, lesquelles relèvent a priori davantage des compétences métiers attendues pour tout chercheur. Nul doute que la mise en place actuelle (2022) de l'écosystème Recherche Data Gouv réponde en partie aux besoins et attentes exprimés.

#### *Le poids du statut et de la discipline*

Une analyse économétrique plus poussée en cours de finalisation corrobore cette première salve de résultats qui montre l'influence de la discipline et du statut, parfois de la classe d'âge et enfin d'un niveau (élevé) de responsabilité. On notera en revanche l'influence quasi inexistante du genre dans les pratiques de données à de rares exceptions près (préférence féminine pour la formation, par exemple)

## Points forts et limites méthodologiques

Les principales limites méthodologiques sont liées à la nature même du questionnaire, en l'occurrence un questionnaire long et adressé à toute une population répondant sur la base du volontariat. On se doit de rappeler ici l'absence d'échantillonnage a priori, ce qui peut induire un biais de sélection (par exemple, lorsque les personnes les plus averties ou les plus concernées ont une propension plus élevée à répondre). Par ailleurs la longueur du questionnaire a induit un taux de réponse assez faible, notamment de la part des doctorants. Notons que le taux de participation était deux fois plus élevé que le taux de réponse, en d'autres termes le nombre de questionnaires non remplis suggère un certain découragement, ou du moins une lassitude de la part des répondants. Il convient aussi de noter une disparité des voies de communication du questionnaire sur les deux établissements cibles, qui explique partiellement un certain déséquilibre des populations interrogées.

En revanche le périmètre de l'établissement est très pertinent : administré avec le soutien de l'université le questionnaire a été sans doute plus mobilisateur qu'une enquête nationale, et sa réplique dans d'autres établissements ou structures (Synchrotron Soleil, INRAé, université de Nantes, etc.) peut permettre des comparaisons utiles et riches d'enseignements. Cependant, le canal de diffusion du questionnaire doit être maîtrisé et surveillé de près pour permettre de telles comparaisons.

Un autre point extrêmement sensible à noter compte tenu de l'objet même du questionnaire, est celui de la granularité des disciplines scientifiques. La liste des disciplines proposées était bien plus large que la simple distinction entre SHS et STM. Cependant elle a été conçue de manière ad hoc, elle n'était ni universelle ni exhaustive et elle permettait plusieurs réponses simultanées, ce qui a rendu quasi impossible son exploitation. Forts de cette expérience, nous envisageons et conseillons à l'avenir d'utiliser une typologie reconnue, par exemple celle de l'ERC / CAT-Opidor, qui présente en outre l'avantage d'être utilisée dans le cas des données de recherche. D'un autre côté, il est hautement vraisemblable que ce niveau de granularité disciplinaire ne soit pas pertinent. Le niveau le plus approprié semble être celui des communautés scientifiques, voire des communautés de pratiques.

## Recommandations

Les évolutions récentes du cadre réglementaire et institutionnel (La loi pour une République Numérique (2016), les Programmes Horizon 2020 et Europe, le Plan National pour la Science Ouverte (2018), ANR (2019)... ) imposent le principe d'ouverture des données de la recherche publique.

Plus récemment l'article L211-2<sup>22</sup> du code de la recherche dédié à l'intégrité scientifique précise : « *Sans préjudice des dispositions du code du patrimoine sur les archives publiques, ils [Les établissements publics contribuant au service public de la recherche...] **conservent les résultats bruts des travaux scientifiques réalisés en leur sein afin de permettre leur vérification** ».*

Ainsi, les incitations voire les obligations au partage et à l'archivage des données de recherche sont fortes. Les raisons scientifiques, sociétales et éthiques semblent évidentes. Pourtant notre enquête montre une forme de paradoxe : une certaine frilosité à diffuser avec cependant des velléités de partage des données de la recherche.

Rappelons ici que cette enquête agrège les réponses de deux universités, mais que les résultats pour chacune des universités sont similaires, marquant ainsi plus encore ce paradoxe puisqu'il n'est pas lié à des situations locales, mais plutôt globales, du moins des situations locales généralisables et peut-être généralisées à d'autres universités françaises.

Il nous semble important de distinguer plusieurs axes pour travailler sur ce paradoxe.

### Appréhender le cycle de vie de la recherche pour et par chacun.

La nature des données, les étapes et les modes d'obtention, de traitement, de sauvegarde, de diffusion des données, le temps de la recherche, ... pour ne citer que ces éléments, sont à contextualiser non seulement par domaine scientifique, mais également par projet de recherche. En effet, entre un projet financé par l'Europe ou des fonds privés, entre un projet réparti sur dix équipes dans le monde ou cinq personnes dans un laboratoire, entre un projet

---

<sup>22</sup> [https://www.legifrance.gouv.fr/codes/article\\_lc/LEGIARTI000042753467](https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000042753467)

qui demande un équipement lourd et couteux ou un projet frugal, entre un projet échelonné sur 1 année ou 3 ans, entre un projet qui analyse de la matière ou un projet qui analyse des comportements humains... les intrants et extrants du cycle de vie de la donnée sont très différents et renvoient à des appréhensions de la notion de « donnée de la recherche » différentes, qui sont en capacité de brouiller ce qui peut, ce qui doit être diffusé ou partagé.

Les « Données brutes », « données sous-jacentes », « données embarquées », « données collectées », « données sensibles », « données froides », « données chaudes »,... bref, les qualificatifs de « données » et ce à quoi ils renvoient sont autant de problèmes méthodologiques et terminologiques pour questionner les chercheurs de différentes disciplines que des interrogations sur la nature de ce qui doit ou peut être partagé. Semer à tout vent ou contraindre à archiver sans un minimum de *FAIRisation* ne sert pas la recherche. Comprendre les besoins et les pratiques autour d'un ensemble que l'on pense homogène (« les données des chercheurs ») semble être un préalable à toute recommandation.

**Une tâche supplémentaire à effectuer.** Le ratio personnel de soutien / chercheur n'a cessé de baisser dans les Universités depuis 2007 et dans les EPST depuis 2012<sup>23</sup>, ce qui explique le besoin exprimé par les répondants. Les besoins sont déjà criants pour gérer les charges dites « administratives » qui s'amoncellent depuis le développement *New Public Management* (reporting, évaluations, recherche de financement, ...), or la Science Ouverte réclame également du temps pour diffuser ses résultats (publications et données).

Si l'écosystème Recherche Data Gouv répond en partie à l'aide réclamée par les chercheurs, notamment en mutualisant des ressources, des outils, de l'accompagnement. Une des recommandations serait de doter les établissements de personnels de soutien à la science ouverte et ce non seulement pour répondre aux exigences légales (cf. décret du 3 dec 2021), mais également répondre à des critères d'excellence, de labellisation que les établissements cherchent à conserver ou obtenir. Sans cet effort, il faut soit admettre un temps plus long pour finaliser les projets de recherche, soit l'abandon ou la simplification des tâches « administratives », soit le non-respect de certaines obligations ou pour le moins revenir sur les principes de partage et de transparence, puis accepter la dispersion, la perte de données et le gaspillage financier associé.

