HAL open science

Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2

Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet

To cite this version:

Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet. Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2 : Rapport ; Annexe 1 : Résultats de l'enquête statistique ; Annexe 2 : Croisements statistiques ; Annexe 3 : Extraits des entretiens ; Synthèse des résultats. . [Rapport de recherche] Université Rennes 2. 2017, 159 p., 47 p., 114 p., 26 p., 23 p. hal-01635186v1

HAL Id: hal-01635186

https://hal.science/hal-01635186v1

Submitted on 14 Nov 2017 (v1), last revised 17 May 2018 (v2)

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

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

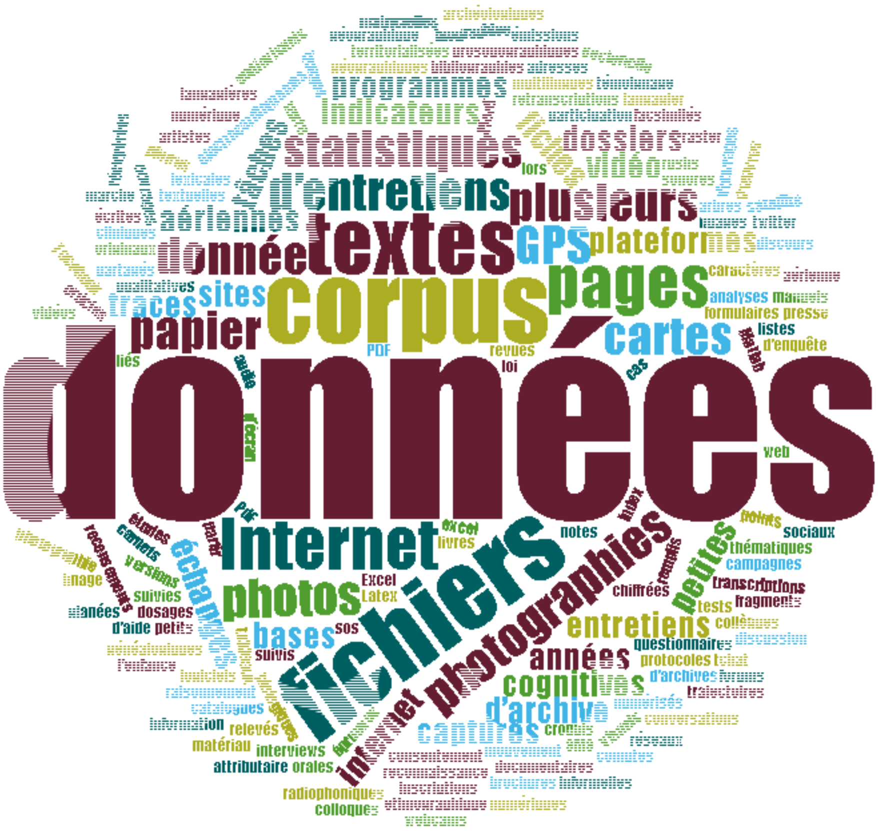
Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

UNIVERSITÉ RENNES 2

Novembre 2017

Données de la recherche en SHS.
Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2

Rapport rédigé par Alexandre Serres (dir.)
Marie-Laure Malingre, Morgane Mignon, Cécile Pierre,
Didier Collet



urfIST Rennes
UNIVERSITÉ RENNES 2 BU
Maison des Sciences de l'Homme en Bretagne

Coordination du rapport :

- **Alexandre Serres**, Maître de conférences en Sciences de l'Information et de la Communication, co-responsable de l'URFIST de Rennes

Equipe des rédacteurs :

- **Marie-Laure Malingre**, Conservatrice des bibliothèques, co-responsable de l'URFIST de Rennes
- **Morgane Mignon**, Ingénieure d'étude, MSHB
- **Cécile Pierre**, Conservatrice des bibliothèques, responsable du Département Recherche, SCD de l'Université Rennes 2
- **Didier Collet**, Technicien de bibliothèque, Département Recherche, SCD de l'Université Rennes 2

Résultats statistiques et graphiques :

- **Michaël Rabe**, étudiant en Master Statistiques et économétrie, Université Rennes 1

Image de couverture :

- **Cécile Pierre**, SCD de l'Université Rennes 2

SOMMAIRE

REMERCIEMENTS **4**

1. INTRODUCTION **5**

2. METHODOLOGIE ET MODALITES DE L'ENQUETE **11**

2.1 L'ENQUETE STATISTIQUE 11
2.2 L'ENQUETE QUALITATIVE 14
2.3 SYNTHESE SUR LA METHODOLOGIE DE L'ENQUETE 17

3. PRESENTATION DES RESULTATS **21**

3.1 OBSERVATIONS SUR LES DONNEES SOCIO-DEMOGRAPHIQUES 21
3.2 QUELLES DONNEES DE RECHERCHE ? 29
3.3 QUELLES PRATIQUES DE STOCKAGE ET D'ARCHIVAGE ? 48
3.4 PRATIQUES DE PARTAGE ET DE DIFFUSION DES DONNEES DE RECHERCHE 64
3.5 QUELS BESOINS, QUELLES ATTENTES ? 107

4. SYNTHESE DES LEÇONS DE L'ENQUETE **130**

CONCLUSION : RECOMMANDATIONS ET PROPOSITIONS **134**

DES RECOMMANDATIONS POUR QUI ? 134
SIX PROPOSITIONS POUR L'UNIVERSITE RENNES 2 135

TABLE DES ANNEXES **139**

TABLE DES FIGURES **140**

PROPOSITIONS POUR L'UNIVERSITE RENNES 2 ..... **142**

LISTE DES UNITES DE RECHERCHE DE L'UNIVERSITE RENNES 2 ..... **148**

LE QUESTIONNAIRE DE L'ENQUETE STATISTIQUE **149**

TABLE DES MATIERES **158**

Remerciements

Nous tenons à exprimer ici nos plus vifs remerciements à toutes celles et ceux qui ont aidé à la réalisation de ce projet d'enquête :

- **Marc Bergère**, Vice-Président Rennes 2 chargé de la Documentation et des Ressources technologiques et **Leszek Brogowski**, Vice-Président Rennes 2 chargé de la Recherche, pour leur indéfectible soutien politique et leur implication personnelle dans l'enquête ;
- **Nicolas Thély**, Directeur de la Maison des Sciences de l'Homme en Bretagne, pour le soutien matériel de la MSHB et pour son accompagnement actif tout au long du projet ;
- **Ottilia Henriet**, Directrice du SCD Rennes 2, pour le soutien total du SCD ;
- les enseignants-chercheurs, notamment les membres de la Commission Recherche, qui ont manifesté leur intérêt pour l'enquête et pour leurs remarques et suggestions sur le questionnaire, et toutes celles et ceux qui ont répondu au questionnaire ;
- les vingt-et-un enseignants-chercheurs et ingénieurs, qui ont accepté les entretiens et nous ont témoigné une totale confiance ;
- le personnel de la Direction de la Recherche et de la Valorisation, pour ses informations et la diffusion de l'enquête ;
- **Louisette Lechaux**, du Pôle Applications de la DSI, qui nous a initiés à LimeSurvey et nous a apporté une aide technique précieuse ;
- **Laurent Rouvière**, enseignant-chercheur au Département MIASHS, qui nous a aidés à trouver un étudiant stagiaire ;
- **Michaël Rabe**, étudiant en Master 1 Statistiques - Econométrie, à l'Université Rennes 1, qui a permis de faire toute l'exploitation des résultats statistiques, les tris à plat, les croisements des variables, la mise en ligne des données, et qui a fait preuve lors de son stage d'une infinie patience et d'une inépuisable bonne volonté pour satisfaire toutes nos exigences.

1. Introduction

Avant de présenter l'enquête menée à l'Université Rennes 2, il n'est sans doute pas inutile de rappeler l'une des définitions standard des données de recherche, car il en existe plusieurs[1] et la notion même de données de recherche est sujette à débats, notamment en SHS. Selon l'OCDE[2], « *les « données de la recherche » sont définies comme des* ***enregistrements factuels*** *(chiffres, textes, images et sons), qui sont utilisés comme* ***sources principales pour la recherche scientifique*** *et sont généralement reconnus par la communauté scientifique comme* ***nécessaires pour valider des résultats*** *de recherche.* » (souligné par nous). Tous les chercheurs de toutes disciplines travaillent donc à partir de et produisent des « données de la recherche », sous différentes formes : données de capteurs, statistiques, corpus de textes, archives, etc. Et comme nous le rappelions dans le texte de présentation de notre enquête, ces données de la recherche sont au cœur du travail de tout chercheur[3].

Pour mieux saisir le contexte général de cette enquête, il faut prendre en compte plusieurs éléments, que nous ne ferons que mentionner pour mémoire. Tout d'abord, l'omniprésence des données : que l'on parle de Big Data, de Small Data, d'Open Data, d'Open Data Research, les données sont partout, elles concernent évidemment toute la société (avec les enjeux du Big Data) mais aussi la recherche et toutes les disciplines. Ensuite, la multiplicité des défis et des questions posées : qu'il s'agisse de leur collecte, de leur traitement, de leur stockage, de leur archivage, de leur description, de leur régime juridique, de leur diffusion…, les données de recherche posent de nombreuses questions, loin d'être résolues.
L'un des enjeux les plus cruciaux est celui de leur ouverture, de leur diffusion, comme l'indiquait le même Rapport de l'OCDE de 2007 : « *La valeur des données réside dans leur exploitation. L'accès total et ouvert aux données scientifiques devrait devenir la norme internationale pour l'échange des données scientifiques issues de la recherche financée sur fonds publics.* »[4]. Cette question de l'ouverture des données de recherche est à replacer dans le **c**ontexte plus large de l'Open Data, mouvement international déjà ancien, très fort, qui concerne la mise à disposition du public des données des gouvernements, des administrations, des collectivités locales, et aussi maintenant des universités. Concernant les données de recherche, le mouvement pour l'ouverture des données date déjà de plusieurs années, et sans faire un historique, on peut juste évoquer trois étapes importantes.
Le premier jalon est précisément ce rapport de l'OCDE de 2007, « *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* », où étaient énoncés tous les principes de l'ouverture des données (13 principes concernant l'ouverture, la flexibilité, la transparence, la conformité au droit, etc.) et qui faisait des recommandations à tous les pays membres de l'OCDE pour développer des politiques d'ouverture des données de recherche. Un autre jalon essentiel est le programme de recherche européen Horizon 2020, qui précisait : « *Dès le programme de travail 2017, l'actuel projet pilote de libre accès aux données de recherche sera étendu de manière à couvrir l'ensemble*

---

[1] Sur la définition des données de recherche, nous renvoyons le lecteur au texte suivant : Joachim Schöpfel, Eric Kergosien, Hélène Prost. « Pour commencer, pourriez-vous définir 'données de la recherche' ? » Une tentative de réponse. *Atelier VADOR : Valorisation et Analyse des Données de la Recherche; INFORSID 2017*, May 2017, Toulouse, France. 2017. 〈hal-01530937〉

[2] OCDE. "*Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics*". 2007. 29 p. Disp. sur : http://www.oecd.org/fr/sti/sci-tech/principesetlignesdirectricesdelocdepourlaccesauxdonneesdelarecherchefinanceesurfondspublics.htm

[3] Pour des raisons de commodité, nous utiliserons dans l'introduction le terme de chercheur de manière générique, pour désigner à la fois les chercheurs des établissements de recherche, les enseignants-chercheurs des universités et les doctorants.

[4] OCDE, *Ibid*.

des domaines thématiques d'Horizon 2020, faisant du ***libre accès aux données de recherche la règle générale***. » (souligné par nous). Ce qui signifie que pour les projets de recherche financés par H2020, l'ouverture des données de recherche relève de l'obligation, sauf cas particuliers à justifier.
La troisième étape est française, avec la Loi du 7 octobre 2016 « Pour une République numérique », loi très importante sur de nombreux aspects touchant au numérique, mais aussi pour les universités (avec l'Open Data par défaut) et pour les laboratoires et les chercheurs, avec la reconnaissance officielle de l'Open Access pour les publications et l'autorisation de l'ouverture des données. Rappelons l'article 30 : « *II. Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche,* ***leur réutilisation est libre****. III. L'éditeur d'un écrit scientifique mentionné au I* ***ne peut limiter la réutilisation des données de la recherche,*** *rendues publiques dans le cadre de sa publication.* » (souligné par nous). Contrairement à H2020, la loi française n'oblige donc pas mais incite et protège les chercheurs, qui souhaitent diffuser leurs données de recherche.

Un autre enjeu important doit être souligné : le lien avec l'intégrité scientifique. Le partage des données est un élément essentiel d'un mouvement plus large, la Science Ouverte (*Open Science*), et un moyen au service de l'intégrité scientifique, comme l'explique un guide du CNRS : « *Cette publication et cet accès facilité doivent permettre à tout chercheur de questionner la scientificité et la véracité des conclusions, en rendant possible le renouvellement des expériences* »[5]. Face au développement de la fraude scientifique et du plagiat, le libre accès est encore l'un des meilleurs moyens pour assurer la transparence et la vérifiabilité des données.

Un dernier élément du contexte académique doit enfin être rappelé : le nécessaire développement de politiques sur les données de recherche dans les établissement, selon les recommandations de l'OCDE, de la Commission Européenne, du Ministère de l'Enseignement Supérieur et de la recherche… Il va falloir, dans les années à venir, mettre en place des politiques à différents niveaux d'échelle, notamment pour le stockage, l'archivage, la description des données, la diffusion des Plans de Gestion de Données, les politiques d'ouverture, etc. Or, on peut faire le constat d'un certain retard des universités par rapport aux organismes de recherche (CNRS, INRIA, INSERM, INRA…). Encore très peu d'universités ont commencé à réfléchir à cette question des données de recherche et à élaborer des politiques. Il convient de saluer ici les initiatives de certaines universités pionnières, comme l'Université de Lille 3, qui a réalisé récemment une enquête et commencé la définition d'une politique globale sur les données de recherche.

En bref, les enjeux liés aux données de recherche sont désormais bien documentés et de mieux en mieux connus : enjeux épistémologiques, avec le modèle discutable d'une « science conduite par les données » ; enjeux scientifiques avec l'ouverture des données de recherche, s'ajoutant à l'Open Data des données administratives et l'Open Access des publications ; enjeux sociétaux avec notamment les questions liées aux données personnelles et à leur exploitation ; enjeux techniques, avec les nouveaux besoins de stockage, d'archivage et de partage ; enjeux juridiques, avec le nouveau cadre juridique ouvert par la Loi sur le numérique d'octobre 2016 ; enjeux économiques, avec l'exploitation des données, le Big Data ; enjeux professionnels avec

[5] CNRS, COMETS, *Guide pour une pratique de la recherche intègre et responsable*, CNRS, septembre 2016

les nouvelles compétences à développer, les nouveaux métiers ou profils professionnels ; enjeux de formation, avec les supports, les tutoriels, les formations à mettre en place… La liste est longue de toutes les questions, de tous les défis posés par les données de recherche, et une littérature de plus en plus abondante est d'ores et déjà disponible pour les appréhender[6].

À tous ces enjeux s'en ajoute un, moins connu, car il concerne avant tout les décideurs, les acteurs des politiques des données de recherche, les professionnels et les chercheurs en IST : il s'agit de l'enjeu de la connaissance des pratiques, des représentations, des positions, des besoins de la première communauté concernée par les données de recherche, celles et ceux qui les collectent, les traitent, les produisent et parfois les diffusent, à savoir la communauté des enseignants-chercheurs et des chercheurs. Ils sont à la base de tous les processus liés aux données, présents dans la totalité du cycle de vie des données. Si les données sont au cœur du travail des chercheurs, les chercheurs sont au cœur de toute politique des données, et s'ils font l'objet de nombreuses attentions, parfois d'injonctions, pour changer leurs pratiques, celles-ci restent encore mal connues. Or, aucune politique de description, de stockage, d'archivage, de partage des données ne pourra se mettre en place sans l'adhésion ou l'implication des chercheurs. Et pour obtenir leur accord et leur implication, encore faut-il mieux connaître leurs pratiques réelles, leurs attentes, leurs craintes, leurs besoins. Cette connaissance est indispensable et constitue même, selon nous, le préalable indispensable à toute action locale pertinente. L'un des enjeux de la connaissance des pratiques des chercheurs porte notamment sur la mesure de l'écart entre les discours (d'injonction, de recommandation...) et les pratiques réelles et les représentations des enseignants-chercheurs, le tout au regard de la complexité des questions posées par les données de recherche, notamment en SHS.

Pourquoi une enquête sur les données de la recherche à l'Université Rennes 2 ?

Tous ces éléments du contexte et tous ces enjeux entremêlés, que nous venons de rappeler brièvement, font que la question des données de la recherche ne cesse de monter en puissance ; on ne compte plus le nombre d'événements scientifiques, d'articles, de communications, qui lui sont consacrés, elle intéresse de plus en plus d'acteurs du monde académique (chercheurs, responsables de la recherche, professionnels de l'IST, ingénieurs, etc.) et touche des thématiques et des domaines de plus en plus importants. En SHS et ALL, le mouvement autour des Humanités Numériques est évidemment investi au premier plan sur cette question et les Maisons des Sciences de l'Homme en sont des acteurs incontournables. De son côté, le monde des bibliothèques de recherche et des Services Communs de Documentation est également fortement impliqué dans cette thématique depuis plusieurs années. Enfin, le Réseau national des URFIST (Unités Régionales de Formation à l'Information Scientifique et Technique) se mobilise également, notamment à travers le projet national de formation à distance, Doranum[7], co-piloté par l'URFIST de Nice et l'INIST.

C'est dans ce contexte général que l'URFIST de Rennes a pris l'initiative, en juin 2016, de lancer un projet d'enquête sur les données de recherche à l'Université Rennes 2, et compte tenu de l'ampleur et de l'intérêt du projet, a sollicité deux partenaires privilégiés, la Maison des Sciences de l'Homme en Bretagne et le Service Commun de Documentation de Rennes 2, qui ont accepté immédiatement. Cette enquête a donc été, dès le départ, le travail collectif d'une équipe qui a mené l'ensemble du projet.

[6] Voir notamment le site de veille « Données de la recherche » : http://www.donneesdelarecherche.fr/
[7] http://doranum.fr/

Quel était le public visé par l'enquête ?

Il s'agissait uniquement des chercheurs, enseignants-chercheurs et ingénieurs de recherche de l'Université Rennes 2. Pourquoi en exclure les doctorants ? Il nous a semblé que les doctorants pouvaient avoir des besoins spécifiques, mais aussi des pratiques ou des représentations différentes, et qu'il faudrait adapter un autre questionnaire. De plus, notre enquête portant notamment sur la question d'une politique d'établissement, nous avons préféré viser plutôt les enseignants-chercheurs, et notamment les responsables de laboratoires.

Quels étaient les objectifs et ont-ils été atteints ?

Trois objectifs principaux étaient visés :

- **Établir un état des lieux :**

Il s'agissait d'établir un panorama le plus complet possible de la situation des données de recherche à Rennes 2, sur trois plans : d'abord des données elles-mêmes (quels types de données sont collectées et produites, quels volumes sont en jeu, etc.), ensuite des pratiques des chercheurs, mais aussi de leurs perceptions, notamment sur la question de l'ouverture et du partage des données, de leurs réticences éventuelles, et enfin de leurs besoins (en matière de services, d'infrastructures, d'accompagnement, etc.). À l'instar de l'Université de Lille 3, dont l'enquête auprès des doctorants nous a beaucoup inspirés, nous avons pensé qu'un tel panorama pouvait présenter un intérêt pour la définition éventuelle d'une politique sur les données de recherche à Rennes 2.
Cet objectif a-t-il été rempli ? Nous pouvons ici, et sans fausse modestie, répondre incontestablement oui : l'enquête, aussi bien statistique que qualitative, nous a permis de faire un tableau assez détaillé et approfondi, même s'il n'est pas exhaustif, des pratiques, des représentations et des besoins des enseignants-chercheurs. La taille même de ce rapport et la richesse des annexes témoignent du soin pris à cet état des lieux.

- **Élaborer une offre de services :**

Le deuxième objectif concernait surtout les acteurs mêmes de l'enquête, l'URFIST, le SCD et la MSHB, qui ont tous trois une mission de soutien à la recherche. À travers cette enquête et notamment le recueil des besoins, il s'agissait pour ces trois services de pouvoir mieux répondre aux besoins exprimés par les chercheurs, en termes d'accompagnement, de formation, de services, de ressources, etc. S'il est encore évidemment trop tôt pour élaborer et présenter une telle offre de services, nous avons d'ores et déjà une bien meilleure connaissance de ces besoins, ce qui nous a permis de faire un ensemble de propositions à la communauté des chercheurs de Rennes 2, notamment en termes de services.

- **Alimenter la recherche :**

Le troisième objectif visé portait sur un double apport possible à la recherche :

- d'une part, pour les travaux sur l'observation des pratiques informationnelles des chercheurs : l'intérêt ici concerne d'abord l'URFIST de Rennes et au-delà, le nouveau GIS (Groupement d'Intérêt Scientifique) « Réseau des URFIST », dont l'un des objectifs de recherche porte précisément sur cette thématique des pratiques informationnelles ; sur ce plan, l'enquête a indubitablement permis de mieux connaître les représentations, les réticences, les points de blocage, d'identifier les points critiques, notamment sur le partage des données. Par ailleurs, la réflexion sur la méthodologie de l'enquête nous a d'ores et déjà permis d'envisager une suite, en termes de projet de recherche du GIS « Réseau des URFIST » ;

- d'autre part, pour la réflexion et le débat scientifique autour de la place des données en SHS : la MSHB est ici au premier plan, et l'enquête permettra d'alimenter le Séminaire Humanités Numériques, qui porte notamment sur cette question des données de recherche.

Deux objectifs complémentaires avaient été également identifiés, au début du projet :

- d'une part, un objectif de sensibilisation des enseignants-chercheurs : par elle-même, l'enquête avait une dimension de "méta-communication", en incitant les chercheurs à réfléchir à leurs pratiques et en les sensibilisant aux enjeux ; sur ce point, l'objectif a été atteint en partie, de l'aveu même de nombreux participants aux entretiens, qui ont reconnu que l'enquête leur a fait mesurer l'importance de cette thématique ;
- d'autre part, un objectif de test et d'élargissement éventuel : le projet d'enquête a intéressé d'autres universités de l'UBL, ainsi que d'autres régions de France. Et ce sera l'objet d'une autre phase.

Quelles étaient les modalités et le déroulement de cette enquête ?

Dès le départ, nous avons eu l'idée de faire marcher cette enquête sur deux jambes : une enquête statistique en ligne et une enquête qualitative, par des entretiens avec des chercheurs volontaires. L'enquête statistique, fondée sur un questionnaire de 32 questions, que l'on trouvera en annexe[8], a été accessible en ligne sur la plateforme LimeSurvey de Rennes 2, en février et mars 2017. Les entretiens se sont déroulés de mars à mi-avril et ont touché 21 chercheurs. La méthodologie des deux volets de l'enquête est présentée en détails dans la partie 2 du rapport.
Une enquête de cette nature demande du temps et surtout des soutiens. Nous avons pris le premier et obtenu les seconds.
Du temps, il en fallut d'abord pour élaborer, peaufiner et tester le questionnaire de l'enquête statistique ; plusieurs mois ont été nécessaires à l'automne-hiver 2016-2017, avec beaucoup d'allers-retours au sein de l'équipe mais aussi un test des questions par un panel de chercheurs volontaires[9]. L'importance de cette phase de test fut décisive quant aux résultats, et nous y revenons plus loin dans la partie 2.1. Si l'enquête proprement dite s'est déroulée pendant environ trois mois, de février à avril 2017, il fallut ensuite deux bons mois pour l'analyse et le dépouillement des résultats statistiques, et deux autres mois pour ceux des entretiens.

Les soutiens, aussi bien politiques que techniques, n'ont pas manqué : soutien politique de la direction de Rennes 2 et notamment des deux Vice-Présidents impliqués sur cette thématique (VP Recherche et VP Documentation), soutien de la Commission Recherche devant laquelle le projet d'enquête fut présenté en novembre 2016, soutien technique de la DSI pour l'utilisation de LimeSurvey, du Master MIASHS pour le recrutement d'un étudiant stagiaire…

Au final, le matériau recueilli a dépassé nos espérances par sa richesse, notamment pour les entretiens. Le présent rapport, réalisé par les cinq membres de l'équipe de l'enquête, s'organise en trois grandes parties, de taille inégale :

- une première partie (partie 2) présente la méthodologie, aussi bien de l'enquête statistique que de l'enquête qualitative. L'explicitation des critères et des choix méthodologiques est évidemment essentielle dans tout travail d'enquête ;

[8] Voir en Annexes ci-dessous, p. 149
[9] Le questionnaire lui-même a fait l'objet de plus de 15 versions entre juillet 2016 et le lancement de l'enquête en janvier 2017.

- la deuxième partie (partie 3) est la plus importante en taille (avec plus de 100 pages), car elle est consacrée à la présentation de tous les résultats. Nous avons suivi l'ordre de l'enquête statistique, en présentant d'abord les données socio-démographiques de l'enquête, puis les résultats sur la typologie des données, sur le stockage et l'archivage, sur le partage, et enfin sur les besoins. Pour chacune de ces sous-parties, nous avons opté pour une présentation successive des résultats de l'enquête statistique, puis des entretiens, pour terminer par une synthèse ;
- la troisième partie regroupe à la fois les principales leçons à tirer de l'enquête et la conclusion, contenant nos propositions pour une politique des données de la recherche à Rennes 2.

Par ailleurs, d'importantes annexes viennent compléter le rapport :

- certaines ont pu être intégrées directement à la fin du rapport, notamment le rappel de nos six propositions et le questionnaire de l'enquête statistique ;
- d'autres annexes, trop volumineuses, ne pouvaient être intégrées au rapport et font l'objet de trois documents autonomes : l'Annexe 1, qui regroupe tous les résultats bruts de l'enquête statistique, l'Annexe 2 qui présente une partie des résultats des croisements statistiques effectués entre différentes questions, enfin l'Annexe 3 fournit de très larges extraits des entretiens (soigneusement anonymisés), classés par grandes thématiques puis par sujets.

Compte-tenu du volume de notre rapport, nous avons réalisé également une synthèse, d'une trentaine de pages, qui permettra aux lecteurs pressés de prendre connaissance de l'essentiel des résultats.

Enfin, partisans de l'ouverture des données, nous ne pouvions pas ne pas nous appliquer les préceptes que nous préconisons et l'on pourra trouver, d'une part sur HAL tous les documents du rapport (les textes et les annexes), d'autre part sur une plateforme de présentation de données, TableauPublic, les graphiques de l'enquête statistique. Quant à nos données « brutes », i.e. les résultats bruts de l'enquête statistique et plus encore les textes des entretiens, elles entrent dans la catégorie des données personnelles, et à ce titre sont juridiquement protégées ; il n'était donc évidemment pas question de les rendre publiques telles quelles ! Sur ce point, notre enquête illustre elle-même la complexité de la question du partage des données, en pratiquant une ouverture partielle et contrôlée !

2. Méthodologie et modalités de l'enquête

L'ambition de ce projet n'était pas de réaliser une véritable enquête sociologique, qui aurait demandé des compétences qu'à l'évidence nous ne possédions pas, mais de mettre en œuvre notre démarche avec toute la rigueur possible et selon une méthode précisément définie. Dans sa réflexion sur le dispositif à élaborer, le groupe de travail constitué autour de cette enquête s'est assez vite orienté vers un processus à double détente, à la fois quantitatif et qualitatif, constitué d'abord par la rédaction et la diffusion d'un questionnaire, puis par la mise en œuvre d'une série d'entretiens individuels. Les entretiens ont été pensés comme un moyen particulièrement adapté, d'une part d'avoir un retour sur le questionnaire, d'autre part de revenir sur certains points méritant des développements plus conséquents, détaillés et nuancés que ce qu'il était possible d'obtenir avec un questionnaire. Par ailleurs, certains aspects qu'il n'était pas possible d'intégrer de prime abord, pouvaient ainsi être abordés par le biais des entretiens. Enfin l'intérêt était aussi d'approfondir un positionnement différencié en orientant les échanges selon les responsabilités des chercheurs à l'université (direction de laboratoire, vice-présidence, etc.).

2.1 L'enquête statistique

2.1.1 Le questionnaire

La gestion des données de la recherche devient, on le voit, une question de plus en plus centrale et les enquêtes mettant en évidence les différentes problématiques liées à leur manipulation se multiplient. Nous avons choisi de nous appuyer sur certaines d'entre elles, récentes et ayant avec la nôtre une proximité d'objectifs et de thématiques. Nous avons en particulier repris et adapté les éléments et la structure de l'enquête menée en 2015 à l'Université de Lille 3 par Hélène Prost et Joachim Schöpfel[10], qui s'adressait aux chercheurs et doctorants en Sciences Humaines et Sociales. Notre périmètre d'investigation étant l'Université Rennes 2, le champ disciplinaire pris en compte est, dans le cas de notre enquête, également très majoritairement SHS. En revanche, nous avons volontairement ciblé les enseignants-chercheurs, chercheurs, ingénieurs de recherche et laissé de côté le public doctorant. En effet, d'une part l'objectif était d'identifier spécifiquement les pratiques, les représentations et besoins des enseignants-chercheurs et des responsables de la recherche, d'autre part il nous semblait préférable que les doctorants fassent ultérieurement l'objet d'une enquête séparée et spécifique, la totalité des questions auxquelles nous pensions n'étant pas forcément transposables à leur activité de recherche et à leur parcours. Enfin, le questionnaire diffusé par Joumana Boustani sur la « *Culture informationnelle liée aux données* »[11], ainsi que les résultats de l'enquête autrichienne sur les données de recherche, menée en 2015 dans le cadre du projet *e-infrastructure Austria*[12] ont également inspiré nos travaux.
Sur la base de ces précédentes initiatives, nous avons élaboré un questionnaire de type directif, en procédant, entre mai et décembre 2016, par étapes successives : un noyau initial de questions

[10] Prost Hélène, Schöpfel Joachim. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. : Rapport final.* [Rapport de recherche] Lille 3. 2015. http://hal.univ-lille3.fr/hal-01198379
[11] Boustany Joumana *Culture informationnelle liée aux données*. http://survey.docinfos.fr/index.php/963778?newtest=Y&lang=fr
[12] Bauer Bruno, Ferus Andreas, Steineder Christian, Gorraiz Juan, et al. *Researchers and their data. Results of an Austria survey*– Report 2015. Version 1.2. Disp. : https://phaidra.univie.ac.at/view/o:409318

a d'abord été soumis à discussion interne pour arriver à une première stabilisation ; il semblait cependant indispensable d'en faire valider la formulation et la cohérence et d'examiner en outre si ne manquaient pas certains aspects importants. Il a donc été décidé de solliciter à plusieurs reprises un petit panel de chercheurs de l'Université Rennes 2, et parmi eux, le Vice-Président chargé de la recherche et le Vice-Président chargé de la documentation et des ressources technologiques, dont le soutien a été déterminant, pour tester, évaluer le questionnaire et chercher au final la plus grande adéquation possible avec les situations réelles de la communauté académique. C'est ainsi que les retours obtenus par ce système de navettes entre les chercheurs et le groupe de travail ont conduit à opérer, par retouches successives, des modifications concernant l'ordre ou l'énonciation de certaines questions (clarification des énoncés), à rendre certaines d'entre elles facultatives, à supprimer ou au contraire ajouter des questions (ou des items), pour arriver, au fil des différentes versions du questionnaire, à un état définitif.
Le questionnaire tel qu'il a été diffusé comporte 32 questions, dont 13 sont obligatoires et 19 facultatives. Outre les éléments relatifs au profil des répondants, il a été structuré selon quatre axes thématiques principaux : la typologie des données collectées et produites (nature et support des données), les pratiques de stockage et d'archivage des données (volume et lieux de stockage, sauvegarde, politique d'archivage, description des données), les pratiques de partage et de diffusion des données, avec en perspective le développement du libre accès aux résultats de la science, enfin les besoins et attentes liés aux données de recherche (connaissance et pratique du Plan de Gestion de Données, dispositifs à mettre en place à l'université, besoins en matière de gestion des données, de formation ou d'assistance). Un accent particulier a été mis sur les questions relatives au partage et à la diffusion des données en en rendant plusieurs obligatoires. En revanche, sur certains aspects (dont des aspects plus techniques comme l'attribution de métadonnées), on a choisi de rendre certaines questions facultatives, afin d'éviter qu'elles ne soient bloquantes. Dans tous les cas, le discours se voulait neutre et non prescriptif.

Structuration et répartition des questions :

	Questions obligatoires	Questions facultatives	TOTAL
1. Vous-même	5	0	**5**
2. Typologie des données	2	4	**6**
3. Stockage, archivage	2	7	**9**
4. Partage, diffusion	3	4	**7**
5. Besoins, attentes	1	4	**5**
TOTAL	**13**	**19**	**32**

Figure 1 : Structuration du questionnaire

Pour des raisons d'efficacité dans le traitement du formulaire, celui-ci comporte très peu de questions ouvertes, les entretiens étant par ailleurs censés permettre des développements plus libres et personnels : 2 questions proposent de répondre dans un champ de texte libre, l'une concerne le stockage et l'archivage des données, l'autre concerne les besoins et attentes du chercheur (une troisième question donne la possibilité d'indiquer son adresse mail pour un entretien ultérieur) ; 29 questions permettent de répondre via une liste d'options (à choix multiples ou à choix unique, avec 22 questions fermées et 7 questions mixtes). Le temps nécessaire pour remplir le questionnaire a été estimé à environ un quart d'heure.
Le questionnaire a été saisi sur l'application LimeSurvey, logiciel d'enquête en open source disponible à l'université et permettant une gestion souple de l'interaction avec les répondants. Il a été diffusé en ligne sur le site web de l'Université Rennes 2, après avoir bénéficié des

conseils et de l'expertise de la DSI (Direction des Systèmes d'Information) : là encore, il a été procédé à des tests avant ouverture officielle. À l'issue d'une campagne de communication autour du dispositif, l'enquête a été lancée le 6 février 2017 pour une durée initiale d'un mois, prolongée d'un mois.

2.1.2 L'échantillon

L'enquête, visant notamment à établir un état des lieux le plus complet possible des pratiques et des représentations dans le périmètre choisi (les acteurs de la recherche dans une université), a retenu le principe d'une exhaustivité de la population ciblée à l'intérieur de ce cadre, s'agissant d'une « population fermée et identifiée »[13]. Elle s'est ainsi adressée à l'ensemble des personnels enseignants-chercheurs, chercheurs et ingénieurs de recherche de l'Université Rennes 2, soit 496 personnes, qui ont reçu un mel d'invitation via LimeSurvey. La totalité des laboratoires de l'université a été visée, soit 19 laboratoires au moment de l'enquête (l'université comporte actuellement 20 unités de recherche), sans considération particulière de la taille des équipes ou du volume et de l'usage supposés des données dans les disciplines.

2.1.3 Le nombre de réponses

Les taux de réponse observables pour les enquêtes en ligne peuvent être inférieurs, mais aussi selon les cas supérieurs à ceux des questionnaires papier, c'est ce qui ressort d'un certain nombre de travaux sur le sujet[14]. L'enquête menée à Rennes 2 a bénéficié d'un taux de réponses relativement élevé, puisqu'elle a eu 143 répondants sur les 496 chercheurs sollicités.

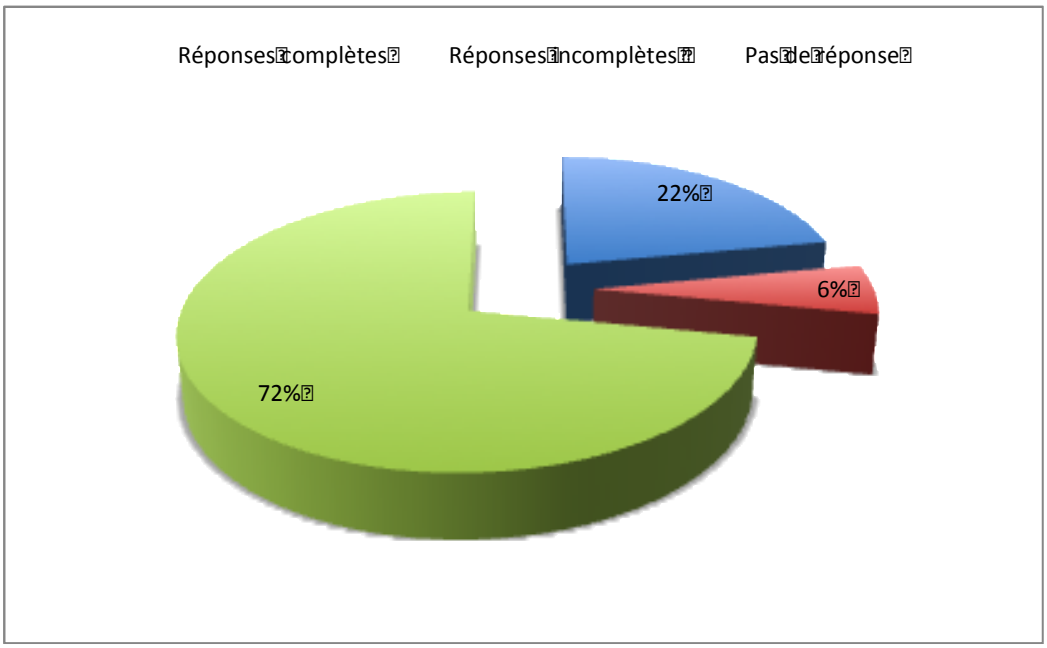


Figure 2 : Taux de participation à l'enquête des enseignants-chercheurs de Rennes 2 (N=496)

Si l'on estime généralement satisfaisant un taux de réponses se situant entre 20 et 30 % pour une enquête en ligne, les résultats obtenus ici seront jugés positifs, puisqu'ils représentent un

[13] Ganassali Stéphane, Moscarola Jean. « Protocoles d'enquêtes et efficacité des sondages par internet ». *Décisions Marketing*, 2004, n° 33, p. 63-75.
[14] Aragon Yves, Bertrand Sandrine, Cabanel Magali, Legrand Hervé. "Méthode d'enquête par internet : Lecons de quelques expériences". *Décisions Marketing*, 2000, n° 19, p. 29-37.

taux de participation de 28 %. Sur les 143 retours de questionnaires collectés, 109 sont complets (22 %), 34 incomplets (6 %), dont 4 inexploitables. Si la totalité des réponses exploitables a servi de base aux traitements statistiques simples, en revanche les croisements de données ont été opérés sur les seules 109 réponses complètes. Il est intéressant de constater par ailleurs que les 19 unités de recherche existant à Rennes 2 au moment où a été lancée l'enquête sont représentées sans exception, même si c'est à des échelles variables. En l'état, le nombre de réponses apportées est déjà susceptible de fournir une image assez précise des pratiques relatives à la gestion des données de recherche. La restitution des tris à plat et la définition des croisements à opérer ont été tout l'enjeu du travail de traitement confié à l'étudiant statisticien recruté par la MSHB. L'enquête qualitative devait en outre permettre d'affiner cette image.

2.2 L'enquête qualitative

En organisant une série de rendez-vous avec des chercheurs de l'Université Rennes 2, il s'agissait, dans le prolongement du questionnaire, d'abord d'avoir des retours sur celui-ci, ensuite de mieux cerner et de préciser les pratiques, les représentations et perspectives concernant la gestion des données de recherche, mais également les positions des uns et des autres sur la question de l'ouverture et du partage des données, sur l'éventualité d'une politique en la matière et sur le niveau auquel celle-ci pourrait être menée ; enfin on souhaitait acquérir une connaissance approfondie des besoins et attentes des chercheurs, ainsi que des services qu'il était souhaitable de mettre en œuvre pour y répondre.
Dans cette optique, il était intéressant de concevoir des approches différentes dans les échanges, selon le niveau de positionnement que l'on voulait privilégier : des réactions fondées sur la pratique et l'expérience personnelle de l'enseignant-chercheur, une expression de nature plus politique émanant d'une unité de recherche, enfin un positionnement institutionnel au niveau de l'établissement, en l'occurrence l'université. C'est pourquoi différents types d'interlocuteurs ont été sollicités et plusieurs scénarios d'entretiens distincts ont été élaborés.

2.2.1 Le canevas des entretiens

Trois canevas ont donc été conçus, à destination de représentants de la direction de l'Université, de directeurs de laboratoires et d'enseignants-chercheurs, à titre individuel. Toutefois, il a été prévu que chaque entretien débute pour tous de façon identique par un retour sur l'enquête en ligne, les échanges sur le fond commençant véritablement, pour les directeurs de laboratoires et les représentants de la direction de l'université, par un questionnement sur leur perception de l'importance des données de recherche, tandis que les enseignants-chercheurs étaient plutôt invités à expliquer concrètement la nature des données qu'ils étaient amenés à gérer et la manière dont ils le faisaient.
On a souhaité introduire ensuite le point capital de la politique des données ; dans le cas d'un entretien avec un directeur de laboratoire, deux questions spécifiques devaient être posées : l'une portait sur l'intérêt et les formes que pourrait prendre une politique des données de recherche au niveau du laboratoire, la seconde sur le positionnement à l'égard d'une politique qui serait portée par l'université, notamment en matière de partage des données. Cette dernière question d'une politique d'établissement en matière de données de recherche, et en particulier de partage des données, a été placée au centre des échanges avec les représentants de l'établissement.
Par ailleurs, il s'agissait aussi, en repartant des éléments de l'enquête en ligne, de clarifier l'approche personnelle du chercheur sur le partage des données précisément, mais en l'abordant cependant du triple point de vue du responsable de laboratoire, des instances dirigeantes de

l'université et du chercheur individuel. Il a été prévu de poser cette question dans les mêmes termes à tous les interlocuteurs.
L'étape suivante des entretiens était constituée par un sondage sur les besoins et attentes générés par la question des données de recherche : individuellement pour les enseignants-chercheurs retenus dans le panel, au niveau du laboratoire pour les directeurs de laboratoire, à l'échelle de l'université pour les institutionnels.
Avec le panel d'enseignants-chercheurs à titre individuel, on a choisi de terminer l'entretien, le cas échant, par une question plus scientifique visant à situer la question des données dans le processus de recherche en Sciences humaines et sociales et en contexte numérique.
Nous avons repris ci-dessous et explicité, dans l'ordre de leur déroulement, l'ensemble des questions formant le canevas des entretiens, en regroupant celles qui étaient communes et en spécifiant les questions propres à chaque type d'interlocuteur sollicité :

Scénarios des entretiens	**Directeurs de labos**	**Institutionnels** (Vice-Présidents)	**Panel d'EC** (choisis et volontaires)
Le questionnaire en ligne	Si vous avez déjà rempli le questionnaire en ligne, quels sont vos retours sur l'enquête : les libellés des questions étaient-ils clairs, quelles sont les questions qui vous ont posé le plus de problèmes, le questionnaire vous a-t-il paru adapté, pertinent ?		
Importance des DR	Comment ressentez-vous l'importance accordée actuellement aux données de recherche ?		
Usage individuel des données			Pouvez-vous repréciser la nature des données collectées ? Les types de données produites ? Quelle gestion en avez-vous ?
Politique des données au niveau d'un laboratoire	Quelles est votre position en tant que directeur de laboratoire, sur une éventuelle politique des données de recherche : est-ce que c'est une question qui a déjà été débattue dans votre labo ? Si non, pensez-vous que votre laboratoire devrait avoir une politique commune, cohérente ? Et sur quoi devrait-elle porter ?		
Scénarios des entretiens	**Directeurs de labos**	**Institutionnels** (Vice-Présidents)	**Panel d'EC** (choisis et volontaires)
Politique d'établissement pour les données de recherche	Pensez-vous que l'Université Rennes 2 ou votre tutelle devrait avoir une politique commune en matière de données de recherche ? Quelles en seraient les priorités, quelle position en matière de partage ?		