**Le cadre de la Science Ouverte pas si bien connu que cela** car il évolue relativement vite (aspect éditoriaux, commerciaux avec les éditeurs, juridique avec les bailleurs de fonds, ...), avec en creux, des enjeux de concurrence, d'évaluation, de carrière qui pèsent fortement quant aux modes de diffusion de résultats de la recherche. Avec les actions initiées par le CoSO (guides, passeports, baromètre, groupes de réflexion, études, rencontres ...) un contexte et des éléments favorables permettent de disséminer aisément les enjeux de la science ouverte. Si un nombre d'actions sont déjà en cours au niveau des établissements, la prise en compte au sein de formations initiales ou continues de la Science Ouverte est à densifier.

Au terme de cette enquête, il est possible d'affiner et consolider quelques résultats par entretiens, de colliger les résultats avec d'autres établissements pour une étude contrastive, de proposer une enquête « flash » (Cf. Annexe 2) en capacité de cibler des indicateurs pour

---

<sup>23</sup> L'état de l'emploi scientifique en France [Rapport 2020]

mesurer le chemin parcouru en matière de « science ouverte » à intervalle régulier. Pour mieux appréhender l'évolution de la SO sur les établissements d'autres outils pourrait être mise en place : l'adaptation du baromètre de la Science Ouverte du CoSO (cf. Université de Lorraine et en cours dans nos universités) ou l'installation du « tableau de bord » SOvisu expérimenté à l'université de Toulon qui propose un baromètre individuel d'ouverture des publications et possiblement des données. Au-delà de ces indicateurs de progressions de la SO il s'agit surtout de faire progresser la notion de partage des résultats de la recherche.

## Annexes

- Annexe 1 : questionnaire de l'enquête amélioré
- Annexe 2 : questionnaire Flash
- Annexe 3 : qualification des données de la recherche (Francis André, 2015)

## Annexe 1 - Le questionnaire de l'enquête « amélioré »

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

# Enquete sur vos pratiques en matière de "données de la recherche" améliorer

Dans le contexte actuel d'ouverture de la science et des données de la recherche (DR), cette enquête vise à mieux connaître les pratiques, perceptions et incitations des chercheurs à partager leurs données. Financée par le GIS Réseau Urfist, elle est réalisée avant tout à des fins de recherche, mais elle contribuera également à éclairer les décisions et les politiques de science et d'innovation. Se fondant sur l'enquête menée à l'université Rennes 2 \*\*, elle est déployée localement et dans d'autres établissements.

Le questionnaire est transdisciplinaire. Il concerne tous les chercheurs et tous les types de données (textes, chiffres, images,...).

Le remplir prend une vingtaine de minutes et vous pouvez interrompre et reprendre le processus à tout moment.

Il restera accessible jusqu'au

Les questions qui sont précédées d'une astérisque \* doivent être renseignées. Seul des résultats agrégés seront présentés.

L'enquête sera suivie d'une phase complémentaire basée sur quelques entretiens individuels. À la fin du questionnaire, une adresse e-mail de contact est demandée à ceux qui accepteraient d'être interviewés.

Quels que soient les travaux produits à partir de cette recherche, les données seront anonymisées.

*La participation à cette recherche est volontaire et il est possible de se retirer à tout moment. Conformément à la législation relative à la protection des données, les informations recueillies (données sur le statut professionnel, le genre, le domaine de recherche et éventuellement adresse de contact) le seront uniquement pour les besoins de la recherche présentée ci-dessus.*

\*\* Serres A. et alii (2017) : "Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2." Rapport, novembre 2017. Disponible sur HAL : <https://hal.archives-ouvertes.fr/hal-01635186/> (<https://hal.archives-ouvertes.fr/hal-01635186/>)

Il y a 67 questions dans ce questionnaire.

## 1ère partie : Informations personnelles

<https://sondagesv3.unistra.fr/index.php/admin/printablesurvey/sa/index/surveyid/567925/lang/fr>

1/30

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Votre établissement de rattachement \*

Veillez écrire votre réponse ici :

## Votre unité de recherche de rattachement ?

Veillez écrire votre réponse ici :

## Votre statut \*

Veillez sélectionner une seule des propositions suivantes :

- Chercheur
- Enseignant-chercheur
- Doctorant
- Ingénieur
- Post-doctorant
- Autre

## Votre genre \*

Veillez sélectionner une seule des propositions suivantes :

- Homme
- Femme
- Non binaire
- Ne souhaite pas répondre

## Votre année de naissance \*

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- 2000 et après
- 1999
- 1998
- 1997
- 1996
- 1995
- 1994
- 1993
- 1992
- 1991
- 1990
- 1989
- 1988
- 1987
- 1986
- 1985
- 1984
- 1983
- 1982
- 1981
- 1980
- 1979
- 1978
- 1977
- 1976
- 1975
- 1974
- 1973
- 1972
- 1971
- 1970
- 1969
- 1968

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

- 1967
- 1966
- 1965
- 1964
- 1963
- 1962
- 1961
- 1960
- 1959
- 1958
- 1957
- 1956
- 1955
- 1954
- 1953
- 1952
- 1951
- 1950
- 1949
- 1948
- 1947
- 1946
- 1945
- 1944
- 1943
- 1942
- 1941
- 1940
- 1939
- 1938
- 1937
- 1936
- 1935
- 1934
- 1933
- 1932
- 1931
- 1930 et avant

<https://sondagesv3.unistra.fr/index.php/admin/printablesurvey/sa/index/surveyid/567925/lang/fr>

4/30

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Exercez-vous actuellement des responsabilités ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

## Lesquelles ?

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

La réponse était 'Oui' à la question '6 [DR1x6]' ( Exercez-vous actuellement des responsabilités ? )

**i** Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

- Responsabilité de filière, de diplôme...
- Direction de département, d'UFR
- Direction ou co-direction d'unité de recherche, ou de composante d'unité de recherche
- Responsable d'équipe, de groupe ou projet de recherche (ANR, H2020...)
- Encadrement de thèses, de post-docs...
- Directeur d'école doctorale
- Responsabilités éditoriales (direction de revue, membre de comité de rédaction...)
- Responsabilités politiques (VP, chargé(e) de mission, élu(e) dans une instance de l'université...)
- Autre:

## Quel est votre domaine disciplinaire ? \*

**i** Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

- Sciences Humaines et Sociales
- Sciences et Technologies
- Vie et Santé

Cette typologie a été proposée par l'ERC (the European Research Council) et est notamment utilisée par Cat OPIDoR. Pour plus de détails sur les disciplines :

Détails typologie

(<https://erc.europa.eu/sites/default/files/document/file/erc%20peer%20review%20evaluation%20panels.pdf>)