Position personnelle sur le partage	Quelle est votre approche personnelle de la question du partage des données ? Êtes-vous plutôt favorable / défavorable au partage des données, dans les conditions juridiques prévues ? peut-on repartir des questions 20 et 21 du questionnaire et les développer ?		
Besoins et attentes	Quels seraient les principaux besoins		
	du labo	à l'échelle de l'université	selon vous
	en termes de services, d'accompagnement, de formation, d'outils techniques, d'infrastructures de stockage ?		
Les données dans la recherche SHS			Comment percevez-vous la question des données dans la recherche en SHS : quelle est l'importance de cette question, quel est le principal impact du numérique dans ce domaine ?

Figure 3 : Canevas des questions pour les différents entretiens

2.2.2 La constitution du panel

Pour constituer la liste des chercheurs à contacter, trois modalités ont été retenues : se baser sur l'enquête en ligne dans laquelle les répondants volontaires pour un entretien avaient donné leur adresse mel (19 initialement) ; s'adresser à des collègues approchés dans divers contextes : groupe de travail Humanités numériques, collègues partenaires de l'enquête, enseignants-chercheurs d'un laboratoire précis et membres de la Commission Recherche (21 initialement) ; contacter systématiquement les directeurs de laboratoires de l'Université Rennes 2 (19 au moment de l'enquête). Les mels de sollicitation, puis d'organisation des rendez-vous ont été envoyés entre mi-février et début mars.

2.2.3 Le nombre et le déroulement des entretiens

Au final, 21 entretiens ont pu avoir lieu entre le 3 mars et le 12 avril, ce qui constitue une base substantielle d'analyse. Y ont participé, à titre individuel, 12 enseignants-chercheurs et un ingénieur d'étude, au titre de leurs responsabilités, 2 vice-présidents de l'université (le Vice-Président chargé de la recherche et le Vice-Président chargé de la documentation et des ressources technologiques), ainsi que 6 Directeurs de laboratoire ou d'organisme de recherche. Treize des dix-neuf laboratoires que comptait l'université au moment des entretiens ont été représentés, douze champs disciplinaires ont été impliqués.
Chaque entretien a été mené par un binôme issu du groupe de travail dédié à l'enquête, sur la base des canevas préétablis, mais en laissant à l'interlocuteur toute liberté de développer certains points ou même d'aborder de nouvelles questions, suivant en cela le principe d'entretiens semi-directifs. Les entretiens - d'une durée totale de douze heures - ont tous fait l'objet d'enregistrements qui ont pu être récupérés sur un outil de transcription et d'analyse.

2.2.4 L'outil de transcription Sonal

Le choix du logiciel Sonal, entre les différents outils envisagés pour le traitement de l'enquête qualitative, résulte de sa puissance et de la richesse de ses fonctionnalités. C'est un outil gratuit

de retranscription et de dépouillement d'enquêtes par entretien[15], qui a facilité à la fois le travail de transcription des interviews et leur analyse. L'ensemble des données des entretiens constituent un corpus, que l'application permet de découper, d'encoder en créant des thématiques à attribuer aux extraits audio avec des codes de couleurs et enfin d'indexer par des mots clés ; il est ainsi possible de filtrer rapidement et finement le contenu des entretiens. C'est l'un des grands intérêts de Sonal que de permettre l'indexation de passages d'entretiens, de favoriser l'identification et le repérage des sujets qui y sont contenus. La retranscription, qui reste manuelle, est synchronisée avec le son, elle s'effectue en suivant des portions définies d'enregistrements sonores. Les fichiers peuvent être décrits par des variables et des modalités qui sont stockées dans une base de données. Les fonctions d'analyse (analyse chronométrique, analyse lexicométrique, etc.) peuvent s'appliquer aux transcriptions et aux encodages. Enfin, c'est un outil qui s'est aisément adapté à une utilisation collective, après concertation sur le choix des thématiques et sur la pratique d'indexation.

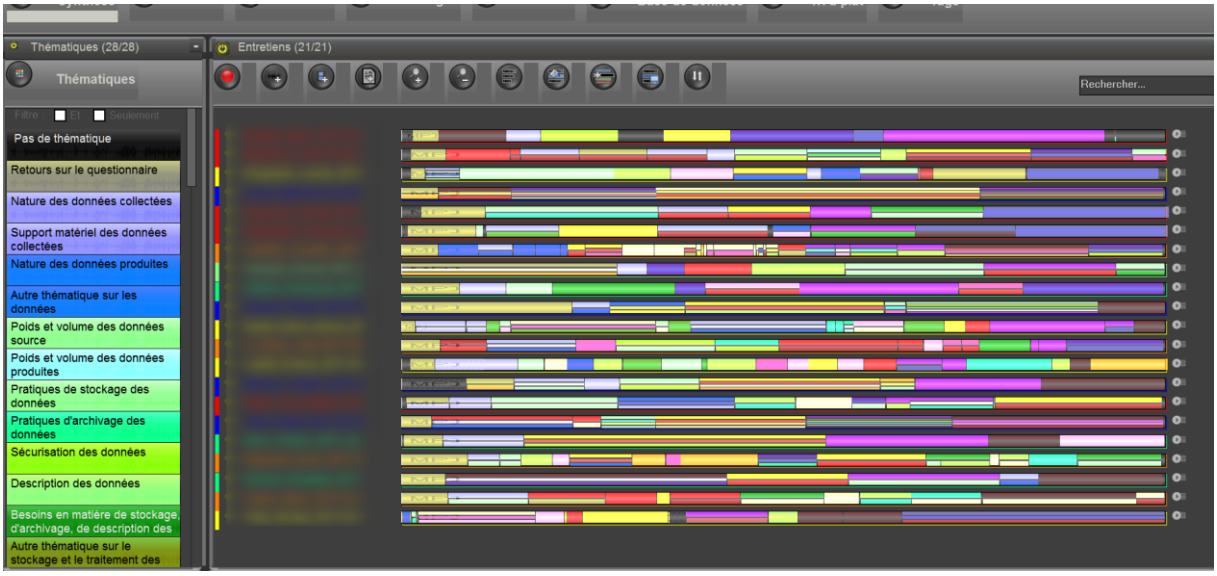


Figure 4 : Affichage des entretiens et de leurs thématiques sur Sonal

2.3 Synthèse sur la méthodologie de l'enquête

Quelques leçons méthodologiques sont à tirer de cette enquête, dont le déroulement a permis de valider différentes procédures, mais a mis également en évidence certaines difficultés et limites de sa mise en œuvre.

2.3.1 L'importance des tests en amont

Les tests réalisés avant le lancement de l'enquête par un panel de chercheurs (cf. 2.1.1) ont révélé toute leur utilité : ils ont suscité un certain nombre de remarques pertinentes, qui ont été transmises par mail et prises en considération après concertation, lorsque cela semblait possible et approprié. Les corrections effectuées sur la base de ces retours ont été de différentes natures. En voici quelques illustrations : affinements et précisions apportées sur les types de données pour toutes les questions qui s'y rapportent ; reformulation de questions, pour distinguer

[15] L'application a été créée par Alex Alber, sociologue, enseignant-chercheur à l'université F. Rabelais de Tours (laboratoire COST-CITERES, UMR6173) et chercheur associé au Centre d'Etudes de l'Emploi. Voir : http://sonal.hypotheses.org/ et http://www.sonal-info.com/.

davantage les questions 9 et 10 sur les données résultats ou pour préciser la question 21 sur l'accès aux données ; ajout d'options pour les réponses, par exemple sur la question 23 relative au stockage des données en libre accès, où trois nouvelles possibilités de réponse ont été ajoutées (« *non car je ne sais pas le faire* », « *non, car je n'ai pas le temps* », « *non car je pense que cette tâche nécessite le soutien d'un personnel qualifié* ») ; changement des modalités de réponse aux questions 25 et 26 (sur les facteurs d'incitation ou les freins au libre accès), initialement obligatoires et rendues facultatives à la demande d'un enseignant-chercheur. Ces quelques exemples montrent bien l'apport qu'a pu constituer la consultation de chercheurs dans l'élaboration finale du questionnaire : détection des imprécisions, ambiguïtés, lacunes, erreurs et énoncés impropres. Enfin cette démarche de concertation a produit un autre bénéfice, qui n'était pas attendu, c'est de conduire les réalisateurs de l'enquête eux-mêmes à mieux comprendre la complexité des problématiques inhérentes à la gestion des données de recherche.

2.3.2 La perception du questionnaire par les chercheurs

Les entretiens qualitatifs ont constitué un atout à la fois pour prolonger le questionnaire, pour l'évaluer et le cas échéant l'amender en direct, car ils débutaient en interrogeant les chercheurs sur la structure, le contenu du questionnaire et sur les difficultés qu'ils avaient éventuellement rencontrées en le remplissant. Ces retours ont été très utiles pour une évaluation réflexive de la méthodologie employée. Ils ont montré un consensus sur l'utilité et la pertinence du questionnaire, un avis positif largement majoritaire sur la facilité à le remplir. Parallèlement, l'approche critique de certains aspects méthodologiques de l'enquête a, elle aussi, été particulièrement riche d'enseignements.

2.3.2.1 La compréhension des questions

Si les chercheurs travaillent avec des données, la terminologie née des questions relatives à leur gestion, en particulier en contexte numérique, ne leur est pas pour autant toujours familière[16] et les réalités recouvertes par les notions de libre accès ou de plan de gestion de données peuvent être diversement comprises. Les tests effectués avant validation du questionnaire en ligne avaient déjà détecté le caractère insuffisamment explicite de certaines questions et cherché à y remédier ; c'est également une difficulté qui a parfois été évoquée lors des entretiens. Il a été indiqué, par exemple, qu'un glossaire, un lexique en fin de questionnaire auraient pu être utiles aux néophytes sur le sujet, soulignant par là la nécessité d'une pédagogie pour accompagner la démarche. Il ressort également qu'il aurait sans doute fallu expliquer plus précisément ce qui était entendu par donnée source et donnée résultat, ou que l'expression « données en libre accès » pouvait prêter à confusion entre données personnelles et données de recherche, ce que nous n'avions pas anticipé. Néanmoins, la majorité des interlocuteurs en entretien a jugé les questions claires et non ambigües ; et la difficulté à répondre rencontrée par certains chercheurs peut être aussi vue paradoxalement comme un bénéfice de l'enquête en termes de sensibilisation, lorsque cette difficulté provient de ce que le chercheur se pose pour la première fois la question à laquelle il doit répondre[17] et qu'il découvre tout l'intérêt de poser une réflexion sur cette question.

[16] En témoigne cette remarque : « Les questions étaient claires, mais elles étaient difficiles, destinées à un public déjà averti de beaucoup de choses. »

[17] Avec des réactions du type : « *Il y a toujours des questions qu'on ne s'est jamais posées, auxquelles on essaie de répondre* » ou « *ce sont des questions sur lesquelles on ne s'est pas vraiment interrogé* », « *on commence à se poser la question.* »

2.3.2.2 Les limites d'un questionnaire directif

Une limite méthodologique révélée par les entretiens concerne les possibilités offertes par les listes de réponses dans les questions à choix multiples ou à choix unique, ainsi que le caractère obligatoire ou non des questions. Les premières réactions au questionnaire émanant du panel de chercheurs testeurs avaient déjà conduit à revoir et compléter les options possibles pour plusieurs questions, cependant les ajustements faits avant le lancement de l'enquête n'ont pas levé toutes les difficultés. Et certains des chercheurs interrogés lors des entretiens sont revenus sur ces aspects. C'est particulièrement patent pour deux types de questions : les questions ayant trait à la typologie des données (sources ou résultats) et les questions relatives au positionnement à l'égard du libre accès aux données.

Les problèmes relevés sont de quatre ordres :

- le chercheur ne se retrouve pas dans les options proposées et ne dispose pas d'une option « Autre » (par exemple question 25 sur le partage des données, cette question a pour cette raison été rendue facultative)[18] ;
- la liste est à choix unique, alors que plusieurs réponses correspondent à la question[19] ;
- la catégorisation proposée n'est pas totalement adaptée à la configuration dans laquelle se trouve le chercheur[20] , ce qui dans certains cas a pu conduire le chercheur à abandonner le questionnaire (sur les 34 rendus incomplets, on notera que 7 chercheurs se sont arrêtés à la question 6 sur la typologie des données sources) ;
- enfin, la structure inévitablement schématique du questionnaire ne permet pas assez de nuances : par exemple, il n'était pas possible de faire varier la réponse selon la nature et les modalités de la recherche (éléments publics ou confidentiels, ce qui change les conditions du partage)[21] ou encore selon les phases de la recherche (impliquant des pratiques et des besoins différents).[22]

Dans tous les cas, si la question est de surcroît obligatoire, l'enquêté, qui hésite à répondre, soit est tenté de passer la question, mais se trouve bloqué et laisse le questionnaire incomplet, soit choisit une réponse qui ne reflètera pas sa situation réelle et faussera de fait les résultats. Le caractère potentiellement dissuasif des questions obligatoires s'est parfois révélé de manière très nette, et sur les 34 résultats de questionnaires incomplets, 16 chercheurs ont buté sur des questions obligatoires auxquelles ils n'ont pas pu ou voulu répondre.

Ces limites sont sans doute inhérentes à toute enquête directive, comme cela a été évoqué lors d'un entretien à propos de la question 25 sur les raisons incitant au partage : « *C'est un des problèmes classiques des enquêtes directives où tu dois choisir parmi des réponses ; les gens qui proposent des réponses peuvent difficilement anticiper, sauf avec une autre méthode, mais ce n'est pas la vôtre, la totalité des réponses que les gens vont vouloir donner, et vient un moment où il n'y a aucune raison qui te convient.* »

---

[18] « *Aucune des options ne correspondait* », « *en première analyse, j'ai l'impression qu'aucune de ces raisons en fait, pour reprendre le libellé de la question, ne m'inciterait à rendre mes données de recherche accessibles*. »

[19] « *Parfois on était entre deux réponses, mais ça, c'est le problème de tous les questionnaires* », « *quand on peut cocher plusieurs choses, en général, ça va, et puis il y a 'Autre'* », « *quand c'est ouvert, ça ne pose pas de problème* »

[20] « *Le fait que les données soient assignées à telle ou telle discipline, par exemple, les données expérimentales, c'est en médecine, alors que ça pourrait très bien être en géographie* », « *On peut discuter sur la classification que vous faites des formats de données, mais on y retrouve nos petits globalement.* »,

[21] « *Il y a différents types de projets de recherche, donc forcément, avoir une réponse homogène, c'était compliqué en fait.* », « *il aurait fallu que j'arrive à mieux nuancer* »

[22] « *En fonction de l'étape à laquelle on est, si on est plutôt dans la collecte, ou dans le traitement ou dans la valorisation, on n'a pas forcément les mêmes espaces de stockage, ou les mêmes besoins*. »

2.3.3 L'intérêt du couplage quantitatif / qualitatif

Le choix d'un sondage directif a en réalité été dicté par l'idée que, compte tenu de la nouveauté de la thématique, une démarche guidée pour un questionnement sur la gestion des données de recherche serait plus efficace. Il semblait difficile de partir d'emblée sur un sondage libre. Le cadrage précis et la précaution prise de réaliser des tests ont permis de traiter de façon sûre un ensemble d'éléments conduisant à dresser un état des lieux relativement complet des pratiques et des représentations. Les options « Autre » et les champs de réponse libre ont été introduits pour pallier la relative rigidité du dispositif et ouvrir les questions et les réponses[23], les entretiens qualitatifs étant à même, par ailleurs, de corriger les limites et biais découlant des aspects plus directifs de l'enquête[24], en présentant des propos développés, détaillés et nuancés.

[23] « *Il y en a pas mal où vous avez proposé une réponse personnelle, Autre, etc. Du coup, ça laisse la souplesse nécessaire*. »

[24] « *À partir du moment où les enquêtes directives sont produites par les personnes qui font la recherche, elles proposent à la confirmation ou à l'infirmation leurs propres réponses*. »

3. Présentation des résultats

La présentation des résultats de l'enquête suit une double logique :
- d'une part, le fidèle suivi de l'ordre du questionnaire de l'enquête statistique, avec une division en cinq parties : les données socio-démographiques sur les répondants, les questions sur la nature des données collectées et produites, les pratiques de stockage et d'archivage des données, les pratiques de partage et les besoins exprimés ;
- d'autre part, dans chacune de ces cinq parties, la distinction des résultats des deux modalités d'enquête : d'abord les résultats statistiques, puis ceux des entretiens, avec en conclusion une synthèse des deux.

Pour accompagner et approfondir cette présentation, on pourra consulter trois documents d'annexes différents :
- l'Annexe 1 sur les résultats statistiques de l'enquête ;
- l'Annexe 2 sur les croisements statistiques ;
- l'Annexe 3 sur les extraits des entretiens.

3.1 Observations sur les données socio-démographiques

3.1.1 Les données socio-démographiques de l'enquête statistique

La première partie du questionnaire s'attache à cerner le profil des répondants selon des données factuelles : la tranche d'âge, le statut, les responsabilités actuelles, le laboratoire d'appartenance, la section CNU concernée.

3.1.1.1 L'âge des répondants

Compte tenu de l'âge moyen du recrutement des MCF, il n'est pas étonnant que l'écrasante majorité des répondants ait plus de 30 ans (98 %)[25], la tranche des 30-50 ans est la plus nombreuse (88, soit 63 %), suivie par celle des plus de 50 ans (50, soit 35 %) (cf Figure 5). Ce dernier chiffre, si l'on tient compte de la pyramide des âges des enseignants, où les plus de 50 ans ne représentent qu'un tiers des effectifs globaux[26], prend tout son relief et conforte une hypothèse pressentie, à savoir qu'on ne peut pas corréler trop schématiquement l'intérêt que suscite la thématique récente de l'enquête et l'âge des participants.

[25] L'âge moyen du recrutement des MCF (le premier corps des enseignants-chercheurs) en Lettres-sciences humaines est de 35 ans et 11 mois. MESR, DGRH n°7, septembre 2016 [consulté le 10/04/17], disponible sur : http://cache.media.enseignementsup-recherche.gouv.fr/file/statistiques/61/1/Note_DGRH_n7_Septembre_2016_654611.pdf

[26] Bilan social 2015 (version longue), Pyramide des âges : enseignants, p. 28. Disponible sur : https://www.univ-rennes2.fr/system/files/UHB/DRH/bilan_social/bilan_social_2015_vdef.pdf

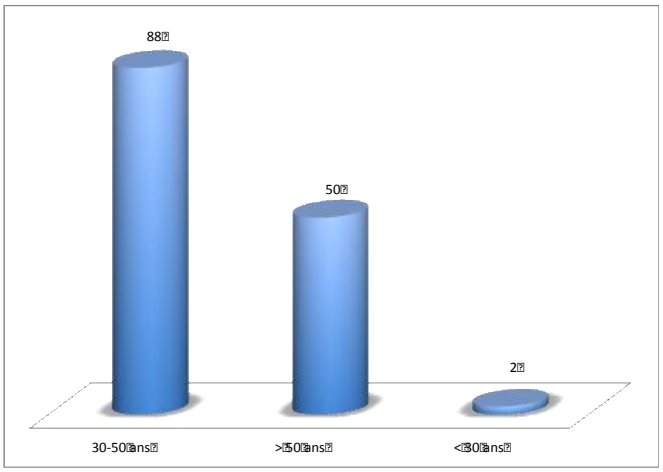


Figure 5 : Répartition des répondants par tranche d'âge (N=140)

3.1.1.2 Les statuts

Le nombre de répondants par statut suit mécaniquement la proportion des effectifs de Rennes 2 : 80 maîtres de conférence, 43 professeurs, 12 HDR, 3 chercheurs, 2 ingénieurs (Figure 6). Mais, à part pour les chercheurs, il est intéressant de noter que le taux de participation augmente à mesure que le statut s'élève et que les effectifs se réduisent. Il est « seulement » de 23 % pour les MCF, pourtant les plus nombreux (365), il monte à 35 % pour les professeurs et atteint 50 % des HDR et 67 % des ingénieurs[27]. Ce n'est pas la taille de l'effectif de départ qui influe le plus sur le taux de participation et le statut du répondant est un élément à prendre en compte tout comme les responsabilités.

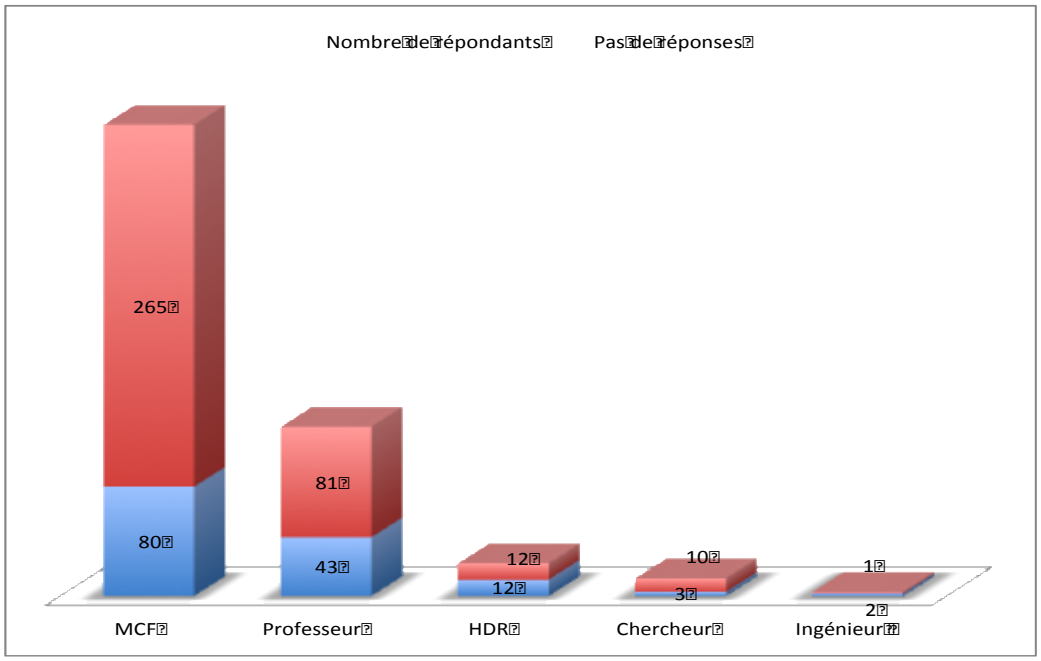


Figure 6 : Taux de participation des répondants par statut

[27] Il faut souligner ici le petit effectif considéré des ingénieurs (3), qui peut constituer un biais.

3.1.1.3 Les responsabilités

Le profil des répondants selon les responsabilités exercées peut nous permettre d'envisager les différents angles sous lesquels les enseignants-chercheurs sont susceptibles d'aborder la problématique des données. L'attention devra porter ici sur les multiples « casquettes » qu'un même agent est en situation de cumuler en gardant à l'esprit que la catégorisation des « responsabilités » pour les besoins du questionnaire ne semble pas couvrir toutes les situations et qu'elles sont en mesure d'évoluer tout au long d'une carrière. C'est peut-être le sens qu'il faut donner à la catégorie « aucune des responsabilités mentionnées » dont le flou permet de récolter un nombre de réponses non négligeable pour les moins de 50 ans (18) et très réduit pour les plus de 50 ans (2). La répartition des répondants par types de responsabilités exercées donne le classement suivant par ordre décroissant : encadrement de thèse (77), éducatives (64), scientifiques (58), éditoriales (51), politiques (36), administratives (14) (Figure 7).

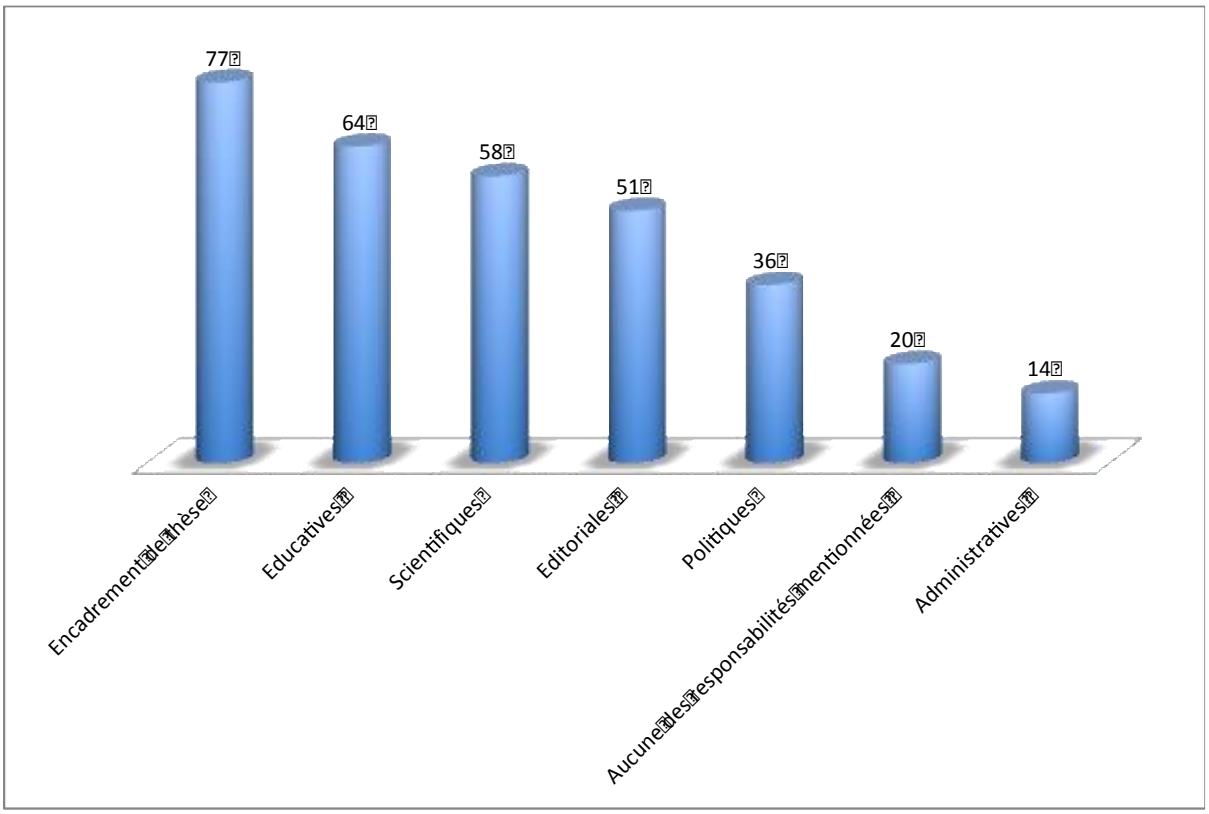


Figure 7 : Répartition des répondants par types de responsabilités exercées (N=139)

Le croisement des responsabilités avec les tranches d'âges donne une nouvelle répartition sans bouleversement. Dans la tranche d'âge des « 30-50 ans », la hiérarchie reste quasiment inchangée, sauf pour les deux premiers items (encadrement de thèses et éducatives) qui intervertissent leur place (Figure 8). Les responsabilités d'encadrement de thèses, éditoriales, politiques et administratives demeurent stables en quantité d'une tranche à l'autre, alors que les effectifs se réduisent pour les plus de 50 ans (les responsabilités ont naturellement tendance à s'accumuler au cours de la carrière). Dans cette tranche, le classement reste donc sensiblement le même à l'exception notable des responsabilités éducatives qui se retrouvent au 4ème rang en divisant ses effectifs par deux (respectivement 43 et 21). À noter que la responsabilité scientifique enregistre une baisse modérée de 10 % dans la tranche des plus de 50 ans.

Responsabilités	*< 30 ans*	*30-50 ans*	*> 50 ans*
Éducatives		43	21
Encadrement de thèses		39	38
Scientifiques		32	26
Éditoriales		27	24
Politiques		19	17
Autres	2	16	2
Administratives		7	7

Figure 8 : Croisement des responsabilités et des tranches d'âge (N = 139)

Par ailleurs, une majorité de répondants se situant dans la tranche d'âge n'excédant pas 50 ans n'a qu'une seule responsabilité (40), ce qui est conforme au déroulement de la première phase d'une carrière. À partir du seuil de trois responsabilités, il y a un rééquilibrage dans les deux principales tranches d'âge[28] et il est intéressant de souligner que plus de 40 % des participants (58) ont de 3 à 6 responsabilités, la multiplicité des points de vue de ce type de profil sur la problématique des données doit faire l'objet d'une attention particulière (Figure 9).

Nombre de responsabilités	*< 30 ans*	*30-50 ans*	*> 50 ans*
1	2	38	11
2		19	11
3		16	16
4		11	7
5		3	4
6			1

Figure 9 : Nombre de responsabilités et tranche d'âge

3.1.1.4 Les laboratoires d'appartenance

On trouvera en Annexes une liste complète des 19 unités de recherche figurant alors à l'Université Rennes 2 (une vingtième a été créée depuis).
Bien que tous représentés, les laboratoires d'appartenance sont répartis de manière très déséquilibrée parmi les répondants : l'axe médian qui sépare en deux le nombre de participants (67 sur un total de 134) est situé entre le 5ème et 6ème laboratoire du classement, qui en compte en tout 19. Il faut noter que ce classement rapporté aux effectifs de ces unités de recherche reste quasi inchangé (les écarts de classement n'excèdent jamais 2 rangs), sauf pour un seul laboratoire, le LIDILE qui passe du 7ème rang au 2ème. Certaines sous-représentations interrogent par rapport aux axes de recherche de ces unités : quelles peuvent être les raisons du faible écho (en apparence) de cette thématique sur ces structures de recherche ? L'enquête n'a pas pu le déterminer.

[28] L'écart d'âge important des tranches choisies pour l'enquête (ex. 20 ans pour celle des 30-50 ans) peut expliquer l'apparent paradoxe avec l'analyse précédente.

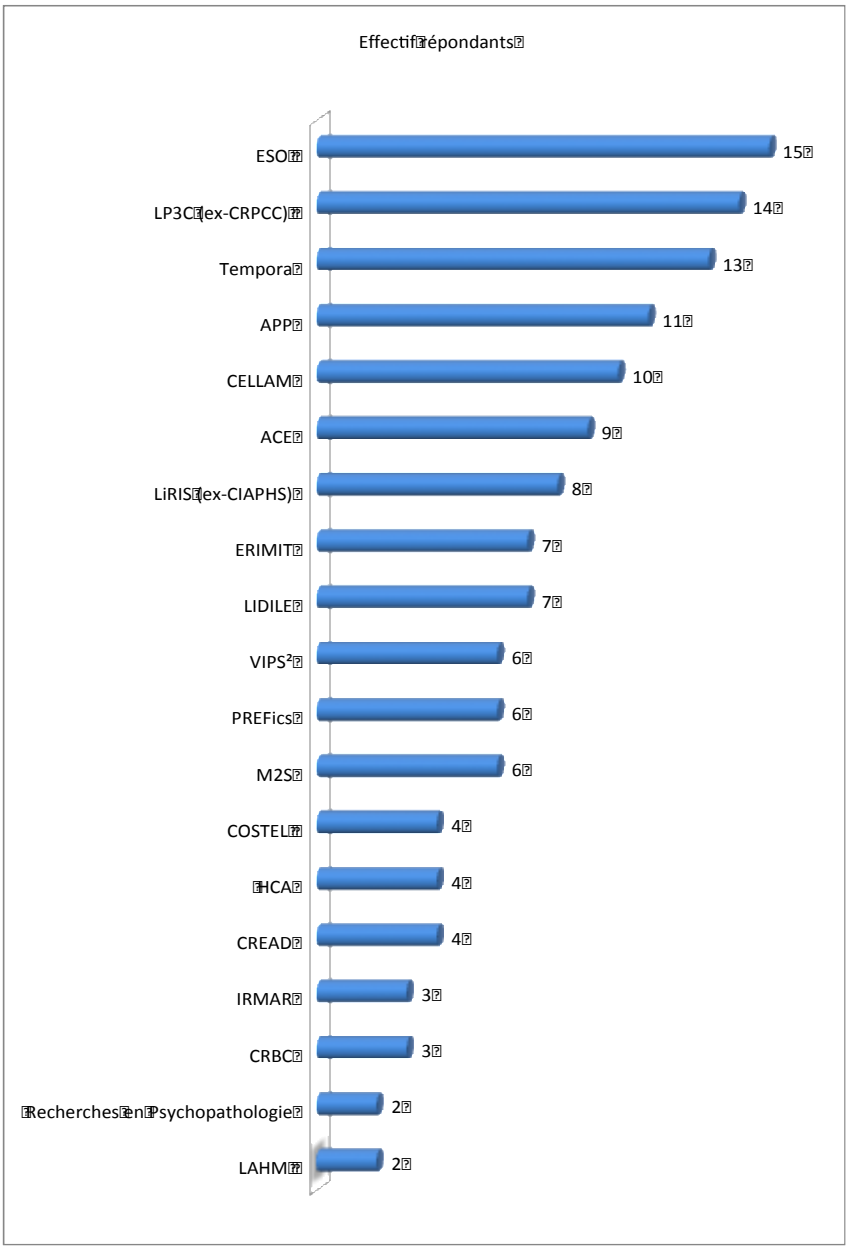


Figure 10 : Nombre de répondants par unité de recherche (N=134)

Enfin, on observe que parmi les laboratoires les mieux représentés par ce classement, 2 possèdent également la plus grande variété de personnels répondants : 5 (le maximum) pour le laboratoire ESO (qui est le seul à posséder 2 ingénieurs) et 4 pour Tempora (dont 2 chercheurs et 2 HDR), ils sont également les seuls à comprendre à la fois un personnel HDR et chercheur (Figure 11). Deux autres laboratoires (LP3C et LiRIS), dont les répondants représentent 3 statuts différents, figurent dans la première moitié de ce classement. Ainsi, parmi les laboratoires d'appartenance des participants, on constate une meilleure représentativité lorsque le statut des personnels est diversifié. Par ailleurs, il s'agit logiquement des unités de recherche qui possèdent les plus gros effectifs.

Unité de recherche	*Chercheur*	*HDR*	*Ingénieur*	*MCF*	*Professeur*	*Total*
ACE				7	2	***9***
APP				5	6	***11***
CELLAM				5	5	***10***
COSTEL				2	2	***4***
CRBC				1	2	***3***
CREAD				2	2	***4***
ERIMIT				5	2	***7***
ESO	1	1	2	7	4	***15***
HCA				2	2	***4***
IRMAR				2	1	***3***
LAHM				2		***2***
LIDILE				6	1	***7***
LiRIS (ex-CIAPHS)		2		5	1	***8***
LP3C (ex-CRPCC)		1		11	2	***14***
M2S		2		2	2	***6***
PREFics				3	3	***6***
Recherches en Psychopathologie					2	***2***
Tempora	2	2		6	3	***13***
VIPS²		2		3	1	***6***
Total	***3***	***10***	***2***	***76***	***43***	***134***

Figure 11 : Répartition des répondants par statut et par unité de recherche

3.1.1.5 Les sections CNU

La majeure partie des répondants n'est rattachée qu'à une seule section CNU (87 %), seuls 10 % possèdent 2 qualifications. La représentativité des répondants par rapport aux sections CNU illustre la disparité des situations à Rennes 2 : on retrouve bien les disciplines liées aux thématiques des laboratoires évoqués précédemment : psychologie (17 répondants), Histoire (15 répondants), et STAPS (14 répondants) par exemple. Mais la sous-représentation de certaines d'entre elles doit faire l'objet d'une analyse particulière : les langues et la littérature, les sciences de l'éducation par exemple (Figure 12).

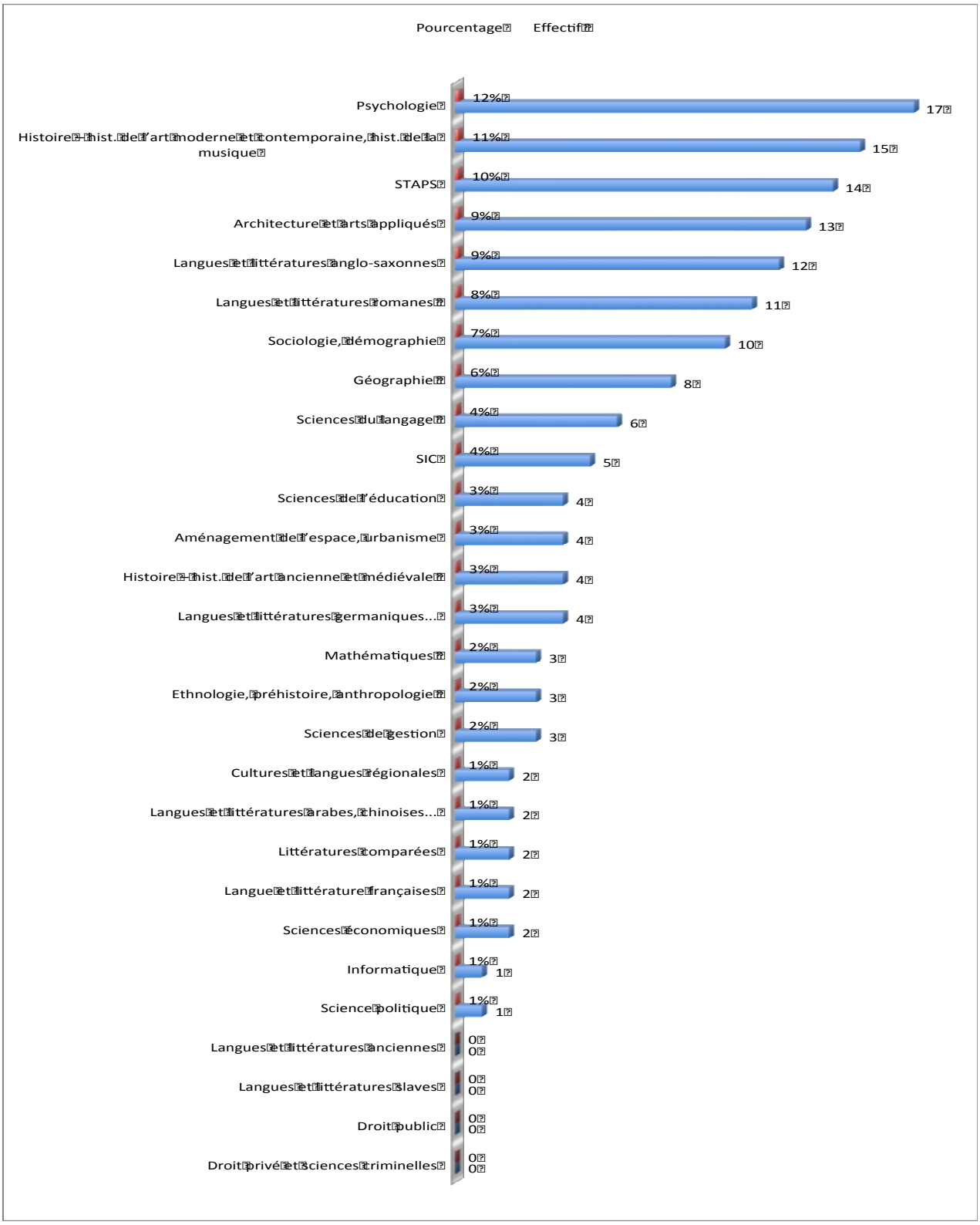


Figure 12 : Participation selon les disciplines CNU (N=138)

3.1.2 Données sur les entretiens

La fin du questionnaire invitait les répondants à prolonger l'enquête sous forme d'un entretien individuel. Cette seconde phase de l'enquête a permis de réaliser 21 rencontres et de recueillir près de 12 heures d'entretiens. Parmi les participants, 13 sont compris dans la tranche d'âge des 30-50 ans et 8 ont plus de 50 ans (une proportion analogue à celle du questionnaire). Cette démarche individuelle ne pouvait prétendre à un échantillon représentatif de l'ensemble des unités de recherche ; cependant les participants couvrent 13 des 19 laboratoires existants à Rennes 2. Les déséquilibres observés suivent logiquement ceux constatés lors du questionnaire : les unités absentes (6) sont principalement celles qui enregistrent le plus faible taux de réponse au questionnaire, alors que les laboratoires les mieux représentés correspondent aux effectifs de répondants les plus nombreux et où l'on trouve la plus grande diversité de statuts (4 membres de Tempora, 3 d'ESO, par exemple). À la différence du questionnaire, les maîtres de conférence et les professeurs sont représentés à parts égales (9), on compte également 2 HDR et un ingénieur. Les responsabilités[29] éducatives et scientifiques sont les mieux représentées (respectivement 7 et 6 répondants), suivent loin derrière les responsabilités politiques et administratives qui, avec l'item « sans catégorie », totalisent chacun deux représentants. À noter la faible part des responsabilités d'encadrement de thèses et éditoriales (un seul représentant chacun) au sein de ce groupe de répondants.

[29] Il s'agit des responsabilités mises en avant par les participants et extraites du logiciel Sonal qui ne permet pas de tenir compte des éventuels cumuls.

3.2 Quelles données de recherche ?

3.2.1 Les résultats de l'enquête statistique

Les questions 6 à 11 sur la typologie des données s'inspirent très largement de l'excellente enquête réalisée en 2015 par Hélène Prost et Joachim Schöpfel à l'université Lille 3[30]. Nous avons adopté la même logique (distinction entre données sources et données résultat) ; certaines typologies ont été reprises verbatim (notamment pour les questions 6 et 10), de façon à permettre facilement une comparaison des résultats avec cette enquête précédente. Il a paru intéressant de rajouter des questions sur le support matériel des données sources ainsi qu'une répartition par grandes catégories, la typologie intellectuelle des données se distinguant du type informatique.