## Quel est votre champ disciplinaire principal ? (Sciences Humaines et Sociales)

\*

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

La réponse était 'Sciences Humaines et Sociales' à la question '8 [DR1x70]' (Quel est votre domaine disciplinaire ? )

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Individus, marchés et organisations (économie, finance et management)
- Institutions, valeurs, environnement et espace (science politique, droit, science de la durabilité, géographie, étude et aménagement du territoire)
- Le monde social, diversité, population (sociologie, psychologie sociale, anthropologie sociale, démographie, éducation, communication)
- L'esprit humain et sa complexité (sciences cognitives, psychologie, linguistique, philosophie de l'esprit)
- Cultures et production culturelle (littérature, philologie, études culturelles, étude des arts, philosophie)
- L'étude du passé humain (archéologie et histoire)

## Quel est votre champ disciplinaire principal ? (Sciences et Technologies) \*

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

La réponse était 'Sciences et Technologies' à la question '8 [DR1x70]' (Quel est votre domaine disciplinaire ? )

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Mathématiques (tous les domaines des mathématiques, pures et appliquées, plus les fondements mathématiques des sciences informatiques, la physique mathématique et les statistiques)
- Constituants fondamentaux de la matière (physique des particules, nucléaire, des plasmas, atomique, moléculaire, des gaz et optique)
- Physique de la matière condensée (structure, propriétés électroniques, fluides, nanosciences, physique biologique)
- Chimie physique et analytique (chimie analytique, théorie chimique, chimie physique/physico-chimie)
- Chimie de synthèse et matériaux (synthèse des matériaux, relations structure-propriétés, matériaux fonctionnels et avancés, architecture moléculaire, chimie organique)
- Sciences informatiques et informatique (systèmes informatiques et d'information, sciences informatiques, calcul scientifique, systèmes intelligents)
- Ingénierie des systèmes et de la communication (ingénierie électrique, électronique, de la communication, optique et des systèmes)
- Ingénierie des produits et des procédés (conception de produits, conception et contrôle des procédés, méthodes de construction, génie civil, systèmes énergétiques, ingénierie des matériaux)
- Sciences de l'Univers (astro-physique/chimie/biologie; système solaire; astronomie stellaire, galactique et extragalactique, systèmes planétaires, cosmologie, science de l'espace, instrumentation)
- Sciences du Système Terre (géographie physique, géologie, géophysique, sciences de l'atmosphère, océanographie, climatologie, cryologie, écologie, changements environnementaux globaux, cycles biogéochimiques, gestion des ressources naturelles)

## Quel est votre champ disciplinaire principal ? (Vie et Santé) \*

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

La réponse était 'Vie et Santé' à la question '8 [DR1x70]' (Quel est votre domaine disciplinaire ? )

📌 Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

- Biologie moléculaire, biochimie, biologie structurale et biophysique moléculaire (synthèse, modification, mécanismes et interaction moléculaires, biochimie, biologie structurale, biophysique moléculaire transduction de signaux)
- Génétique, "omiques", bioinformatique et biologie des systèmes (génétique moléculaire, génétique quantitative, épidémiologie génétique, épigénétique, génomique, métagénomique, transcriptomique, protéomique, métabolomique, glycomique, bioinformatique, biologie computationnelle, biostatistique, biologie des systèmes)
- Biologie cellulaire et du développement (biologie cellulaire, physiologie cellulaire, transduction du signal, organogenèse, génétique du développement, formation de motifs et biologie des cellules souches, chez les végétaux et les animaux ou les microorganismes)
- Physiologie, physiopathologie et endocrinologie (physiologie des organes, physiopathologie, endocrinologie, métabolisme, vieillissement, cancérogenèse, maladies cardio-vasculaires, syndromes métaboliques)
- Neurosciences et troubles neurologiques (signalisation et fonction neuronale, neuroscience des systèmes, bases neurales des processus cognitifs et comportementaux, troubles neurologiques et psychiatriques)
- Immunité et infection (le système immunitaire et les troubles associés, biologie des agents infectieux et de l'infection, bases biologiques de la prévention et du traitement des maladies infectieuses)
- Technologies médicales appliquées, diagnostics, thérapies et santé publique (développement d'outils de diagnostic, surveillance et traitement des maladies, pharmacologie, médecine clinique, médecine régénérative, épidémiologie et santé publique)
- Ecologie, évolution et biologie environnementale (démécologie, synécologie, écologie des écosystèmes, biologie de l'évolution, écologie comportementale, écologie microbienne)
- Sciences de la vie appliquées, biotechnologie, et ingénierie moléculaire et des biosystèmes (sciences appliquées des végétaux et des animaux, sylviculture, sciences des aliments, biotechnologie appliquée, biotechnologie environnementale et marine, bioingénierie appliquée, biomasse et biocarburants, dangers biologiques)

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

**Vous pouvez préciser ici votre domaine, sous-domaine disciplinaire ou champ de recherche.**

Veillez écrire votre réponse ici :

## 2ème partie : Quels types de données de recherche ?

Selon les disciplines et domaines de recherche, le terme « données de recherche » donne lieu à des acceptions différentes. Cette partie vise à connaître « votre » acception des données, qu'elles soient « brutes », « produites », « collectées », « analysées », « publiées ».

Nous distinguons les données qui nourrissent un travail de recherche des données qui résultent de ce travail.

Intéressons-nous d'abord aux **données qui nourrissent vos travaux de recherche.**

**Comment nommez-vous habituellement les “éléments” qui vont nourrir vos recherches ?**

Veillez choisir toutes les réponses qui conviennent :

- Données brutes
- Matériaux de recherche
- Corpus
- Observables

Autre:



20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## En règle générale comment obtenez-vous les données qui nourrissent vos recherches ? \*

Veillez choisir toutes les réponses qui conviennent :

- L'observation (remontées de terrain : données de capteurs, images satellites, enquêtes, entretiens...)
- L'expérimentation (données obtenues en laboratoire : séquençage ADN, chromatographie, eye-tracking, données comportementales...)
- Simulations et modélisations (générées par des modèles informatiques...)
- Données dérivées (issues du traitement et de la compilation de données brutes : bases de données, text/data mining...)
- Données de référence (préalablement organisées, gérées, voire publiées : banques de données de séquences ADN, de structures chimiques, portails de données spatiales, données INSEE...)
- Autre:

## De quel(s) type(s) sont ces données ? \*

Veillez choisir toutes les réponses qui conviennent :

- Données textuelles (issues de corpus de textes, réseaux sociaux, archives...)
- Données audiovisuelles (enregistrements audio, vidéo, films...)
- Données d'images (photographies, dessins... d'objets, de paysages, d'architectures, spectroscopies, 3D...)
- Données chiffrées (tableaux de données statistiques, logs de serveurs...)
- Autre:

## En règle générale, est-ce que vous produisez vous-même les données qui nourrissent vos recherches (constitution d'un jeu de données original) ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Comment ?