3.2.1.1 Les données sources

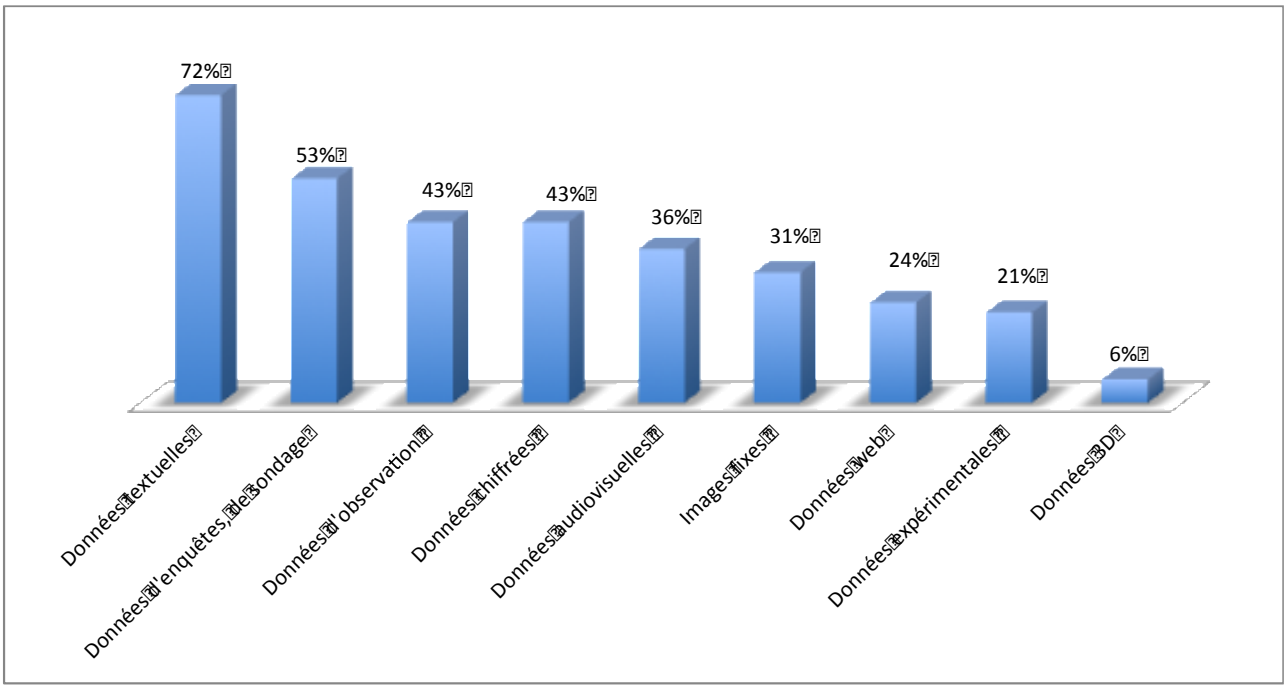


Figure 13 : Catégories des données sources (N = 127)

Quels types de données initiales sont collectées ou créées par les chercheurs pour servir de socle à leur recherche ? Sur les 127 réponses obtenues à la question 6 (catégories des données sources), ce qui frappe d'abord, c'est la diversité des données manipulées (plus de 20 % pour chaque catégorie, à part la 3D). Il s'agit principalement de données textuelles, mais les données d'observation ou d'enquête (non reproductibles) sont citées par presque la moitié des répondants. Le multimédia est également assez présent (1/3 travaillent sur des données image ou multimédia).

[30] Hélène Prost, Joachim Schöpfel. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. : Rapport final*. [Rapport de recherche] Lille 3. 2015. <hal-01198379>

On retrouve en trio de tête la même hiérarchie que dans l'enquête menée en 2015 à Lille 3 : textes, très largement en tête (72 % contre 64 % à Lille), enquêtes et entretiens (53 % contre 47 % à Lille) puis observations (43 % contre 41 % à Lille). La suite du classement montre des variations probablement liées aux spécificités des champs de recherche des équipes : importance des données chiffrées, audiovisuelles et image, et plus faible représentation des données expérimentales qu'à Lille.

La question permettant des réponses multiples, il est intéressant de constater que sur les réponses exprimées, 83 % des répondants manipulent plus d'un type de données, dont 42 % 4 types ou plus, ce qui conforte l'intuition précédente concernant la variété des données collectées.

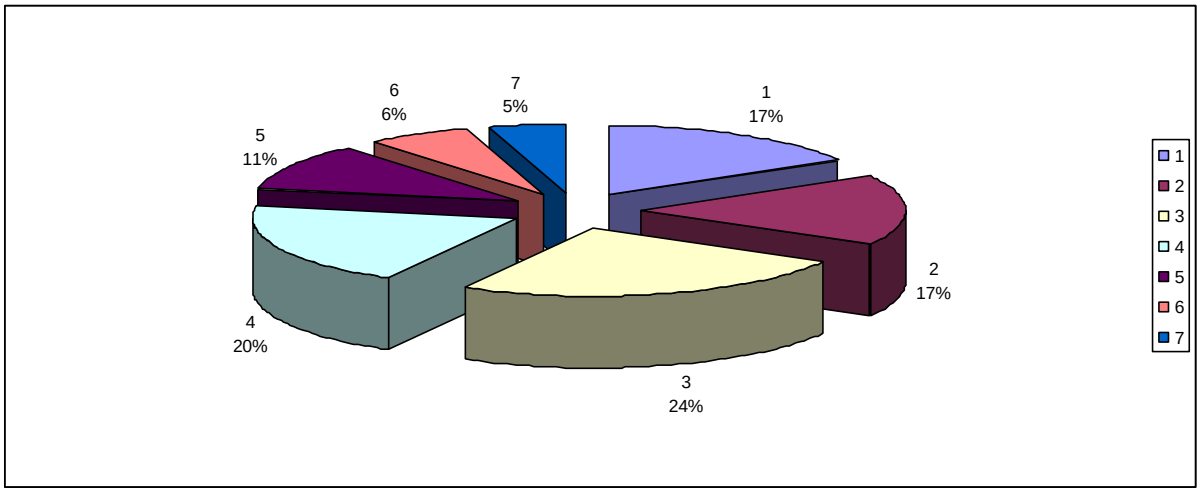


Figure 14 : Nombre de réponses par répondant sur les catégories de données sources

On constate un lien étroit entre certains types de données, citées ensemble dans les réponses : les données d'observation sont quasi-systématiquement associées aux données d'enquête et aux données textuelles. Les images fixes sont presque toujours corrélées à des données textuelles. Il y a aussi un lien fort entre données chiffrées et données d'enquête.

L'appartenance à un laboratoire ou à une discipline CNU joue-t-elle sur le type de données sources collectées ? Il faut tout d'abord noter qu'un même laboratoire de l'université Rennes 2 peut rassembler en son sein des chercheurs de disciplines différentes (géographes, sociologues, psychologues, ou encore linguistes, historiens et spécialistes de littérature par exemple).
Malgré cela, on observe une quasi-adéquation entre le nombre de répondants et le nombre de réponses maximal pour la catégorie privilégiée dans chaque laboratoire : *a minima*, les chercheurs manipulent au moins une catégorie de données identique (ainsi 11 répondants sur 11 ont du texte comme données sources principales au laboratoire Arts : pratiques et poétiques, 14 sur 15 des données d'enquête/sondage pour Espaces et Sociétés, 13/14 des données expérimentales pour le Laboratoire de Psychologie : Cognition, Comportement, Communication).

La hiérarchie des réponses globales à la question 6 n'est que partiellement respectée si on regarde plus finement les réponses par laboratoire[31] :

[31] Voir : Annexe 2, Croisements statistiques, partie 1.1.1.1 Synthèse 1 : Laboratoires et types de données sources, p. 5

- Les données textuelles qui arrivaient en tête globalement avec 72 % des réponses, sont en tête dans 9 laboratoires sur 19, principalement de langues ou Arts lettres communication. Les laboratoires d'arts et lettres ont comme source complémentaire l'image (fixe ou audiovisuelle). En langues, le texte est complété par l'image mais aussi par les enquêtes/sondages et les données web ;
- Les laboratoires de sciences sociales (à l'exception du laboratoire d'histoire Tempora) et de sciences humaines citent en premier d'autres formes de données que le texte : données d'enquête / sondage (CREAD, LIRIS, ESO, VIPS), données d'observation (LIRIS, M2S, Costel), données expérimentales (LP3C, Psychopathologie). Pour deux laboratoires (en STAPS et en psychopathologie), les répondants ne mentionnent d'ailleurs pas les données textuelles ;
- Les données chiffrées apparaissent également plus souvent dans le trio de tête en sciences humaines et sociales, notamment pour les unités de recherche de psychologie et de géographie. Les données expérimentales sont l'apanage des laboratoires de sciences humaines et de STAPS (74 % des réponses), les données d'observation comparativement plus fréquentes en sciences sociales (50 % des réponses) et en STAPS.

Toutefois, cette vision majoritaire d'une forte orientation disciplinaire recouvre un paysage assez nuancé, puisque les répondants de 13 laboratoires sur 19 ont cité au moins une fois 7 types de données sur les 9 que comportait l'enquête. Les autres laboratoires sont ceux qui ont le moins de répondants (dont 4 avec moins de 3 réponses).
Certains répondants en langue ou arts utilisent aussi des données chiffrées ou d'observation, les mathématiciens des données 3D… Le maître-mot est donc la diversité.

L'analyse du croisement des données sources et des catégories CNU recoupe les observations sur les laboratoires :

- prédominance du texte pour les langues, sciences du langage, arts, histoire et SIC ;
- des données expérimentales pour la psychologie ;
- des données d'observation pour les sections ethnologie, géographie, urbanisme et STAPS ;
- des données d'enquête et de sondage pour la sociologie et les sciences de l'éducation.

Avec là aussi de multiples types de données utilisés pour chaque discipline.

3.2.1.2 Support matériel

Si l'on s'intéresse au support matériel des données (question 7), il s'avère que les données nativement numériques sont encore très peu importantes en quantité (10 % seulement).
La création ou la collecte de données numériques ou numérisées représente 55 % des données sources manipulées. Il reste encore un peu moins de la moitié des données initialement en format non numérique.

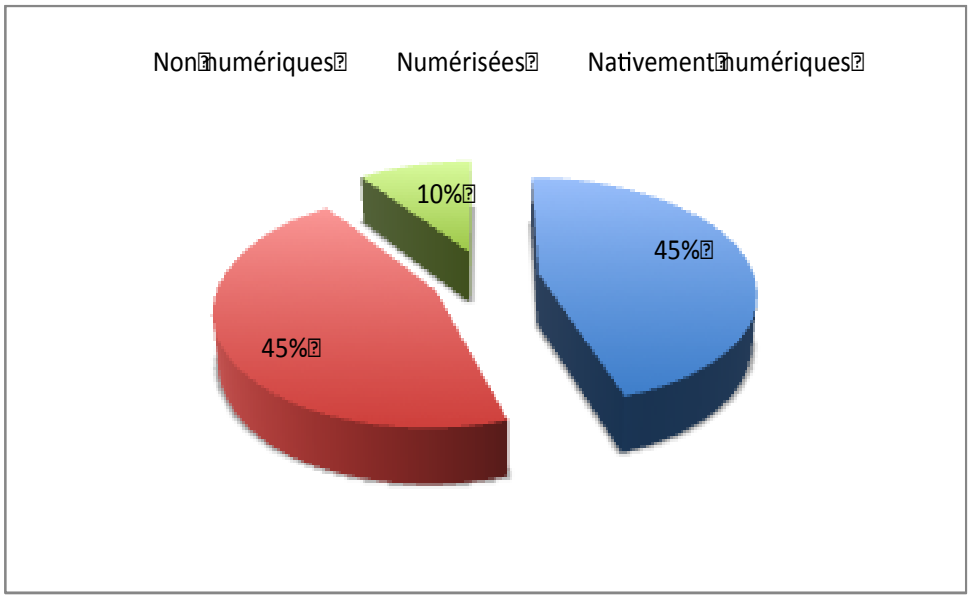


Figure 15 : Support matériel des données sources (N = 127)

Pour les données indiquées précédemment comme non numériques, 44 % des répondants numérisent *a posteriori* les données non numériques récoltées.

Oui	Non	Sans réponse
25	21	11

Si l'on rapproche cette question de la question précédente, on aboutit donc à plus de 75 % de données numériques au final (soit nativement numériques, soit numérisées avant la collecte, soit numérisées par le chercheur après collecte), ce qui rend importante la question de leur gestion, leur préservation, leur archivage et leur partage.
On peut remarquer aussi le nombre assez important de « Sans réponse » (20 %) sur une question qui semble simple. Est-ce parce que la pratique des chercheurs varie suivant les cas et les projets, et qu'il leur est difficile de se prononcer sur un choix binaire ?

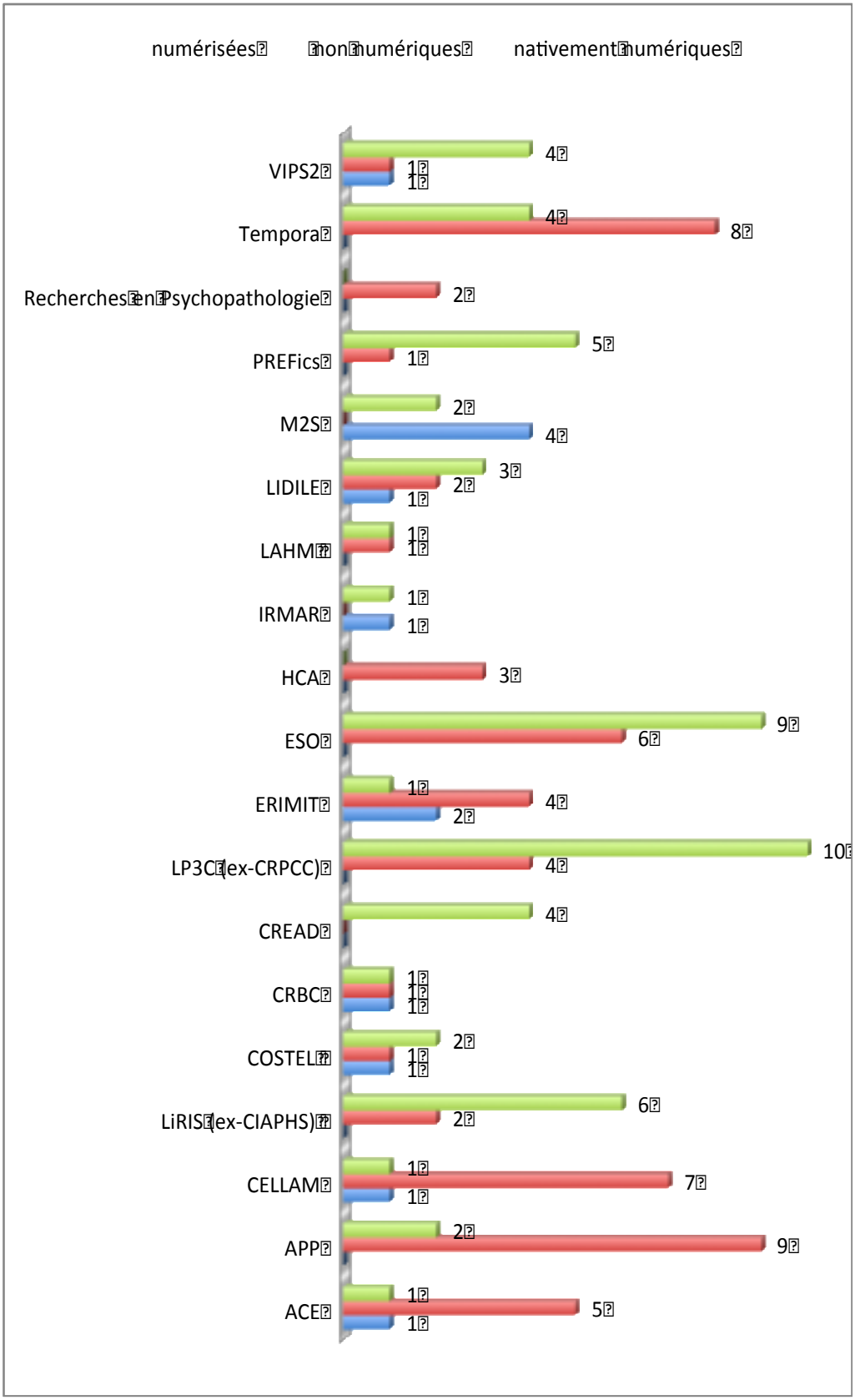


Figure 16 : Support matériel des données sources par laboratoire (N = 127)

Une analyse croisée des réponses par laboratoire confirme la tendance globale (majorité de données non-numériques) mais montre aussi des disparités entre les unités :

- les données nativement numériques sont minoritaires sauf pour un laboratoire de STAPS (M2S). Cependant presque la moitié des unités mentionnent ce support (9 sur 19, dans des domaines divers : STAPS, mathématiques, linguistique, langues et littérature, géographie). Il y a donc des usages diversifiés au sein d'un même laboratoire ;
- les données numérisées sont majoritaires pour 8 laboratoires : les deux laboratoires de géographie (Costel et Eso), deux laboratoires de psychologie (LP3C, LIRIS), le VIPS (STAPS), le CREAD (sciences de l'éducation), le PREFics (Infocom), le Lidile (linguistique) ; mais la quasi-totalité des unités (17 sur 19) citent ce support ;
- les données non numériques sont en tête pour 7 laboratoires : Tempora (histoire), la psychopathologie, HCA, APP (arts) Erimit et ACE (langues), le CELLAM (langue et littérature). Là aussi une large majorité (16 laboratoires sur 19) cite ce support.

Le paysage qui se dégage est donc celui d'une différence de pratiques entre SHS et Arts-Lettres-Langues, même si l'on observe au sein d'un même laboratoire des pratiques hétérogènes, et ce même lorsque le nombre de répondants est faible (LAHM, IRMAR, CRBC). Certaines tendances se dégagent nettement sur les laboratoires avec le plus de répondants.

L'analyse des croisements avec les disciplines CNU confirme cette hétérogénéité des pratiques. Sur les 23 disciplines représentées à Rennes 2, 19 mentionnent deux supports au moins (pour un total de 122 réponses, soit 89 % des réponses).
On retrouve la distinction entre SHS et Arts-Lettres. Par contre, au niveau plus fin par discipline, il est difficile de tirer des conclusions, car il y a peu de réponses dans certaines disciplines. Pour les disciplines comportant plus de 10 répondants, en dehors des arts et de l'histoire où les données non numérisées sont largement majoritaires, les répondants se répartissent de façon moins tranchée (la catégorie principale représentant au mieux la moitié des réponses). En sociologie et en STAPS, les données numérisées dominent, en psychologie, langues romanes et anglo-saxonnes, ce sont les données non numérisées. On peut noter aussi une différence en histoire suivant la période étudiée, les antiquisants et les médiévistes ayant davantage recours aux sources numérisées (du fait probablement de la rareté des sources)
Contrairement à ce qui est observé pour le type de données sources, le laboratoire et la discipline CNU ne semblent donc pas discriminants pour le support matériel des données collectées, même si une logique globale se dégage au niveau des grands champs disciplinaires.

3.2.1.3 Les données « résultat », produites

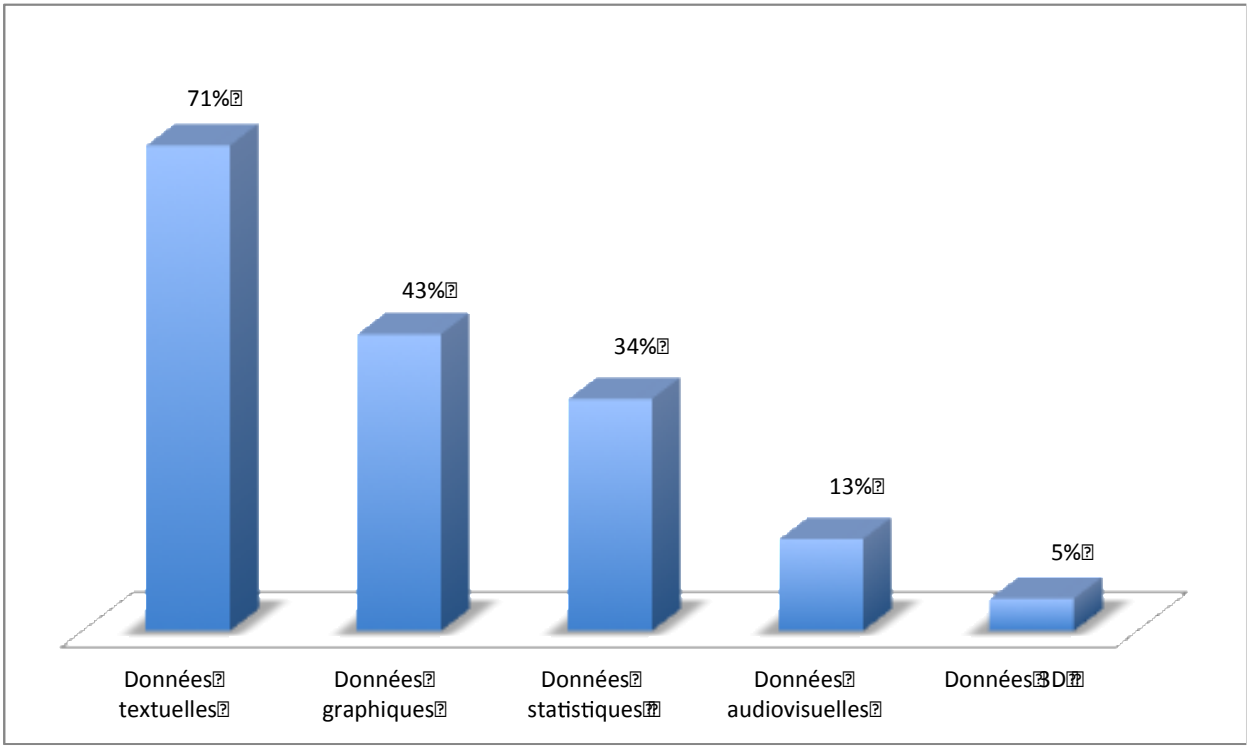


Figure 17 : Catégories de données produites (N = 127)

Comme pour les données sources, c'est le texte qui vient en premier. Cela n'est pas surprenant, d'autant que les entretiens ont permis de montrer un flou dans la définition du terme donnée, la publication scientifique étant englobée et considérée parfois comme la principale donnée produite.
Le nombre de réponses multiples est moins important que pour les données sources : si les chercheurs collectent une grande variété de données, leur production semble de prime abord plus homogène (49 % des répondants ayant répondu à la question n'ont coché qu'une catégorie). 2 chercheurs ne se sont pas reconnus dans les catégories proposées, mais ont coché uniquement la catégorie « autres ».

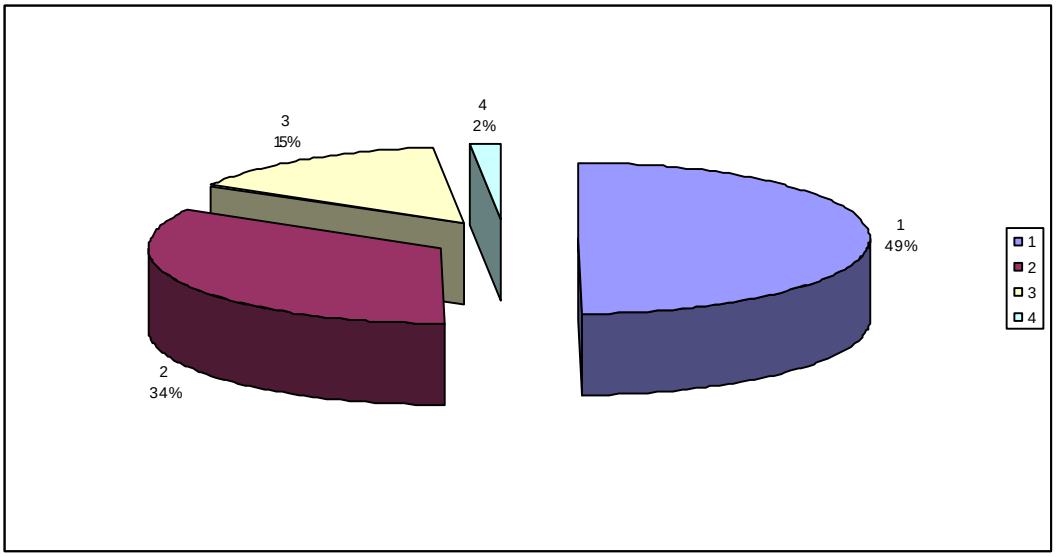


Figure 18 : Nombre de catégories de données produites indiquées par les répondants

L'analyse des données produites ventilées par laboratoire confirme d'une part la prédominance du texte pour une majorité d'entre eux (pour 13 unités sur 19), et d'autre part trace, comme pour les données sources, une différence entre champs disciplinaires.
Langues, arts et littératures produisent surtout du texte (autour de 90 % des répondants l'ont sélectionné dans ces laboratoires) et des données graphiques (notamment au PREFics) et audiovisuelles (APP…). Les sciences sociales respectent la hiérarchie globale (données textuelles, puis graphiques, puis statistiques). Les sciences humaines produisent plus de données statistiques.

Les données audiovisuelles, qui ne représentent que 13 % des réponses, sont réparties assez largement et de façon transversale aux champs disciplinaires, entre 11 unités (plus de la moitié des laboratoires). Par contre les données 3D restent l'exception (3 laboratoires seulement).

On voit encore un grand éclatement : 14 unités produisent au moins 3 types de résultats différents, et 7 en produisent 4 (APP, Erimit, CREAD, ESO, Tempora, M2S, VIPS).
On peut noter aussi des particularités pour certaines unités : la forte production de données statistiques et graphiques au LP3C (psychologie), l'importance de la 3D pour M2S (STAPS).
La ventilation par discipline CNU confirme ce constat de singularités identifiables : prédominance des données statistiques en psychologie et en mathématiques, et des données graphiques en géographie et en SIC. Pour la plupart des disciplines, on voit la prédominance d'une catégorie (plus de 50 % des réponses), avec éventuellement une seconde mineure. Trois disciplines font preuve d'une répartition moins tranchée (trois catégories se partageant entre 20 et 30 % des réponses) : les sciences de l'éducation, les SIC et les STAPS.

La question 10 permet de préciser la nature des données produites.
Les données textuelles sont là encore privilégiées pour plus de ¾ des répondants. Le 2e item semble différer au premier abord de la question précédente (données graphiques contre tableaux et bases de données en 2e et 3e position). Cependant c'est logique car le libellé associé aux données « graphiques » incluait les tableaux à côté des schémas et visualisations.
On peut remarquer également que 42 répondants produisent des bases de données et/ou des programmes et applications (soit un pourcentage cumulé de 42,5 % des réponses). Ces types de données complexes et longs à produire posent des questions particulières en termes d'archivage

et de partage (dépendance de l'environnement technique avec une obsolescence rapide, nécessité d'une réflexion sur les migrations et l'interopérabilité au moment de la conception même).

Enfin ces résultats recoupent ceux de l'enquête de Lille 3, à la notable exception des données de visualisation et 3D, beaucoup moins présentes à Rennes 2 qu'à Lille, consolidant une certaine homogénéité des pratiques en SHS à l'échelle macro, même si elles varient beaucoup au niveau micro d'une discipline et d'un laboratoire, voire d'un individu à l'autre.

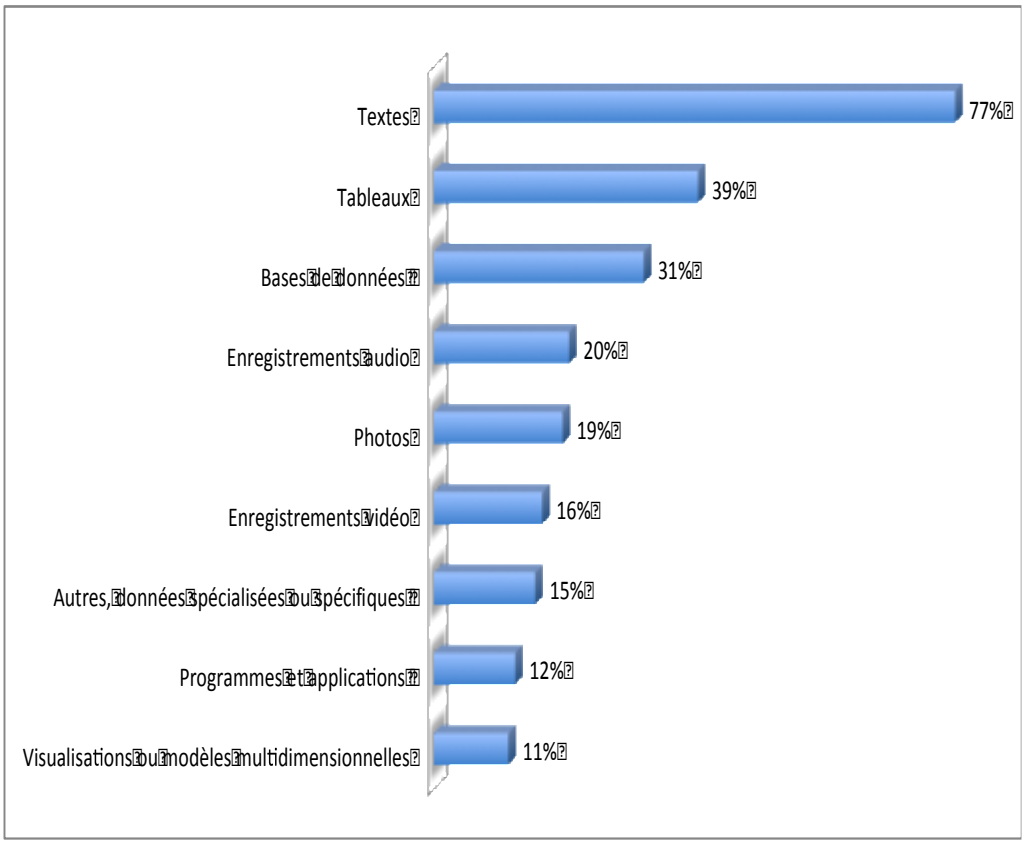


Figure 19 : Nature des données produites (N = 125)

Par rapport à la question 9, le nombre de réponses multiples est supérieur, nuançant la vision « monolithique » de la production des données : si les données appartiennent globalement à la même catégorie, leur nature n'est pas identique. L'enquête ne permet pas de savoir si cette différence est liée à l'interprétation du contenu des catégories proposées, au fait que la typologie plus détaillée proposée pour la nature des données suscite des réponses pour des productions plus ponctuelles omises dans la 1ere question (données les plus fréquentes citées uniquement), ou à d'autres facteurs. Certains types de données vont systématiquement de pair : quasiment toutes les réponses mentionnant des photos ou de l'audio mentionnent également des données texte ; la production de bases de données s'accompagne souvent de textes et/ou de tableaux, et celle de programmes s'appuie sur des tableaux ou des bases de données.

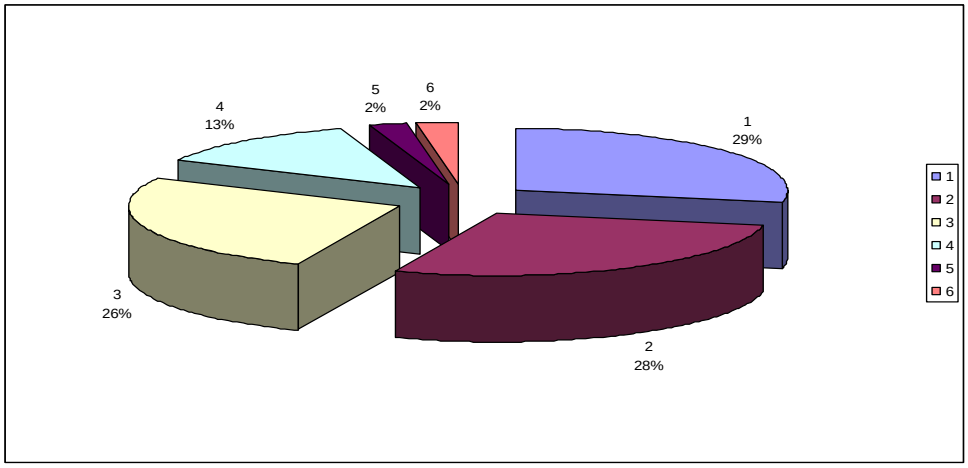


Figure 20 : Nombre de réponses par répondant sur la nature des données produites

Seules les 19 personnes ayant choisi « Autres, données spécialisées ou spécifiques » pouvaient répondre à la question 11, qui permettait de détailler la nature des données produites, et toutes ont répondu. Cette question semble avoir été utilisée la plupart du temps pour préciser les catégories plus vastes sélectionnées dans la question 9 (seuls 2 répondants sur 19 ne sélectionnant pas de réponse à la question 10). Les pourcentages étant peu significatifs sur ce faible échantillon, il semble préférable d'afficher les chiffres bruts. On y retrouve notamment des données géographiques (cartes et plans, géolocalisation, images satellites, télédétection), des pratiques sociologiques (questionnaires et transcription d'entretiens).

Nature des données	Nombre de réponses	Pourcentage
Analyses statistiques	9	47 %
Cartes et plans	7	37 %
Questionnaires	6	32 %
Corpus de texte / annotations	6	32 %
Transcriptions d'entretiens	5	26 %
Données de géolocalisation (SIG)	5	26 %
Séries de mesures	3	16 %
Cartographies de réseaux ou de corpus	3	16 %
Images satellite	1	5 %
Données de télédétection	1	5 %
Spectres		
Modélisations du climat		
Données de santé		

Figure 21 : Différentes natures de données produites

3.2.2 Les résultats des entretiens

Les entretiens en présentiel avec les chercheurs et responsables d'unité ont permis d'affiner et de compléter la vision quantitative de l'enquête en ligne par une approche plus qualitative. Et en premier lieu d'interroger la notion même de donnée, qui était au cœur du questionnaire mais dont les frontières s'avèrent difficiles à cerner.

3.2.2.1 Typologie et nature des données : une grande diversité de données collectées et produites

Les échanges confirment l'extrême variété des données, tant collectées que produites, qu'il s'agisse de leur nature ou des formats, et de la manière d'en parler, variété illustrée dans le petit inventaire à la Prévert qui suit. Les échanges révèlent aussi l'utilisation de nombreux logiciels et formats de données propriétaires, problématiques pour un archivage à long terme.

A/ Quelles sont les données collectées ?

« du matériau langagier [extrait d'] un terrain ethnographique [investi] sur la durée, sur la profondeur, avec une compréhension très contextualisée de la situation ; des conversations ; du témoignage ethnographique ; des entretiens ; des biographies langagières orales ou écrites ; des documents d'archive ; des documents audiovisuels ; des émissions radiophoniques ; des corpus de presse ; des programmes, les discours des inspecteurs généraux, des manuels ; des corpus de dossiers d'aide à l'enfance ; des corpus de textes de loi ; des échanges avec les artistes ; des interviews avec des collègues ; des relevés GPS ; des données d'enquête ; des images ; des textes ; des corpus de textes ; des corpus visuels ; des adresses ; des ouvrages ; des articles ; des facsimilés d'ouvrages ; des documents ; des brochures ; des photos ; des fragments de pierres, des inscriptions ; de la donnée géographique numérique ; des cartes papier ; des données de type raster, matriciel, tout ce qui est image, photographie aérienne, image satellite ; de la donnée vectorielle, avec une composante géométrique (la donnée c'est de l'objet) ; de la donnée attributaire (ce qui vient décrire l'objet ; des tables) ; des information informelles glanées lors de colloques ; des données récupérées sur Internet ; des SMS ; des enregistrements de webcams ; des vidéos (youtube ou autres plateformes) ; des pages de forums Internet, de discussion, sites web, ; des pages de tchat sos ; des pages ou échanges de réseaux sociaux, comptes twitter, Facebook, ; des PdF d'œuvres sur Internet ; des index et listes ; des photos aériennes. ; des données numériques, statistiques ; des cartes ; des données archéologiques ; des captures d'écran ; des traces GPS ; des trajectoires de marche ; des indicateurs liés à de petites tâches cognitives ; des données territorialisées… »

B/ Et les données produites ?

« des données textuelles ; des bases de données ; des enregistrements audio ou vidéo d'entretiens ; des notes sur des entretiens ; des questionnaires ; des formulaires papier remplis ; des tableaux Excel ; des fichiers Word ; plusieurs versions de textes ; des fiches à partir de documents d'archive ; des sites ou plateformes internet ; des retranscriptions d'entretiens ; des analyses statistiques lexicales ; des dossiers PDF, des textes, essentiellement des textes ; des livres ; des revues ; des données excel ; des données chiffrées ; des données qualitatives ; des études de cas cliniques ; des logiciels ; des tâches ; des petits programmes ; de petites épreuves cognitives ; des tests de raisonnement ; des fichiers texte avec reconnaissance de caractères, des transcriptions ; des documents numérisés ; des photographies de documents originaux ; des enregistrements vidéo ; des rushs ; des

documentaires ; des catalogues prosopographiques ; des arbres généalogiques ; de recensements d'archives ; des campagnes de photos ; des fiches thématiques ; des bases de données géographiques ; des traces GPS ; des indicateurs ; des données biologiques, des dosages ; des électrocardiogrames ; des consentement de participation (papier) ; des protocoles ; des fichiers Latex ; des fichiers Matlab ; des fichiers partagés ; des bibliographies ; des fichiers statistiques ; des nuages de points, des captures de mouvement ; des photographies aériennes ; des cartes ; des croquis ; des enregistrements sonores suivis sur plusieurs années ; des photographies suivies sur plusieurs années ; des corpus multilingues ; des carnets papier… »

Cette diversité s'impose aux chercheurs, ce qui a été explicité au cours de plusieurs entretiens et corrobore les résultats de l'enquête quantitative : *« La donnée que je manipule est diverse »*.

Quels facteurs explicatifs avancent-ils ? Les données peuvent changer selon les projets de recherche. La variable disciplinaire est plusieurs fois citée :
- *« Après, y'a des pratiques qui sont différentes selon les disciplines ; principalement en géographie et en sociologie ; et selon les disciplines, on mobilise des matériaux qui sont pas forcément les mêmes, et avec des niveaux d'appropriation des outils numériques variables d'un collègue à l'autre »*

Cependant, cette variable ne suffit pas à tout expliquer. À l'intérieur de la même discipline, la langue et la période étudiée et les écoles de pensée induisent le choix des sources, des méthodes et des données manipulées.
Inversement certaines communautés de pratiques peuvent transcender la discipline, comme l'explicite un chercheur en sciences du langage :
- *« Ce n'est pas tant une question de discipline, mais de paradigme de recherche avec des méthodes qui ne sont pas les mêmes ; nous ici, on partage des méthodes avec les gens […de] géographie sociale, […] ou certains sociologues ou ethnologues, etc. bien qu'on soit pas dans la même discipline ; mais il y a des grandes catégories de méthodes, c'est quelque chose de structurant ; et puis il y a aussi les cultures disciplinaires. »*

3.2.2.2 La donnée, une notion difficile à cerner dans le continuum de la recherche

Qu'est-ce qu'une donnée ? La difficulté des spécialistes à donner une définition de cette notion protéiforme fait écho à l'interrogation des producteurs eux-mêmes sur l'ambiguïté du terme et ce qu'il recouvre.
À côté des données de recherche, d'autres types peuvent coexister, et poser des questions de mise à disposition et d'archivage pour les laboratoires, ce qui est relevé par un directeur d'unité :
- *« Je distingue deux types de données : on a les données de recherche et les données de gestion et de pilotage [du labo]. Donc on a nos données de gestion et ensuite il y a la question des données de recherche, je crois que ce n'est pas le même statut. »*

Une fois ce distinguo posé, si l'on se concentre sur les données de la recherche, à partir de quand une information ou un matériau collectés deviennent-ils une donnée ? Certains identifient informations collectées, veille et données :
- *« Moi j'utilise Zotero, mais est-ce que ça fait partie des informations de la recherche, sachant que c'est plutôt des pdf, des textes, enfin si ce n'est pas des corpus que je constitue ?» ;*

Inversement, un chercheur souligne le caractère exceptionnel et délibéré de la collecte :
- *« Les données de recherche, on ne va pas en collecter tous les jours non plus »*.

Les termes utilisés pour en parler sont d'ailleurs divers : *« matériaux », « informations », « choses », « documents », « observables », « sources », « boîtes noires »*.
La distinction entre donnée brute et donnée produite n'est pas non plus aussi tranchée que le présente le questionnaire et que le voudrait un monde idéal. La donnée n'est pas figée dans le temps ou dans des états successifs bien identifiés, et cette catégorisation n'est pas forcément parlante pour les chercheurs :
- *« C'est difficile de savoir où commence le traitement. On a le fichier brut, on a le fichier traité, et après en sortie, on a des positions au cours du temps, des gros tableaux, c'est comme ça que ça se lit, et ensuite on fait nos calculs avec MATLAB »*.

Passer de « la donnée » abstraite, évoquée dans les questions, au travail sur des « données » concrètes manipulées dans un laboratoire impliquerait donc une série de choix et de renoncements.
De la même façon, la frontière entre donnée de travail, hypothèse et résultat est poreuse.
Une chercheuse évoque ses fichiers word préparatoires à la saisie dans une base de données, qui lui prennent beaucoup de temps à établir, comme des « *brouillons* » dont la sauvegarde ne semble pas la préoccuper. La donnée, pour elle, n'est pas le début mais le résultat final d'un travail, une hypothèse évolutive, ce qui est le cas aussi de l'édition de textes pour un historien, et pose la question de la mise à jour continue des publications.
- *« Si tout d'un coup je retrouve un fragment et je complète le nom... les dates de mes sources elles peuvent changer tout d'un coup vingt ans après, on décale tout de cinquante ans, on reprend, donc il faut inventer un système où on puisse gérer cette mobilité […] c'est toutes ces étapes de déconstruction en fait des données que j'essaye de rendre apparentes... »*
La publication, résultat des données, est-elle elle-même une donnée ? *« Et pour nous, où est la limite entre un article et les données ? »*. Cette problématique, soulevée notamment par des historiens et philosophes dont la principale production est les textes, se retrouve aussi dans la bouche d'un géographe, qu'on pourrait penser *a priori* plus enclin à distinguer données et publications : *« Comment situer les données, par rapport à ce processus d'élaboration dont je parlais un peu ? Est-ce que les données, c'est les données brutes, ou est-ce que les résultats on les appelle aussi données ? »*.

La difficulté à différencier la donnée du résultat a un impact sur le positionnement concernant le partage. Lorsque l'interviewer parle « partage de données », certains entendent et répondent partage des publications en Open Access : « *Moi je serais même assez partisan de mettre en ligne des choses raisonnées qui sont plus proches des résultats de recherche que des données elles-mêmes »*.

3.2.2.3 La donnée comme construction

Deux chercheurs en SIC et linguistique vont plus loin en commentant et remettant en cause le terme utilisé, qui induit en erreur. La donnée est un construit, un regard sur le monde, choix délibéré et fruit du travail du chercheur. Ce n'est pas un départ mais bien déjà un premier résultat.
- *« Le problème, on dit des données, mais en fait pour moi, c'est pas des données, c'est des construits. […] le mot donnée, concrètement ça m'embête […] ça veut dire que les choses sont données. Si vous avez construit tout un appareil méthodologique qui permet de prélever des éléments, et donc en aucun cas c'est donné, mais c'est construit, prélevé, et donc du coup ça pose problème déjà, parce que les données, on a l'impression que c'est quelque chose qui est donné, physique, factuel, que les choses sont comme ça, en fait, non »*
- *« Contextualiser, c'est sélectionner »*.