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

La réponse était 'Oui' à la question '17 [DR2x4]' (En règle générale, est-ce que vous produisez vous-même les données qui nourrissent vos recherches (constitution d'un jeu de données original) ? )

Veillez écrire votre réponse ici :

Nous nous intéressons maintenant aux **données qui résultent de vos recherches.**

## Quels sont les principaux types de données qui résultent de vos recherches ?

Veillez écrire votre réponse ici :

## Comment les classeriez-vous selon la typologie ci-dessous ? \*

Veillez choisir toutes les réponses qui conviennent :

- Données textuelles (issues de corpus de textes, d'archives...)
- Données audiovisuelles (enregistrements audio, séquences filmiques...)
- Données d'images (photographies, dessins... d'objets, de paysages, d'architectures, spectroscopies, 3D...)
- Données chiffrées (tableaux de données statistiques, données web...)
- Autre:

<https://sondagesv3.unistra.fr/index.php/admin/printablesurvey/sa/index/surveyid/567925/lang/fr>

11/30

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## 3ème partie : Pratiques de gestion de l'ensemble de vos données (stockage, archivage, descriptio

Note explicative :

- **données chaudes** : données liées au projet de recherche, dont le contenu peut évoluer au cours de la recherche, données faisant l'objet de **stockage** ;
- **données froides** : données produites à l'issue du projet de recherche, dont le contenu est stabilisé, données pouvant faire l'objet d'un **archivage**.

### À propos de vos pratiques de stockage (données chaudes)

**Rappel** : données chaudes : données liées au projet de recherche, dont le contenu peut évoluer au cours de la recherche.

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

\*

Veuillez sélectionner une seule des propositions suivantes :

- Moins de 50 Go
- Entre 50 et 500 Go
- Entre 500 Go et 1 To (1 Tera-octet = 1000 Go)
- Entre 1 To et 1 Po (1 Peta-octet = 1000 To)
- Plus d'un Po
- Ne peut être estimé

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Où stockez-vous vos données de recherche en cours de traitement (données "chaudes") ? \*

Veillez choisir toutes les réponses qui conviennent :

- Sur votre ordinateur personnel
- Sur votre ordinateur professionnel
- Sur un serveur de votre unité de recherche ou de votre établissement (université, EPST,...)
- Sur un serveur externe d'un autre établissement (université, EPST,...)
- Sur une plate-forme privée dans le "cloud" (Google Drive, Dropbox, etc.)
- Sur le serveur d'une infrastructure de recherche (grid, TGIR, plateforme d'instrumentation ou de données...)
- Autre:

## Veillez préciser quelle infrastructure de recherche

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

La réponse était à la question '24 [DR3Ax2]' ( Où stockez-vous vos données de recherche en cours de traitement (données "chaudes") ? )

Veillez écrire votre réponse ici :

## Lors du travail sur vos données, faites-vous (vous ou votre équipe) régulièrement des copies de sauvegarde de vos données "chaudes" \*

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Ces copies de sauvegardes sont-elles automatisées ?

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

La réponse était 'Oui' à la question '26 [DR3Ax3]' (Lors du travail sur vos données, faites-vous (vous ou votre équipe) régulièrement des copies de sauvegarde de vos données "chaudes" )

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

## À propos de vos pratiques d'archivage (données froides)

**Rappel** : Données froides : donnée produites à l'issue du projet de recherche, dont le contenu est stabilisé.

## Existe-t-il une politique ou une pratique d'archivage des données "froides" de recherche (critères, modalités, nommage des fichiers...) :

### 1) Dans votre établissement ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non  
 Je ne sais pas

### 2) Dans votre unité de recherche ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non  
 Je ne sais pas

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

### 3) À votre niveau individuel ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Selon les projets
- Non

Concernant les données produites par des chercheurs non-permanents de votre unité de recherche (doctorants, chercheurs associés, partenaires...), est-ce que votre unité de recherche assure la conservation et la pérennisation de ces données ?

\*

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas

### À propos de vos pratiques de description

Savez-vous ce qu'est un Plan de Gestion de Données (PGD), ou Data Management Plan (DMP) ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Avez-vous déjà participé à l'élaboration d'un Plan de Gestion de Données ? \*

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

La réponse était 'Oui' à la question '34 [DR3Cx1]' ( Savez-vous ce qu'est un Plan de Gestion de Données (PGD), ou Data Management Plan (DMP) ? )

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

## Ajoutez-vous habituellement des informations (métadonnées) décrivant vos jeux de données chaudes ou froides ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

## Précisez le type d'informations \*

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

La réponse était 'Oui' à la question '36 [DR3Cx2]' (Ajoutez-vous habituellement des informations (métadonnées) décrivant vos jeux de données chaudes ou froides ? )

Veillez choisir toutes les réponses qui conviennent :

- Des informations administratives (auteur, date de création, etc.)  
 Des informations techniques (format des fichiers, taille, logiciel, etc.)  
 Des informations scientifiques (origine des données, projet de recherche, mots-clés, etc.)  
 Autre:

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Comment procédez-vous ? \*

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

La réponse était 'Oui' à la question '36 [DR3Cx2]' (Ajoutez-vous habituellement des informations (métadonnées) décrivant vos jeux de données chaudes ou froides ? )

Choisissez la réponse appropriée pour chaque élément :

	Oui	Selon les projets	Non
<b>Vous élaborez une documentation qui vous est propre</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Vous utilisez un cahier de laboratoire (ou de recherche)</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Vous vous référez à un Plan de Gestion de Données (PGD / DMP : Data Management Plan)</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Vous utilisez un standard de métadonnées (Dublin Core, OAI-PMH...)</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## 4ème partie : Pratiques de partage des données

Cette Partie ne concerne que les données froides.

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## À votre connaissance, qui peut accéder librement à vos données "froides" ?

\*

Veuillez choisir toutes les réponses qui conviennent :

- Personne
- Mon groupe de travail
- Mon unité de recherche
- Mon institut
- L'ensemble de l'établissement
- Les chercheurs de la discipline
- Tout le monde
- Je ne sais pas

 Autre public spécifique ::

Si vous ne pouvez pas répondre en général, prenez l'exemple d'un projet de recherche représentatif de vos pratiques

## Avez-vous déjà téléchargé des données d'autres **chercheurs** disponibles sur le web ?

\*

Veuillez sélectionner une seule des propositions suivantes :

- Oui, c'est indispensable pour mes recherches
- Oui, régulièrement
- Oui, au moins une fois
- Non, pas pour le moment
- Non, je ne le ferai pas
- Non, je ne savais pas que c'était possible

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Ces données imposent-elles un travail de curation (nettoyage, sélection, correction, ....) \*

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

La réponse était 'Oui, au moins une fois' ou 'Oui, régulièrement' ou 'Oui, c'est indispensable pour mes recherches' à la question '40 [DR4x2]' (Avez-vous déjà téléchargé des données d'autres chercheurs disponibles sur le web ? )

Veillez sélectionner une seule des propositions suivantes :

- Toujours
- Souvent
- Rarement
- Jamais

## De manière générale, pensez-vous que les données diffusées par vos pairs sont de :

Veillez sélectionner une seule des propositions suivantes :

- Bonne qualité
- Moyenne qualité
- Faible qualité

## Vous-même, avez-vous déjà diffusé vos données "froides" sur le web ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Combien de fois ?

\*

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

La réponse était 'Oui' à la question '43 [DR4x5]' ( Vous-même, avez-vous déjà diffusé vos données "froides" sur le web ? )

Veillez sélectionner une seule des propositions suivantes :

- Au moins une fois
- Régulièrement

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

### 1) Sur les plateformes "Science Ouverte" : \*

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

La réponse était 'Oui' à la question '43 [DR4x5]' ( Vous-même, avez-vous déjà diffusé vos données "froides" sur le web ? )

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)
- Vous n'utilisez pas les plateformes "sciences ouvertes"

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## 2) Sur d'autres plateformes :

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

La réponse était 'Oui' à la question '43 [DR4x5]' ( Vous-même, avez-vous déjà diffusé vos données "froides" sur le web ? )

Veillez choisir toutes les réponses qui conviennent :

- 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)
- Vous n'utilisez que les plateformes de "Science Ouverte"

Autre:

## Pour quelle(s) raison(s) ?

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

La réponse était 'Non' à la question '43 [DR4x5]' ( Vous-même, avez-vous déjà diffusé vos données "froides" sur le web ? )

Veillez choisir toutes les réponses qui conviennent :

- Je ne savais pas que c'était possible
- Je n'ai pas le droit
- Je n'y avais pas pensé
- Je ne le ferai pas
- Je ne sais pas le faire
- Je n'ai pas le temps
- Je pense que cette tâche nécessite le soutien d'un personnel qualifié

Autre raison:

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

**De manière générale, êtes-vous d'accord pour partager vos données "froides" selon les principes de la science ouverte (diffusables, et réutilisables librement et gratuitement) ? \***

Veillez sélectionner une seule des propositions suivantes :

- Oui, sans délai
- Oui, après un délai
- Plutôt oui
- Plutôt non
- Non, pas du tout
- Je ne sais pas

**Pensez-vous que vos données pourront être facilement réutilisables ?**

\*

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

La réponse était 'Plutôt oui' ou 'Oui, après un délai' ou 'Oui, sans délai' à la question '48 [DR4x6]' ( De manière générale, êtes-vous d'accord pour partager vos données "froides" selon les principes de la science ouverte (diffusables, et réutilisables librement et gratuitement) ? )

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

**Parmi les raisons suivantes, qu'est-ce qui vous inciterait à partager vos données froides selon les principes de la Science Ouverte (*diffusables, et réutilisables librement et gratuitement*) ? \***

Veillez choisir toutes les réponses qui conviennent :

- Une obligation faite par le financeur de votre recherche (dans le cadre des projets européens, ANR...)
- La reconnaissance de la communauté scientifique
- Le souhait d'éviter de coûteuses duplications de travaux de recherche
- L'occasion de nouveaux contacts, de nouvelles coopérations scientifiques
- L'attachement à l'intégrité scientifique (transparence...)
- Une meilleure visibilité de vos travaux de recherche
- Le souci de pouvoir faire valider vos résultats à partir des jeux de données
- La volonté de développer la "science citoyenne" (science participative)
- Une meilleure prise en compte de vos travaux de recherche dans l'évaluation
- De possibles retombées financières (crédits de recherche...)
- Le souhait de voir se développer d'autres recherches à partir de vos jeux de données
- La volonté altruiste de partager ses recherches
- Une obligation de votre unité ou de votre établissement
- Autre:

**Quelles seraient les principales raisons qui vous freineraient pour partager vos données de recherche ?**

**1) Des freins liés à la nature de vos données : \***

Veillez choisir toutes les réponses qui conviennent :

- Données trop spécifiques à un contexte pour être réutilisées
- Données à caractère personnel (permettant d'identifier un individu)
- Données sous droit (text mining de bases commerciales,...)
- Données réglementées et/ou sensibles (liées à la Défense, à l'environnement, etc.)
- Pas de frein à ce niveau

## 2) Des raisons économiques et contractuelles : \*

Veillez choisir toutes les réponses qui conviennent :

- Protection à des fins de valorisation de la recherche (brevet , autres propriétés industrielles, ...)
- Contraintes liées aux partenariats de recherche
- Pas de frein à ce niveau

## Précisez le type de partenaire : \*

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

La réponse était à la question '52 [DR4x8B]' ( 2) Des raisons économiques et contractuelles : )

Veillez choisir toutes les réponses qui conviennent :

- Public
- Privé
- ONG
- Sans réponse

## Pouvez-vous préciser le type de problème ?

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

La réponse était à la question '52 [DR4x8B]' ( 2) Des raisons économiques et contractuelles : )

Veillez écrire votre réponse ici :

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

### 3) Des raisons stratégiques institutionnelles et personnelles : \*

Veillez choisir toutes les réponses qui conviennent :

- La peur de favoriser une équipe concurrente
- La constitution d'un terrain de recherche exploitable sur le long terme
- La crainte du plagiat
- Ma stratégie personnelle de publication
- Un retour sur investissement insuffisant
- Une politique de labo ou d'établissement (ex : mon labo ne veut pas diffuser les données)
- Pas de frein à ce niveau

### 4) Des raisons de manque de ressources : \*

Veillez choisir toutes les réponses qui conviennent :

- Le manque de temps
- Le manque de ressources humaines
- Le manque de moyens financiers
- Le manque de connaissances générales sur l'ouverture des données
- Le manque de compétences (attribution de métadonnées, rédaction PGD/DMP, modalités de dépôt, ...)
- Pas de frein à ce niveau

### 5) Autre frein, précisez :

Veillez écrire votre réponse ici :

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

### Connaissez-vous les data journals ou d'autres publications destinées à décrire un ou plusieurs jeux de données ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

### Avez-vous déjà publié dans ce type de revue ? \*

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

La réponse était 'Oui' à la question '58 [DR4x9]' (Connaissez-vous les data journals ou d'autres publications destinées à décrire un ou plusieurs jeux de données ? )

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

### Avez-vous déjà utilisé ce type de revue pour effectuer de la veille sur les jeux de données ? \*

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

La réponse était 'Oui' à la question '58 [DR4x9]' (Connaissez-vous les data journals ou d'autres publications destinées à décrire un ou plusieurs jeux de données ? )

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

## 5ème partie : Vos besoins, vos attentes

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

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

Choisissez la réponse appropriée pour chaque élément :