Le travail et le regard du chercheur donnent de la valeur à la donnée, parfois insignifiante autrement :
- *« Ma collecte de données, c'est justement ce dont je parlais […] c'est à dire tous les échanges avec les artistes, mais qui ne sont pas méthodiques, ce n'est pas des formes d'enquêtes, des formes d'interviews, c'est divers types d'échanges à diverses occasions qui peuvent avoir l'air de rien du tout mais qui sont des données sur lesquelles je m'appuie réellement ».*

Cette valeur est aussi mise en exergue par la conscience du long travail de construction que représentent la collecte et le traitement des données, soulignée dans de nombreux entretiens :
« J'ai passé 2 ans de ma vie à traiter, exploiter une base d'entretiens ».
Les termes de construction, fabrication sont d'ailleurs récurrents (33 occurrences), ainsi que les références au temps.
La pénibilité de ce travail est soulignée par le vocabulaire :
- *« Nos données qu'on s'est **embêtés** à produire » ;*
- *« Au fond, ça a **coûté** des sous, de l'énergie à collecter » ;*
ainsi que par les comparaisons religieuses auxquelles ont recours deux chercheurs.
- *« Rentrer des données prosopographiques c'est juste l'enfer[...] c'est un travail de bénédictin en fait » ;*
- *« ..., je travaille en fait assez "en moinesse", en moniale ».*

Le temps de collecte peut être aussi un temps d'élaboration, ce qui peut justifier le choix de certaines méthodes de recueil, et conduire à favoriser certains types de données (transcription manuelle, reprise de notes) plutôt que la collecte « brute » :
- *« Je garde les photos pour des documents exceptionnels ou alors pour des documents, comment dirais-je, de nature sérielle : des listes, des tableaux... Mais j'aime bien travailler quand même intellectuellement la source. »*
- *« Moi j'écris dans des carnets, des cahiers, des bouts de papier etc., ce qui fait que le numérique, ou alors c'est plutôt l'informatique, c'est un moment de reprise et de ressaisie et c'est pour ça que mes premières données, ce sont mes premiers documents de traitement de texte, ce sont des ressaisies de mes textes. ».*

3.2.2.4 Valeur de la donnée : une reconnaissance scientifique et affective

Ce travail de constitution des données représente un lourd investissement personnel, y compris au détriment de la vie privée. On sait, et plusieurs entretiens l'expriment, que la lourdeur des tâches administratives laisse peu de temps pour la recherche.
D'où la tentation avouée par un historien de constituer un pré carré de données, qui crée de fait un frein au partage.
- *« Voilà il y a un moment donné où on s'engourdit, et c'est bien légitime, et où on a envie de faire quelque chose d'un petit peu moins envahissant sur le plan personnel. Donc voilà, donc du coup quand on se constitue un terrain comme ça, c'est quelque chose de très précieux car c'est aussi la condition du maintien dans un niveau d'échange scientifique assez élevé »*

Cette permanence des données et du terrain de recherche ne prévaut pas dans toutes les disciplines, puisque d'autres chercheurs évoquent au contraire des données diverses suivant les projets de recherche.
- *« C'est très variable, je vais vous parler de plusieurs projets si vous voulez. »*

Cependant, quelle que soit la discipline, la question de la reconnaissance du travail accompli par la propriété intellectuelle revient de façon récurrente dans les échanges, de façon

transversale aux disciplines (infocom, histoire, géographie, linguistique, littérature…). Elle s'exprime comme une revendication positive, tout en donnant à entendre par les formulations employées une méconnaissance du statut légal qui s'applique aux données :
- *« Il faut qu'il y ait une forme de, non pas de droit d'auteur, mais au moins de propriété intellectuelle, ne serait-ce que morale ».*
Cette revendication s'exprime au niveau des chercheurs, mais aussi au niveau de l'équipe :
- *« Les gens sont inscrits dans une équipe, et pas seulement des agents de Rennes 2, ils sont membres d'une équipe et leurs données, leurs recherches font partie de l'équipe ».*

La recherche collective, qu'il s'agisse de la vie des équipes de recherche internes au laboratoire ou de projets collaboratifs entre équipes, complique d'ailleurs cette question des droits d'auteur, allant parfois jusqu'à bloquer l'exploitation des données, comme en témoigne un chercheur en STAPS :
- *« Quand vous travaillez à plusieurs sur un projet, ces données appartiennent à qui ? […] Pour l'instant on manque d'un cadre] pour savoir si systématiquement les gens qui ont récolté les données ont le même statut que ceux qui vont ensuite les analyser, les traiter et finalement produire une analyse scientifique » […] « on a été confronté au problème il y a quelque temps. En l'occurrence j'ai écrit un article que je n'ai jamais pu soumettre parce que je risquais de m'exposer à des problèmes, j'ai préféré laisser mes résultats de côté, c'est un peu dommage parce que c'était un projet chronophage, intéressant et scientifiquement porteur »*

Clairement, le droit d'auteur constitue un terrain sur lequel l'université pourrait mieux accompagner les chercheurs :
- *« Sur la propriété, je crois savoir à peu près comment ça se passe, et puis quand j'ai un doute, comme j'ai un fils qui est avocat je vérifie les choses ; et puis il y avait un stage à Rennes 2 sur la propriété intellectuelle ; je n'ai pas beaucoup de recours à l'université »*

Si la propriété intellectuelle est un enjeu identifié, c'est plus pour une question de droit moral que de valeur économique des données. Cette dernière question surgit seulement en linguistique :
- *« On sait qu'elles sont plébiscitées par des entreprises […], ça me dérangerait par exemple, que ce soit réexploité sans le dire ou sans bénéfice pour l'université Rennes 2. »*

Les données utilisées sont parfois payantes, et dans ce cas l'argument financier justifie une priorité d'exploitation pour deux géographes ; dans tous les cas les données représentent une richesse, un trésor à exploiter :
- *« C'est les photos sur lesquelles on va travailler, celles qui sont payantes, il y a une priorité pour la personne qui les a achetées, sur un contrat de recherche et ainsi de suite »*
- *« Tant qu'on n'a pas fini d'exploiter ces données-là, les collègues ont un peu de mal à les ouvrir à d'autres ; au fond, ça a coûté des sous, de l'énergie à collecter »*
- « *On ne va pas filer notre stock de données pour que d'autres les travaillent.* »

Le paradoxe souligné par certains, c'est que le travail de construction et de partage des données est peu reconnu institutionnellement :
- *« Notre métier n'est pas de créer des bases de données pour créer des bases de données, c'est de publier donc on produit les résultats, pas la boîte noire ; donc rendre publique la boîte noire, c'est pas une habitude, on n'est pas valorisé scientifiquement pour avoir passé du temps à publier nos boîtes noires. »*
La publication vient comme une « *légalisation* », une « *forme officielle* » de travaux qui ont demandé sacrifices et investissements :

- *« Mes écrits, chacun dit que ça vaut quand même, c'est mon enfant en quelque sorte ».*
Dès lors, la dimension affective et l'attachement au terrain de recherche ne doivent pas être négligés pour l'établissement d'une politique de données. Mais ce sentiment d'intimité et de propriété s'adresse souvent plus aux publications qu'aux données :
- *« J'avais jusqu'à maintenant une sorte de vision instrumentale et utilitaire de mes propres données, ce qui m'intéressait, c'était le résultat de mes données. Mais la donnée elle-même était une sorte de structure un peu neutre que j'interrogeais. Une sorte d'impensé, un peu » ;*

3.2.2.5 La perte, révélateur de la valeur

Paradoxalement, l'attachement aux données s'exprime davantage en creux, dans la perte, leitmotiv récurrent des entretiens, qu'elle soit réelle avec « des expériences douloureuses de perte des données » ou fantasmée. Quatre chercheurs évoquent des incidents matériels subis par eux-mêmes ou des connaissances :

- *« Dans les manipulations, une sauvegarde de l'ordinateur a écrasé ce qui était déposé sur le disque, donc c'était une perte vraiment considérable pour moi. C'est comme après un incendie, je repartais à zéro » ;*

Quatre autres soulèvent le spectre du vol :

- *« Si vous me volez mon ordi portable et mon disque dur externe, qui est là et celui qui est à la maison, j'aurai perdu 6 ans de recherches avec aucun moyen de récupérer ces données là. Donc c'est plus de ça qu'on a peur. » ;*

Cinq chercheurs enfin évoquent des données, ou même des outils, devenues inutilisables :

- *« Après j'ai eu des déboires d'enseignant-chercheur, […] qui légitiment une réflexion sur les pratiques. […] j'ai perdu toutes les données de ma thèse […] à l'époque […] j'avais enregistré et sauvegardé ça sur disquette. Et aujourd'hui, je n'ai plus les moyens de convertir ça. Données pas perdues physiquement, mais inaccessibles. Jamais pensé à les convertir » ;*
- *« Je dirai que globalement avec le temps, on perd le matériau issu du recueil initial »*

Cette thématique de la perte semble donc un argument à retenir pour sensibiliser à la nécessité de la sauvegarde. Mais autant la perte de la donnée « vive » et active est ressentie vigoureusement et brutalement (un « incendie »), autant la perte de la donnée « dormante » de projets de recherche antérieurs est constatée mais vécue sans drame. Trois chercheurs (sensibilisés car historiens ou auteurs de recherches chronologiques sur la langue ou le paysage) évoquent également des trésors de données dormantes dans leurs laboratoires, rêvant à une possible numérisation… par d'autres *(« Tu te rends compte de ce que ça représente comme travail ! Comment tu fais ça ?* »).
On pourrait d'ailleurs rattacher certaines réticences au partage à cette thématique d'une perte ou d'une dépossession fantasmée.

3.2.2.6 Evolutions des pratiques liées aux données

A/ Poids et volume : le foisonnement des données

Quand on interroge les enquêtés sur le volume de leurs données, on a des situations contrastées. Sur 21 chercheurs interrogés, cinq évoquent des données très volumineuses. Il s'agit essentiellement de chercheurs de STAPS, géographie et infocom (à l'exception d'un historien). La masse et la diversité des données suscitent aussi quelques commentaires. Ce foisonnement est diversement perçu, suivant le point de vue où l'enquêté se place : producteur, auteur… À noter que la thématique de la difficulté d'exploitation et de gestion n'est pas abordée par les

chercheurs, mais uniquement par un personnel en soutien à la recherche. Les chercheurs se satisfont de leur méthode personnelle, et ne semblent pas en demande sur ce point :

- *« Cela ne veut pas dire pour autant qu'on n'a pas énormément de données - pas en poids mais plutôt en terme de foisonnement.» ;*
- *« Et quand je discute avec mes collègues […], dans toutes les disciplines, on voit bien que l'on nous demande de plus en plus de données, ça devient de plus en plus volumineux à stocker, les enquêtes sont de plus en plus encombrantes. Il faut qu'on apprenne à travailler ensemble et à traiter ensemble cette masse de données » ;*
- *« Je pense que les données, il y en a toujours eu une multitude, une diversité, je ne pense pas qu'il y en ait plus aujourd'hui. Par contre, elles sont pour beaucoup numériques, et là, ça change énormément la donne. […] Mais après pour moi, ça n'augmente pas forcément les données sur lesquelles je vais pouvoir travailler, moi, dans la mesure où une donnée numérique, elle n'a pas de sens si elle ne peut pas être mise en lien. ».*

B/ Support matériel des données

Le format des données et leur disponibilité en ligne ont changé la donne, voire les pratiques :
- « *J'ai toujours plaisir à aller dans des Archives et à travailler sur des documents papiers. Après, peut-être de moins en moins envie de faire des retranscriptions. C'est-à-dire que je prends beaucoup de photos maintenant et je fais ça chez moi.* »

Plusieurs chercheurs issus de disciplines différentes saluent la facilité d'accès et la simplification de leur travail apportée par la mise en ligne, mais aussi les avancées liées à l'open access et l'open data (disponibilité, gratuité) :

- *« On est très contents que les collectivités territoriales rendent publiques leurs données. J'ai souvenir d'une époque où l'on essayait de monter des bases de stockage d'informations, de mutualiser les données qu'on arrachait à l'INSEE, à l'époque où l'information était rare et chère. Maintenant comme tout est plus ou moins public, ça a beaucoup facilité notre travail* » ;
- « *Il y a certaines choses comme Gallica, ou d'autres... RetroNews pour la numérisation des périodiques, qui sont des outils absolument essentiels. Qui sont mis gratuitement en plus au service du public. Et ça, ça fait avancer la recherche.* » ;

Cette disponibilité des données en ligne a un impact sur les pratiques et les lieux de travail. Toutefois, tout n'est pas encore numérique, et si certains ont basculé entièrement vers le numérique, d'autres travaillent encore exclusivement sur des sources papier (deux interviewés). Il faut nuancer aussi par le fait que les personnes qui ont accepté un entretien étaient probablement plus sensibilisées aux données et à l'informatique, avec aussi un effet générationnel souligné par deux de nos interlocuteurs :

- *« Toutes les générations de chercheurs n'évoluent pas à la même vitesse. Il y a des collègues plus âgés qui travaillent encore à l'ancienne » ;*
- *« Il y a peut-être une question aussi générationnelle, c'est à dire que là on est en train de travailler pour des chercheurs qui ont 50- 40 ans, 35 ans aujourd'hui ils ne sont pas très... et puis ceux qui arrivent n'auront même plus besoin de ces services ».*

Certains font cependant écho d'un besoin d'orientation face à la profusion des sources. Il y a là aussi un point de vigilance pour qui veut favoriser le partage et la réutilisation des données, voire alléger la nécessité de stockage : *« on a accès à tellement de choses, ça devient tellement énorme, qu'en fait tu te noies ; […] et en plus ça va très très vite. »*

C/ Le renouvellement de la science par les données

L'émergence des humanités numériques bouleverse la façon dont se fait la science en SHS, et plusieurs chercheurs de Rennes 2 témoignent de cet impact :

- *« En tout cas, pour moi effectivement les données permettent de faire des thèses différentes. » ;*
- *« Et puis après effectivement tous ces logiciels comme Gephi, Paraview, […] ça m'a sorti de ce qu'on a appelé la lecture de près pour aller vers d'autres pratiques de géographe, vers la spatialisation des données. Donc ça a bouleversé ma façon de travailler ; ça a radicalisé ma façon de travailler » ;*
- *« J'aime bien faire de la recherche en plein texte. Ca c'est important pour moi. Ce sont des choses qu'on ne pouvait pas faire avant » ;*
- *« De plus en plus, on peut diffuser des ensembles de données plus importants et de plus en plus publier des choses qui ont moins d'impact scientifiquement mais qui sont plus intéressantes méthodologiquement ».*

Mais les données sont-elles pour autant une problématique largement partagée et débattue ?

3.2.2.7 Les unités et la thématique des données

Certains expriment leur conviction de l'importance du sujet :

- *« C'est une question que l'on se posait moins il y a 10 ans : la mise à disposition des données de la recherche. Aujourd'hui, c'est aussi une demande à la fois institutionnelle et de nos étudiants ».*

Mais la conscience individuelle ne semble pas encore relayée par un débat plus large
Les données constituent encore largement un « *impensé* », contrairement aux archives ouvertes qui commencent à être abordées dans les assemblées de laboratoires :

- *« La question du numérique, de la gestion des données, de ce qu'on doit faire ne se pose même pas. [… Enfin c'est assez surprenant, ce n'est pas un élément structurant, une question structurante de l'équipe, alors qu'en fait il y a des chercheurs qui s'intéressent au numérique. Donc ce n'est pas si évident que ça. »*

3.2.3 Synthèse sur les données collectées et produites

3.2.3.1 Complexité du paysage

Le paysage des données qui se dessine à travers l'enquête et les entretiens est celui d'une grande diversité.
Si, à un niveau macro, les pratiques des champs disciplinaires et des chercheurs présentent des similitudes, un chercheur individuel manipule des données variées.
La masse des données, la difficulté à distinguer les informations à archiver dans un projet, et les questions relatives à la propriété intellectuelle ne doivent pas être sous-estimées.
La problématique des données n'est pas pour l'instant un objet de débat dans les unités ou de questionnement pour les chercheurs ; le terme lui-même ne recouvre pas forcément la même réalité pour tous les répondants.

3.2.3.2 Comment adopter un discours de sensibilisation efficace sur la question des données ?

Les échanges et résultats du questionnaire permettent de cerner les préoccupations des chercheurs, et le peu de temps qu'ils ont à consacrer à ces nouvelles problématiques. Ils permettent aussi d'identifier les thématiques qui ont du sens pour eux, et sur lesquelles on peut s'appuyer.

A/ Tenir compte de la valeur affective en adaptant le discours aux problématiques des chercheurs :

- plutôt que de commencer par le partage, montrer l'intérêt pour le chercheur lui-même de veiller à préserver la recherche dans laquelle il a investi son temps et son énergie, pour pouvoir continuer à l'exploiter et à s'y référer ;
- s'appuyer sur les pratiques de publication de certaines disciplines (dépôt des données avec l'article…) et tenir compte des spécificités disciplinaires.

B/ Proposer un recours et des informations pratiques et non une injonction théorique (en interne ou avec des partenaires) et notamment sur les problématiques qui sont ressorties comme sensibles :

- sur la propriété intellectuelle ;
- sur les conditions pour un archivage à long terme (formats, nommage, description…) ;
- pour des solutions de stockage adaptées à un mode de travail mobile et collaboratif ;
- mettre en place une information simple, rapide et facilement accessible sur les plateformes, les outils et les labels de validité des données.

C/ Cibler les interventions prioritaires :

- former en priorité les étudiants, les doctorants et les chercheurs entrant à l'université ;
- intervenir au début des projets de recherche pour éviter l'effet « masse » (sensibilisation des personnels en soutien au montage de projets et des porteurs de projets collaboratifs avant demande de financement) et non pas en rétrospectif, sauf exception ;
- sensibiliser les responsables d'unité et les « influenceurs ».

D/ Démontrer par l'exemple la nécessité de garder les données vivantes :

- éventuellement réaliser un ou deux projets pilotes identifiés comme à forte valeur ajoutée par les laboratoires et utiles au grand public sur des données dormantes (numérisation rétrospective de données ou migration de bases à forte valeur ajoutée).

E/ Faire des données un objet de débat dans les instances universitaires

3.2.3.3 Un paysage en évolution rapide

Comme en ont témoigné certains chercheurs, les données transforment la pratique même de la science. Les pressions des financeurs font évoluer le paysage rapidement. Il est donc indispensable de faire et de partager une veille sur ce sujet.

3.3 Quelles pratiques de stockage et d'archivage ?

Après un examen des types de données collectées et produites, l'enquête a également permis d'interroger les pratiques des enseignants-chercheurs de Rennes 2 en matière de stockage et d'archivage. Ces dernières sont-elles normalisées ou relèvent-elles davantage de méthodologies personnelles ? Des politiques d'archivage sont-elles déjà mises en place ? Et si oui, quelles sont leur portée et leur efficacité ? En abordant la question de l'archivage, l'enquête vise par ailleurs à mieux évaluer comment les enseignants-chercheurs se sont emparés des enjeux de sécurisation et de documentation des données de recherche.

3.3.1 Les résultats de l'enquête statistique

Les pratiques de stockage et d'archivage des données, collectées et produites, font l'objet des questions 12 à 20 de l'enquête en ligne. Ces neuf questions abordent les thématiques suivantes : le stockage des données (volume, lieu et fréquence de sauvegarde), la politique d'archivage et la responsabilité de cette politique, ainsi que les pratiques de description des données dans une visée d'archivage à long terme.

3.3.1.1 Le stockage des données

A/ Le volume de stockage

Les enseignants-chercheurs ont exprimé des besoins très diversifiés en matière d'espace de stockage. À la question 12 « À combien estimez-vous l'espace nécessaire au stockage de vos données ? », environ un tiers (34 %) des répondants a sélectionné la même réponse parmi les cinq items proposés, soit « entre 20 et 100 Go ».

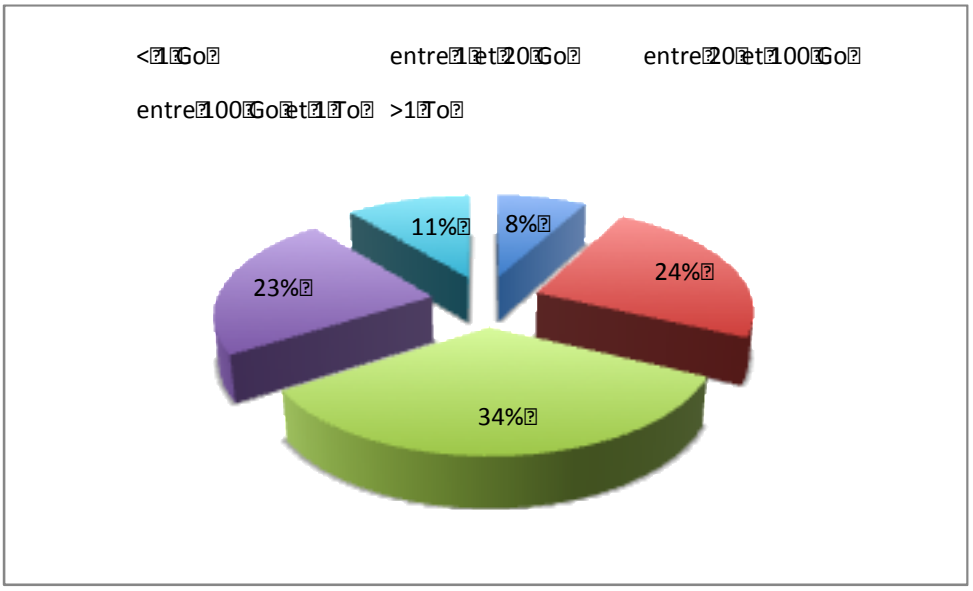


Figure 22 : Espace de stockage requis par les chercheurs (N = 95)

Cette diversité semble se retrouver également en grande partie à l'échelle des unités de recherche. Par exemple, pour le laboratoire Tempora, les dix répondants ont couvert à eux seuls l'ensemble des réponses possibles : on recense deux réponses pour l'item « < 1 Go », deux

réponses « entre 1 et 20 Go », une réponse « entre 20 et 100 Go », quatre réponses « entre 100 Go et 1 To » et une réponse « > 1 To ». Le constat est le même pour les treize répondants de l'unité de recherche LP3C.

	< 1 Go	Entre 1 et 20 Go	Entre 20 et 100 Go	Entre 100 Go et 1 To	>1 To	Total
ACE	1			1		**2**
APP	1	2	4		1	**8**
CELLAM			3	3		**6**
LiRIS (ex-CIAPHS)		2	3			**5**
COSTEL			2	1	1	**4**
CRBC		1				**1**
CREAD			2			**2**
LP3C (ex-CRPCC)	2	5	2	3	1	**13**
ERIMIT		2	2	1		**5**
ESO		3	7	5		**15**
HCA				1		**1**
IRMAR	1					**1**
LAHM					1	**1**
LIDILE	1	2		1		**4**
M2S		2	1		3	**6**
PREFics		1	2	2	1	**6**
Recherches en Psychopathologie						**0**
Tempora	2	2	1	4	1	**10**
VIPS²		1	3		1	**5**
Total	**8**	**23**	**32**	**22**	**10**	

Figure 23 : Espace de stockage requis par les chercheurs selon leur laboratoire (N = 95)

Ainsi, cela montre qu'il est en réalité assez difficile de présumer des besoins réels d'espace de stockage pour chaque laboratoire tant les pratiques de chaque enseignant-chercheur sont variables. Ce résultat renforce ainsi l'intuition selon laquelle il serait plus cohérent, bien que plus complexe en termes de gestion, de mettre en place un service personnalisé de stockage qui s'adapte aux différents espaces réellement requis. Par exemple, si l'on se réfère à l'offre actuelle de l'université qui attribue à chaque enseignant-chercheur le même espace de stockage, il est évident que l'espace non utilisé par certains pourrait être d'une grande utilité à d'autres, dont les données sont plus volumineuses.

Compte tenu de l'émergence et de la multiplication de nouvelles méthodes de recherche intrinsèquement liées au numérique, le nombre d'enseignants-chercheurs en SHS, ayant besoin d'espaces de stockage de plus en plus conséquents, est vraisemblablement amené à croître au fil des années. En effet, le pourcentage de répondants estimant le volume de leurs données supérieur à un To est plus élevé à Rennes 2 (11 %) qu'à Lille 3 (6 %). Toutefois, il faut rester prudent avec ces résultats dans la mesure où la question 12 de notre enquête, directement reprise de l'enquête de Lille, ne comporte pas le même nombre de réponses possibles. L'enquête lilloise proposait de fait un item supplémentaire : « ne sais pas, pas d'avis ».

Par ailleurs, même si la question 12 ne contenait effectivement pas de proposition de réponse de type « ne sais pas », le taux de réponse est significatif de ce point de vue. Seulement deux personnes sur trois ont répondu à cette question, ce qui représente un ratio assez faible de 66 %. Ce chiffre pourrait alors témoigner d'une difficulté éprouvée par certains chercheurs pour évaluer le volume de leurs données. L'enquête de Lille avait déjà permis de pointer cette difficulté : à la même question, ils étaient 39 sur 210 à avoir répondu « ne sais pas, pas d'avis » soit 18 %. Plusieurs raisons à cette incapacité d'évaluer ses données sont possibles : l'éparpillement des données sur plusieurs supports, ce qui rend difficile l'estimation globale, une simple méconnaissance des unités de mesure informatiques ou encore l'instabilité de la définition des « données de recherche » qui peut rendre les frontières poreuses entre ce qui relève effectivement des données de recherche et ce qui n'en relève pas.

B/ Le choix du lieu de stockage

À la question 13 « Où stockez-vous les données de recherche ? », une grande majorité des répondants a opté pour un stockage local et personnel : le disque dur externe arrive en tête pour 77 % des répondants, puis l'ordinateur professionnel (62 %) et l'ordinateur personnel (57 %). Ces supports sont également les plus utilisés dans chacune des différentes unités de recherche de Rennes 2. C'est un résultat qui se place dans la continuité de l'enquête lilloise où le stockage local et personnel était déjà le plus plébiscité, et de loin, choisi par 207 enquêtés sur 212.

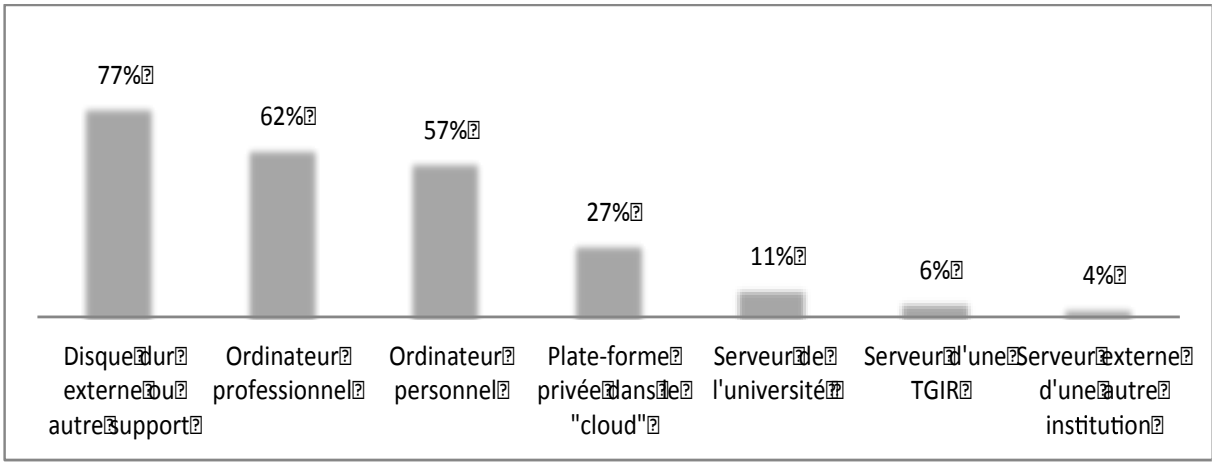


Figure 24 : Lieux de stockage (N = 122)

En revanche, les services de stockage institutionnels, proposés par l'université ou par les TGIR, ne remportent qu'une faible adhésion. Seulement un répondant sur dix environ a recours au serveur de l'université pour y stocker ses données. Quant aux services de la TGIR, seulement 6 % des répondants les utilisent. Le stockage institutionnel et en réseau constitue pourtant un des moyens les plus sécurisés de conservation des données pour les enseignants-chercheurs. Si cela doit être en partie dû à une méconnaissance de l'offre, d'autres facteurs sont certainement à prendre en considération. C'est notamment ce qu'il faudra tenter de cerner ici.

Le recours au stockage sur le *cloud* semble progressivement se développer : ils sont 27 % à l'utiliser, alors qu'ils étaient 19 % dans l'enquête de Lille. Et ce bien que les entretiens individuels aient révélé une certaine méfiance de la part des enseignants-chercheurs face à ces nouveaux services.

Enfin, 87 % des répondants cumulent plusieurs types de lieux de stockage. La majorité en utilisent deux (51 répondants sur 122), voire trois (43 répondants). Seulement 16 enquêtés ne stockent leurs données que sur un seul et même support.

Lieux de stockage sélectionnés	Nombre de réponses	Pourcentage
1	16	13 %
2	51	42 %
3	43	35 %
4	10	8 %
5	2	2 %
6	0	0 %
7	0	0 %

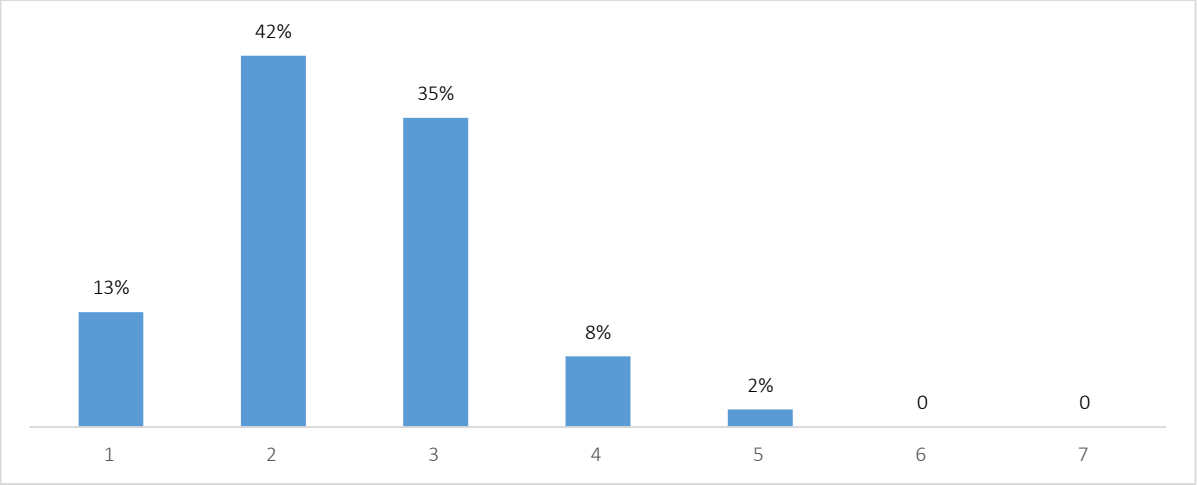


Figure 25 : Types de lieux de stockage par répondant (N = 122)

C/ La fréquence des copies de sauvegarde

Les résultats de la question 14 « À quelle fréquence faites-vous des copies de sauvegarde de vos données ? » témoignent encore une fois d'une grande disparité des pratiques.

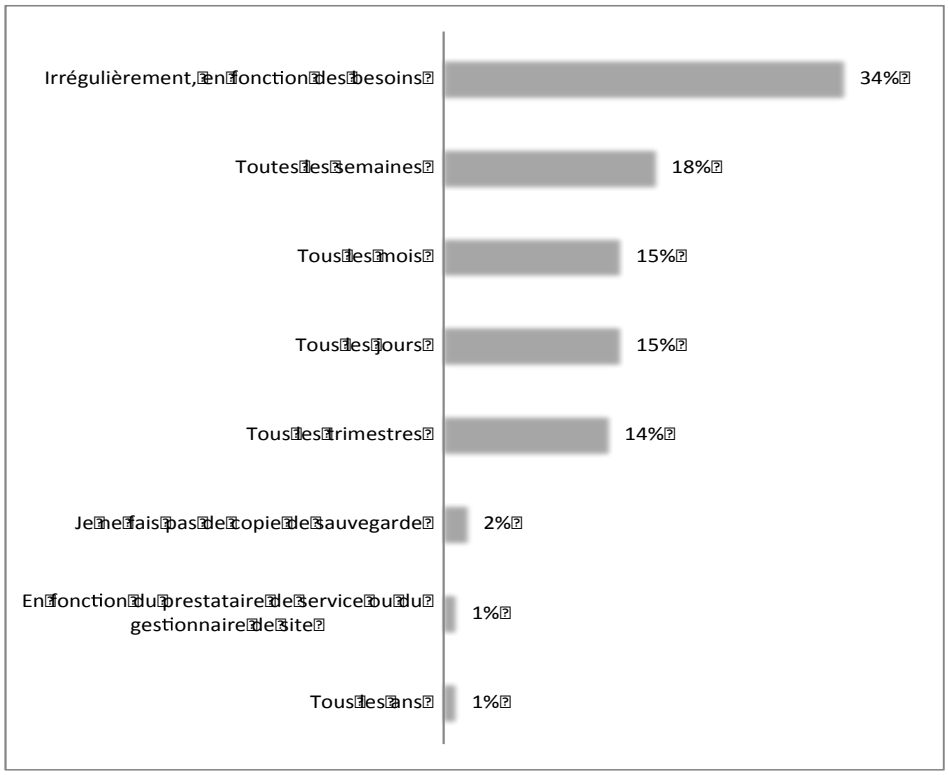


Figure 26 : Fréquence de sauvegarde des données (N = 121)

L'item qui a reçu le plus grand nombre de réponses est « irrégulièrement, en fonction des besoins » avec 34 % des réponses. Ceci est tout à fait révélateur d'une absence globale de politique concrète, qu'elle soit collective ou personnelle, de gestion des données. Il s'agit le plus souvent d'une gestion au cas par cas, mais aussi au plus pressé et non systématique.

Loin derrière, les items « toutes les semaines », « tous les jours » et « tous les mois » reçoivent à peu près tous le même nombre de réponses (entre 18 % et 15 % des réponses pour chacun). Il reste enfin trois répondants (2 %) qui déclarent ne faire aucune copie de sauvegarde.

Les enseignants-chercheurs de Rennes 2 paraissent ici faire preuve de moins de prudence que leurs collègues de Lille 3, qui sont 28 % à sauvegarder quotidiennement leurs données et 26 % à le faire en fonction des besoins. À noter que, comme dans l'enquête lilloise, ces résultats doivent être interprétés avec précaution en raison des significations variables attribuées par les chercheurs au terme de « sauvegarde », pouvant aller de l'enregistrement quotidien des modifications d'un document à la copie régulière de jeux de données sur plusieurs supports.

3.3.1.2 Les politiques d'archivage

La mise en place et le suivi d'une politique d'archivage font l'objet des questions 15 à 18.

Lors de la question 15, une écrasante majorité des répondants a notifié l'absence de politique d'archivage à l'échelle de leur laboratoire : on compte 75 réponses « non » et 38 réponses « je ne sais pas », soit un total de 113 réponses sur 122 (93 %). Certes, les réponses « je ne sais pas » ne permettent pas de véritablement attester l'existence ou non d'une politique d'archivage mais elles n'en demeurent pas moins éloquentes : si une telle politique est mise en place mais qu'elle demeure méconnue des chercheurs, cela signifie qu'elle n'est pas appliquée et donc non efficace.

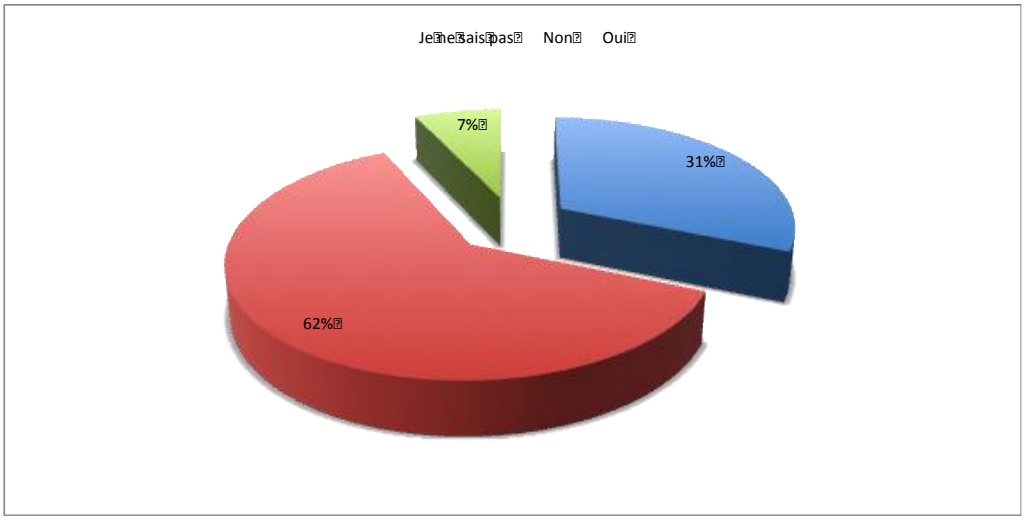


Figure 27 : Élaboration d'une politique d'archivage dans le laboratoire

On note une faible progression des pratiques depuis l'enquête de Lille 3 dans laquelle les répondants étaient 97 % à se déclarer eux-mêmes responsables de la sauvegarde de leurs données, à défaut d'une politique mise en place par leur laboratoire respectif.

Ces réponses et le rattachement des répondants aux différentes unités de recherche sont globalement cohérents. Deux cas de figures principaux se retrouvent dans chacun des laboratoires : d'une part, une majorité de « oui » avec quelques « je ne sais pas » ou, d'autre part, un ensemble de « non » et de « je ne sais pas ». On note qu'il y a davantage de réponses « je ne sais pas » lorsque c'est le « non » qui l'emporte plutôt que le « oui ». L'importance du nombre de réponses « je ne sais pas », même s'il est moindre pour les unités de recherche ayant bel et bien une politique d'archivage, indique qu'il subsiste un flou autour de cette question.

	Oui	**Non**	**Je ne sais pas**	**Total**
ACE		6		**6**
APP		9	2	**11**
CELLAM		5	2	**7**
LiRIS (ex-CIAPHS)		5	3	**8**
COSTEL	3		1	**4**
CRBC		2	1	**3**
CREAD		4		**4**
LP3C (ex-CRPCC)		9	5	**14**
ERIMIT		3	4	**7**
ESO		7	8	**15**
HCA		2	1	**3**
IRMAR		1		**1**
LAHM		1	1	**2**
LIDILE	1	4	1	**6**
M2S	5		1	**6**
PREFics		5	1	**6**

	Oui	Non	Je ne sais pas	Total
Recherches en Psychopathologie		2		**2**
Tempora		7	5	**12**
VIPS²		3	2	**5**
Total	**9**	**75**	**38**	

Figure 28 : Perceptions des politiques d'archivage dans les laboratoires (N = 122)

Les seules réponses contradictoires à signaler sont pour le laboratoire LIDILE où l'on relève notamment un « oui » et quatre « non ». Cela pourrait également signaler que la notion de « politique d'archivage » n'a pas la même signification pour tous et qu'elle devrait être davantage explicitée.

À l'exception du LIDILE, seulement deux laboratoires se distinguent comme ayant développé une politique d'archivage : l'UMR Costel et l'équipe d'accueil M2S. Il s'agit de deux unités qui sont confrontées au traitement et au stockage de données considérées comme lourdes. En effet, les enseignants-chercheurs rattachés à ces deux laboratoires font partie de ceux qui ont estimé les besoins les plus importants d'espace de stockage à la question 12 : une personne sur les quatre répondants enseignants-chercheurs de Costel et trois personnes sur les six répondants du laboratoire M2S ont évalué leurs besoins à plus d'un To[32]. Ainsi, il est évident que le poids des données à stocker a un impact non négligeable sur l'urgence et, donc, sur la mise en œuvre effective d'une politique adaptée.

La question 16 « Si oui, qui est responsable de l'archivage des données ? » dépend directement de la question précédente. Parmi les quatre items proposés, les neuf réponses recensées sont unanimes : lorsqu'une politique d'archivage est en place dans le laboratoire, c'est à un membre de l'équipe de recherche qu'en reviendrait la responsabilité.

Membre de l'équipe de recherche	9
Membre d'un service de soutien à la recherche	
Un prestataire de service externe	
Pas de responsable identifié	

Figure 29 : Personne responsable des données dans une politique d'archivage du laboratoire (N = 9)

En revanche, lorsqu'une politique d'archivage n'a pas été décidée à l'échelle de l'unité de recherche, il semblerait qu'il n'y ait pas non plus de mise au point d'une politique personnelle équivalente dans la majorité des cas. C'est ce que révèlent les résultats de la question 17, dépendant également de la question 15, « Si non, avez-vous mis au point vous-même une politique d'archivage de vos données ? ». Deux tiers des répondants ont répondu à cette question par la négative.

Non	Oui
41	23

Figure 30 : Élaboration d'une politique personnelle en cas d'absence de politique d'archivage du laboratoire (N = 64)

[32] Cf. ci-dessus Figure 23 : Espace de stockage requis par les chercheurs selon leur laboratoire (N = 95).

La problématique de conservation des données des chercheurs non-permanents par les unités de recherche fait l'objet de la question 18.

Reformulation des réponses	Nombre de réponses	Pourcentage (par rapport aux 54 répondants)
Archivage individuel par les chercheurs :	**44**	**77 %**
- données perdues	38	
- stockées sur des plateformes en ligne	1	
- stockées dans les locaux	2	
- stockées sur des supports de stockages externes	2	
- stockées sur un serveur en réseau	1	
Politique commune d'archivage :	**10**	**18 %**
- accessible à l'ensemble des membres d'un même laboratoire	9	
- accessible à l'ensemble des membres d'un même groupe de recherche	1	

Figure 31 : Gestion et conservation des données produites par des chercheurs non-permanents du laboratoire (doctorants, chercheurs associés, partenaires…) (N = 54)

Il semble que l'archivage étant majoritairement individuel (pour 44 répondants sur 54), la plupart de ces données sont perdues pour le laboratoire. Lorsqu'une politique d'archivage est appliquée par l'unité de recherche, on remarque sans surprise que les données ont plus de chance d'être conservées, bien que cela concerne seulement moins d'un cinquième des réponses.

3.3.1.3 La description des données

Les questions 19 et 20 portent enfin sur la documentation des données par les enseignants-chercheurs. Environ la moitié des répondants à la question 19 « Attribuez-vous habituellement des informations supplémentaires (métadonnées) à vos données de recherche ? » attribue des métadonnées à leurs données. Ces métadonnées sont de type majoritairement scientifique (55 réponses), puis administratif (29 réponses) et enfin technique (14 réponses).

Métadonnées	**Nombre de réponses**	**Pourcentage**
N'attribue pas d'informations supplémentaires	57	50 %
Des informations scientifiques	55	48 %
Des informations administratives	29	25 %
Des informations techniques	14	12 %

Figure 32 : Types de métadonnées attribuées par les chercheurs (N = 115)

Certains enseignants-chercheurs affirment documenter leurs données de plusieurs manières : 19 répondants utilisent deux types de métadonnées différents parmi ceux proposés ici et 10 répondants cumulent les trois types.
D'après la question 20 « Si vous attribuez des informations supplémentaires, comment procédez-vous ? », lorsque les répondants associent à leurs données des métadonnées, ils élaborent le plus souvent une documentation personnelle (95 %). Ils sont peu nombreux en

revanche à utiliser un cahier de laboratoire (16 %) ou à s'appuyer sur un standard de métadonnées (11 %).

	Oui	Non	Pas de réponse
Élaboration d'une documentation propre	52 (95 %)	1	2
Utilisation d'un standard de métadonnées	6 (11 %)	23	26
Utilisation d'un cahier de laboratoire (ou de recherche)	9 (16 %)	22	24

Figure 33 : Méthodes d'attribution d'informations supplémentaires aux données (N = 55)

En outre, près de la moitié des répondants n'ont coché aucune case pour les deux items "Utilisation d'un standard de métadonnées" et "Utilisation d'un cahier de laboratoire". Cela montre d'une part que ces méthodes sont minoritaires, mais également qu'elles restent globalement méconnues en SHS. Bien qu'essentielle au partage et au bon archivage des données, l'attribution de métadonnées ne s'est pas encore imposée comme une pratique systématique, loin de là. De fait, la documentation des données recouvre un ensemble de méthodologies, de compétences et d'outils spécifiques que très peu d'enseignants-chercheurs affirment maîtriser. Les réponses à cette question sont ainsi révélatrices de l'importance du travail de sensibilisation et de formation à réaliser en la matière auprès des chercheurs de l'université.