	Tout à fait d'accord	Plutôt d'accord	Sans opinion	Plutôt pas d'accord	Pas du tout d'accord	Mon institution offre déjà une solution adéquat
J'aurais besoin d'informations sur les enjeux de la science ouverte	<input type="radio"/>					
Je souhaiterais que mon établissement propose un espace de stockage de mes données, sécurisé, partageable et adapté à mes besoins, pendant la durée d'une recherche	<input type="radio"/>					
Je souhaiterais que mon établissement propose une solution d'archivage des données au-delà de la durée de vie d'une recherche	<input type="radio"/>					
Je souhaiterais que mon établissement dispose d'un plan de gestion de données (PDG / DMP)	<input type="radio"/>					

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Sur quels aspects souhaiteriez-vous bénéficier d'une aide, et quel type d'aide choisiriez-vous en premier ? \*

Choisissez la réponse appropriée pour chaque élément :

	Aucun besoin	Documentation / Templates	Conseil	Formation	Personnel dédié
Sur la gestion des données de recherche	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sur la rédaction des Plans de Gestion de Données	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sur les questions techniques (par exemple: description jeu de données, métadonnées, normes, archivage à long terme)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sur la publication et la citation des données de recherche	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sur les aspects juridiques (par exemple les restrictions d'accès, le traitement des données sensibles, l'utilisation de licences)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sur la recherche de jeux de données disponibles dans ma discipline	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Si vous souhaitez bénéficier d'une aide sur un point non listé ci-dessus, précisez lequel :

Veuillez écrire votre réponse ici :

<https://sondagesv3.unistra.fr/index.php/admin/printablesurvey/sa/index/surveyid/567925/lang/fr>

28/30

20/12/2020

LimeSurvey - Enquete sur vos pratiques en matière de "données de la recherche" améliorer

## Souhaiteriez-vous des outils/espaces de travail collaboratifs pour vos projets de recherche ?

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

## Avez-vous déjà consulté : \*

Choisissez la réponse appropriée pour chaque élément :

	Oui	Non
DoRANum	<input type="radio"/>	<input type="radio"/>
OPIDoR (DMP OPIDoR ou Cat OPIDoR)	<input type="radio"/>	<input type="radio"/>

## Nous souhaiterions compléter cette enquête par quelques témoignages individuels. Accepteriez-vous d'être contacté pour un entretien ?

Veillez sélectionner une seule des propositions suivantes :

- Oui  
 Non

## Merci de nous indiquer votre adresse mail :

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

La réponse était 'Oui' à la question '66 [DR5x6]' ( Nous souhaiterions compléter cette enquête par quelques témoignages individuels. Accepteriez-vous d'être contacté pour un entretien ? )

Veillez écrire votre réponse ici :

Merci pour votre contribution.

04/01/2020 – 14:29

<https://sondagesv3.unistra.fr/index.php/admin/printablesurvey/sa/index/surveyid/567925/lang/fr>

29/30

## Annexe 2 - Questionnaire Flash

# Enquête sur vos pratiques en matière de "données de la recherche"

[NB : questionnaire adressé aux personnels ayant des activités de recherche]

## 1ère partie : Informations personnelles

### 1 - Quel est votre établissement de rattachement ? \*

Veillez écrire votre réponse ici :

### 2 - Votre statut ?\*

Veillez sélectionner une seule des propositions suivantes :

- Chercheur
- Enseignant-chercheur
- Doctorant
- Ingénieur
- Postdoc
- Autre

### 3 - Quel est votre champ disciplinaire ? \*

Veillez sélectionner une seule des propositions suivantes :

- Sciences Humaines et Sociales
- Sciences et Technologies
- Vie et Santé

Cette typologie a été proposée par l'ERC (*the European Research Council*) et est notamment utilisée par Cat OPIDoR. Pour plus de détails sur les disciplines :

[Détails typologie](#)

#### 3.1 - Quel est votre domaine disciplinaire principal ? (Sciences Humaines et Sociales) \*

[Répondre à la question seulement si la réponse à la question 3 était « Sciences Humaines et Sociales » ]

Veillez sélectionner une seule des propositions suivantes :

- SH1 Individus, marchés et organisations (économie, finance et management)
- SH2 Institutions, gouvernance et systèmes juridiques (sciences politiques, relations internationales, droit)
- SH3 Le monde social et sa diversité (sociologie, psychologie sociale, anthropologie sociale, sciences de l'éducation, études de communication)
- SH4 L'esprit humain et sa complexité (sciences cognitives, psychologie, linguistique, philosophie théorique)
- SH5 Cultures et production culturelle (études littéraires, études culturelles, étude des arts, philosophie)
- SH6 L'étude du passé humain (archéologie et histoire)
- SH7 Mobilité humaine, environnement et espace (Géographie humaine, démographie, santé, sciences de la durabilité, aménagement du territoire, analyse spatiale)

### 3.2 - Quel est votre domaine disciplinaire principal ? (Sciences et Technologies)

\*

[Répondre à la question seulement si la réponse à la question 3 était : « Sciences et Technologies »]

Veuillez sélectionner une seule des propositions suivantes :

- PE1 Mathématiques (tous les domaines des mathématiques, pures et appliquées, plus les fondements mathématiques des sciences informatiques, la physique mathématique et les statistiques)
- PE2 Constituants fondamentaux de la matière (physique des particules, nucléaire, des plasmas, atomique, moléculaire, des gaz et optique)
- PE3 Physique de la matière condensée (structure, propriétés électroniques, fluides, nanosciences, physique biologique)
- PE4 Chimie physique et analytique (chimie analytique, théorie chimique, chimie physique/physico-chimie)
- PE5 Chimie de synthèse et matériaux (nouveaux matériaux et nouvelles approches de synthèse, relations structure-propriétés, chimie de l'état solide, architecture moléculaire, chimie organique)
- PE6 Sciences informatiques et informatique (systèmes informatiques et d'information, sciences informatiques, calcul scientifique, systèmes intelligents)
- PE7 Ingénierie des systèmes et de la communication (ingénierie électrique, électronique, de la communication, optique et des systèmes)
- PE8 Ingénierie des produits et des procédés (Conception de produits et de processus, génie chimique, génie civil, ingénierie environnementale, génie mécanique, ingénierie automobile, processus énergétiques et les méthodes de calcul correspondantes)
- PE9 Sciences de l'Univers (Astro-physique/-chimie/-biologie; système solaire; systèmes planétaires; astronomie stellaire, galactique et extragalactique, cosmologie, sciences de l'espace, instrumentation et données astronomiques)
- P10 Sciences du Système Terre (géographie physique, géologie, géophysique, sciences de l'atmosphère, océanographie, climatologie, cryologie, écologie, changements environnementaux globaux, cycles biogéochimiques, gestion des ressources naturelles)
- PE11 Génie des matériaux (développement de matériaux avancés : amélioration des performances, modélisation, préparation à grande échelle, modification, adaptation, optimisation, nouvelles applications des matériaux, utilisation combinée des matériaux, etc.)