3.3.2 Les résultats des entretiens

Les entretiens réalisés auprès de 21 enseignants-chercheurs et ingénieurs volontaires ont également permis de faire un état des lieux des pratiques de stockage et d'archivage.

3.3.2.1 Le stockage des données

Dans la continuité de l'enquête quantitative, les entretiens ont mis à jour des pratiques et des besoins diversifiés en matière de stockage. Les domaines et les méthodes de recherche ont une influence sur les données récoltées ou produites, et donc sur le volume de données à stocker.

A/ Des volumes inégaux de données

Certains enseignants-chercheurs sont amenés à traiter et stocker des volumes importants :
- *« Je travaille avec de l'audiovisuel, je filme sur le terrain, et du coup c'est des teras et des teras, et c'est vraiment lourd. »*
- *« Parce que nos données explosent, on a de plus en plus de vidéos, […] les fichiers deviennent de plus en plus lourds, en plus avec les années qui passent, moi je stocke des choses maintenant depuis 20 ans. »*
- *« On a des masses de données, on sait plus quoi faire pour les stocker. [...] Chaque fois on a des centaines de giga de données par jour. »*
- *« Et puis j'ai un énorme stock de photos que j'utilise à la fois sous forme d'archives et [...] pour distribuer dans mes cours. C'est essentiellement pour ça que j'ai des problèmes de données et que j'ai reçu des mails me disant que mon quota était dépassé. »*

Tandis que d'autres n'ont pas de besoins d'espace très importants et semblent moins préoccupés par la question :
- *« Ce n'est pas très lourd. »*
- *« Par rapport à la donnée que j'exploite et que je manipule, mon ordinateur suffit. Ce n'est pas des gros gros gros volumes. »*
- *« C'est beaucoup de fichiers textes, un peu de tableurs, mais pour l'instant ça va, je n'ai pas de difficulté à stocker les données »*
- *« Mais effectivement ça, ça pèse un peu, mais ce n'est pas non plus... je vis très bien avec mon disque dur. ».*
- *« Pas de données massives pour l'instant. Plutôt du little data, mais très varié. »*

À noter que le développement du partage est une solution pouvant alléger ces besoins de stockage. En effet, lorsque les données sont disponibles en ligne en accès ouvert et pérennisé, les enseignants-chercheurs n'ont plus besoin de les conserver, en tout cas à long terme, de manière individuelle :
- *« La SNCF, on ne cherche plus à stocker leurs données puisqu'elles sont accessibles. […] Rennes, l'agglomération, la Métropole, on ne cherche plus à développer des espaces de stockage puisqu'on a accès directement à leurs données. Pour nous, c'est une logique différente : on va se servir quand on a besoin. »*
- *« Alors pour l'instant en France, les photos aériennes, elles sont en accès quasiment toujours gratuit sur Géoportail, sauf les plus récentes, mais ça, ça va. Par contre dans d'autres pays, il faut les acheter, les scanner et tout de suite, ça occupe de la place. »*
- *« Pendant un moment, j'ai archivé les données de houle, maintenant elles sont en accès public, donc je n'ai plus besoin de le faire... »*

B/ Un stockage principalement local et individuel

Si les enseignants-chercheurs ont des besoins très variés en matière d'espace de stockage, ils semblent davantage unanimes sur le type de support utilisé. En effet, tel que l'a montré l'enquête quantitative, l'emploi d'un stockage à la fois local et personnel reste prééminent :
- *« Toutes mes données sont sur trois disques : le disque dur de mon portable, le disque dur externe qui est chez moi et [le disque S] » ;*
- *« J'utilise des espaces sur le disque S de la DSI et un disque dur externe » ;*
- *« C'est sur disque dur externe et sur l'ordinateur » ;*
- *« Je stocke sur mon ordinateur, sur mon disque dur et après, il y a des documents qui sont partagés via le Drive ».*

Ce type de stockage est préféré par les chercheurs bien qu'il ne permette qu'une collaboration restreinte et malgré les risques de pertes de données qu'il présente.

Plusieurs personnes interrogées pointent une même lacune de ce type de support, qui est la difficulté d'établir une synchronisation des versions de données sauvegardées entre les différents espaces de stockage :
- *« Il y a des questions qui ne sont pas forcément faciles pour moi dans le stockage de versions »*
- *« J'ai mon ordinateur portable qui est personnel [...] cet ordinateur-là (ordinateur fixe de bureau) c'est vraiment un lieu de stockage, je ne modifie jamais un fichier sur cet ordinateur. Enfin quand je le fais comme cet après-midi, là par exemple, je me l'envoie tout de suite par internet pour que je le réenregistre, pour que je n'aie qu'un seul disque dur complètement à jour. [...]. Je suis obligé de faire tout ce travail de sauvegarde régulière. Ce qui serait pour moi formidable, c'est s'il y avait une procédure [automatique] »*
- *« J'ai un disque dur externe sur lequel je stocke, je ne stocke pas tout parce que je ne suis jamais à jour ».*

Cela nécessite en effet des opérations manuelles, qui sont réalisées de manière systématique ou, le plus souvent, en fonction des besoins. La synchronisation est le point fort d'un autre type de stockage : le *cloud*. Or, plusieurs personnes interrogées disent faire preuve d'une grande réserve face aux services de stockage en ligne, et plus particulièrement à Google Drive ainsi qu'à l'ensemble des outils développés par la firme américaine :
- *« On a des besoins d'outils collaboratifs pour arriver à travailler en projet avec des données multiples qui viennent s'agréger. On est obligés de passer par des outils type Google ou autres et cela pose de gros problèmes. »*

C/ Avantages et limites du stockage collaboratif

Si les chercheurs en SHS continuent à travailler de manière solitaire, ils ont néanmoins de plus en plus tendance à se regrouper en équipes, par incitation ou par choix, que ce soit au sein de leur unité de recherche ou dans le cadre de projets de recherche. Or, un autre atout du *cloud* est justement son aspect collaboratif. Au moins deux répondants identifient le besoin d'avoir à leur disposition des outils de travail et de stockage collaboratif :
- *« Je suis obligé de collaborer avec des collègues en Suisse et ailleurs »*
- *« Cette demande d'espace collaboratif est très forte et régulière ».*

Plus précisément, ces outils doivent être spécifiquement dédiés à la recherche. Par défaut ou par méconnaissance de l'offre existante, les enseignants-chercheurs sont nombreux à se rabattre sur des services de type « grand public », comme Google Drive ou Dropbox, alors que les données traitées peuvent être sensibles. Le manque de transparence quant à la sécurisation des données dont font preuve les entreprises qui proposent ces services est problématique et n'est pas sans alerter les chercheurs :
- *« La fiabilité et la sécurité, c'est quelque chose de fondamental, mais dont je ne maitrise pas bien les éléments. [...] On commence à parler aussi de stockage dans le cloud, j'avoue qu'avec tout ce qu'on peut lire par ailleurs sur les capacités de pillage, ou de captation par autrui, ça me laisse plus perplexe. »*

De fait, certains enquêtés déclarent ne pas utiliser le *cloud* pour cette raison spécifique :
- *« Tu n'utilises pas des infrastructures sur le web, ou Google Drive ? » « Non, non, alors moi je milite pour la dégooglisation d'internet, alors je n'utilise pas Google. Je pourrais utiliser un équivalent, mais non je n'utilise pas ça. »*
- *« Vous n'utilisez pas de stockage comme une Dropbox ou Google Drive ? » « Je ne connais pas et j'évite. On m'a toujours dit de ne rien y mettre. Alors depuis, Dropbox, même quand on m'a envoyé des photos quand j'étais en colloque, on m'a dit vas sur Dropbox, j'ai dit vous me les envoyez, je n'y vais pas. ».*

Quant aux chercheurs qui utilisent ce type de service, ils ont pour la plupart bien conscience des limites, voire des risques, qu'il présente :
- *« J'ai une Dropbox, un site distant, mais sur lequel je ne mets que des fichiers que je dépose temporairement, soit pour donner l'accès à d'autres, comme à mes étudiants, soit parce que je vais me déplacer et au cas où j'ai un problème avec mon ordinateur, que je puisse retrouver un Powerpoint, ou quelque chose dont je vais avoir besoin. Et dès que j'en ai plus besoin, je le retire exprès, pour pas que quelqu'un d'autre puisse y avoir accès hors de ma vigilance. »*
- *« On utilise des espaces collaboratifs très ponctuellement, mais plutôt entre collègues pour les cours ou dans des programmes de recherche où il y a une Dropbox par exemple. Mais pour moi, c'est quelque chose à éviter, la Dropbox, pour des raisons juridiques. Et plusieurs fois, j'ai des collègues qui ont ouvert une Dropbox, on a commencé à l'utiliser et puis on s'est ravisé, on a changé d'avis. On l'utilise pour la coordination des TD entre collègues de première année mais je n'irai pas plus loin. »*

D/ Les principaux risques de perte de données

Les défaillances de stockage mènent à des accidents de pertes de données – des *« histoires dramatiques »* comme les qualifie l'un des chercheurs interrogés – qui demeurent fréquentes. Les enseignants-chercheurs en identifient deux causes principales. D'une part, la perte ou le vol du support de stockage :
- *« La collègue avec laquelle je travaille depuis maintenant dix ans, et qui, elle, est chargée des mises à jour, etc. mais aussi des mises en page des divers documents, on lui a volé son ordinateur. Donc toutes les données sont parties. »*
- *« Si vous me volez mon ordi portable, mon disque dur externe qui est là et celui qui est à la maison, j'aurais perdu six ans de recherches, avec aucun moyen de récupérer ces données là. »*

et d'autre part, les incompatibilités liées à une migration des formats non assurée :
- *« J'ai perdu par exemple, je sais où elles sont mais elles ne sont pas convertibles, j'ai perdu toutes les données de ma thèse. Des données relativement lourdes, j'avais fait des bases de données. »*
- *« J'ai deux types de sources vidéo. J'en ai qui datent de la fin des années 1990, début 2000 et je crois que je n'arrive même plus à les lire maintenant parce qu'il y a les codecs qui ne fonctionnent plus. »*

Le cumul des supports de stockage est alors la principale précaution prise par les enseignants-chercheurs afin d'anticiper ces risques de perte :
- *« Je sauvegarde [...] en trois endroits différents, et même en deux endroits spatialement différents. »*
- *« J'ai tellement l'expérience, je fais gaffe, j'ai plein de sauvegardes. »*

3.3.2.2 L'archivage des données

Quant à l'archivage des données, il semble encore quasi-inexistant dans les pratiques des enseignants-chercheurs. Aucune des personnes interrogées n'a présenté de stratégie personnelle d'archivage sur le long terme. Il s'agit le plus souvent de solutions au cas par cas :
- *« Il n'y a pas de politique de laboratoire pour l'instant sur la façon dont on procède, ni pour la collecte, ni pour l'archivage, ni pour la valorisation. Chacun fait comme il peut, ou comme il sait faire ou pas faire. »*
- *« Sur l'archivage des données une fois numérisées, c'est pareil. C'est un peu chaque collègue qui archive, soit sur son poste, soit sur un lecteur externe. Il n'y a pas d'espace d'archivage à l'échelle du labo. »*
- *« Sur l'archivage, moi je me rends compte que je suis nulle en archivage. Il me faudrait un cours de rangement. Je mets des choses de côté et je mets "À trier". »*
- *« Après, par contre, la question de l'archivage... On ne la traite pas pour l'instant. »*

Néanmoins, bien que les enseignants-chercheurs n'aient pas développé de pratiques d'archivage, cela ne signifie pas pour autant qu'ils n'ont pas conscience d'un manque criant en la matière. Ils sont plusieurs à le mettre en évidence lors des entretiens :
- *« Ce n'est pas tellement le côté diffusion peut-être, dont on manque le plus, c'est le côté archivage à long terme. Et pour ça par contre, on n'a vraiment pas grand chose qui existe. »*

De fait, cette lacune a une incidence forte sur la réutilisation et la valorisation *a posteriori* de leurs données :
- *« Je me rends compte que j'archive mes projets de façon pas très ordonnée. C'est-à-dire qu'un projet se termine, on passe à un autre, je l'archive tant bien que mal. En fait, l'archivage c'est comme un enterrement. C'est-à-dire que si on doit y revenir, c'est presque une exhumation, c'est*

vraiment très dur. Et donc ça, ça me peine un peu parce qu'en fait on n'est pas dans une logique de valorisation des travaux antérieurs. »
- *« On refait la même chose parce qu'on ne sait plus ce qu'on a fait. Et on ne sait plus ce qu'on a fait parce que c'est mal archivé, mal valorisé, mal documenté. »*
Par ailleurs, le problème lié à l'utilisation des plateformes de stockage privées (*cloud*) ressurgit là encore :
- *« Il y a aussi un enjeu qui n'est peut-être pas tellement apparent dans ce que je vous dis, qui est celui de la pérennité des données à long terme. Parce que justement, quand on stocke ses données sur Dropbox, dès que la Dropbox est fermée, il n'y a plus rien »*

Ainsi, les entretiens individuels ont été l'occasion pour les enquêtés d'identifier les freins empêchant un bon archivage de leurs données et de mettre en évidence leurs besoins.

A/ Ergonomie des outils et des services proposés

Près d'un tiers des personnes interrogées expriment la nécessité d'avoir à leur disposition de nouveaux outils et davantage d'espace de stockage. La nature de ces outils n'est pas toujours précisée. En revanche, les enseignants-chercheurs insistent plutôt sur l'importance de la souplesse d'utilisation et de l'ergonomie des nouveaux outils qui sont mis à leur disposition :
- *« Il faut aussi qu'il y ait aussi une forme de souplesse, d'ergonomie » ; « Il faut que ce soit interactif, ergonomique »*
Ces qualités étant essentielles pour garantir leur bonne prise en main, et par conséquent tant leur efficacité que leur utilisation réelle, il paraît important que ces éléments soient davantage pris en considération lors de l'élaboration d'un service :
- *« Pourtant, il y avait vraiment eu des efforts faits en terme de sensibilisation, de formation auprès des collègues. Mais cela n'a pas pris. Et je pense qu'il y avait aussi un problème de technologie, à l'époque… ça manquait de souplesse, d'ergonomie. ».*

B/ Aspect chronophage de l'archivage

Hormis l'outillage technique qui serait lacunaire, l'autre frein le plus souvent évoqué par les enseignants-chercheurs est l'aspect chronophage de l'archivage des données. Ils sont plusieurs à exprimer un « manque de temps » pour assurer ces tâches de gestion de données. Par ailleurs, ils peuvent considérer cette activité comme une tâche supplémentaire qui leur incombe mais pour laquelle ils n'ont pas forcément les compétences techniques ni le matériel adapté :
- *« L'archivage, l'entretien d'une année sur l'autre, la réactualisation des outils informatiques, là ça dépasse un peu les compétences et le temps que les chercheurs ont à consacrer à ça. »*
Bien que considéré comme crucial pour leur recherche, l'archivage demeure donc relégué au second plan.

Les entretiens ont également permis de mettre en lumière l'impuissance pouvant être ressentie par les ingénieurs devant la charge imposante que représente l'archivage des données d'une unité de recherche :
- *« Penser par exemple que les techniciens et les ingénieurs qui sont dans le traitement peuvent aussi gérer à eux seuls l'administration de la donnée, je pense que c'est une erreur. On ne peut plus. Ici au niveau du labo, le labo a grossi, grossi, grossi sans vergogne. Par contre, en termes de personnels, on est resté au même stade. Donc l'étendue de nos missions a explosé, en termes de volumes de données aussi, donc c'est très compliqué. On a besoin de soutien du point de vue logistique, matériel et humain, tout simplement. »*

C/ Demande d'un accompagnement technique par du personnel dédié et spécialisé

Au moins un tiers des personnes interrogées sollicitent explicitement l'appui d'ingénieurs pour les accompagner sur le plan technique dans le stockage et l'archivage de leurs données :
- *« Je crois vraiment que l'on est maintenant dans une période où on a besoin de professionnels de l'informatique qui soient dédiés à ça, mais qui soient dédiés aussi, pas aux équipes, je sais bien que chaque équipe ne peut pas avoir un ingénieur informatique, ça serait très exagéré certainement, qui soit dédié de façon pérenne. Qu'on ne change pas d'ingénieur tous les ans. Qu'on n'ait pas quelqu'un pour trois semaines et puis le temps de lui expliquer... Quelqu'un qui soit aussi un peu partie prenante qui pourrait nous permettre d'améliorer pourquoi pas les procédures qu'on utilise. ».*
- *« La plupart des collègues chercheurs, s'ils ont pas l'appui d'un ingénieur pour ça, ils ne s'en sortent pas. »*
- *« Et puis après, un soutien technique qui manque quand même cruellement. Enfin, dans le labo ou à l'université, mais avec un vrai soutien. Et après ce sont des questions aussi personnelles, c'est-à-dire qu'il faut trouver la personne qui soit et compétente et disponible. »*

Il se pose alors une première question concernant le rattachement de ces ingénieurs. Plusieurs niveaux sont évoqués : l'université, l'UFR, l'unité de recherche ou l'équipe d'un projet. Ensuite, le rôle et la place de l'ingénieur au sein d'une équipe resteraient également à définir. Enfin, la question des compétences requises est à préciser pour déterminer s'il faut privilégier, en fonction des postes, soit des ingénieurs hyperspécialisés (en informatique ou IST), soit des ingénieurs plus polyvalents. Toutefois, bien que le soutien d'un personnel qualifié soit nécessaire, il semble aussi important que les enseignants-chercheurs soient autonomes dans le traitement de leurs données, ou qu'ils détiennent au moins quelques bases, pour ne pas être entièrement dépendants d'autrui dans ce domaine.

Ainsi, il faudrait réfléchir à l'élaboration d'une chaîne d'archivage des données dont les chercheurs, en tant que producteurs de ces données, seraient les premiers maillons et assureraient les tâches basiques – sous réserve d'une formation aux bonnes pratiques en la matière. Ils seraient ensuite idéalement accompagnés et/ou relayés, en fonction des besoins, par un personnel spécialisé et dédié aux tâches de documentation, de stockage et de sécurisation des données.

3.3.3 Synthèse sur les pratiques de stockage et d'archivage

Alors que les enseignants-chercheurs font part de l'importance qu'ils accordent à la question du stockage et de l'archivage de leurs données, que ce soit par le fait même d'avoir répondu à cette enquête ou explicitement lors des entretiens individuels, leurs pratiques réelles pourraient laisser sous-entendre le contraire. Comment expliquer dès lors l'écart entre la théorie et la pratique ? L'utilisation de Google Drive par les chercheurs est par exemple une bonne illustration des contradictions qui traversent leurs pratiques. Ils sont une majorité à se méfier de ce service mais s'y résignent bien souvent par défaut.

Les enseignants-chercheurs ont indéniablement besoin d'être, au minimum, guidés dans la conservation de leurs données de recherche. Cela implique de prendre la mesure de la diversité de leurs pratiques effectives, de leurs besoins et de leurs préoccupations. Quelques priorités ont pu être identifiées : le stockage d'importants volumes de données (quantifiés en To), le stockage des données dans un espace collaboratif pour simplifier le travail en équipe et la sécurisation des données.

Il ressort de l'enquête que les enseignants-chercheurs de Rennes 2 assurent globalement un stockage artisanal de leurs données sans application d'une stratégie au long cours. C'était déjà le constat qui émanait de l'enquête de Lille 3 : les pratiques de stockage et d'archivage sont globalement personnelles et réalisées avec les seuls moyens directement à disposition (stockage local sur un support privé, avec une documentation minimale et non standardisée) bien qu'essentiellement faillibles et soumis à de nombreux risques (perte ou vol de matériel, illisibilité des formats). Certaines précautions de base sont tout de même prises, notamment le cumul des supports de stockage.

Les lacunes en matière d'archivage se révèlent liées à deux principales contradictions inhérentes au paysage actuel de la recherche en France :

- la première se trouve entre les ressources, aussi bien humaines que financières, des équipes de recherche et leurs besoins réels. Les laboratoires n'ont que trop rarement la possibilité de se doter chacun d'un matériel de stockage conséquent et sécurisé. Et quand bien même c'est le cas, encore faut-il du personnel qualifié et en nombre suffisant pour assurer le bon fonctionnement de ce matériel et la conservation des données qui sont stockées.

- une seconde contradiction subsiste entre les différentes temporalités de la recherche. Les chercheurs sont encouragés à assurer un archivage pérenne de leurs données. Or, les jeux de données qu'ils produisent sont souvent liés à différents projets de recherche dont les durées sont limitées et dont l'enchaînement n'est pas propice à un bon archivage. Les données d'un projet précédent se trouvent bien souvent abandonnées, là encore parce que l'archivage ne constitue pas une priorité pour les chercheurs au regard de leurs autres activités.

Les enseignants-chercheurs sont encouragés à faire preuve de plus en plus d'autonomie dans les tâches les plus élémentaires du stockage et de l'archivage de leurs données. Cela recouvre un ensemble de bonnes pratiques, auxquelles ils doivent être sensibilisés et formés, relatives notamment au choix des formats de données (en privilégiant les formats ouverts), au nommage de leurs fichiers, à la diversification des supports de stockage, à la sélection des données à conserver (le tri faisant partie intégrante de l'archivage) et à la normalisation des métadonnées. Cette autonomie est importante et doit être préservée le plus possible car c'est ce qui permet aux enseignants-chercheurs de garder le contrôle sur leurs données. Par ailleurs, le stockage et l'archivage sont deux problématiques complémentaires, qui traversent tout le cycle de vie des données. C'est pourquoi la conservation des données ne doit pas être pensée comme une étape finale mais plutôt comme un processus, dont la logique d'ensemble doit être prévue dès la production de ces mêmes données.

Toutefois, il ne s'agit ici que d'une première étape et il est impératif que cette question de l'archivage soit par la suite traitée à une échelle supérieure. Sur le plan national, les Très Grandes Infrastructures de Recherche en SHS se sont déjà emparées de cette préoccupation avec la mise en œuvre d'une offre de service adaptée en matière de stockage et d'archivage. Pourtant, l'utilisation de ces dispositifs reste minoritaire d'après l'enquête, laissant suggérer notamment que des actions de sensibilisation et de formation sont là encore nécessaires. Elles ne semblent pourtant pas manquer et on peut se demander alors si elles sont suffisantes.

En effet, la méconnaissance de l'offre de service ou le manque de compétences des enseignants-chercheurs en matière de gestion de données ne constituent pas les seuls facteurs faisant obstacle à la mise en place d'un archivage systématique des données de recherche. Deux autres éléments explicatifs se sont détachés au cours de l'enquête. D'une part, dans certains cas, les entretiens ont révélé qu'il serait pressant de renforcer les activités de formation par l'accompagnement et la collaboration à long terme d'un personnel spécialisé et dédié, qui soit

sur le terrain – à l'échelle de l'établissement ou des unités de recherche – et en prise avec le quotidien des enseignants-chercheurs. D'autre part, les services de soutien à la recherche déjà mis en place pourraient encore être améliorés afin de dépasser les contraintes qu'ils imposent – non à dessein – aux chercheurs. Parmi elles, on peut citer par exemple les diverses démarches impliquées par l'archivage des données vécues comme une charge supplémentaire pour les enseignants-chercheurs, une ergonomie des outils qui n'est pas toujours transparente ou adéquate pour les utilisateurs ou encore l'impression déplaisante d'une perte de contrôle sur les données. Autant d'éléments qui peuvent paraître parfois minimes, parfois irréductibles, mais qu'il s'agirait néanmoins de prendre en compte et d'identifier plus précisément pour tenter de les dissiper.

3.4 Pratiques de partage et de diffusion des données de recherche

3.4.1 Résultats de l'enquête statistique

Sur le partage des données, l'enquête statistique a cherché à cerner trois éléments : les pratiques réelles, les opinions et représentations, les souhaits et besoins exprimés. Ce qui frappe le plus est sans doute l'écart, parfois les contradictions, entre ces trois aspects.

3.4.1.1 Quelles pratiques du partage des données ?

Trois questions visaient à cerner la réalité des pratiques du partage des données : quelle est l'accessibilité réelle des données des chercheurs, quelles sont les pratiques d'accès aux données des autres chercheurs, quelles sont les pratiques de stockage des données en libre accès.

A/ Sur l'accessibilité des données des chercheurs

Sur l'accessiblité, nous avons repris la question de l'enquête lilloise, « *En dehors de vous, qui peut accéder directement à vos données ?* » : 116 personnes sur 143, soit un taux de 81 % de réponses, ont répondu à la question 21, optionnelle. Et comme à Lille 3[33], le résultat est « simple et sans équivoque » : pour 54 % des répondants, personne d'autre ne peut accéder aux données des chercheurs (à Lille 3, le chiffre est plus élevé : 64 %). En bref, les données de recherche, collectées ou produites, sont largement inaccessibles et réservées, soit au chercheur lui-même (54 %), soit à son groupe de travail (42 %). Mais il faut relever que le total des enseignants-chercheurs qui partagent leurs données (avec leurs collègues de la discipline, de l'université, d'autres publics spécifiques ou tout le monde) est cependant loin d'être négligeable (22 %), et nettement supérieur au pourcentage de Lille 3 (14,8 %).

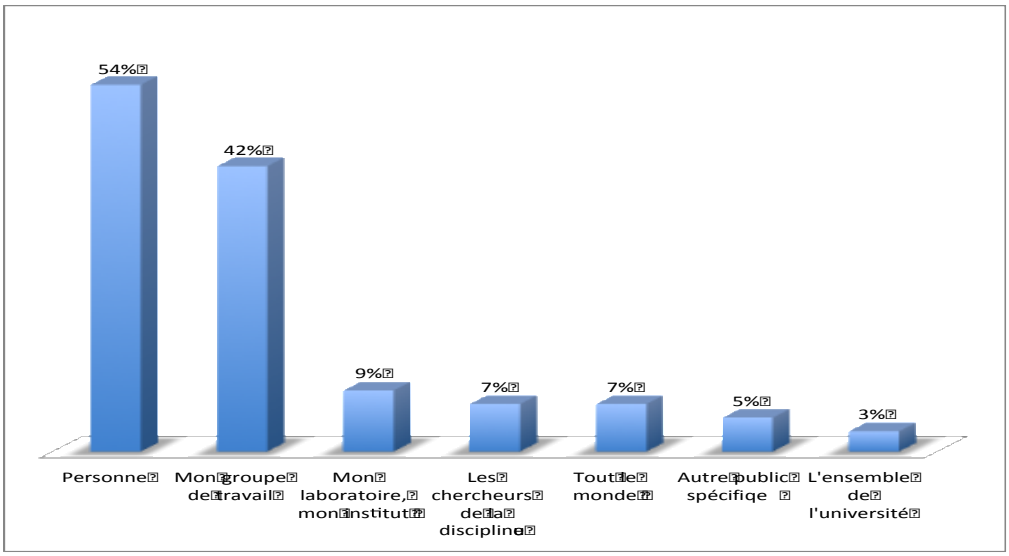


Figure 34 : Accessibilité des données des chercheurs (N = 116)

[33] Hélène Prost, Joachim Schöpfel. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3*., p. 18. http://hal.univ-lille3.fr/hal-01198379v1

B/ Sur les pratiques d'accès aux données d'autres chercheurs

Nous avons repris, là encore, une question de l'enquête lilloise (« *Avez-vous déjà téléchargé des données à partir d'une archive ouverte ?* »), en élargissant un peu la formulation : « *Avez-vous déjà téléchargé les données d'autres chercheurs ?* ». Cette question n° 22, obligatoire, a reçu 116 réponses sur 143, soit un taux de réponse de 81 % (nettement supérieur à celui de Lille, de 58 %).
Si la tendance reste la même dans les deux universités, à savoir que le téléchargement des données d'autres chercheurs reste une pratique minoritaire, les chiffres sont un peu différents :
- à Rennes 2, 56 % de Non contre 44 % de Oui ;
- à Lille 3, 62 % de Non contre 38 % de Oui.
On observe donc une pratique de partage de données légèrement supérieure à Rennes 2, dans la continuité de la question précédente.

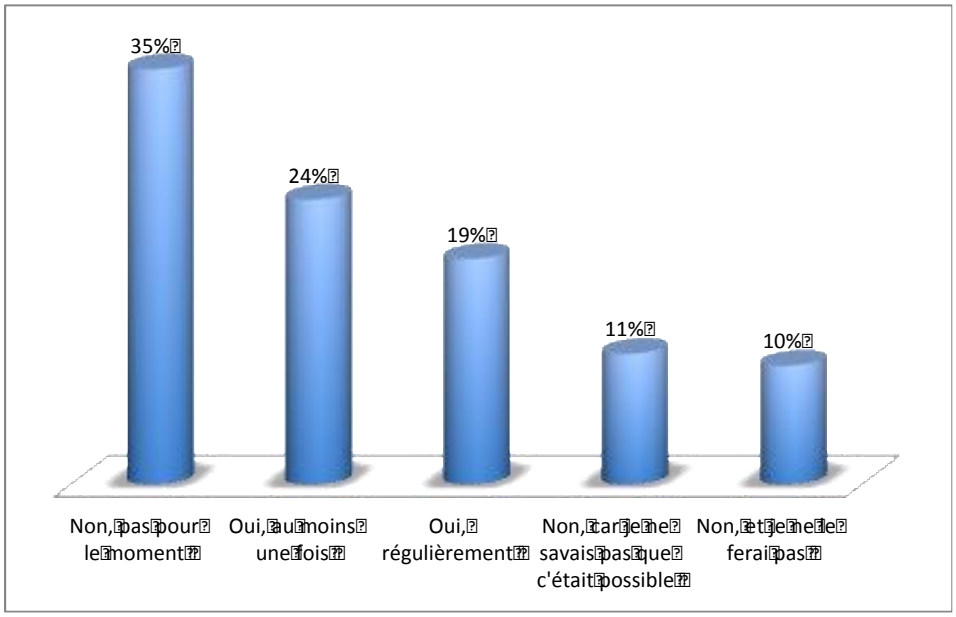


Figure 35 : Téléchargement des données d'autres chercheurs (N = 116)

C/ Sur la pratique du libre accès

Pour compléter la question sur l'accessibilité des données, nous avons voulu en avoir le cœur net sur la pratique de stockage des données en libre accès. Une question obligatoire, avec une seule réponse possible, a donc été introduite (question n° 23 : « *Vous-même, avez-vous déjà stocké vos données en libre accès sur le web ?* »), qui a recueilli 116 réponses sur 143, soit un taux de réponses de 81 %.
Le total des Oui est de 24 %, en cohérence avec le chiffre de 22 %, relevé plus haut dans la question 21[34], celui des Non de 76 %.

[34] La différence s'expliquant par les différentes réponses possibles de la question 21.

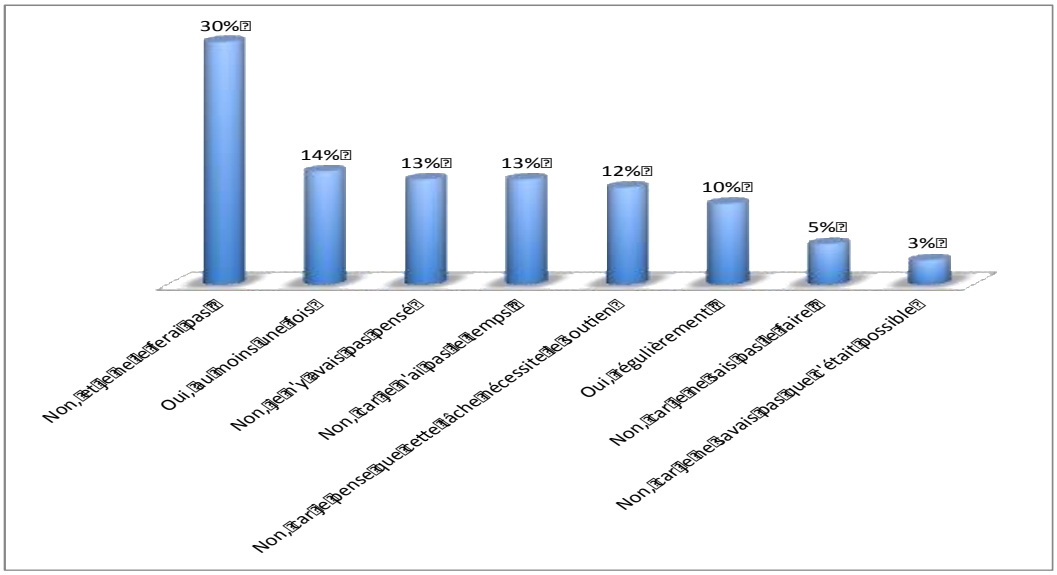


Figure 36 : Avez-vous déjà stocké vos données en libre accès sur le web ? (N = 116)

Au total, on peut résumer ainsi les pratiques de partage des données de recherche à l'Université Rennes 2 :
- si l'on voit le verre à moitié vide, on relèvera que les ¾ des répondants n'ont jamais déposé leurs données de recherche en libre accès, que plus de la moitié n'ont jamais téléchargé les données d'autres chercheurs (56 %), ou donné l'accès à quiconque (54 %) ;
- à l'inverse, si l'on considère le verre à moitié plein, on observera que plus de 40 % ont déjà partagé leurs données avec d'autres chercheurs (46 %), ou téléchargé les données des autres (44 %), et qu'environ un quart des répondants (24 %) ont déjà déposé leurs données de recherche en libre accès.

Si l'on compare avec les résultats de Lille 3, les pratiques de partage des données semblent un peu plus développées à Rennes 2. Mais dans les deux universités de SHS, le partage de ses données est encore loin de faire partie de la pratique ou de la préoccupation courante des chercheurs.

3.4.1.2 Quelles représentations, quelles opinions sur le partage des données de recherche ?

Quels peuvent être les facteurs d'incitation ou de frein au partage des données ? Quelle est la position des enseignants-chercheurs sur le libre accès en général, et sur la libre diffusion des données de recherche en particulier ? Sans forcément respecter leur ordre de classement, nous présentons ici les résultats de plusieurs questions portant sur ce qu'on pourrait appeler les « représentations », les valeurs de référence, les opinions.

A/ Sur les facteurs d'incitation au libre accès de ses données de recherche

La question n° 25 (« *Parmi les raisons suivantes, qu'est-ce qui vous inciterait à rendre vos données de recherche accessibles en libre accès ? »),* et la question symétrique n° 26 (« *Quelles seraient les principales raisons qui vous empêcheraient ou vous freineraient pour rendre vos données de recherche accessibles en libre accès ?* »), visaient à cerner les représentations du partage des données chez les enseignants-chercheurs, plus précisément les raisons inclinant au

partage, ou au contraire le dissuadant. Ces deux questions, d'abord obligatoires, ont ensuite été rendues optionnelles.
La question 25 a recueilli 107 réponses sur 143, soit un taux de réponse de 75 % (36 personnes n'y ont pas répondu). Plusieurs réponses possibles étaient proposées, que rappelle le graphique ci-dessous :

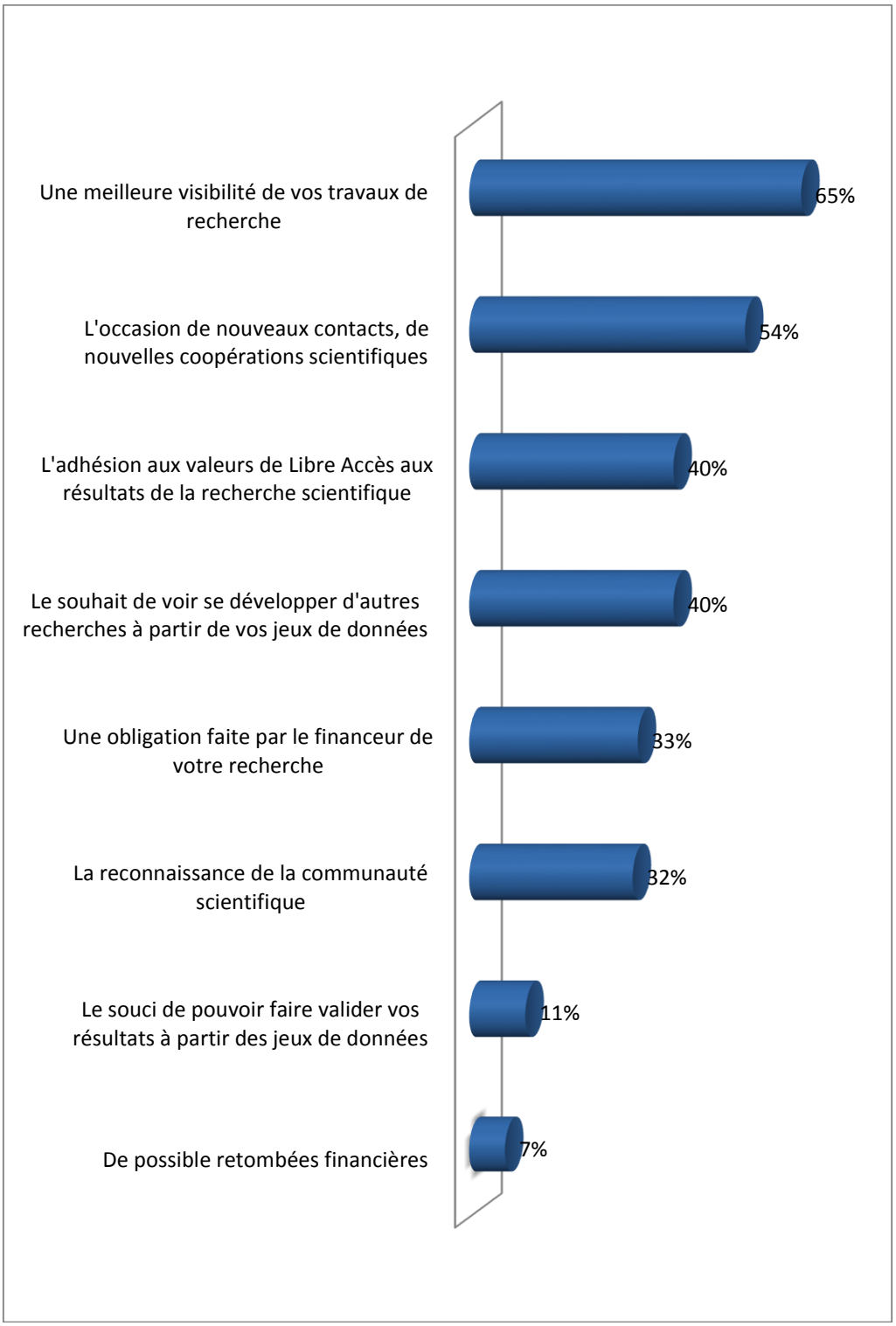


Figure 37 : Facteurs d'incitation au libre accès de ses données de recherche (N = 107)

On y voit la prédominance très nette de deux facteurs, classiques et tout à fait complémentaires : d'une part, la recherche de visibilité de ses travaux de recherche (65 %), d'autre part la recherche de contacts et de coopérations scientifiques (54 %). À ces deux premières réponses s'ajoute la reconnaissance de la communauté scientifique (32 %). Autrement dit, la dimension sociale de la science l'emporte largement, dans les facteurs d'incitation au partage. Mais la dimension plus scientifique occupe également une place importante : le souhait de voir se développer d'autres recherches à partir de ses jeux de données (40 %), le souci de validation des résultats (11 %), ou l'adhésion aux valeurs du Libre Accès (40 %), si l'on classe ce motif dans les raisons scientifiques. Le facteur institutionnel (l'obligation faite par le financeur de la recherche) est également non négligeable, en recueillant un tiers des réponses (33 %). On relèvera enfin la faiblesse du facteur d'incitation financière (7 %).

B/ Sur les facteurs de frein au libre accès de ses données de recherche

La question 26 « *Quelles seraient les principales raisons qui vous empêcheraient ou vous freineraient pour rendre vos données de recherche accessibles en libre accès ?* », également optionnelle, a mobilisé 108 répondants sur 143**,** soit un taux de réponse de 75,5 %. Plusieurs réponses étaient possibles, et le total des items cochés est de 254.

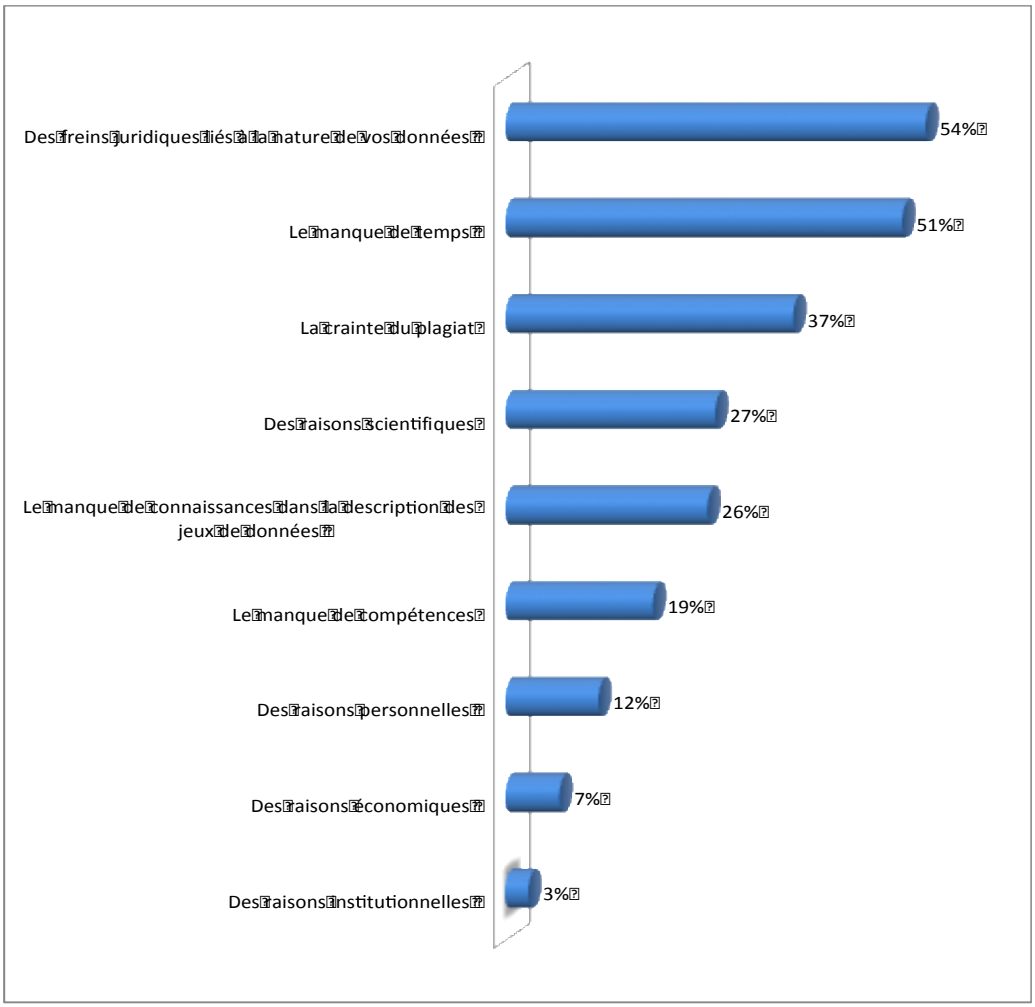


Figure 38 : Facteurs de frein au libre accès de ses données de recherche (N = 108)

Par ailleurs, douze réponses ont été cochées « Autre », et permettent de préciser les représentations des enseignants-chercheurs sur les freins au libre accès[35].