### 3.3 - Quel est votre domaine disciplinaire principal ? (Vie et Santé) \*

[Répondre à la question seulement si la réponse à la question 3 était : « Vie et Santé » ]

Veuillez sélectionner une seule des propositions suivantes :

- LS1 Molécules de la vie : Mécanismes biologiques, structures et fonctions (*pour tous les organismes* : biologie moléculaire, biochimie, biologie structurale, biophysique moléculaire, biologie de synthèse et chimique, conception de médicaments, méthodes innovantes et modélisation)
- LS2 Biologie intégrative : des gènes et des génomes aux systèmes (*pour tous les organismes* : génétique, épigénétique, génomique et autres études " omiques ", bioinformatique, biologie des systèmes, maladies génétiques, modification de gènes, méthodes et modélisation innovantes, " omiques " pour la médecine personnalisée)
- LS3 Biologie cellulaire, du développement et régénérative (*pour tous les organismes* : structure et fonction de la cellule, communication intercellulaire, embryogenèse, différenciation des tissus, organogenèse, croissance, développement, évolution du développement, organoïdes, cellules souches, régénération, approches thérapeutiques)
- LS4 Physiologie de la santé, de la maladie et du vieillissement (physiologie des organes et des tissus, physiologie comparée, physiologie du vieillissement, physiopathologie, communication entre les organes et les tissus, endocrinologie, nutrition, métabolisme, interaction avec le

microbiome, maladies non transmissibles dont le cancer - à l'exception des troubles du système système nerveux et les maladies liées à l'immunité)

- LS5 Neurosciences et troubles du système nerveux (développement, homéostasie et vieillissement du système nerveux, fonctionnement et dysfonctionnement du système nerveux, neurosciences des systèmes et modélisation, bases biologiques des processus cognitifs et du comportement, troubles neurologiques et mentaux)
- LS6 Immunité, infection et immunothérapie (le système immunitaire, les troubles associés et leurs mécanismes, biologie des agents infectieux et de l'infection, bases biologiques de la prévention et du traitement des maladies infectieuses, outils et approches immunologiques innovants, y compris les thérapies)
- LS7 Prévention, diagnostic et traitement des maladies humaines (technologies et outils médicaux pour la prévention, diagnostic et traitement des maladies humaines, approches et interventions thérapeutiques, pharmacologie, médecine préventive, épidémiologie et santé publique, médecine numérique)
- LS8 Biologie environnementale, écologie et évolution (*pour tous les organismes* : Écologie, biodiversité, changements environnementaux, biologie de l'évolution, écologie comportementale, écologie microbienne, biologie marine, écophysiologie, développements théoriques et modélisation)
- LS9 Biotechnologie et ingénierie des biosystèmes (biotechnologie utilisant tous les organismes, biotechnologie pour les applications environnementales et alimentaires, sciences appliquées aux plantes et aux animaux, bioingénierie et biologie synthétique, biomasse et biocarburants, risques biologiques)

## 2ème partie : Pratiques de gestion de l'ensemble de vos données (stockage, archivage, description)

Note explicative :

Données « chaudes » : données liées au projet de recherche, dont le contenu peut évoluer au cours de la recherche, données faisant l'objet de stockage.

Données « froides » : données produites à l'issue du projet de recherche, dont le contenu est stabilisé, données pouvant faire l'objet d'un archivage.

### À propos du stockage de vos données « chaudes »

#### 4 - Où stockez-vous vos données de recherche en cours de traitement (données « chaudes ») ? \*

Veillez choisir toutes les réponses qui conviennent :

- Sur votre ordinateur personnel (et/ou disque dur associé)
- Sur votre ordinateur professionnel (et/ou disque dur associé)
- Sur le cloud privé (espace personnel sur Google Drive, Dropbox,...)
- Sur le cloud institutionnel (Onedrive, Seafile,...)
- Sur un serveur institutionnel (de votre unité de recherche, de votre établissement ou d'une infrastructure de recherche)
- Je ne stocke pas de données chaudes
- Autre:

## **5 - Existe-t-il une politique de stockage des données « chaudes » de recherche (critères, modalités, nommage des fichiers...) :**

### **5.1 - Dans votre établissement ? \***

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas

### **5.2 - Dans votre unité de recherche ? \***

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas

## À propos de l'archivage de vos données « froides »

## **6- Existe-t-il une politique d'archivage des données « froides » de recherche (critères, modalités, nommage des fichiers...) :**

### **6.1 – Au niveau de votre établissement ? \***

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas

### **6.2 – Au niveau de votre unité de recherche ? \***

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas

## **7 - Où archivez-vous vos données de recherche après traitement (données « froides ») ? \***

Veillez choisir toutes les réponses qui conviennent :

- Sur votre ordinateur personnel (et/ou disque dur associé)
- Sur votre ordinateur professionnel (et/ou disque dur associé)
- Sur le cloud privé (espace personnel sur Google Drive, Dropbox,...)
- Sur le cloud Institutionnel (Onedrive, Seafile,...)
- Dans un entrepôt institutionnel (de votre unité de recherche, de votre établissement, ou national (Recherche Data Gouv) )
- Dans un entrepôt international (multidisciplinaire ou disciplinaire)
- Je n'archive pas de données froides
- Autre :

À propos de vos pratiques de description

**8 - Savez-vous ce qu'est un Plan de Gestion de Données (PGD), ou Data Management Plan (DMP) ? \***

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

**8.1 Avez-vous déjà participé à l'élaboration d'un Plan de Gestion de Données ? \***

*[Répondre à la question seulement si la réponse à la question 8 était : « Oui » ]*

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

## **3ème partie : Pratiques de partage des données**

Cette Partie ne concerne que les données froides.

**9 – Savez-vous ce que sont les principes FAIR ?**

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non

**10 - Avez-vous déjà téléchargé des données d'autres chercheurs sur le web ? \***

Veillez sélectionner une seule des propositions suivantes :

- Oui, c'est indispensable pour mes recherches
- Oui, régulièrement
- Oui, au moins une fois
- Non, pas pour le moment
- Non, je ne le ferai pas
- Non, je ne savais pas que c'était possible

**11 - Vous-même, avez-vous déjà diffusé vos données « froides » sur le web ? \***

Veillez sélectionner une seule des propositions suivantes :

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

**11.1 Si oui, par quel(s) moyen(s) diffusez-vous vos données « froides » ?**

*[Répondre à la question seulement si la réponse à la question 11 était : « Oui, régulièrement » ou « Oui, au moins une fois » ]*

**11.1.1- Sur les plateformes de Science Ouverte :**

Veillez choisir toutes les réponses qui conviennent :

- Un entrepôt de votre (vos) établissement(s)
- L'entrepôt national Recherche Data Gouv

- Un entrepôt de votre discipline
- Autre plateforme de Science ouverte, laquelle ?
- Vous n'utilisez pas les plateformes de Science Ouverte

### 11.1.2 - Sur d'autres plateformes :

Veillez choisir toutes les réponses qui conviennent :

- 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 / d'un éditeur (jeu de données associé à une publication)
- Vous n'utilisez que les plateformes de Science Ouverte
- Autre, laquelle :

### 11.2 Si non, pour quelle(s) raison(s) ?

[Répondre à la question seulement si la réponse à la question 11 était : « Non » ]

Veillez choisir la réponse qui convient le mieux :

- Je ne savais pas que c'était possible
- Je ne savais pas que c'était « obligatoire sauf exception »
- Je n'ai pas le droit
- Je n'en vois pas l'intérêt
- Mes données ne sont pas partageables
- Je ne sais pas le faire
- Je n'ai pas le temps
- Autre raison, laquelle ?