Au final, quels sont les principaux obstacles à la diffusion des données de recherche en libre accès, selon les chercheurs ? Sans surprise, ce sont les freins juridiques liés à la nature des données qui arrivent en tête avec 58 réponses, soit 54 % des répondants, et 22,8 % des réponses. Les chercheurs en SHS travaillent souvent sur des données personnelles, protégées, et les raisons éthiques et juridiques ont été souvent avancées comme le premier facteur empêchant une libre diffusion des données. Nous sommes ici au cœur d'un problème bien connu des spécialistes : la dimension juridique est l'une des plus complexes des multiples facettes des données de recherche.

Le manque de temps (qu'il soit réel ou qu'il corresponde à une impression subjective, peu importe) arrive immédiatement en deuxième position avec 55 réponses, soit 51 % des répondants. Nous avons relevé que sur ces 55 réponses, 18 répondants ont cité cet item en réponse unique, ce qui témoigne de la force de ce ressenti. Cette réponse exprime le sentiment courant, largement partagé chez les enseignants-chercheurs, d'être constamment submergé par la diversité des tâches. L'organisation de la diffusion de ses données de recherche est perçue d'abord comme une nouvelle contrainte, s'ajoutant aux nombreuses autres et son caractère prioritaire n'étant pas d'une évidence aveuglante, il est assez logique que le manque de temps soit évoqué par un peu plus de la moitié des répondants.

La troisième réponse est à la fois assez prévisible et néanmoins problématique : en effet, la crainte du plagiat, qui recueille 40 réponses, soit 37 % des répondants, est largement partagée par de nombreux enseignants-chercheurs, notamment en SHS, mais elle témoigne aussi d'une incompréhension, ou d'une méconnaissance, des mécanismes actuels de protection des données et des publications, dans un régime d'open access. Rappelons qu'une publication ou un jeu de données, diffusé en libre accès sous un régime de licence Creative Commons (ou autre), est beaucoup mieux protégé contre le plagiat qu'un fichier ou un texte mal stocké sur une plateforme de partage... Par ailleurs, la notion même de plagiat des données est discutable au plan juridique, étant donné la nature, très diversifiée, de la propriété intellectuelle des données[36]. Mais il est néanmoins intéressant de relever que cette crainte du plagiat, qui relève d'une inquiétude excessive ou d'une méconnaissance du droit d'auteur, arrive en troisième position dans les facteurs de frein au partage des données.

Le quatrième frein invoqué, cité par 29 personnes, soit 27 % des répondants, est sans doute le plus intéressant à relever : les obstacles scientifiques à la diffusion des données constituent une raison de fond, une véritable objection au partage des données. Si le questionnaire ne précisait pas de quels obstacles il s'agissait, l'une des réponses libres, dans les réponses « Autres », nous a mis sur la voie : « *hors de leur contexte de production, des données ethnographiques / qualitatives n'ont plus de sens.* ». La principale raison scientifique invoquée, qui a été longuement exposée dans certains entretiens, porte sur la nécessaire contextualisation des données, notamment des données d'entretiens (en géographie sociale et en socio-linguistique). Nous y revenons plus loin.

[35] On en trouvera la liste dans l'Annexe 1, Résultats statistiques, p. 31.
[36] Rappelons simplement qu'au plan juridique, le droit des données dépend avant tout de la nature de celles-ci. Voir sur Doranum « Aspects juridiques et éthiques » : http://doranum.fr/logigramme-aspects-juridiques-ethiques/

Une répartition différente des réponses peut donner à voir une autre interprétation intéressante de ces résultats. En effet, les neuf facteurs de frein, proposés dans le questionnaire, peuvent se répartir, selon nous, en deux ensembles :
- d'une part, les raisons « externes », liées aux aspects juridiques, scientifiques, économiques, institutionnels... des données de recherche ; les obstacles à la libre diffusion sont à relever ici du côté des données de recherche elles-mêmes, ou du contexte scientifique ; ce premier ensemble regroupe quatre des cinq réponses proposées ;
- d'autre part, un ensemble de facteurs que nous qualifierons d'« internes », dans la mesure où ils sont davantage liés aux perceptions, aux représentations personnelles des enseignants-chercheurs. Cinq réponses sont regroupées ici.

Ensembles des facteurs	Réponses proposées	Nombre de réponses	% // répondants (N = 108)	% // réponses (N = 254)
Facteurs « externes »				
	Freins juridiques	58	54 %	22,8 %
	Raisons scientifiques	29	27 %	11,4 %
	Raisons économiques	8	7 %	3,1 %
	Raisons institutionnelles	3	3 %	1,2 %
Sous-total :		**98**		**38,5 %**
Facteurs « internes »				
	Manque de temps	55	51 %	21,6 %
	Crainte du plagiat	40	37 %	15,7 %
	Manque de connaissance dans la description des données	28	26 %	11,0 %
	Manque de compétences	20	19 %	7,8 %
	Raisons personnelles	13	12 %	5,1 %
Sous-total :		**156**		**61,4 %**
Total :		**254**		

Figure 39 : Tableau de synthèse sur les facteurs de frein au libre accès de ses données de recherche

Ce tableau permet de compléter le premier graphique, centré sur chaque réponse spécifique, en mettant en évidence des faisceaux de freins au libre accès. Il montre clairement que la majeure partie des obstacles à la diffusion des données de recherche provient des représentations, des perceptions des enseignants-chercheurs, et non des contraintes extérieures (plus de 61 % des réponses contre 38 %).
En clair, l'enquête confirme une intuition assez courante, visible également dans les résultats d'autres enquêtes (comme celle de Lille 3) : les principaux freins à la diffusion des données de recherche sont à chercher chez les producteurs des données eux-mêmes (pour diverses raisons), plus que dans les contraintes objectives liées aux données.

C/ Sur l'opinion générale sur le partage des données de recherche et ses enjeux

Il est intéressant de mettre en regard les résultats des questions 25 et 26 (sur les incitations et les freins) avec la question 24, plus globale, qui portait sur la position de principe sur la diffusion

des données (« *Seriez-vous d'accord pour diffuser vos données de recherche en libre accès (à condition qu'elles soient diffusables et protégées juridiquement) ? »).* Cette question, obligatoire, a recueilli 115 réponses sur 143, soit un taux de réponse de 80,4%. Et elle montre un premier écart intéressant entre les pratiques, les représentations et les opinions, puisque 66 % des répondants se déclarent en faveur du partage des données (total des « Oui, tout à fait », « Oui, après un délai » et « Plutôt oui »), contre 27 % qui s'y opposent, et 8 % qui ne savent pas.

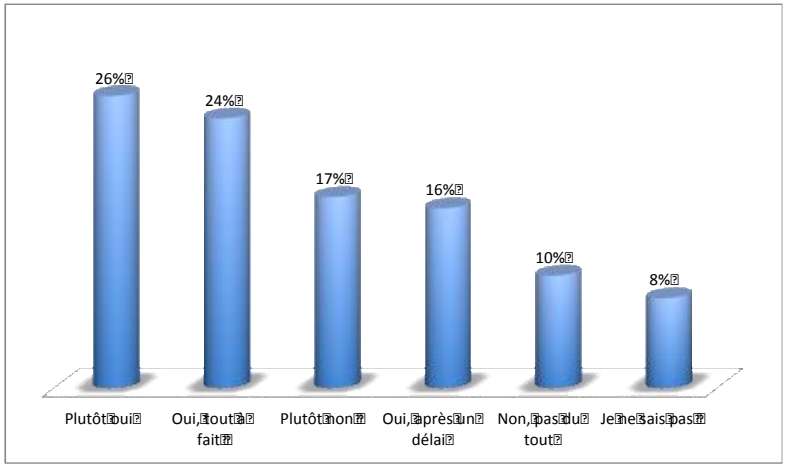


Figure 40 : Position sur la diffusion en libre accès de ses données de recherche (N = 115)

Trois items d'une autre question (la question 29 : « *Dans quelle mesure êtes-vous d'accord ou pas d'accord avec ce qui suit ?* ») confirment cette position de principe nettement favorable au partage des données :

- l'item « *Je suis prêt à partager mes données de recherche avec les autres* », qui a obtenu 71 % de participation, recueille une nette majorité (67 réponses, soit plus de 65 %) en faveur du partage. Mais on trouve ici un phénomène bien connu d'adhésion spontanée à un énoncé se voulant « positif » (il est toujours délicat de se prononcer contre le partage).

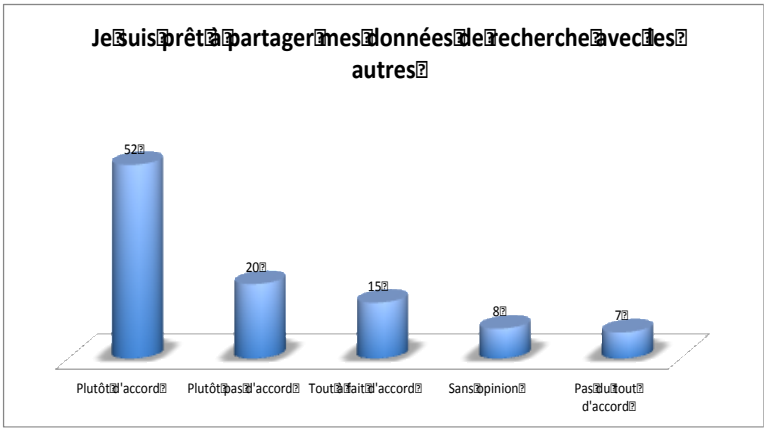


Figure 41 : Sur le partage de ses données de recherche (N = 102)

- un autre item atteste également, sinon d'un positionnement de principe en faveur du partage, du moins d'une sensibilité à l'importance d'une déontologie des données. En effet, une très nette majorité se déclare d'accord avec l'énoncé suivant (83 réponses, soit plus de 84 %) :

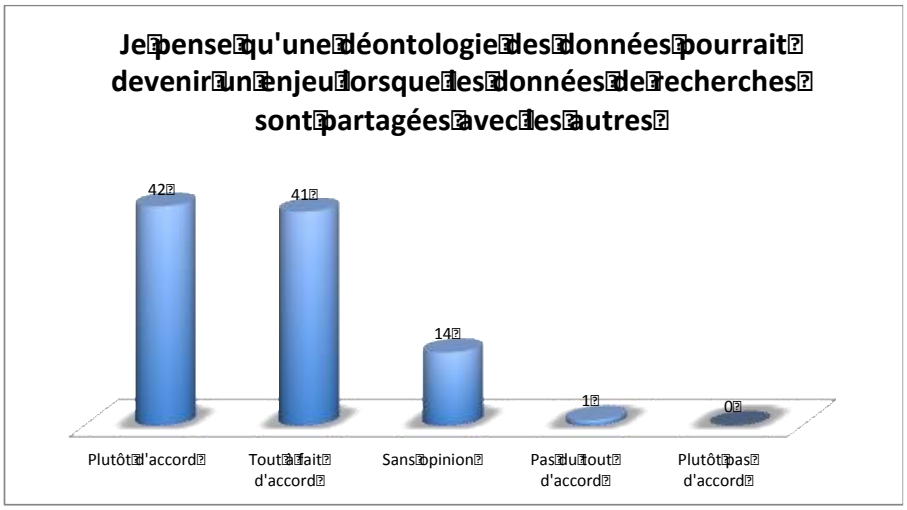


Figure 42 : Sur l'enjeu déontologique dans le partage des données (N = 98)

- enfin, le troisième énoncé portait sur la familiarisation avec les enjeux de l'Open access : les résultats sont beaucoup plus nuancés, car sur les 97 réponses obtenues (soit 68 % de participation), un peu moins de la moitié (48, soit 49,4 %) se déclarent en accord avec l'énoncé, contre 37 (soit 38 %) en désaccord.

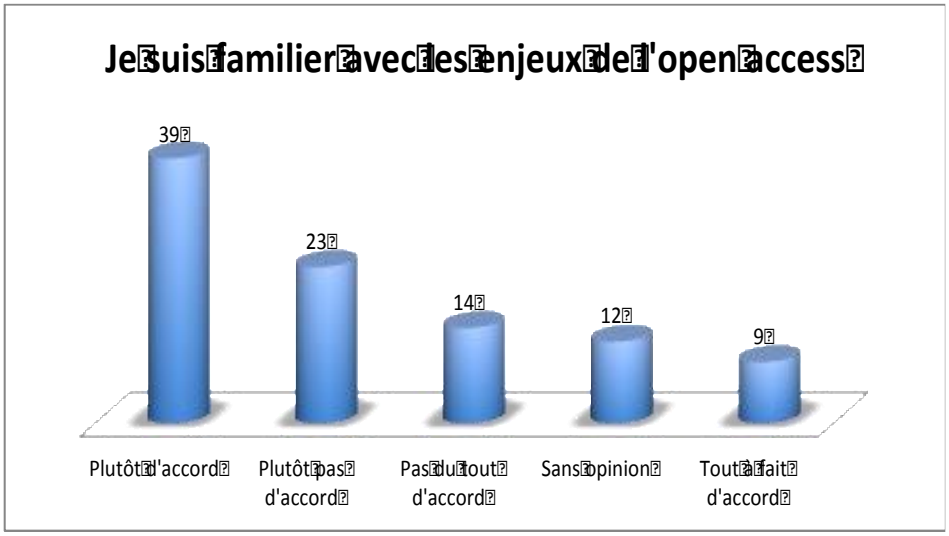


Figure 43 : Sur la connaissance des enjeux de l'Open Access (N = 97)

Au final, l'enquête statistique nous montre une articulation complexe entre pratiques, représentations et opinions, que l'on pourrait résumer ainsi : des pratiques de libre accès minoritaires, qui s'expliquent par différentes perceptions négatives de la diffusion des données, mais qui sont contredites par une position de principe favorable au partage des données.

3.4.1.3 Quels besoins, quels souhaits en matière de partage des données ?

Quels sont les principaux souhaits exprimés par les enseignants-chercheurs sur le partage des données ? Une question spécifique et l'item d'une autre question donnent quelques éléments de réponse.

A/ Sur le support à privilégier pour le dépôt des données

La question 27 portait sur le support à privilégier pour un dépôt des données en libre accès : « *Si vous souhaitez diffuser vos données en libre accès, quel est (ou serait) votre support principal ?* ». Cette question optionnelle a recueilli 96 réponses sur 143, soit un taux de réponses de 67 %.

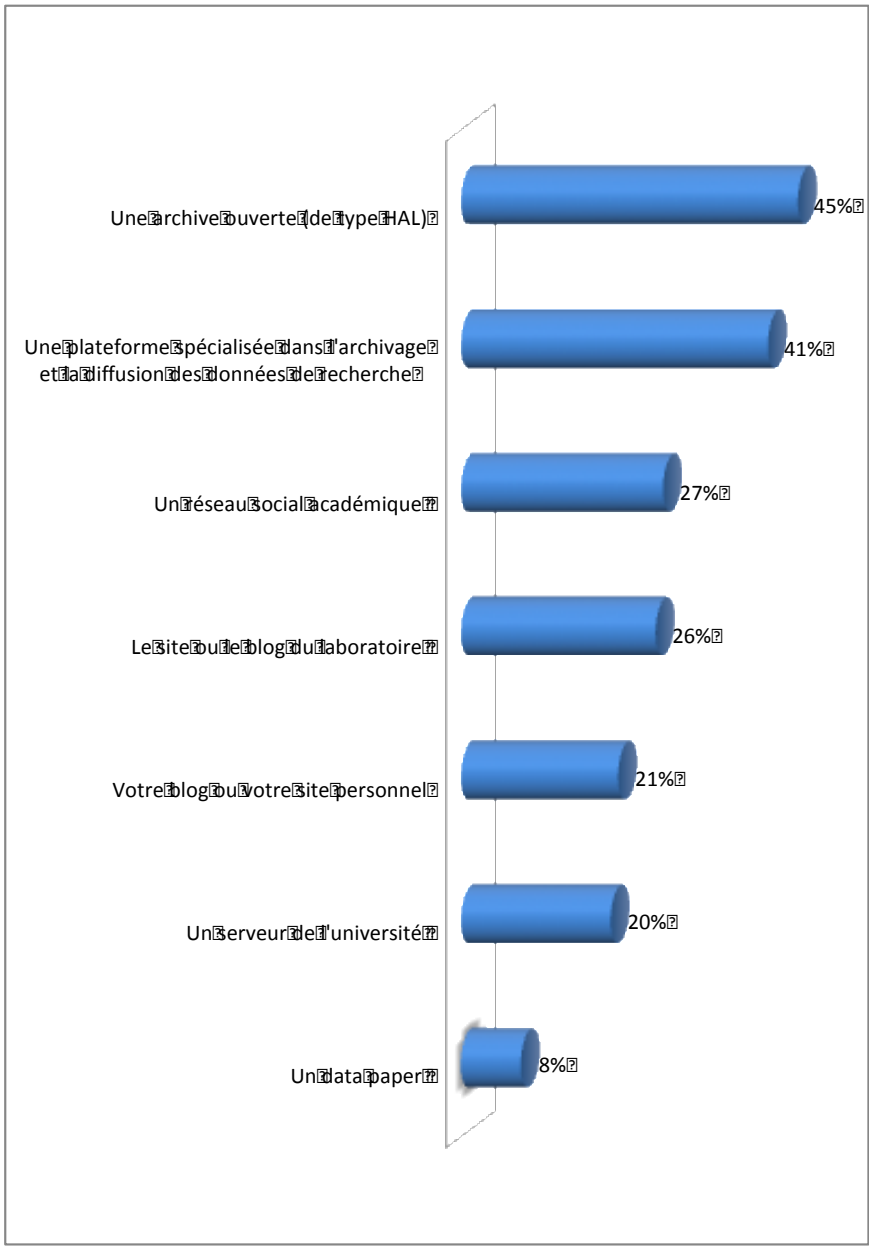


Figure 44 : Support à privilégier pour un dépôt en libre accès (N = 96)

Les résultats placent nettement en tête des dépôts souhaités les archives ouvertes (comme HAL), avec 43 réponses (soit 45 %) et les plateformes spécialisées (de type infrastructure de recherche, comme Nakala en SHS), avec 39 réponses (soit 41 %). Les autres supports envisagés (réseaux sociaux académiques, blogs, serveur de l'université...) arrivent plus loin derrière. Au final, sur le total des 180 réponses, les supports spécialisés dans la diffusion des données de recherche, archives ouvertes et infrastructures, sont plébiscités à 45,5 %. Notons aussi le très

faible score du serveur de l'université (19 réponses, soit 20 % des répondants et 10,5 % des réponses).
Au-delà de ces résultats intéressants (et encourageants pour les tenants des archives ouvertes et des plateformes spécialisées), il convient néanmoins de pointer, là encore, l'écart entre les souhaits et les pratiques réelles : ainsi, si 41 % des enseignants-chercheurs souhaiteraient déposer leurs données sur des entrepôts spécialisés, seulement 6 % d'entre eux les stockent réellement sur le serveur d'une TGIR, comme le montrent les résultats de la question 13 « Où stockez-vous les données de recherche ? ».

B/ Sur les besoins exprimés, liés au partage des données

La question 30, qui visait à recenser les besoins (« *Sur quels points souhaiteriez-vous bénéficier d'une aide (de votre unité de recherche, de la DSI, du SCD, de l'URFIST, de la MSHB...) ?* »), comportait un item portant sur le partage et la publication des données de recherche.

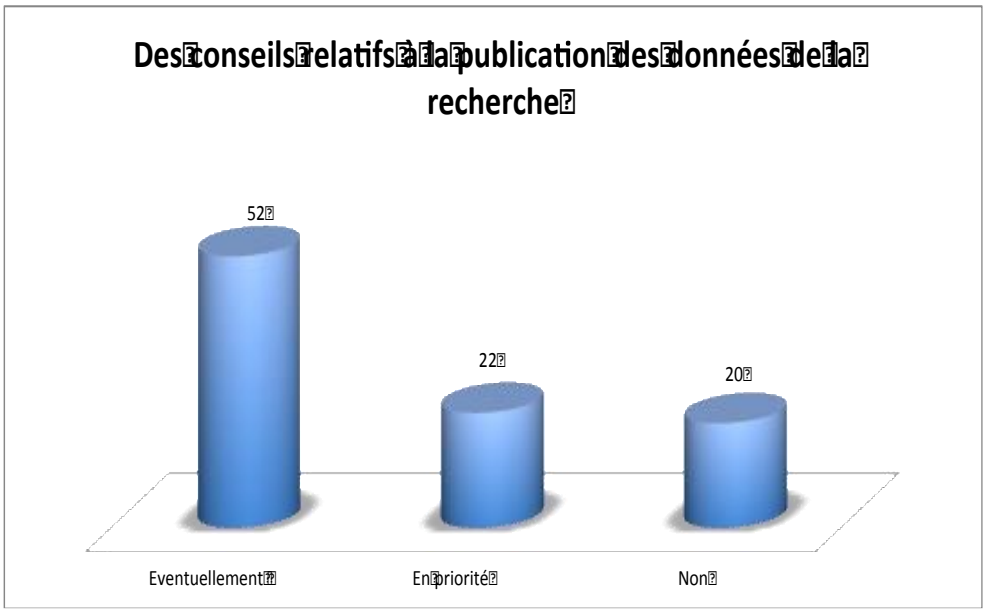


Figure 45 : Besoin de conseils sur la publication des données de recherche (N = 94)

Cet item a obtenu 94 réponses, soit un taux de 66 %, et le graphique montre clairement qu'il ne s'agit pas d'une priorité pour les enseignants-chercheurs, puisque plus de la moitié (55 %) souhaitent *éventuellement* des conseils relatifs à la publication des données, 21 % les refusent, et seulement 23 % les considèrent *prioritaires*. Et dans le classement des items par ordre décroissant de priorité[37] les conseils relatifs à la publication des données de recherche arrivent en dernier.

3.4.2 Quelles variables sur les pratiques de partage des données ?

Plusieurs croisements ont été effectués entre diverses questions, pour chercher à savoir si des variables (et lesquelles) intervenaient dans les pratiques et les représentations des enseignants-chercheurs sur le partage des données. Trois ensembles de variables ont été déterminées : les variables « socio-démographiques » (l'âge, le statut et les responsabilités), scientifiques (le laboratoire, la discipline) et celles liées à la nature des données.

[37] Voir Annexe 1, Résultats statistiques, figure 34, p. 39

3.4.2.1 La variable de l'âge

D'après les croisements effectués sur la question 1 (la tranche d'âge des répondants), il semble bien que la variable générationnelle intervienne, certes de manière tendancielle, dans le positionnement des enseignants-chercheurs face au partage des données. Le résultat des croisements montre en effet que les dispositions les plus favorables au partage et à la diffusion des données, notamment la sensibilité aux valeurs du Libre Accès, se trouvent plutôt chez les plus de 50 ans. Voici un tableau de synthèse, qui reprend les résultats positifs (i.e. le sous-total des réponses Oui, D'accord, ou En priorité), des deux tranches d'âge[38], à huit questions portant sur le partage des données. Les pourcentages sont calculés sur le rapport entre le nombre de réponses dans une tranche d'âge et le total des répondants de cette tranche (ainsi, 46,5 % des plus de 50 ans ayant répondu à la question 22 sur le téléchargement des données ont coché des réponses positives). On trouvera dans l'Annexe 2 tous les chiffres et les résultats détaillés des croisements.

N° des questions	Questions	30-50 ans	> 50 ans	Diff.
Q 22	J'ai déjà téléchargé les données d'autres chercheurs	40 %	46,5 %	+ 6,5
Q 23	J'ai déjà stocké mes données en libre accès sur le web	20,8 %	27,9 %	+ 7,1
Q 24	Je suis d'accord pour diffuser mes données de recherche en libre accès (à condition qu'elles soient diffusables et protégées juridiquement)	63,3 %	69,7 %	+ 6,4
Q 25	L'adhésion aux valeurs de libre Accès aux résultats de la recherche scientifique (comme facteur d'incitation au partage des données)	32,3 %	52,6 %	+ 20,3
Q 29	Je suis familier avec les enjeux de l'open access	44,2 %	57,1 %	+ 12,9
Q 29	Je suis prêt à partager mes données de recherche avec les autres	63,4 %	68,4 %	+ 5
Q 29	Je pense qu'une déontologie des données pourrait devenir un enjeu lorsque les données de recherches sont partagées avec les autres	85,2 %	83,3 %	- 1,9
Q 30	Je souhaiterais bénéficier de conseils relatifs à la publication des données de la recherche (en priorité)	20,3 %	29,4 %	+ 9,1

Figure 46 : Synthèse de la variable de l'âge dans le partage des données

Dans la quasi-totalité des items (sauf celui sur la déontologie des données), on relève des valeurs plus élevées chez les plus de 50 ans, avec des différences variables, de + 5 points (pour la prédisposition au partage des données) à + 20,3 points, écart le plus important, portant sur l'adhésion aux valeurs du Libre Accès comme facteur d'incitation au partage.
Ainsi, aussi bien sur les pratiques de diffusion que sur les opinions, les représentations ou les besoins exprimés, les enseignants-chercheurs de plus de 50 ans sont proportionnellement plus

[38] Nous avons volontairement ignoré la troisième tranche d'âge des moins de 30 ans, qui n'avait qu'un seul répondant.

nombreux que leurs collègues plus jeunes à exprimer des positions favorables au partage des données.

Les différences entre les tranches d'âge peuvent s'exprimer de deux manières :
- soit dans les pourcentages des réponses par rapport au nombre total de réponses de la tranche d'âge, comme ci-dessus (autrement dit, combien de 30-50 ans ont répondu à tel ou tel item d'une question) ;
- soit dans les écarts à la représentation de chaque tranche d'âge, *i.e.* dans le calcul du pourcentage de réponses à un item, et dans sa comparaison avec le pourcentage des répondants de la tranche d'âge (autrement dit, est-ce que les 30-50 ans répondent plus, ou au contraire moins, que leur part dans l'ensemble des répondants). Les deux tableaux suivants expriment ce mode de calcul.

A/ Sur les facteurs d'incitation au libre accès

Croisement des questions 1 et 25 : « Parmi les raisons suivantes, qu'est-ce qui vous inciterait à rendre vos données de recherche accessibles en libre accès ? » (107 répondants et 303 réponses). Le tableau ci-dessous, qui paraît un peu compliqué, se lit de la manière suivante :
- 68 personnes de 30-50 ans ont répondu à la question 25, soit 63,5 % des 107 répondants ;
- pour le premier item (l'obligation du financeur de la recherche), 23 personnes de 30-50 ans ont coché cette réponse, soit 65,7 % du total des 35 réponses à cet item ;
- pour cet item, l'écart de représentation (entre le pourcentage de la tranche d'âge sur l'item et son échantillon global) est donc de + 2,2 %.

Items	**< 30 ans (1)**	**30-50 ans (68)**		**> 50 ans (38)**		**Total (107)**
		% // total réponses	Ecart // échantillon **63,5 %**	% // total réponses	Ecart // échantillon **35,5 %**	
Une obligation faite par le financeur de votre recherche		23 *65,7 %*	+ 2,2 %	12 *34,2 %*	- 1 ,3 %	**35**
La reconnaissance de la communauté scientifique	1	18 *52,9 %*	- 10,6 %	15 *44,1 %*	+ 8,6 %	**34**
une meilleure visibilité de vos travaux de recherche	1	39 *55,7 %*	- 7,8 %	30 *42,8 %*	+ 7,3 %	**70**
L'occasion de nouveaux contacts, de nouvelles coopérations...	1	38 *65,5 %*	+ 2 %	19 *32,7 %*	-2,8 %	**58**
Le souci de pouvoir faire valider vos résultats ...		9 75 %	+ 11,5 %	3 25 %	- 10,5 %	**12**
Le souhait de voir se développer d'autres recherches à partir de vos jeux de données	1	25 *58,1 %*	- 5,4 %	17 *39,5 %*	+ 4 %	**43**

Items	< 30 ans (1)	30-50 ans (68)		> 50 ans (38)		Total (107)
		% // total réponses	Ecart // échantillon **63,5 %**	% // total réponses	Ecart // échantillon **35,5 %**	
L'adhésion aux valeurs du Libre Accès ...	1	22 *51,1 %*	- 12,4 %	20 *46,5 %*	+ 11 %	**43**
De possible retombées financières		5 62,5 %	- 1 %	3 37,5 %	+ 2 %	**8**
Total des réponses	**5**	**179**		**119**		**303**

Figure 47 : Variable de l'âge dans les facteurs d'incitation au partage des données

Du côté des similitudes entre les tranches d'âge, on relève quatre items, pour lesquels les écarts entre le pourcentage de réponses et celui de la tranche d'âge sont faibles ou très faibles :
- **les retombées financières** : cet item, qui arrive en dernier avec un très faible nombre de réponses, ne marque presque aucun écart entre les deux tranches d'âge (- 1 % et + 2 %) ;
- **l'obligation du financeur** : + 2 % ou – 1,3 % ;
- **l'occasion de nouveaux contacts** : + 2 % chez les 30-50 ans, – 2,8 % chez les > 50 ans ;
- **le souhait de voir se développer d'autres recherches** : les écarts de représentation sont un peu plus forts (+ 4 % chez les > 50 ans, – 5,4 % chez les 30-50 ans), mais sans doute pas encore suffisamment significatifs.

Sur ces quatre facteurs d'incitation au partage des données, on ne décèle donc pas de véritable impact de la variable de l'âge sur les réponses.

Du côté des différences, on trouve les quatre autres items :
- **une meilleure visibilité** : classé premier facteur d'incitation dans les deux tranches d'âge, cet item recueille cependant une part plus grande chez les plus de 50 ans, avec un écart de + 7,3 % par rapport à leur total ; chez les 30-50 ans, l'écart est à l'inverse de - 7,8 % ;
- **la reconnaissance de la communauté scientifique** : plus forte chez les > 50 ans, avec un écart de + 8,6 % par rapport à leur part dans le total des répondants (44,1 % contre 35,5 %) ; à l'inverse, les 30-50 ans, qui représentent 63,5 % des 107 répondants, ne constituent que 52,9 % des réponses sur cet item ;
- **le souci de pouvoir faire valider vos résultats à partir des jeux de données** : en revanche, cet item est plus fort chez les 30-50 ans, avec un écart de + 11,5 % par rapport à leur total (tandis que l'écart est négatif, chez les > 50 ans, avec - 10,5 %) ; mais les chiffres sont assez faibles et il faut se méfier des sur-interprétations ;
- **l'adhésion aux valeurs du Libre Accès** : elle est nettement plus forte chez les > 50 ans, avec un écart de + 11 % par rapport à leur total, contre un écart négatif de - 12,4 % chez les 30-50 ans ; les chiffres absolus étant significatifs (avec un total de 43 réponses), c'est sur cet item que les différences semblent les plus fortes entre les deux tranches d'âge, comme nous l'avons vu sur le premier tableau.

B/ Sur les facteurs de frein au libre accès

Croisement des questions 1 et 26 : « *Quelles seraient les principales raisons qui vous empêcheraient ou vous freineraient pour rendre vos données de recherche accessibles en libre accès ?* » (108 répondants et 254 réponses)
Nous avons refait le même type de tableau et de calcul des écarts de représentation dans les réponses des deux tranches d'âge.

	< 30 ans (1)	**30 - 50 ans (66)**		**> 50 ans (41)**		**Total (108)**
	0,9 %	% // total réponses	Ecart // échantillon **61,1 %**	% // total réponses	Ecart // échantillon **37,9 %**	
des freins juridiques...		38 *65,5 %*	+ 4,4 %	20 *34,4 %*	- 3,5 %	**58**
des raisons économiques		5 *62,5 %*	+ 1,4 %	3 *37,5 %*	- 0,4 %	**8**
des raisons scientifiques		18 *62 %*	+ 0,9 %	11 37,9 %	0 %	**29**
la crainte du plagiat	1	25 *62,5 %*	+ 1,4 %	14 *35 %*	- 2,9 %	**40**
des raisons institutionnelles		3 *100 %*	+ 38,9 %			**3**
des raisons personnelles	1	8 *61,5 %*	+ 0,4 %	4 *30,7 %*	- 7,2 %	**13**
le manque de temps	1	31 *56,3 %*	- 4,8 %	23 *41,8 %*	+ 3,9 %	**55**
le manque de connaissances dans la description des jeux de		15 *53,5 %*	- 7,6 %	13 *46,4 %*	+ 8,5 %	**28**
le manque de compétences		7 *35 %*	- 26,1 %	13 *65 %*	+ 27,1 %	**20**
Total	**3 1,1 %**	**150 59 %**	**- 2,1 %**	**101 39,7 %**	**+ 1,8 %**	**254**

Figure 48 : Variable de l'âge dans les facteurs de frein au partage des données

Du côté des similitudes, on peut relever :
- **les raisons économiques, scientifiques et la crainte du plagiat** : on ne voit pas vraiment de différence dans les deux tranches d'âge, les écarts de représentation sont très faibles (entre 0 % et 2,9 %) ;
- **les freins juridiques** : l'écart est un peu plus grand, mais ne semble pas vraiment significatif : + 4,4 % chez les 30-50 ans, - 3,5 % chez les plus de 50 ans ;
- **le manque de temps** : cet item semble un peu plus marqué chez les plus de 50 ans, avec un écart de + 3,9 %, et un écart négatif de - 4,8 % chez les 30-50 ans.

Pour ces cinq items, il nous semble donc que la variable de l'âge ne joue pas vraiment dans les résultats. Notons cependant que les raisons juridiques sont citées comme premier facteur de

frein chez les 30-50 ans, alors que c'est le manque de temps qui arrive en tête chez les plus de 50 ans.
Du côté des différences :
- **les raisons personnelles** : l'écart négatif chez les plus de 50 ans est de - 7,2 %, ce qui laisserait penser que cet item est moins important dans cette tranche d'âge ; mais les chiffres des réponses (total de 13) sont trop faibles et ne peuvent être significatifs ;
- **le manque de connaissances dans la description des jeux de données** : pour cet item, les écarts nous paraissent plus intéressants, car ils se fondent sur des valeurs un peu plus élevées (total de 28 réponses), et les écarts de représentation sont plus marqués : - 7,6 % chez les 30-50 ans, + 8,5 % chez les plus de 50 ans ;
- **le manque de compétences** : de même, sur cet item très proche, les écarts sont encore plus grands : - 26,1 % chez les 30-50 ans, + 27,1 % chez les plus de 50 ans ;
- **les raisons institutionnelles** : enfin, ce dernier item semble marquer également une nette différence, puisqu'il est absent chez les plus de 50 ans, et représente un écart de représentation de + 38,9 % chez les 30-50 ans. Mais là encore, la faiblesse de la valeur de référence (3 réponses) incite à la plus grande prudence !

Sur la question des compétences (informationnelles, numériques, etc.), il semble donc que la variable de l'âge intervienne bien, les enseignants-chercheurs entre 30 et 50 ans se déclarant plus compétents que les « seniors ».

3.4.2.2 Les variables du statut et des responsabilités sur les pratiques de partage

A/ La variable du statut

Est-ce que le statut des enseignants-chercheurs (i.e. les différences entre Maître de conférences, MCF ayant une HDR, Professeur d'université...) joue un rôle quelconque dans les opinions, les pratiques, les positionnements face au partage des données ? Nous avons refait le même tableau de synthèse sur les mêmes huit questions, en examinant les différences entre MCF et Professeurs (le nombre de HDR étant trop réduit pour pouvoir être pris en compte). Les taux de réponses correspondent aux réponses positives (Oui, D'accord, En priorité...).

N° des questions	Questions	MCF	Pr	Diffé-rence
Q 22	J'ai déjà téléchargé les données d'autres chercheurs	40,6 %	40,5 %	- 0,1
Q 23	J'ai déjà stocké mes données en libre accès sur le web	18,7 %	24,3 %	+ 5,6
Q 24	Je suis d'accord pour diffuser mes données de recherche en libre accès (à condition qu'elles soient diffusables et protégées juridiquement)	63,4 %	67,5 %	+ 4,1
Q 25	L'adhésion aux valeurs de libre Accès aux résultats de la recherche scientifique (comme facteur d'incitation au partage des données)	36,6 %	40,6 %	+ 4
Q 29	Je suis familier avec les enjeux de l'open access	42,5 %	51,6 %	+ 9,1
Q 29	Je suis prêt à partager mes données de recherche avec les autres	63,7 %	62,5 %	- 1,2

Q 29	Je pense qu'une déontologie des données pourrait devenir un enjeu lorsque les données de recherches sont partagées avec les autres	85,9 %	83,8 %	- 2,1
Q 30	Je souhaiterais bénéficier de conseils relatifs à la publication des données de la recherche (en priorité)	12,7 %	37,9 %	+ 25,2

Figure 49 : Synthèse de la variable du statut dans le partage des données

Comme on peut l'observer, le statut ne semble pas avoir d'influence particulière sur les réponses concernant le partage et la diffusion des données. Les quelques variations observées paraissent analogues à celles observées pour la variable de l'âge, et comme le nombre de professeurs est sans doute plus élevé chez les plus de 50 ans que chez les 30-50 ans, on peut probablement mettre ces variations sur le compte de l'âge et non sur celui du statut. Ainsi, la différence la plus notable porte sur la demande de conseils relatifs à la publication des données de recherche, plus forte chez les professeurs que chez les maîtres de conférence, comme chez les plus 50 ans.

B/ Statut et choix du support de diffusion en libre accès

La question 2 sur le statut a été croisée avec la question 27 « *Si vous souhaitez diffuser vos données en libre accès, quel est (ou serait) votre support principal ?* » et les résultats[39] semblent montrer ici un possible impact de la variable du statut. En effet, on observe des différences sensibles dans le choix du support de diffusion des données sur le web :
- les professeurs choisissent plutôt les archives ouvertes (33,3 % contre 18,8 % des MCF) et les plateformes spécialisées (22,8 % contre 19,8 % chez les MCF) ;
- les maîtres de conférence semblent préférer plutôt un réseau social académique (à 15,8 % contre 10,5 % chez les professeurs), ainsi que les supports de l'université (le blog du laboratoire, à 11,8 % et surtout un serveur de l'université : 14,8 % contre 5,2 % seulement chez les professeurs).

Mais les valeurs de référence restent peu importantes et il faut donc rester prudent dans l'interprétation.

C/ La variable des responsabilités

Nous avons cherché à savoir si l'exercice de responsabilités avait une incidence quelconque sur les pratiques et les représentations des enseignants-chercheurs sur le partage de données. La question 3 sur les responsabilités a donc été croisée avec plusieurs questions sur le partage, mais les résultats ne nous semblent pas vraiment probants (on les trouvera dans l'Annexe 2). On peut tout juste faire les observations suivantes :

- les enseignants-chercheurs ayant des responsabilités administratives semblent les moins enclins au téléchargement des données d'autres chercheurs (75 % de total des réponses Non, dans la question 22) ; mais l'hypothèse reste très fragile compte tenu de la faiblesse des valeurs de référence ;
- celles et ceux qui ont des responsabilités éducative stockent le moins leurs données en libre accès (total de 80,3 % de réponses Non à la question 23) ;

[39] Voir les résultats complets dans : Annexe 2, Croisements statistiques, 3.2.8 Statut et support privilégié pour la diffusion, p. 46

- celles et ceux qui ont des responsabilités éditoriales (71,4 % de Oui) et administratives (75 % de Oui) seraient les plus favorables pour diffuser leurs données en libre accès (question 24) ;

Mais au final, sur l'ensemble des questions, les écarts entre les différents types de responsabilités ne sont pas vraiment significatifs, et il est plus prudent, en l'état, de renoncer à voir un effet de la responsabilité exercée sur les pratiques de partage de données.

3.4.2.3 Les variables scientifiques sur les pratiques de partage

Variable paraissant essentielle pour déterminer les positions et les pratiques sur le partage des données, la variable scientifique n'a pourtant pas fourni de réponses très probantes et claires. Il est en effet très difficile de tirer des conclusions des croisements effectués sur deux questions : la question n° 4 sur le laboratoire d'appartenance et la n° 5 sur la section du CNU. La raison vient probablement de la grande dispersion des réponses : 19 items pour le laboratoire, 29 pour la section CNU. La majorité des résultats pour plusieurs réponses sont donc très faibles et ne permettent aucune leçon généralisable. Pour autant, quelques observations intéressantes restent possibles.

A/ La variable de l'appartenance au laboratoire

Le croisement de la question 4 « *Quel est votre laboratoire ?* », obligatoire, avec les questions sur le partage des données, donne à voir un résultat global assez contrasté, avec quatre laboratoires seulement dépassant en moyenne 10 répondants, et neuf autres donnant un nombre de répondants inférieur à 10. Nous avons donc choisi de mettre l'accent sur ces quatre laboratoires et nous présentons ci-dessous le même type de tableau que précédemment, à savoir le total des réponses positives (Oui, D'accord, En priorité) à sept questions. Nous indiquons entre parenthèses le total moyen des réponses de chaque laboratoire.

N°	**Questions**	**APP[40] (11)**	**ESO[41] (15)**	**LP3C[42] (12)**	**Tempora[43] (10)**
Q 22	Téléchargement des données d'autres chercheurs	45,4 %	46,6 %	46,1 %	63,6 %
Q 23	Stockage de ses données en libre accès	18 %	13,3 %	15 %	36,3 %
Q 24	Accord pour diffuser ses données de recherche en libre accès	72,7 %	53,3 %	46,1 %	54,5 %
Q 25	Adhésion aux valeurs du libre Accès...	27,2 %	57,1 %	45,4 %	40 %
Q 29	Familiarisation avec les enjeux de l'open access	40 %	71,4 %	66,6 %	33,3 %

[40] APP : Arts : pratiques et poétiques ; regroupait au moment de l'enquête la plupart des disciplines artistiques, y compris les Arts plastiques, sauf l'Histoire des arts ;
[41] ESO : Espaces et Sociétés ; regroupe des géographes (géographie humaine et sociale), des sociologues, des psychologues…
[42] LP3C : Laboratoire de Psychologie : Cognition, Comportement, Communication (ex-CRPCC)
[43] Tempora (ex-CERHIO) : unité de recherche en Histoire.

N°	Questions	APP[44] (11)	ESO[45] (15)	LP3C[46] (12)	Tempora [47] (10)
Q 29	Disposition à partager ses données de recherche	70 %	57,1 %	58,3 %	55 %
Q 30	Conseils relatifs à la publication des données (en priorité)	44,4 %	7,6 %	20 %	25 %

Figure 50 : Synthèse de la variable du laboratoire dans le partage des données

Ce tableau montre des résultats intéressants par quelques dissonances :

- d'un côté, les membres d'APP semblent les plus enclins au partage des données, avec des écarts parfois significatifs, comme sur les questions 24 (72,7 % d'accord pour la diffusion en libre accès), 29 (70 % d'accord avec la disposition au partage), et 30 (44,4 % d'accord sur la priorité à donner aux conseils relatifs à la publication des données) ;
- d'un autre côté, le laboratoire ESO paraît en pointe dans l'adhésion aux valeurs (57,1 %) et surtout dans la familiarisation avec les enjeux du Libre Accès (71,4 %), et il aurait le moins besoin de conseils relatifs à la publication des données (7,6 %.) ;
- de leur côté, les historiens de Tempora sont les plus nombreux à télécharger les données d'autres chercheurs (63,6 %) et à stocker les leurs en libre accès (36,3 %), mais sont les moins familiers avec les enjeux du Libre Accès ;
- quant aux psychologues de LP3C, ils semblent se situer dans une bonne moyenne pour chaque question, avec moins d'écarts entre leurs réponses.