### 12 - De manière générale, êtes-vous d'accord pour partager vos données « froides » selon les principes de la Science Ouverte (diffusables, et réutilisables librement et gratuitement) ? \*

Veillez sélectionner une seule des propositions suivantes :

- Oui, sans délai
- Oui, après un délai
- Plutôt oui
- Plutôt non
- Non, pas du tout
- Je ne sais pas

### 13 - Parmi les raisons suivantes, qu'est-ce qui vous inciterait à partager vos données froides selon les principes de la Science Ouverte (diffusables, et réutilisables librement et gratuitement) ? \*

Veillez choisir au maximum 3 réponses parmi les propositions ci-dessous :

- Une obligation faite par le financeur de votre recherche (dans le cadre de projets européens, ANR...)
- La reconnaissance de la communauté scientifique
- Le souhait d'éviter de coûteuses duplications de travaux de recherche
- L'occasion de nouveaux contacts, de nouvelles coopérations scientifiques
- L'attachement à l'intégrité scientifique (transparence...)
- Une meilleure visibilité de vos travaux de recherche
- La volonté de développer la "science citoyenne" (science participative)
- Une meilleure prise en compte de vos travaux de recherche dans l'évaluation
- De possibles retombées financières (crédits de recherche...)
- Le souhait de voir se développer d'autres recherches à partir de vos jeux de données

- La volonté altruiste de partager ses recherches
- La généralisation de cette pratique au sein de votre proche communauté scientifique
- Autre, préciser
- Rien ne m'inciterait

## **14 - Quelles seraient les principales raisons qui vous freineraient pour partager vos données de recherche ?**

### **14.1 - Des freins liés à la nature de vos données : \***

Veillez choisir toutes les réponses qui conviennent :

- Données trop spécifiques à un contexte pour être réutilisées
- Données à caractère personnel (permettant d'identifier un individu)
- Données sous droit (*text mining* de bases commerciales,...)
- Données réglementées et/ou sensibles (liées à la Défense, à l'environnement, etc.)
- Pas de frein à ce niveau

### **14. - Des raisons économiques et contractuelles :\***

Veillez choisir toutes les réponses qui conviennent :

- Protection à des fins de valorisation de la recherche (brevet, autres droits de propriété industrielle,...)
- Contraintes liées aux partenariats de recherche
- Pas de frein à ce niveau

### **14.3 - Des raisons stratégiques institutionnelles et personnelles : \***

Veillez choisir toutes les réponses qui conviennent :

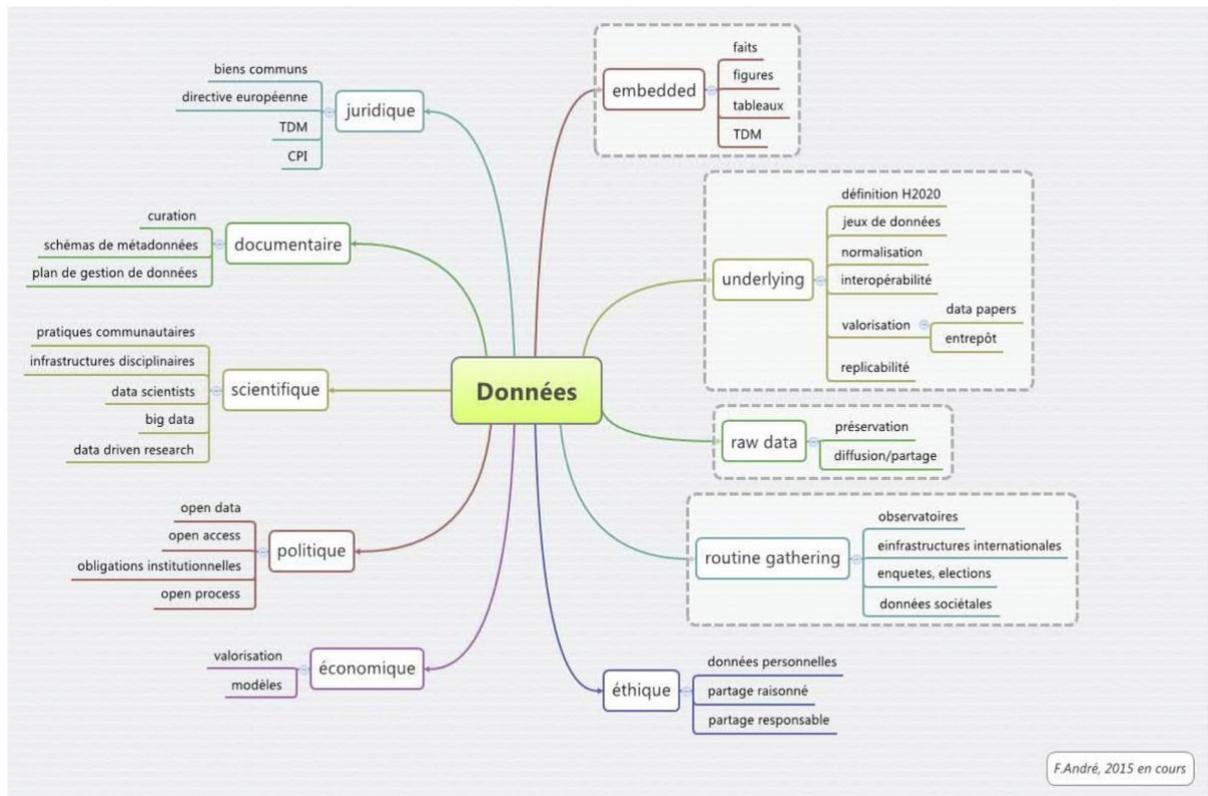
- La peur de favoriser une équipe concurrente
- La constitution d'un terrain de recherche exploitable sur le long terme
- La crainte du plagiat
- Ma stratégie personnelle de publication
- Un retour sur investissement insuffisant
- Une politique de laboratoire ou d'établissement
- Pas de frein à ce niveau

### **14.4 - Des raisons de manque de ressources : \***

Veillez choisir toutes les réponses qui conviennent :

- Le manque de temps
- Le manque de ressources humaines
- Le manque de moyens financiers
- Le manque de connaissances générales sur l'ouverture des données
- Le manque de compétences (attribution de métadonnées, rédaction PGD/DMP, modalités de dépôt, ...)
- Pas de frein à ce niveau

## Annexe 3 – Qualification des données de la recherche



23 nov 2015, réunion projet BSN9  
formations données