Ainsi ce croisement donne à voir des informations intéressantes sur certains laboratoires (on peut également voir les résultats complets en Annexe 2), mais ne permet pas de répondre à la question de l'impact de l'appartenance à tel ou tel laboratoire sur les pratiques de partage. D'autres calculs et d'autres croisements, que nous n'avons pu faire, auraient été sans doute nécessaires ici.

B/ La variable de la discipline (section CNU)

Qu'en est-il de l'affiliation à une section CNU ? La discipline (d'appartenance ou de qualification) a-t-elle un impact sur les pratiques et représentations concernant le partage des données ? L'intuition, de même que l'enquête qualitative présentée un peu plus loin, donnent une réponse positive à cette question. Pourtant, les croisements de notre enquête statistique ne fournissent pas, là non plus, de résultats très probants et irréfutables. Et ce en raison de la dispersion des répondants et du faible nombre de réponses sur de nombreuses disciplines. Compte-tenu du nombre d'items possibles (29 sections du CNU), nous n'avons procédé au croisement que sur deux questions (23 sur le stockage en libre accès de ses données, et 24 sur l'accord pour une diffusion en libre accès). Voici une synthèse des résultats pour les six sections du CNU ayant recueilli au moins 10 réponses. Même avec ces sections les mieux représentées dans notre échantillon, les chiffres restent très faibles et ne permettent de tirer aucune

[44] APP : Arts : pratiques et poétiques ; regroupait au moment de l'enquête la plupart des disciplines artistiques, y compris les Arts plastiques, sauf l'Histoire des arts ;
[45] ESO : Espaces et Sociétés ; regroupe des géographes (géographie humaine et sociale), des sociologues, des psychologues…
[46] LP3C : Laboratoire de Psychologie : Cognition, Comportement, Communication (ex-CRPCC)
[47] Tempora (ex-CERHIO) : unité de recherche en Histoire.

conclusion. Le chiffre entre parenthèses correspond au nombre de répondants (et non au numéro de la section CNU !) et les résultats correspondent au sous-total des réponses Oui, aux deux questions 23 et 24. Pour des raisons de place, nous avons dû tronquer les intitulés complets des sections[48].

Questions	**L & L romanes (11)**	**Psycho. (16)**	**Architecture… (12)**	**Sociologie… (10)**	**Histoire… (10)**	**STAPS (12)**
Stockage de ses données en libre accès	5 (45,4 %)	3 (18,7 %)	3 (25 %)	2 (20 %)	3 (30 %)	4 (33,3 %)
Accord pour diffuser ses données en libre accès	10 (90,9 %)	8 (50 %)	9 (75 %)	4 (40 %)	4 (40 %)	8 (66,6 %)

Figure 51 : Synthèse de la variable de la section CNU dans le partage des données

L'une des rares observations possibles tient sans doute à la correspondance avec les croisements présentés plus haut sur les quatre laboratoires : entre Psychologie et LP3C, Architecture et arts appliqués et APP, Sociologie et ESO, Histoire et Tempora, bien que les chiffres des réponses soient différents. Mais par ailleurs, les sections Langues et littératures romanes ou STAPS relèvent d'autres laboratoires.
Ainsi la variable disciplinaire s'avère-t-elle également décevante dans les croisements statistiques. L'enquête qualitative fournira sur ce point des leçons plus pertinentes.

3.4.2.4 Les variables liées à la nature des données sur les pratiques de partage

Les derniers croisements effectués ont porté sur la nature des données : est-ce que le type de données collectées (textuelles, expérimentales, d'observation, etc.), le support (nativement numérique ou non…), le type de données produites (statistiques, textuelles, etc.) ont une incidence quelconque sur les pratiques de partage ? Trois questions sur le partage des données (questions 21, 23 et 24) ont été croisées avec les questions 6, 7 et 9.

A/ La variable de la nature des données collectées

Voici un tableau de synthèse des croisements avec la question 6[49], dans lequel nous n'avons indiqué que le taux de réponses positives :

[48] Voici les intitulés officiels complets et le rappel du numéro de section :
- Langues et littératures romanes, section 11 ;
- Psychologie, section 16 ;
- Architecture et arts appliqués, section 18 ;
- Sociologie, démographie, section 19 ;
- Histoire – histoire de l'art moderne et comtemporaine, histoire de la musique, section 22 ;
- STAPS (Sciences et Techniques des Activités Physiques et Sportives), section 74.

[49] Question 6 : « *Dans quelle catégorie(s) suivantes classeriez-vous vos données sources (i.e. vos données collectées ?* ». Il y avait 9 items possibles : données d'observation (remontées de terrain, données de capteurs...) ; données expérimentales (en médecine, psychologie…) ; données textuelles (issues de corpus de textes, d'archives...) ; données audiovisuelles (enregistrements audio, séquences filmiques…) ; images fixes d'objets, de paysages, d'architectures, scans de textes ; données chiffrées (tableaux de données statistiques…) ; données d'enquêtes, de sondages (entretiens...) ; données web (données d'usages, cartographies de sites, données de réseaux sociaux...) ; données 3D, de simulation.

- pour la question 21[50], il s'agit du sous-total des six items sur sept, indiquant une possibilité d'accès ;
- pour les questions 23[51] et 24[52], il s'agit du sous-total des réponses Oui.

N°	Questions	D. obs.	D. exp.	D. text.	D. av	Im. fixes	D. chif.	D. enq.	D. web	D 3D
Q 21	Accessibilité des données	68,9 %	66,6 %	55 %	51,9 %	50 %	62,3 %	57,1 %	74,5 %	85,7 %
Q 23	Stockage de ses données en libre accès	25 %	20 %	25 %	16,6 %	29,4 %	23,5 %	21,5 %	30 %	66,6 %
Q 24	Accord pour diffuser ses données de recherche…	68,6 %	56 %	70,2 %	61,9 %	58,8 %	56,8 %	63 %	83,3 %	83,3 %

Figure 52 : Synthèse de la variable de la nature des données collectées dans le partage des données

Quelles observations peut-on tirer de ces croisements ? Il semble bien que le support matériel et le type de données collectées peuvent avoir une incidence sur la pratique de partage :

- les données web (i.e. les données d'usages, les cartographies de sites, les données de réseaux sociaux...) et les données 3D sont les plus accessibles, de très loin ; elles favorisent également davantage l'accord pour une diffusion en libre accès (à 83,3 %) ; par ailleurs, les données 3D ont été les plus stockées en libre accès sur le web (66,6 %), et les données web arrivent en deuxième position ; mais les valeurs sont assez faibles et il est difficile de généraliser ;
- à l'inverse, les images fixes (50 %), les données audiovisuelles (51,9 %) et les données textuelles (55 %) sont les moins accessibles en dehors du chercheur ;
- les données audiovisuelles (16,6 %) et expérimentales (20 %) ont été les moins stockées en libre accès sur le web ;
- les données expérimentales semblent susciter le moins de prédisposition à être diffusées en libre accès (56 %).

Il convient toutefois de rester prudent dans l'interprétation, car hormis le cas des données web et données 3D, les écarts entre les autres types de données ne sont pas très probants.

B/ La variable du support des données collectées

Les résultats des trois mêmes questions (21, 23 et 24) ont été croisés avec les résultats de la question 7[53] sur le support matériel :

[50] Question 21 : « *En dehors de vous, qui peut accéder directement à vos données ?* » ; 116 répondants et 522 réponses.
[51] Question 23 : « *Vous-même, avez-vous déjà stocké vos données en libre accès sur le web ?* » ; 116 répondants et 389 réponses.
[52] Question 24 : « *Seriez-vous d'accord pour diffuser vos données de recherche en libre accès (à condition qu'elles soient diffusables et protégées juridiquement) ?* » ; 115 répondants et 388 réponses.
[53] Question 7 : « *Concernant leur support matériel, vos données sources (ou données collectées) sont-elles majoritairement : non numériques (par exemple, des sources archivistiques imprimées, des entretiens non enregistrés en numérique, etc.), numérisées (par exemple, des corpus numériques, des données statistiques…), nativement numériques (par exemple les données web) ?* »

N°	Questions	D. non numériques	D. numérisées	D. nativement numériques
Q 21	Accessibilité des données	55 %	55,7 %	70,5 %
Q 23	Stockage de ses données en libre accès	24,4 %	16,3 %	58,3 %
Q 24	Accord pour diffuser ses données de recherche…	56,2 %	70,9 %	83,3 %

Figure 53 : Synthèse de la variable du support matériel des données collectées dans le partage des données

Selon ces croisements, le support matériel semble être une variable plus forte que la nature même des données, car les écarts sont plus nets.
Levons tout d'abord une interrogation sur la pertinence même d'une partie des résultats : que signifie en effet stocker sur le web des données non numériques ? Comment fait-on techniquement ? Il faut rappeler que la question 7 sur le support matériel précisait bien : « vos données sources sont *majoritairement*… ». Les chercheurs ayant coché « non numériques » peuvent donc collecter aussi des données à numériser ou nativement numériques. Une première observation permet de voir d'abord des écarts importants entre les données nativement numériques et les autres :

- dans l'accessibilité des données en-dehors du chercheur (question 21) : les données nativement numériques sont accessibles, d'une manière ou d'une autre, à un peu plus de 70 %, alors que les données non numériques ou numérisées ne sont accessibles qu'à 55 ou 56 % ;
- dans le stockage en libre accès sur le web (question 23) : elles arrivent également largement en tête (avec 58,3 %) ;
- dans la prédisposition à une diffusion en libre accès (question 24) : elles font aussi le meilleur score (avec 83,3 %)

Faut-il voir dans ces résultats l'indice d'une facilité technique pour le partage des données, une sorte de « déterminisme technique » du support numérique, invitant par lui-même à la diffusion ? La faiblesse des valeurs de référence incite, là encore, à une certaine prudence. Par ailleurs, une autre observation semble venir contredire cette hypothèse de la facilité technique : pourquoi les données numérisées, également faciles à diffuser en libre accès, font-elles des résultats nettement inférieurs (dans l'accessibilité et surtout le stockage en libre accès, où elles arrivent curieusement derrière les données non numériques) ? Àl'inverse des données nativement numériques, faut-il voir là le signe d'une certaine difficulté « humaine » ? Numériser des données peut demander en effet un travail, un effort personnel parfois très important, propre à freiner les velléités de partage. Mais il est difficile, voire impossible, de trancher.

C/ La variable de la nature des données produites

Le dernier croisement des trois questions a porté sur la nature des données produites, à partir des résultats de la question 9[54] et le tableau ci-dessous présente la synthèse des réponses positives :

[54] Question 9 : « *Dans quelle(s) grande(s) catégorie(s) classeriez-vous vos données produites, après exploitation, i.e. vos "données résultat" ?* ». Il y avait cinq items possibles : données statistiques (données chiffrées) ; données textuelles (extraits, champs textuels, extraits d'entretiens…) ; données graphiques (schémas, visualisations, cartographies, tableaux, scans de textes…) ; données audiovisuelles (films…) ; données 3D, de simulation.

N°	Questions	D. stat.	D. text.	D. graph.	D. a.v.	D. 3D
Q 21	Accessibilité des données	64,7 %	51,9 %	56,7 %	80 %	100 %
Q 23	Stockage de ses données en libre accès	26,8 %	48,7 %	23 %	33,3 %	60 %
Q 24	Accord pour diffuser ses données de recherche…	62,5 %	66,6 %	68,6 %	80 %	80 %

Figure 54 : Synthèse de la variable de la nature des données produites dans le partage des données

Quel est le rôle de la nature des données produites, dans l'inclination au partage et quelles différences peut-on relever avec les données collectées ? On peut faire les observations suivantes :

- la spécificité des données 3D, de simulation… : comme pour les données collectées, les données de simulation produites apparaissent comme les plus enclines à la diffusion et au partage : complètement accessibles, stockées en libre accès sur le web à 60 %, et induisant, à 80 %, un accord des chercheurs pour une diffusion en libre accès ; on retrouve là encore une sorte de « logique technique », liée à la nature même et au support des données ; mais les valeurs très faibles doivent toujours nous garder de toute généralisation ;
- pour les données audiovisuelles, une différence sensible entre données collectées et produites : peu accessibles (51,9 %) et très peu stockées en libre accès (16,6 %), lorsqu'il s'agit de données collectées (ce qui peut s'expliquer sans doute par des raisons juridiques), les données audiovisuelles produites par les chercheurs sont largement accessibles (80 %) et davantage stockées en libre accès sur le web (33,3 %) ; de même, elles suscitent un taux d'accord très important (80 %, contre 61,9 %) pour une diffusion en libre accès ;
- pour les données textuelles, l'accessibilité des données produites semble un peu plus faible (51,9 %) que celle des données collectées (55 %) ; mais curieusement, la part des données produites, stockées en libre accès sur le web, est nettement plus forte : 48,7 % contre 25 % pour les données collectées. Mais l'interprétation de cette contradiction, qui n'est peut-être qu'apparente, n'est guère aisée.
- Quant aux données statistiques et graphiques, elles sont les moins stockées en libre accès.

Au final, la nature des données collectées et produites, et plus encore le support matériel, semblent bien avoir un effet sur les pratiques de partage et de diffusion des données, mais qui reste difficile à interpréter. Les effets les plus probants paraissent liés d'abord à la nature nativement numérique des données, qui semble favoriser la diffusion et le partage. On discerne également la place plus importante des facteurs juridiques attachés à certains types de données (comme les images, les photographies, les films, les textes…), qui peuvent freiner la mise en libre accès des données. De même, les données chiffrées, statistiques, paraissent moins favoriser la diffusion en libre accès, aussi bien en tant que données collectées que données produites. Mais la complexité des croisements, ainsi que la faiblesse de certaines valeurs, incitent à une grande prudence dans l'interprétation de ces résultats.

3.4.3 Résultats des entretiens sur le partage des données

L'analyse des 21 entretiens donne des éclairages beaucoup plus fins sur les positions des enseignants-chercheurs face à la question du partage des données. Après avoir présenté les positionnements généraux, nous exposerons le détail des facteurs d'incitation (la question 25) et celui des facteurs de frein (question 26), pour conclure par quelques interrogations formulées par les enseignants-chercheurs. Pour alléger la lecture, nous avons mis en Annexes (Annexe 3) la liste complète des extraits d'entretiens concernant le partage des données, et nous ne mentionnons ici que quelques fragments.

3.4.3.1 Quelle position générale sur le principe du partage des données de recherche ?

Dans l'ensemble, la quasi-totalité des collègues interviewés déclarent leur accord de principe avec l'idée générale d'un partage, d'une diffusion des données. Deux sortes de variables, souvent imbriquées, semblent se dégager des entretiens : d'une part, les adhésions que l'on peut qualifier d'individuelles, faites au nom de valeurs, d'opinions, de réflexions d'abord personnelles, d'autre part, l'adhésion au partage selon une logique, une variable disciplinaire très forte. Nous verrons plus loin qu'à l'inverse, la variable disciplinaire explique aussi de nombreuses réticences, voire des oppositions très nettes, au partage.

A/ L'adhésion personnelle au principe du partage des données

L'expression de l'adhésion aux principes du libre accès et du partage des données, au nom de différents motifs, se retrouve dans plusieurs entretiens, comme le montrent ces extraits :
- *« Concernant les résultats, moi je suis entièrement pour le partage » ;*
- *« Oui, plutôt favorable sur l'idée générale ; pas au nom de principes libéraux, mais plutôt en termes de valorisation, j'y vois aussi un outil de valorisation de la recherche, sous d'autres formes que nos écrits » ;*
- *« Quand on travaille dans le public, on sait qu'on est financé par l'Etat, c'est un peu logique de rendre accessible à tous le fruit de nos recherches.* » ;
- *« Je suis pour la mise à disposition de mes données (...), il peut y avoir d'autres analyses, des analyses qui vont contre les miennes etc. C'est l'objectif de la science » ;*
- « *On ne peut être que favorable sur le fond en tout cas. Et nous on en bénéficie aussi.* ».

B/ La variable disciplinaire

S'ajoutant à des positionnements plus personnels, la variable disciplinaire peut incliner fortement au partage des données, comme l'indique clairement ce chercheur : « *Je pense que philosophiquement, on est tous favorables au partage de données dans les sciences expérimentales* ».
Sans surprise, la nature des travaux de recherche dans certaines sciences incite au, voire nécessite, le partage des données de recherche. Ainsi dans l'étude de la locomotion et du mouvement humain, une chercheuse nous a expliqué qu'elle avait « *besoin de beaucoup de données.* ».
Par ailleurs, les évolutions récentes de certaines disciplines, notamment sous l'effet du numérique, jouent un rôle essentiel dans le positionnement face à la diffusion des données, et ce en lien étroit avec la question de l'intégrité scientifique, comme l'explique cet enseignant-chercheur : *« On a une crise de réplicabilité en ce moment en psychologie. Pas mal de champs de recherche en psychologie ont de gros problèmes à être répliqués, des résultats classiques*

qu'on ne retrouve pas quand on essaie de refaire les mêmes expériences. (...) *La psycho est en train de rentrer doucement dans l'ère du "on enregistre nos expériences avant de les faire, en déclarant à l'avance combien de participants on va collecter, quelle méthode on va utiliser, quelles analyses on va faire et comment on va traiter nos données", dans l'idée que si on déclare ça avant, ça nous empêchera de trafiquer notre analyse après pour trouver des résultats qui nous intéressent. Et c'est à cause de ça aussi que la demande des revues de stocker les données se développe de plus en plus. C'est quelque chose qui n'existait pas il y a 5 ans par exemple.* ».
Le même collègue souligne le retard français, pris en psychologie, par rapport aux équipes américaines : « *Les équipes internationales avec lesquelles on collabore, on voit la différence, parce que quand on bosse avec des équipes américaines par exemple, elles ont des plateformes de données avec archivage à long terme, elles ont des protocoles, des plans de gestion des données justement, donc c'est une différence qu'on voit et on voit que les demandes des revues sont calées plus sur ce que font les équipes américaines que sur ce qu'on fait, nous. Donc je pense que c'est un retard qu'on est en train de prendre aussi.* »

C/ Des positions favorables au partage, mais par obligation ou sur fond de questionnement, de réticences personnelles

Ce positionnement majoritairement favorable au partage est souvent nuancé par différentes interrogations, par des doutes, sinon de sérieuses réserves.
Une première position, que l'on pourrait qualifier de « fataliste », au sens où il s'agit d'accompagner un mouvement jugé irréversible, ressort ainsi de quelques entretiens :
- « *Sur le partage, en-dehors du labo (sur la mise en open access), on nous encourage à y aller ; après, dans quelques années, on ne nous encouragera plus mais on nous obligera ; donc mieux vaut s'y préparer en interne* » ;
Les chercheurs peuvent se trouver également tiraillés entre deux mouvements, deux tendances contradictoires : « *Il y a toujours une ambivalence entre les deux. Favorable scientifiquement, défavorable dans le sens où ça permet aux autres d'avoir un regard sur ce que je fais et donc de me voler des idées sur le contenu de mes expériences.* ».
D'autres chercheurs, tout en exprimant leur accord général, formulent également des interrogations sur les modalités du partage et sur les types de données à ouvrir :
- « *Sur le principe j'y suis plutôt favorable, l'idée paraît séduisante (...). Après c'est la manière dont ça va se faire, se mettre en place qui m'interroge un peu.* » ;
- « *Quand les données sont partageables, c'est très bien qu'elles soient partagées, mais après la question, c'est : partagées avec qui, comment ? C'est les modalités du partage.* » ;
- « *Oui, moi je distingue vraiment les deux* (publications et données brutes) *; pour moi, à partir du moment où les recherches ont été validées, publiées, elles doivent être libres d'accès.* (...) *J'ai plus de réticence par rapport aux données brutes, en tout cas par rapport à certaines données brutes, et encore plus pour les données qualitatives.* » ;
- « *Sur les données qualitatives j'ai plus de mal à voir comment les chercheurs pourraient s'en emparer, qu'est-ce qu'ils pourraient en faire.* ».
Enfin, certains expriment de fortes réserves sur l'idée du « tout-partage », i.e. la perspective d'un partage des données généralisé, « universel » et systématique :
- « *Le positionnement absolu du tout disponible, tout en ligne, tout public me paraît être utopique. Comme toutes les utopies, il a des effets pervers qui peuvent être dangereux.* » ;
- « *je pense qu'il faut nuancer en fait cette position du partage absolu. Il y a des types de données, et du coup je pense qu'on ne peut pas tout partager.* » ;
- *« Ça dépend du cadre légal du partage des données ; ça c'est essentiel ».*

Au final, l'ensemble des positionnements recueillis sur le partage des données semble refléter une large part du spectre complet des questions posées par cette évolution majeure de la science. Nous allons examiner de plus près ces différents points, en revisitant notamment les questions 25 et 26 du questionnaire de l'enquête quantitative, sur lesquelles nous sommes revenus au cours des entretiens.

3.4.3.2 Les facteurs d'incitation au partage des données

Nous avons cherché à mieux cerner les raisons qui inciteraient à rendre les données de recherche accessibles en libre accès en interrogeant les chercheurs sur les différents items proposés dans la question 25. Quand cela a été possible, il nous a semblé intéressant de comparer les résultats des deux modalités de notre enquête, statistique et qualitative. En gardant à l'esprit que les échantillons sont très différents : plus d'une centaine de répondants d'un côté, un panel de 21 entretiens de l'autre.

A/ L'obligation faite par le financeur de la recherche

Sur cet item, la comparaison avec l'enquête statistique montre une assez grande distorsion, puisque l'obligation du financeur a recueilli 33 % des 107 répondants, alors que dans les entretiens, elle n'est pas considérée comme un facteur important d'incitation au partage des données, pour plusieurs raisons :
- soit cette obligation ne concerne pas vraiment la discipline ;
- soit elle n'a pas (encore ?) touché les chercheurs : « *Je n'ai jamais travaillé sur des projets de recherche où on nous a obligés* », précise un chercheur ;
- soit cette contrainte ne joue pas de rôle-clé pour les tenants du partage des données : « *C'est incitatif pour faire avancer, mais je pense que quand on est chercheur, c'est d'emblée pour mettre à disposition ses travaux, ses résultats...* » ;
- enfin, l'obligation même du financeur est récusée au nom d'une certaine conception de la liberté académique du chercheur : « *Ni l'obligation faite par le financeur de la recherche, parce qu'il n'a pas le droit de m'y obliger* ».

B/ La reconnaissance de la communauté scientifique

Si l'on compare avec l'enquête statistique (32 % de réponses sur 107 répondants), on relève également un certain écart, mais pas de véritable contradiction. Dans les entretiens, ce facteur d'incitation a suscité des réactions partagées en deux catégories distinctes, mais d'inégale importance. Assez peu de chercheurs semblent accorder un rôle incitatif (par ailleurs relatif) à cet item :
- « *La reconnaissance de la communauté scientifique, oui, c'est un argument, mais ça je pense que ça rejoint la question de la diffusion, sauf si vous entendez par reconnaissance les enjeux de carrière* » ;
Pour la majorité des chercheurs, elle ne semble pas être un facteur suffisant ou pertinent pour partager les données :
- « *Je ne suis pas sûr que ça nous fasse beaucoup en reconnaissance ou en visibilité, sur le principe.* » ;
Pour plusieurs chercheurs, la reconnaissance de la communauté scientifique passe avant tout par les publications : « *Ce n'est pas par la capacité à accumuler des données, c'est par l'article.* ».

C/ Une meilleure visibilité des travaux de recherche

Sur cet item, l'enquête qualitative semble contredire nettement les résultats de l'enquête quantitative : en effet, dans les résultats statistiques, la visibilité arrive nettement en tête des réponses (65 % des 107 répondants), alors que les entretiens montrent des positions beaucoup plus nuancées.
Là encore, les réactions sont partagées en deux catégories opposées. Certains chercheurs accordent de l'importance à la visibilité, comme incitation à la diffusion de ses données :
- « *D'autres incitations à rendre vos données accessibles ? Réponse : Oui, la visibilité* »
- « *Les blogs, c'est plus pour rendre une visibilité des travaux de recherche et informer, puisqu'on est dans une logique de réseau, il faut constituer des champs, des repères* », selon un chercheur qui a utilisé les blogs comme support de diffusion de ses données.
À l'opposé, la visibilité conférée par les données de recherche ne constitue pas un facteur pertinent pour d'autres chercheurs, qui comptent davantage sur les publications : « *Pour nous ça ne passe pas par les données, ça passe par les articles. Ce n'est pas parce que tu as plein de données que tu vas être plus visible, si tu es visible c'est parce que tu fais quelque chose de tes données* ».

D/ L'occasion de nouveaux contacts, de nouvelles coopérations scientifiques

Les deux enquêtes paraissent plus concordantes sur l'occasion de nouveaux contacts, comme facteur incitatif au partage. L'item arrive en deuxième position dans l'enquête quantitative (avec 54 % de répondants sur 107), et il recueille dans les entretiens une majorité de réponses positives, tout en suscitant également quelques réserves.
Citons quelques uns des avis positifs :
- « *Ça je pense que c'est extrêmement important pour le coup* » ;
- « *Ce que je défends, c'est que les travaux doivent alimenter le débat social et le débat scientifique et c'est sûr que je préfère publier un papier qui sera lu, même si ça dépasse la communauté scientifique que de publier dans une revue scientifique qui sera très peu lue (...), donc cette question de l'ouverture, elle est fondamentale pour moi* » ;
- « *Ce n'est pas la norme mais ça peut arriver* ».

Du côté des avis réservés, l'accent est mis de nouveau sur le couple données de recherche / publications, en donnant la priorité à ces dernières :
- « *Les données en soi, je ne suis pas certain que cela créée de nouveaux contacts ou génère de nouvelles coopérations scientifiques » ; « c'est un enjeu car plus les gens peuvent avoir accès à nos publications, plus facilement on développe des réseaux et des collaborations, ça c'est intéressant, je ne suis pas persuadé que mettre nos données brutes telles quelles accessibles à tous va faire que l'on va avoir plus de visibilité, que l'on va se développer un réseau* ».

E/ Le souci de pouvoir faire valider ses résultats à partir de ses jeux de données

Avant-dernier item dans l'enquête statistique, cité par 11 % seulement des 107 répondants, le souci de pouvoir faire valider ses résultats à partir de ses jeux de données obtient encore moins de succès dans les entretiens, puisqu'aucun enseignant-chercheur interviewé ne reconnaît ce facteur comme une incitation au partage. Cet extrait résume assez bien une position partagée :
- « *Je pense que de fait, on est toujours obligé de garder ses jeux de données, ses logiques de démonstration, si on les demande ; mais de là à les rendre visibles, ça pour moi, c'est une question qui est beaucoup plus à la marge*. ».
Pour certains, cette perspective serait même contre-productive, voire totalement déplacée :

- « *Ce n'est pas une préoccupation que j'ai, plutôt l'inverse en fait quelque part. On a plutôt confiance en nos résultats quand on les publie, par définition, si on n'a pas confiance, on ne les publie pas.* » ;
- « *Je ne vois pas pourquoi on les ferait valider par d'autres. On les valide nous même et si l'article est pris par un comité de lecture, c'est que ça marche* ».

F/ Le souhait de voir se développer d'autres recherches à partir de ses jeux de données :

Sur cet item, les deux enquêtes paraissent également cohérentes dans leurs résultats : placé en quatrième position, avec 40 % des 107 répondants, le souhait de voir se développer d'autres recherches à partir de ses jeux de données recueille la grande majorité des réponses dans les entretiens. Différentes raisons peuvent être convoquées pour justifier ce facteur d'incitation :
- le souhait de la filiation : « *C'est toujours plaisant de voir qu'on peut « faire des petits » en science, dans les domaines, c'est-à-dire avoir une forme de filiation, créer, susciter de nouveaux projets, de nouvelles recherches* » ;
- le souci de la réciprocité dans l'utilisation des données : « *Quand ça sera complètement ouvert, le but du jeu, c'est que tout le monde puisse utiliser et puis importer les données, parce que moi j'en ai besoin par ailleurs* » ;
- la perspective de la discussion, de la confrontation scientifique : « *Scientifiquement, c'est intéressant, parce que justement, ça permet aussi d'avoir d'autres lectures de ce qu'on avait fait comme corpus et qu'on n'aurait pas eues* » ;
- une forme de générosité, visant à permettre d'autres recherches complètement extérieures au domaine : ainsi en géographie physique, « *Quand on exploite des données, on n'exploite pas tout. Par exemple moi dans mes carottages, je ne fais pas de micro-palé (...) donc si quelqu'un veut prendre la carotte et s'occuper des mini coquillages et des machins comme ça, moi ça ne me dérange pas.* » ; ou encore dans l'étude du mouvement : « *J'ai déjà envoyé par mel des données pour des questions complètement différentes, pour une chercheuse en mathématiques en Israël, qui avait des questions sur d'autres choses sur la marche (...) et elle a fait ses analyses, qui n'avaient rien à voir avec les questions que je me posais.* ».
Une réserve forte, voire une opposition, est toutefois exprimée sur les données à diffuser et les réutilisations possibles, au nom de raisons scientifiques, sur lesquelles nous revenons plus loin :
- « *Par contre, sur des données qui sont beaucoup plus qualitatives, qui sont liées à la relation d'enquête, je trouve que ça n'a pas de sens, on dénature en fait le matériau en le donnant comme ça sous la forme de jeu de données.* » ;
- « *Le type de recherches qu'on fait nécessite des corpus tellement contextualisés que personne d'autre que la personne qui connaît le contexte ne peut les exploiter ; du coup ça ne marche pas* ».

G/ L'adhésion aux valeurs du Libre Accès aux résultats de la recherche scientifique

Placé également en troisième ou quatrième position dans l'enquête statistique, à égalité avec l'item précédent, avec 40 % de réponses, l'adhésion aux valeurs du Libre Accès a suscité en revanche moins de réactions dans les entretiens, même si une attitude favorable au libre accès des résultats de la science est assez largement partagée, comme l'ont montré les premiers extraits cités ci-dessus (cf partie 3.4.3.1). Certains collègues expriment toutefois des nuances dans cette adhésion aux valeurs du Libre Accès :
- « *La question du libre accès, je trouve que c'est très important, (...) c'est quelque chose que je défends. Mais de l'autre côté (...), on ne protège pas forcément du coup nos sources (...) ça expose aussi les situations et on doit garder la maîtrise de cette exposition en lien avec l'accès libre.* ».

Deux autres réponses sont intéressantes du point de vue de la méthodologie de l'enquête, car elles portent sur la formulation de la question et ses présupposés, notamment la définition (ou plutôt son absence) des valeurs du Libre Accès :
- « *Il faudrait définir ce que l'on entend par valeurs du libre accès aux résultats de la recherche scientifique, il faudrait être peut-être plus explicite.* » ;
- « *Vous ne dîtes pas quelles sont les valeurs du Libre Accès, donc on ne sait pas à quoi vous nous demandez si on veut y adhérer* ».

H/ De possibles retombées financières (crédits de recherche...)

Dans les deux enquêtes, les possibles retombées financières n'incitent guère les enseignants-chercheurs à ouvrir leurs données : dans l'enquête statistique, il s'agit du dernier item, cité par seulement 7 % des répondants ; et dans les entretiens, personne n'a mentionné cet item comme facteur d'incitation.

I/ Autres raisons

D'autres facteurs, ne figurant pas dans le questionnaire, et pouvant constituer aussi des incitations au partage des données, ont été cités par quelques collègues interviewés :
- **l'obligation faite aux chercheurs par les revues** : « *Et il manque un item : par obligation. Ce qui est vraiment de plus en plus le cas chez nous. Je pense que dans les trois ou quatre derniers articles que j'ai publiés, il y avait une exigence de la revue qu'on rende les données accessibles.* » ;
- **l'enjeu de conservation, de sécurisation des données** : « *Je vois bien l'avantage que ça peut avoir (...) en termes d'enjeux de conservation : si j'avais déposé quelque part mes données, si quelqu'un avait géré la convertibilité de mes données, peut-être qu'aujourd'hui, je pourrais récupérer mes données. Il y a cet avantage, cet enjeu de conservation.* » ;
- **le partage des connaissances théoriques,** permis par le partage des données, est cité par une chercheuse, qui évoque une base de concepts, pouvant être ouverte aux étudiants : « *Je pense que ce qu'on peut partager tout à fait bien, c'est tout ce qui a trait aux concepts fondamentaux d'identité, mémoire et territoire.* ».

Enfin, pour un enseignant-chercheur, aucune raison ne peut l'inciter à partager ses données :
- « Question : *Quelle aurait été l'autre raison possible qui t'aurait incité à partager les données ?* Réponse *: Aucune ! Y'en a pas* ». On ne saurait être plus clair...

3.4.3.3 Les facteurs de frein au partage des données

Les items de la question 26 (« *Quelles seraient les principales raisons qui vous empêcheraient ou vous freineraient pour rendre vos données de recherche accessibles en libre accès ?* ») ont fait l'objet, lors des entretiens, de commentaires et de développements plus longs que les facteurs d'incitation, ce qui constitue déjà, en soi, un indicateur intéressant de l'importance des réserves des enseignants-chercheurs face au partage des données. Par ailleurs, les neuf items de la question 26 ont été enrichis par plusieurs autres facteurs de frein, qui ne figuraient pas dans le questionnaire.
Nous présentons ici tous ces obstacles au partage des données, selon la catégorisation déjà utilisée lors de la présentation de l'enquête statistique, en distinguant d'une part les facteurs « externes », liés aux données, aux raisons scientifiques, etc., d'autre part les facteurs « internes », liés davantage aux enseignants-chercheurs eux-mêmes. Bien entendu, cette catégorisation ne vise qu'à clarifier la présentation d'un ensemble de raisons, par ailleurs inextricablement mêlées.

3.4.3.3.1 Facteurs « externes »

L'enquête statistique proposait quatre types de facteurs « externes », pouvant constituer des freins au partage des données : les freins juridiques, les raisons économiques, scientifiques et institutionnelles. Les entretiens en ont mis en lumière deux autres : l'écosystème de publication, et les pratiques scientifiques disciplinaires.

A/ Les freins juridiques liés à la nature des données

Item partagé avec l'enquête statistique, où il arrive largement en tête des obstacles au partage (avec 54 % des 108 répondants), le frein juridique a été également abondamment évoqué lors des entretiens par de nombreux chercheurs. On peut distinguer plusieurs types de raisons juridiques, constituant des freins, voire des impossibilités, au partage.
Tout d'abord, la question des données personnelles, avec au moins deux sortes de données juridiquement sensibles :
- **les données personnelles, de nature historique, et/ou judiciaire** : « *J'ai souvent travaillé sur des dossiers de justice ; un certain nombre d'affaires ont été jugées, c'est des éléments importants, même si l'historien a droit à la citation, y compris les noms ; mais de là à les partager après, ça ne me serait pas venu à l'époque de mettre ces données en partage* » ;
- **les données personnelles, liées à la protection de la vie privée** : cas fréquent en SHS, ces données personnelles nécessitent généralement une anonymisation, avec toutes les limites connues : « *On les anonymise, mais ce n'est pas si simple que ça (...). L'anonymisation absolue, ça n'existe pas. Du coup, comment on fait pour partager ça, si on n'arrive pas à les anonymiser ?* ».
La protection des données personnelles ou sensibles est d'autant plus complexe qu'elle s'applique ou non à certains projets, voire à certaines parties de projets de recherche : « *Il y a des projets sur lesquels on peut communiquer de manière grand public, mais il y a (...) des parties qui ne doivent pas sortir, parce que ça peut même être dangereux pour les personnes, pour les organismes, etc.* ».
La question de la confiance et des garanties données, dans la relation entre le chercheur et les populations ou les individus observés, est évidemment centrale, comme le montrent ces deux extraits, concernant des disciplines différentes :
- « *Il peut y avoir des réticences sur l'ouverture de ces données, d'une part parce qu'il y a des données un petit peu confidentielles ; (...) on est quand même tenus par un secret* » ;
- « *Pour les supports de conférence, une vidéo de « comment ça s'est passé » (...) : qu'est-ce que je peux montrer, est-ce que je floute les visages ? c'est vrai qu'on fait signer aux personnes les consentements de participation...* ».
Un autre type de frein juridique a été plusieurs fois évoqué, concernant le **respect du droit d'auteur des données**. Ainsi la question des images :
- « *L'image traitée, elle a une propriété, elle a un propriétaire au sens de droit d'auteur (...) ces données là, il faut qu'il les stocke mais elles ne sont pas nécessairement publiques* ».
Nous reviendrons plus loin sur l'un des aspects les plus sensibles de ce problème, avec le plagiat.
Enfin, plusieurs enseignants-chercheurs expriment une incertitude sur leurs droits et devoirs dans la réutilisation ou l'ouverture des données, voire leur méconnaissance des aspects juridiques des données de recherche :
- « *C'est sûr que les freins juridiques, ça pourrait être quelque chose sur lequel je ne suis pas vraiment au point. (...) C'est vrai que c'est des choses sur lesquelles je n'ai pas vraiment d'informations. Mais c'est une méconnaissance personnelle* ».

Mais si la nature juridique des données est un obstacle au partage pour une nette majorité, elle n'en est pas pour quelques uns :
- « *Freins juridiques, non. Non, je ne pense pas que ça nous concerne beaucoup* » ;
- « *Je pense que tant qu'il n'y a pas les textes, non, parce c'est des données qu'on a établies nous-mêmes*. ».
Et selon les terrains de recherche et les types de données, l'anonymisation totale des données personnelles reste possible : « *Elles sont sensibles au moment où je les collecte, mais elles ne le sont plus au moment où je les diffuse*. ».

B/ Les raisons économiques

Les raisons économiques (comme la protection par rapport à la concurrence) arrivaient en avant-dernière position dans l'enquête statistique, avec 7 % des 108 répondants. Les entretiens confirment cette très faible part des aspects économiques, dans les obstacles à la diffusion des données.
Une seule raison économique a été évoquée par un enseignant-chercheur, le caractère gratuit ou payant des données collectées, qui va conditionner leur éventuelle diffusion : « *Les données payantes, à partir du moment où il faut payer pour les avoir, la notion de les mettre à la disposition de tout le monde pose problème* ».

C/ Les raisons épistémologiques et scientifiques

Placé en quatrième position dans l'enquête statistique, avec 27 % des 108 répondants, l'item des raisons scientifiques comme frein au partage des données a été souvent cité dans les entretiens, et parfois abondamment commenté par quelques enseignants-chercheurs, pour qui il constitue le principal, voire le seul, véritable frein au partage des données.
Deux aspects, étroitement imbriqués, participent de cette dimension scientifique des données de recherche en SHS, et lui donnent toute sa complexité :
- d'une part, la question épistémologique de la définition même des « données de recherche » collectées, i.e. leur caractère construit ;
- d'autre part, la question scientifique, méthodologique, de leur indispensable contextualisation.

- **« Obtenues » *vs* « données » :**

La question n'est pas nouvelle et a fait l'objet d'une littérature théorique et scientifique, que nous ne pouvons reprendre ici[55]. Rappelons seulement cette évidence qu'en SHS, la notion de « donnée brute » est très discutable : les données collectées sont toujours obtenues selon un appareil méthodologique, une problématique, des outils, etc., préalablement déterminés. En bref, les « données » de la recherche ne sont justement pas « données » mais construites, comme le rappelle, de manière imagée, ce chercheur : « *Vous n'allez pas à la cueillette, il n'y pas des pâquerettes qui sont déjà là, c'est vous qui les produisez.* » ; « *Les données, c'est toujours des productions, (...) et j'enseigne ça à mes étudiants, et c'est pour ça qu'on les appelle pas des données, on les appelle des observables, et on dit "les observables, vous les construisez".* »
Un autre chercheur confirme cette position constructiviste, essentielle en SHS : « *Si vous, vous avez construit tout un appareil méthodologique qui permet de prélever des éléments, donc en aucun cas c'est donné, mais c'est construit, prélevé* ».
Pour le premier chercheur, la nature construite des données de recherche a une conséquence juridique directe : « *Ça veut dire aussi qu'on a un droit intellectuel sur ces productions, parce*

[55] Nous renvoyons le lecteur au billet de blog : Sylvie Fayet, « Données » de la recherche, les mal-nommées », *UrfistInfo*, 15 novembre 2013. Disp. sur : http://urfistinfo.hypotheses.org/2581

que c'est un vrai travail (...), ça pose le problème du droit intellectuel sur les observables qui ont été produits ».
Aspects épistémologiques et juridiques sont ainsi inextricablement mêlés, notamment en SHS, et la propriété intellectuelle peut, de fait, s'appliquer à certains types de données collectées et produites, dans la mesure où elles répondent au critère d'originalité des œuvres de l'esprit[56].

- **La nécessaire contextualisation :**

Dans le prolongement de cette dimension « constructiviste » des données collectées en SHS se pose la question de la contextualisation des jeux de données. Une enseignante-chercheuse rappelle fort justement le cadre général nécessaire à une diffusion des données : « *Il faut qu'on sache qui a fait quoi, et puis aussi qu'il y ait un vademecum, qu'on sache comment on peut utiliser ces données, comment elles ont été collectées* ».
D'autres chercheurs précisent les aspects disciplinaires de cette contextualisation, et expriment de sérieux doutes sur l'intérêt d'une diffusion et d'une réutilisation des données de recherche.
En linguistique :
- « *Dans ma discipline, les sciences du langage (...), dès lors qu'on travaille sur l'ancrage social et sur les interactions entre les pratiques langagières et les pratiques sociales dans lesquelles elles sont produites et qu'elles contribuent à produire, tu ne peux plus partager*. (...) ; *ça se partage mais avec des gens qui partagent le même contexte ; ou alors quand on présente nos travaux, y compris au séminaire du labo, on partage avec les gens qui vont poser des questions ; ils demandent toujours de la contextualisation.* » ;
- « *Oui, il y a quand même des freins à ce que le matériel ne soit pas dépouillé en quelque sorte de son objectif premier, même si c'est pour une autre utilisation à laquelle on n'a pas forcément pensé au départ.* ».
En sociologie :
- « *Je trouve que cela n'a pas forcément de sens de mettre à disposition ce genre de matériau, parce qu'en fait, il y a toute l'interaction sociale, l'historicité sociale qu'il y a dans cette relation. Seul celui qui a conduit l'entretien peut maîtriser et il peut le retransmettre s'il y a des questions, mais redonner tout ce contexte, c'est un lourd travail, pour après le partager.* » ;
En géographie :
- « *Quand on n'a pas participé au montage du projet, à la conception, à la définition de la problématique, récupérer derrière des données, ça me semble compliqué. Je ne dis pas que c'est impossible, mais (...) ça me semble compliqué de réutiliser des données produites par d'autres, en SHS* » **;**
- « *C'est surtout que c'est inexploitable, il y a un côté y compris un peu démagogique dans l'idée de faire croire que ce que l'on a intéresse tout le monde, c'est exploitable, parce que ce n'est pas le cas. C'est très compliqué* » ;
- « *Une donnée n'a de sens que dans le contexte où elle a été donnée et qu'en maîtrisant bien les métadonnées, c'est à dire les éléments qui renseignent sur la création et la valeur de la donnée.* (...) *Les biais sont multiples, en statistiques ou en géographie, on est toujours sensibles aux biais, c'est à dire de mal exploiter un corpus de données en ne tenant pas compte des effets de contexte qui font que voilà, telle corrélation on ne peut pas l'établir parce que les données ne permettent pas d'établir...* ».
En Sciences de l'Information et de la Communication :

[56] INRA, INSERM, IRSTEA, CIRAD, « *Ouverture des données de recherche. Guide d'analyse du cadre juridique en France* », 2016, 48 p. Disp. sur : http://www.bibliothequescientifiquenumerique.fr/wp-content/uploads/2017/01/Guide_analyse_Cadre_Juridique_Ouverture_donness_Recherche_V1.pdf

- « *Quand bien même je donnerais un corpus à quelqu'un, j'imagine l'effort qu'il me faudrait à moi pour le structurer pour le rendre appréhendable en fait par quelqu'un d'extérieur.* ».
Un chercheur résume cette diversité disciplinaire de la question du contexte :
- « *Je trouve que la question est complexe, je ne pense pas qu'on puisse avoir une règle qui serait la même en tout et pour tout, et qu'il faut imaginer plusieurs cas d'espèce.* (...) *C'est vraiment des situations, je ne pense pas qu'il y ait de règle générale, en tout cas universelle, et qu'il faut garder des spécificités propres à des choix méthodologiques* ».

D/ Les raisons institutionnelles

Les raisons institutionnelles, avec l'exemple d'un laboratoire ne voulant pas diffuser les données de recherche, arrivent bonnes dernières dans l'enquête statistique, ne recueillant que 3 % des 108 répondants. De même, dans les entretiens, personne n'a reconnu cet item comme frein possible au partage. Au contraire :
- « *Raisons institutionnelles, non pas du tout. Mon labo n'a aucun problème avec le fait de diffuser les données. Je pense qu'on le ferait si on avait un support de stockage pratique en fait.* ».

E/ Les facteurs liés à l'écosystème de publication de la discipline

Un autre frein au partage, non prévu dans le questionnaire, a été évoqué par quelques enseignants-chercheurs : le souci de l'économie des revues, des éditeurs, la question de l'embargo dans le libre accès, etc. Nous sommes ici à la frontière, particulièrement poreuse en SHS, entre le libre accès des publications et celui des données de recherche :
- « *Donc à partir du moment où on diffuse trop facilement, là je pense à des données non seulement produites, mais produites sous forme de livres, donc faciles d'accès, encore plus que les articles, on se pose la question de la survie de tout ce secteur de l'édition scientifique, semi-scientifique. » ;*
- *« Plus il y a de diffusion en ligne d'extraits de ces revues, moins il y a d'abonnés*. »
Un autre enseignant-chercheur insiste sur la primauté des ouvrages, dans les modes de publication scientifique :
- « *Je pense que toutes les productions ne doivent pas ou ne méritent pas d'être mises à disposition, parce qu'après il y aura peaufinement, récupération ou bien rassemblage d'un certain nombre de travaux pour que ce soit mis par exemple sous forme officielle à la disposition du public, mais au moins que ce soit légalisé sous forme d'un ouvrage ou sous forme d'un ouvrage collectif au nom d'une équipe* ».
De façon plus globale, le poids des revues, de la publication d'articles, d'ouvrages, de communications, etc., transparaît très nettement, et ce sans surprise, dans la quasi-totalité des entretiens. La diffusion des données de recherche apparaît soit comme secondaire par rapport aux publications, soit comme une possibilité, plus ou moins difficile à réaliser, parfois impraticable. En aucun cas, sauf sans doute dans les disciplines plus proches des sciences expérimentales (psychologie, géographie physique, étude du mouvement...), le partage des données de recherche n'est une priorité pour les enseignants-chercheurs de SHS, d'abord soucieux de publier articles et ouvrages.

F/ Les pratiques scientifiques disciplinaires : l'artisanat de la recherche

Le témoignage d'un enseignant-chercheur nous a paru emblématique d'un sixième facteur de frein à la diffusion des données de recherche : le poids des pratiques scientifiques disciplinaires, notamment le caractère artisanal, très individuel, sinon individualiste, de la recherche en SHS.

Ce n'est certes pas une nouveauté : la recherche dans de nombreuses disciplines des sciences sociales, des arts et des lettres, reste une pratique individuelle, voire solitaire, comme nous l'a expliqué ce collègue historien :
- « *On peut avoir encore un rapport très artisanal à la recherche. Donc c'est à dire aussi très individualiste, on est encore dans la culture du livre, voilà. Qui s'écrit seul, de bout en bout, ça reste encore le nec plus ultra de la profession* » ; « *chez nous, il y a encore cette idée que la part de travail individuel est très très déterminante* » ; « *je pense que nos collègues ont tous dans leurs cartons chez eux, dans leurs greniers en fait, enfin maintenant dans leurs disquettes, dans leurs disques durs, les résultats de leurs dépouillements ; ça tient au fait qu'on mène peu de projets collectifs. Je crois que c'est en grande partie ça*. ».

Cette solitude du chercheur, notamment en histoire, a évidemment des incidences sur son positionnement face au partage des données : d'abord en termes de « culture disciplinaire », puisque la diffusion des données n'est pas du tout inscrite dans la culture collective de la discipline, contrairement à d'autres ; mais aussi en termes d'incitation, de motivation au partage : « *C'est-à-dire qu'on n'est pas financé pour notre recherche. Donc on ne sent pas le devoir de partager puisqu'au fond, on fait nos recherches à nos frais, pour une grande partie* ». Et ces pratiques disciplinaires dominantes, qui vont conditionner fortement l'attitude des chercheurs sur le partage de leurs données, se doublent d'un autre facteur, lié au déroulement de carrière de nombreux chercheurs en SHS.
Le collègue historien donne ainsi une explication cohérente, à la fois aux plans scientifique et social, de ses réticences (car il ne s'agit pas non plus d'un refus délibéré) au partage des données. Tout d'abord, il faut repartir de la primauté de la revue, comme facteur dominant d'évaluation de la carrière des chercheurs en SHS : « *je trouve important d'essayer de placer des textes dans des revues exigeantes*. ». En histoire, la possibilité de publier dans une revue est d'abord liée à l'originalité du terrain de recherche : « *Et chez nous, c'est clair, dans les revues exigeantes, on ne peut placer de papiers que si on a un terrain vraiment original* ». Qu'est-ce qui constitue un « *terrain vraiment original* » en histoire ? Pour ce chercheur, c'est d'abord l'accumulation et l'originalité des données collectées, au prix d'un effort certain : « *Parce que se constituer un terrain vraiment original, c'est un effort. (...), c'est vraiment lourd. Quand on se constitue un terrain comme ça, c'est quelque chose de très précieux car c'est aussi la condition du maintien dans un niveau d'échange scientifique assez élevé* ». Se constituer un terrain original évoque un peu l'exploitation d'un riche filon dans une mine de données, qui sera exploité ensuite pendant plusieurs années pour pouvoir en tirer plusieurs publications : « *Donc en fait, quand moi j'accumule pendant des années de l'information comme ça, ça va permettre la publication d'une grosse publication. Mais ça me donne aussi des billes pour publier pendant 15 ans.* ». Au final, cette pratique de recherche, à la fois individuelle, parfois difficile et de long terme, ne prédispose pas vraiment au partage : « *C'est un terrain que vous vous êtes constitué, vous n'avez pas forcément envie de partager car vous perdriez un peu ce qui fait votre spécificité* ».

Cette pratique scientifique disciplinaire, largement individuelle, est confirmée également par une collègue historienne, qui indique, dans les autres facteurs de frein au partage des données, le poids de la discipline et certains de ses effets : « *J'avais noté le monde des disciplines, parce que dans notre discipline, je pense que toutes les SHS sont comme ça, mais chez les historiens c'est peut-être encore plus accentué, il y a quand même un attachement très fort à l'indépendance du chercheur et surtout ne pas trop s'immiscer dans ses pratiques et surtout ne pas lui imposer des choses*. »

On le voit : tous ces facteurs de frein au partage des données, que nous venons d'évoquer, et que nous avons qualifiés d'externes, se conjuguent, s'imbriquent plus ou moins étroitement, pour susciter des réticences, sinon des oppositions parfois très fortes à l'idée de partage des données, réticences qu'il importe de comprendre et de prendre en compte.

3.4.3.3.2 Facteurs « internes »

Pour avoir un aperçu complet des obstacles au libre accès des données de recherche, il faut ajouter les facteurs que nous avons nommés « internes », car ils nous semblent plutôt liés aux individus eux-mêmes, à leurs représentations, leurs craintes, leur ressenti, leurs compétences, leurs valeurs... Dans notre catégorisation, nous avions regroupé ainsi cinq facteurs : la crainte du plagiat, le manque de temps, le manque de connaissances dans la description des données, le manque de compétences, et les raisons personnelles. Nous les présentons ici selon leur ordre d'apparition dans la question 26 du questionnaire.

A/ La crainte du plagiat

Nous avons vu que la crainte du plagiat arrivait en troisième position dans l'enquête statistique, avec 37 % des 108 répondants. Ce qui est tout à fait significatif. L'enquête qualitative confirme ici la quantitative, car quasiment toutes les personnes interrogées ont mentionné, à un moment ou un autre, la peur, voire la hantise du plagiat, du vol de données... Les enseignants-chercheurs en SHS tomberaient-ils dans une sourde paranoïa ? Il est vrai que le plagiat est le principal type de fraude scientifique en SHS et une réalité assez largement documentée. Que disent les enseignants-chercheurs sur ce point ?
D'abord, le plagiat est identifié comme un facteur de frein important au libre accès des données de recherche :
- « *Si l'on met des retranscriptions d'entretiens brutes, se les faire détourner, voler, c'est des choses que j'ai déjà eues sans qu'il y ait de données en libre accès. On n'est pas à l'abri de ce genre de pratique, ça peut être un frein.* » ;
- « *C'est la crainte, primo du plagiat de la part des étudiants et des jeunes chercheurs, qu'ils prennent, qu'ils utilisent à leur manière, ou d'autres chercheurs ou d'autres collègues, peu importe* ».

Certains soulignent avec honnêteté ce qui est perçu comme une contradiction entre l'adhésion au libre accès et la « crispation sur les droits d'auteur » :
- « *Il y a une espèce de tiraillement là, entre peut-être notre crispation sur les droits d'auteur, la propriété intellectuelle, c'est-à-dire que si on met tout ça en ligne (...), tout le monde va nous voler nos idées, nos résultats, et puis en même temps la nécessité de rendre les choses plus accessibles et ça je crois que tout le monde sait bien que c'est une nécessité aussi et c'est même un devoir.* ».
D'autres évoquent le régime de concurrence auquel sont soumis *de facto* les chercheurs, et qui justifie de prendre des précautions par rapport au pillage :
- « *Le libre accès, faut faire attention ; c'est vrai qu'on est en concurrence entre guillemets, quand on reste sur un sujet de thèse, ou de recherche, ou un bouquin (...). Faut arriver à trouver un système qui nous permette à la fois de partager pour avancer sans se faire plagier ou piquer une partie de la réflexion.* » ;
- « *L'idée de partage en science devrait couler de source. Et en même temps, cela pose un certain nombre de problèmes à la fois pour protéger ses propres données et pour pas non plus se faire piller son propre travail. La recherche est malheureusement un univers où les chercheurs sont en concurrence entre eux, pas qu'au niveau d'un labo mais au niveau national*

ou international. Il y a des collègues peu scrupuleux ailleurs qui pourraient s'emparer facilement de données récupérées d'un travail produit dans un autre laboratoire de recherche ».
Enfin, un autre argument est avancé, lié au déséquilibre entre enseignants-chercheurs investis de responsabilités administratives ou éducatives, et d'autres qui peuvent se consacrer à la recherche à plein temps, et qui pourraient être tentés de plagier les premiers :
- « *Après, se faire voler les données, oui c'est toujours un risque ; le problème, c'est que tous les enseignants-chercheurs ne sont pas logés à la même enseigne, c'est-à-dire vous avez des enseignants-chercheurs comme moi, qui sont responsables de choses, et qui croulent sous les charges administratives, et leur temps de recherche est très faible par rapport à ce qu'il devrait être ; et d'autres qui eux, pour le coup, en fait refusent toute activité administrative et qui ont tout le temps pour faire ce qu'il faut* ».

Il faut relever aussi que, pour quelques chercheurs, la crainte du plagiat est beaucoup moins forte et ne constitue pas vraiment un facteur de frein au partage, compte tenu de la nature de leurs données :
- « *Ces données-là ne sont pas à plagier, cela n'a pas d'intérêt particulier, en revanche on sait qu'elles sont plébiscitées par des entreprises, par ce monde qu'on connaît moins.* »
Pour l'un des chercheurs interrogés, la peur du plagiat reposerait en fait sur un travail insuffisant de description des jeux de données :
- « *Normalement, quand on publie un jeu de données, on est censé être exhaustif dans la publication de ce jeu de données. Si on faisait bien notre travail, on n'aurait même pas à avoir peur du plagiat puisqu'on serait censé les décrire à fond de toute façon. C'est ça je pense, ce qui freinerait le plus mais ce n'est pas à mon sens une bonne raison en tout cas*. ».
Enfin, une enseignante-chercheuse a clairement récusé le plagiat comme frein possible au partage des données : « *La crainte du plagiat, non je ne crois pas, le but du jeu c'est que ce soit utilisé.* ».

B/ Les raisons « personnelles »

L'item « *Des raisons personnelles (je ne veux pas montrer mon « arrière-cuisine », mes données sont à moi)* » a été coché par 12 % des 108 répondants et l'enquête qualitative semble bien confirmer ces faibles résultats. Très peu de chercheurs ont reconnu ce facteur comme un frein important au partage, en faisant preuve, là encore, d'une grande sincérité :
- « *On est toujours un peu gênés par le fait de rendre nos données accessibles parce qu'on a toujours un peu peur que les autres viennent mettre le nez dans ce qu'on a fait et dire que peut-être on a mal fait notre travail. Et on a toujours un peu peur d'être passé à côté de quelque chose d'important ou que quelqu'un ait des idées qu'on n'a pas eues*. »
Au contraire, d'autres récusent ce facteur en plaidant pour ouvrir les « arrières-cuisines » :
- « *Non, l'arrière-cuisine ça me paraît hyper important de la montrer, je trouve qu'on ne la montre pas assez* (...)*, le résultat ne veut rien dire sans l'arrière-cuisine, donc toute la base est construite pour que l'arrière-cuisine soit visible, utilisable, exportable et qu'on puisse faire une autre recette avec les mêmes ingrédients* » ;
- « *Non ça je pense que au contraire, c'est important que les gens aient accès finalement à cette arrière-cuisine. C'est-à-dire qu'eux pourront peut-être faire autre chose avec les données que ce que moi je peux apporter comme traitement, donc il faut avoir accès au matériel, aux matières premières*. »

C/ Le manque de temps

Si le manque de temps arrive en deuxième position dans l'enquête statistique, avec 51 % des 108 répondants, il est beaucoup moins cité dans les entretiens, et ne semble pas vraiment constituer un frein au partage des données. Et les perceptions de ce manque de temps sont diverses :
- « *Ça prend forcément du temps à vraiment formaliser, à décrire, de façon très fine les données, sans pour autant savoir si un jour ça va être utilisé par quelqu'un ; donc y'a quand même le rapport coût-bénéfice* » ;
- « *Je ne pense pas que ce soit très coûteux en temps pour nous honnêtement. Normaliser un fichier avec des intitulés de colonnes un peu clairs, éventuellement diffuser un petit fichier avec pour dire la colonne 1 veut dire ça, la colonne 2 veut dire ça, c'est l'affaire de 15 minutes. Je pense que les bénéfices sont largement plus importants que le coût en temps.* » ;
- « *Le manque de temps : ben non, à partir du moment où je fais la base de données, c'est ça qui est long, ce n'est pas l'accessibilité* ».

D/ Le manque de connaissances dans la description des jeux de données, les modalités de dépôt

Si 26 % des 108 répondants ont cité le manque de connaissances comme facteur de frein au partage des données, presque aucun enseignant-chercheur interrogé ne l'a mentionné comme un obstacle. Cela ne signifie pas, *a contrario*, que tous les chercheurs maîtrisent parfaitement la description des jeux de données ou les modalités de leur dépôt ! Ce relatif silence ne témoignerait-il pas plutôt d'une non-prise en compte, ou d'une sous-estimation de l'importance de la description des données ? Quoi qu'il en soit, quelques chercheurs ont signalé que ce n'était pas un problème pour eux :
- « *Non c'est pareil, je ne pense pas qu'il y ait d'enjeux très complexes par rapport à nos données.* » ;
- « Le manque de connaissances dans la description des jeux de données et les modalités de dépôt *: Ben non je ne pense pas, parce qu'après, ce n'est pas moi qui gère (...), donc il n'y a pas de problème.* ».

Deux extraits sont intéressants à signaler, dans lesquels les collègues interrogés reconnaissent leur manque de connaissance des licences de dépôt des données de recherche et des règles de la propriété intellectuelle :
- « *Le fait est que, en tout cas à l'université, on n'est pas tellement au courant de tout cela, loin de là, et des évolutions très rapides* » (en parlant des licences Creative Commons, etc.) ;
- « *En fait, les freins c'est plus le temps à passer pour faire ça bien* (...)*, on manque d'outils et de soutien pour le faire bien, dans les règles de l'art.* (...) *j'ai déjà exprimé plusieurs fois que j'aurais besoin d'être formé sur tout ce qui relève de la propriété intellectuelle, parce que je sens qu'on n'est pas non plus armé pour en parler en connaissance de cause et être efficace* ».

E/ Le manque de compétences

Le manque de compétences (numériques, techniques, informationnelles...) a été cité par 19 % des 108 répondants ; dans les entretiens, cet item n'a pas été évoqué, sauf pour démentir qu'il puisse constituer un frein :
- « *Le manque de compétences, pareil.* », pour indiquer que ce n'est pas un problème ;

- « *Et le manque de compétences non plus puisque je travaille avec (...). Alors toute seule, je n'aurais fait rien du tout, évidemment* ».

F/ Autres raisons

Parmi les facteurs « internes », propres aux représentations personnelles des enseignants-chercheurs, on peut signaler également deux autres réponses faites par deux collègues, et qui témoignent de la diversité des représentations du partage des données :

- **la réticence à toute visibilité personnelle sur internet**, sentiment beaucoup plus partagé qu'on ne le croit chez beaucoup d'enseignants-chercheurs : « *(je suis) réfractaire à toute mise en scène, à la visibilité. J'aime bien contrôler ce qui est en ligne* » ;
- **une sorte d'excès de modestie scientifique,** avec le doute sur l'intérêt possible de ses données : « *Après, le frein, ça peut être aussi finalement notre secteur de recherche. C'est-à-dire est-ce que ça intéressera... est-ce que (...) des choses comme ça peuvent intéresser d'autres collègues ?* ».

3.4.3.4 Témoignages, interrogations, souhaits

En plus des positionnements personnels, favorables ou non, sur le partage des données, les entretiens ont permis de recueillir également de nombreuses interrogations, des témoignages, parfois des propositions, qui permettent de mieux cerner les représentations et les positionnements des enseignants-chercheurs.

A/ Témoignages sur les pratiques de partage des données

Quelques extraits d'entretiens nous éclairent sur les pratiques réelles de diffusion et de partage des données, notamment sur la primauté des équipes et des projets de recherche par rapport au laboratoire :

- « *Moi, l'expérience que j'ai* (du partage)*, c'est au sein des équipes ; au sein des programmes de recherche, on trouve une espèce de partage, mais autrement au sein du laboratoire, non.* » ;
- « *Oui, je l'ai déjà fait, dans le cadre d'un projet ANR, et sur le site du projet, on avait mis à disposition des jeux de données brutes* » ;
- « "seriez-vous d'accord pour diffuser vos données de recherche en libre accès à condition qu'elles soient diffusables et protégées juridiquement" : *encore une fois, ça dépend, oui, mais en fait ça dépend des projets* » ;
- « *Nous, le problème* (...) *c'est de trouver des moyens de partager nos données à l'intérieur d'une équipe qui mène un même projet. On a déjà du mal actuellement à avoir des outils qui nous permettent de mettre en commun nos données, même à petite échelle, à échelle intersite* ».

B/ Le besoin de garanties pour la diffusion des données

Chez de nombreux collègues, l'une des questions les plus sensibles concerne indubitablement les aspects de propriété intellectuelle, les besoins d'un cadre réglementaire permettant d'assurer la reconnaissance de paternité et du droit d'auteur. Nous avons vu plus haut à quel point la hantise du plagiat est forte, mais au-delà du plagiat, la question de la reconnaissance du travail de recherche, que constituent la collecte, le traitement et la mise en forme des données, est centrale. Avec les données de recherche, nous sommes bien au cœur du travail du chercheur et la volonté, parfaitement légitime, de celui-ci de voir reconnaître ce travail est largement partagée :

- « *Comment reconnaître le travail de celui qui s'est coltiné des heures et des heures pour aller chercher tel manuscrit au fin fond des dépôts d'archives ? Il faut qu'il y ait une forme, non pas de droit d'auteur, mais au moins de propriété intellectuelle* » ;
- « *Ça ne me pose pas de problème de les partager, à condition d'en être crédité et que ça fasse partie aussi, au même titre que la bibliographie papier, de l'évaluation, de l'estimation de la charge de travail réalisé, etc. C'est pour ça qu'il faut que cela apparaisse dans un cadre réglementaire et sur des sites qui ne soient pas privés, particuliers.* » ;
- « *Moi, je serais même assez partisan de mettre en ligne des choses raisonnées, qui sont plus proches des résultats de recherche que des données elles-mêmes. (...) Et puis il y a quand même un élément de confidentialité, c'est-à-dire des heures et des heures qui ont été passées à monter une donnée, je ne sais pas où commence et où s'arrête le droit d'auteur mais il y a toute cette question, quand même* » ;
- « *En même temps, on est vraiment sur des données communes ; qui a la propriété des données ? Et là, je ne parle pas des résultats scientifiques, je parle des données brutes. Là on avait une masse monstrueuse de données (entretiens, archives, etc.), on s'est retrouvés bloqués pour les utiliser parce qu'on n'avait pas de réponse sur ce qu'on avait le droit de faire, avec ces données.* ».

Au-delà de cette importance cruciale de la reconnaissance de paternité et des garanties juridiques qui doivent entourer la diffusion des données, on perçoit, à travers ces extraits, un autre élément, qui apparaît « en creux » : la méconnaissance, qui semble assez largement partagée, de ces garanties juridiques et des règles de propriété intellectuelle sur les données. Les besoins de formation sur les aspects juridiques apparaissent ici dans toute leur ampleur.

C/ Ouverture des données aux étudiants

Le souhait d'ouvrir les données de recherche aux étudiants a été formulé par deux enseignants-chercheurs :
- « *Je crois quand même que l'université, c'est le lien entre l'enseignement et la recherche ; j'aimerais bien que certaines choses qu'on fait* (...) *en recherche, puissent servir de support au niveau enseignement* (...) ; *qu'il y ait une partie de ce travail-là qui puisse être accessible, parce que les étudiants, c'est une partie du travail qu'ils ne voient pas* » ;
- « *Je me vois bien travailler avec des étudiants en leur disant, voilà vous avez des données, qu'est-ce que vous arrivez à produire à partir de ça ?* ».

D/ Interrogations sur le Libre Accès des publications

L'on sait à quel point ces deux aspects de la « Science ouverte », le Libre Accès des publications scientifiques et le partage des données de recherche, sont proches, sinon inséparables. Il était donc normal que les enseignants-chercheurs de SHS évoquent, à plusieurs reprises, la question du Libre Accès des publications, et formulent différentes interrogations :
- sur l'économie des revues et de l'édition, parfois perçue comme menacée par le Libre Accès : « *En particulier si tout chercheur commence à mettre le contenu de ses publications de façon intégrale sur internet, ça pose le problème de la survie des revues, qui vivent, elles sont subventionnées évidemment* », « *Je n'ai pas de solution toute faite, mais je sais que l'open access, ça pose une série de problèmes.* » ;
- sur le « versionning » des textes : « *Ensuite sur les questions des textes et des variantes, c'est des problématiques nouvelles qui se posent à moi. Je travaille sur les dépôts des textes sur les archives ouvertes et en préparant plusieurs textes plutôt anciens, je me rends compte qu'il y a quand même des problématiques complexes ; donc la version, est-ce que c'est la version qui a*

été déposée ou est-ce qu'au bout de 15 ans, il faut plutôt revoir la version parce qu'on n'a pas toujours la même... Ce n'est pas tout à fait les données mais en tout cas je découvre une série de problématiques qui sont liées, c'est quand même à la fois le stockage et la visibilité. ».

3.4.4 Synthèse sur les pratiques de partage des données

3.4.4.1 Quelles leçons ?

Quelles leçons principales retenir des deux enquêtes, quantitative et qualitative, sur la question du partage des données en SHS ?

A/ La complexité de la question

La première leçon à retenir est sans doute un truisme, mais qu'il nous faut redire : c'est la complexité même de la question du partage des données en SHS, s'exprimant à travers le spectre des nuances et des positions à la fois épistémologiques, scientifiques, éditoriales, informationnelles, disciplinaires, individuelles... Toute « politique des données de recherche » devrait d'abord se fonder sur cette diversité, et prendre en compte cette complexité, à rebours des discours d'injonction trop « militants ». Si rien n'est simple en matière de données de recherche, quelle que soit la discipline, la complexité est peut-être redoublée en sciences humaines et sociales et dans les disciplines littéraires.

B/ Un certain hiatus

L'enquête statistique a montré d'abord un certain hiatus entre les déclarations de principe, dans l'ensemble favorables au partage des données et au Libre accès, les représentations, parfois négatives, parfois positives mais souvent inquiètes, et enfin les pratiques réelles, attestant une relative faiblesse de la mise en ligne des données. Les entretiens ont permis de nuancer le tableau, d'approfondir le détail des arguments (pour ou contre) et des représentations, comblant ainsi en partie l'impression d'écart entre pratiques et représentations que produit l'enquête statistique, forcément plus schématique. Mais les entretiens ont confirmé également, en leur donnant plus d'ampleur, l'interrogation, voire l'inquiétude, des enseignants-chercheurs devant la libre diffusion des données de recherche, souvent perçue comme un risque. Nous avons été frappés par la force et la diversité des réticences sur le partage des données, et ce constat, qui n'est pas nouveau, évoque cette citation, assez imagée, de la chercheuse américaine Felicia LeClere : « *Partager ses données, c'est un peu comme aller chez le dentiste. Nous pouvons tous convenir que c'est une bonne chose à faire et que c'est lié, de manière intrinsèque, à une bonne pratique scientifique. En réalité, toutefois, les chercheurs ont tendance à voir le partage de données avec un mélange de peur, de dédain et d'effroi.* »[57].

C/ Le poids des facteurs internes dans les réticences au partage

Dans ces réticences face au libre accès des données de recherche, les deux enquêtes concordent finalement sur le poids de ce que nous avons nommé les facteurs "internes" : nous avons vu l'importance des inquiétudes sur le droit d'auteur et sur la crainte du plagiat, tant dans les

[57] Voici la citation d'origine : « *Data sharing is a bit like going to the dentist. We can all agree that it is a good thing to do and intrinsic to good scientific practice. In reality, however, researchers tend to view data sharing with a mix of fear, contempt, and dread.* ». LeClere, F. (2010, août). Too Many Researchers Are Reluctant to Share Their Data. *The Chronicle of Higher Education*. Disp. sur : http://www.chronicle.com/article/Too-Many-Researchers-Are/123749)

entretiens que dans les résultats statistiques. Mais au-delà de ces questionnements sur la propriété intellectuelle, explicables et légitimes, ce qui nous semble la leçon principale des deux enquêtes, dans les réticences dues aux représentations, tient sans doute à la perception et à la connaissance du sujet, encore lointaine et souvent imprécise : comme l'a noté l'un des enseignants-chercheurs interrogés, les données de recherche constituaient jusqu'alors une sorte « d'impensé » du travail de recherche. Omniprésentes, à la base même de la plupart des travaux de recherche, les données (collectées et produites) restent souvent, et paradoxalement, « invisibles », transparentes, aux yeux des chercheurs, tant elles font partie de leur quotidien, de leur environnement. C'est peut-être d'ailleurs l'un des intérêts de cette enquête, la première du genre à l'Université Rennes 2, que d'avoir provoqué une sorte de prise de conscience de l'importance et de l'autonomie de cette question. Quoi qu'il en soit, cet « impensé des données de recherche » transparaît dans de nombreux résultats, et la nouveauté de cette question des données de recherche explique certainement la plupart des méconnaissances des évolutions en cours, des enjeux, des conditions juridiques...

D/ Le facteur disciplinaire

Une troisième leçon porte sur le poids des cultures et des pratiques disciplinaires : nous savions déjà à quel point la dimension disciplinaire est essentielle dans la thématique des données de recherche. Deux aspects, selon nous, illustrent ce facteur : la discipline elle-même et les pratiques collaboratives à l'intérieur des champs disciplinaires.

- **Un déterminisme disciplinaire ?**

L'enquête a non seulement confirmé mais amplifié ce postulat de base : la discipline, la sous-discipline, et même le domaine de recherche, sont les premiers vecteurs conditionnant le partage des données, soit pour le favoriser (voire le rendre obligatoire), comme en géographie physique, en psychologie, en biomécanique ; soit pour le permettre ou l'encourager, comme en géographie humaine et sociale, en linguistique (de corpus), en SIC, en histoire ancienne ; soit encore pour le freiner, voire le décourager, comme parfois en arts et lettres, en langues, en histoire moderne et contemporaine, en sociolinguistique... Ce tableau ne tient pas compte, évidemment, des positions individuelles de plusieurs chercheurs, qui peuvent démentir cette répartition, forcément schématique, mais qui nous semble refléter certaines tendances disciplinaires. Ce poids disciplinaire renvoie bien sûr aux raisons scientifiques, avancées comme principales objections à une diffusion des données (cf. ci-dessous).

- **Des pratiques collaboratives inégales :**

Au-delà de la dimension scientifique de chaque discipline, de son écosystème éditorial (qui joue un rôle considérable, comme nous l'avons vu pour la psychologie), une autre dimension nous semble importante à prendre en compte, davantage liée aux pratiques sociales, aux « habitus » disciplinaires : le caractère plus ou moins collectif, collaboratif, du travail de recherche. On sait à quel point certaines disciplines sont aux antipodes l'une de l'autre, si l'on met en regard les articles signés par des centaines de physiciens et les travaux individuels en histoire ou en lettres. Aucune révélation n'a été apportée par notre enquête sur ce point, dans la mesure où la solitude du chercheur en SHS est un phénomène bien connu. Mais certains entretiens ont permis de mieux cerner le poids de ce facteur dans le positionnement face au partage des données, de donner chair à cet « artisanat de la recherche », qui transparaît dans plusieurs entretiens, et qui explique une large part des réticences face à une libre diffusion des données. La problématique du partage (des informations, des données, des publications...), la pratique du travail collaboratif, très forte dans certaines disciplines (notamment les disciplines expérimentales), est étrangère à d'autres cultures, où l'emporte largement le travail individuel. Outre la solitude

du chercheur en SHS, qui ne prédispose pas vraiment au travail collaboratif et au partage des données, un autre facteur apparaît, surtout dans les entretiens, qui explique un certain nombre de réticences : une perception, peut-être exacerbée, de la concurrence entre chercheurs, qui n'encourage pas non plus au travail collaboratif et à la culture du partage.
Il nous faut nuancer ce propos sur le lien entre facteur disciplinaire et appétence pour le partage des données, car nous ne croyons nullement à une sorte de « déterminisme disciplinaire », qui enfermerait les chercheurs et déterminerait complètement leur position sur la question du libre accès. Il est évident que la culture du partage, du travail collaboratif, est très inégalement répartie d'abord entre les individus, entre les enseignants-chercheurs eux-mêmes. On peut être « partageur » dans le cadre d'une discipline traditionnellement individualiste ou au contraire très « fermé » au sein d'une discipline, collaborative par nécessité. Mais qu'elle soit favorisée par le cadre social de la discipline, ou acquise individuellement par le chercheur, la « culture du partage », la pratique du travail et des outils collaboratifs, nous paraissent être des facteurs essentiels, voire déterminants, pour le partage des données.

E/ Une culture minoritaire du Libre Accès

Une autre leçon de l'enquête porte sur la culture et l'adhésion aux valeurs de l'Open Access. Rappelons-en les principaux éléments, puisqu'une remarque nous a été faite de considérer ces valeurs comme implicites, en reprenant cette définition de l'UNESCO : « *L'information scientifique est un apport essentiel des chercheurs et la principale ressource pour l'innovation technologique. L'UNESCO soutient et encourage le libre accès – la mise à disposition en ligne pour tous de l'information savante, sans la plupart des barrières imposées par le droit d'auteur et le copyright – afin de favoriser la circulation du savoir, l'innovation et le développement socio-économique à l'échelle planétaire*. »[58]. On peut citer également cette définition de la littérature en libre accès, donnée par Peter Suber, l'un des acteurs historiques de l'Open Access : « *La littérature en accès ouvert est numérique, en ligne, gratuite et exempte de la plupart des restrictions en termes de droit d'auteur et de licence d'utilisation.* »[59]. Ces définitions dessinent en creux les principales valeurs philosophiques au fondement du libre accès : la transparence, l'ouverture et la mise à disposition de tous des résultats de la science, considérée comme un « bien commun » de la connaissance, en vue de la diffusion la plus large des savoirs[60]. Si l'enquête statistique montre que 40 % des enseignants-chercheurs voient l'adhésion à ces valeurs comme une incitation au partage des données, elle établit aussi qu'un peu moins de la moitié des répondants se déclarent familiers des enjeux du Libre Accès. Plusieurs entretiens ont laissé voir une certaine méconnaissance, voire une absence complète, de la « culture du libre accès », au sens d'une culture personnelle faite de connaissances déclaratives, de savoir faire, d'adhésion aux valeurs, de sensibilisation aux enjeux, etc., liés au mouvement du Libre Accès. Ainsi, la crainte ou l'obsession du plagiat, l'opposition parfois exacerbée entre éditeurs et archives ouvertes, entre revues scientifiques et libre accès, la méconnaissance des licences Creative Commons, certaines remarques sur les évolutions de la communication scientifique... témoignent d'une perception souvent lacunaire, partielle, de la réalité du Libre Accès, des mutations de la publication scientifique. Comme pour la culture collaborative, dont elle est

---

[58] UNESCO, *Libre accès à l'information scientifique*. Disp. sur : http://www.unesco.org/new/fr/communication-and-information/access-to-knowledge/open-access-to-scientific-information/

[59] SUBER, Peter. *1. Une définition* In : *Qu'est-ce que l'accès ouvert ?* [en ligne]. Marseille : OpenEdition Press, 2016. Disp. sur : http://books.openedition.org.distant.bu.univ-rennes2.fr/oep/1605

[60] Sur les valeurs et les principes philosophiques du libre accès, voir : Patrick Juignet, Web et libre accès, In *Philosophie, Science et Société*, 20 avril 2015. Disp. sur : https://philosciences.com/Pss/philosophie-et-societe/economie-politique-societe/34-web-et-libre-acces

proche mais avec laquelle on ne saurait la confondre, la culture du libre accès connaît d'importantes variations disciplinaires, et se trouve par ailleurs déterminée par l'expérience personnelle. Ainsi, l'adhésion au Libre Accès nous a paru plus affirmée chez différents cadres et responsables de la recherche, ce qui souligne là encore la part importante de la dimension individuelle, reposant sur des attachements personnels aux valeurs du libre accès.

F/ Des objections fondées contre le partage des données

Dans les objections au partage des données, deux ensembles d'arguments, construits et légitimes, dominent largement, comme nous l'avons vu :
- d'une part, et sans surprise, les facteurs juridiques, avec notamment la protection des données personnelles, ce qui conforte une fois de plus la place cruciale des aspects juridiques des données de recherche ; nous y reviendrons, dans les recommandations ;
- d'autre part, et les entretiens se sont révélés ici particulièrement précieux pour mieux comprendre cet autre argument contre le partage, l'item un peu vague de l'enquête statistique, les « raisons scientifiques ».

Sur ce point, plusieurs entretiens ont permis d'expliciter cette question essentielle de la contextualisation des données, qui représente, selon nous, l'objection scientifique la plus forte au discours d'injonction d'un partage universel et systématique des données de recherche.
Nous touchons ici aux spécificités des SHS, où l'on sait que la notion même de « données » ne va pas de soi. Nous verrons les conséquences de cette spécificité, en termes de politiques possibles des données de recherche.

G/ Une culture de l'IST à développer

Enfin, une autre leçon à tirer de ces deux enquêtes porte sur une autre culture, qui selon nous englobe celle du Libre Accès : la culture de l'information scientifique, au sens large du terme, i.e. la connaissance, déclarative et pratique, de tous les aspects de l'information scientifique et technique (schématiquement, toute la chaîne production-traitement-publication-diffusion-évaluation... de l'information scientifique, quelle que soit la discipline). Certes, notre enquête ne portait pas sur les compétences et connaissances des enseignants-chercheurs en matière d'IST en général, mais uniquement sur la thématique des données de recherche. Mais ce qui apparaît explicitement, ou transparaît implicitement, tant dans l'enquête statistique que dans les entretiens, c'est une culture de l'IST inégalement répartie et insuffisante, avec au moins trois points critiques de méconnaissance, chez de nombreux enseignants-chercheurs :

- les aspects juridiques : qu'il s'agisse du droit des données de recherche, des notions de la propriété intellectuelle, de la protection contre le plagiat, des contrats d'éditeurs et du dépôt sur les archives ouvertes, des licences pour la diffusion des données, des évolutions du contexte, etc., les besoins de conseil, d'accompagnement et de formation des chercheurs et des personnels de recherche, en matière juridique, sont immenses, et l'enquête rennaise ne fait que confirmer ici d'autres constats ;
- les métadonnées et la description des jeux de données : ignorance massive des PGD, faiblesse de la description des jeux de données, des pratiques de nommage de fichiers, etc., cette dimension, aujourd'hui de plus en plus importante en IST, reste encore insuffisamment connue et maîtrisée ;
- les entrepôts de données : la connaissance des différents entrepôts de données, des grandes infrastructures de recherche, semble également assez minoritaire, aussi bien selon les résultats statistiques que dans les entretiens.

3.5 Quels besoins, quelles attentes ?

3.5.1 Les résultats de l'enquête statistique

L'enquête statistique a consacré quatre questions (les questions 28 à 31) à l'investigation des besoins et des attentes des chercheurs sur les différents aspects impliqués par la gestion de données de recherche, en les sondant à la fois sur leurs pratiques, leurs convictions et leurs souhaits : qu'estiment-ils important de mettre en œuvre en matière de sauvegarde, stockage, archivage, description et partage des données ? À quel niveau serait-ce pertinent ? Dans quelle mesure sont-ils familiarisés avec les procédures de gestion de données de recherche et éprouvent-ils le besoin d'être accompagnés, formés, conseillés sur toutes ces opérations ? Il s'agissait en définitive de cerner le niveau et les perspectives d'appropriation, ainsi que de donner des pistes pertinentes pour la possible mise en place de services en la matière.

3.5.1.1 Quelle connaissance du Plan de Gestion de Données ?

La première question dans cette partie concernait non pas l'usage confirmé d'un Plan de Gestion de Données mais, plus modestement, la connaissance plus ou moins précise de la notion de PGD : « Savez-vous ce qu'est un Plan de Gestion de Données (PGD), ou Data Management Plan (DMP) ? ». La question était obligatoire et 109 personnes sur 143 y ont répondu, soit un taux de réponse de 76,2%.

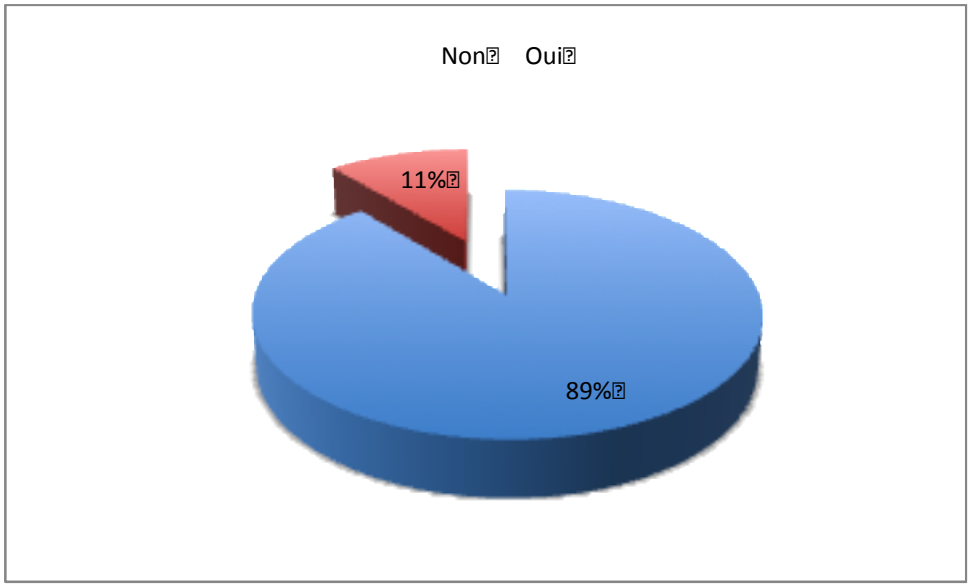


Figure 55 : Connaissance du Plan de Gestion des Données (N = 96)

Le résultat des réponses données est tout à fait clair et montre une méconnaissance très partagée, puisque 89 % des répondants (97 personnes sur 109) déclarent en effet ne pas savoir ce qu'est un Plan de Gestion de Données. Ceci ne veut pas dire pour autant que les chercheurs n'ont pas mis en place de procédure personnelle concernant la gestion de leurs données, mais en revanche qu'ils ignorent le fonctionnement et / ou l'existence de cet outil d'assistance et de normalisation.

Les résultats obtenus montrent-ils une différence selon l'âge ou le statut des répondants ? On compte parmi les répondants 66 personnes qui ont entre 30 et 50 ans (soit 60,5 % des répondants) et 42 personnes (38,5 %) qui ont plus de 50 ans. À la question « Savez-vous ce qu'est un PGD ? », les 12 réponses « oui » sont le fait de 10 personnes de 30 à 50 ans, qui représentent 15 % de leur tranche d'âge et 83 % des réponses positives, les deux dernières réponses positives venant de personnes de plus de 50 ans (soit 17 % des oui et 5 % de la tranche d'âge). Inversement, la réponse « non » a été sélectionnée par 56 personnes entre 30 et 50 ans (représentant 58 % des non et 85 % de cette tranche d'âge), tandis que pour les plus de 50 ans, on trouve 40 réponses « non » à la question (soit 41 % des non, et 95 % de la tranche d'âge). Ainsi certes le pourcentage de réponses négatives est moins fort chez les plus de 50 ans (41 % contre 58 %), mais il est le fait de presque toute cette tranche d'âge (95 % contre 85 % pour les 30-50 ans). Même si cela ne joue qu'à la marge, les chiffres indiquent une méconnaissance du PGD qui peut s'accentuer pour les répondants les plus âgés.

< 30 ans	**30-50 ans**	**> 50 ans**
1	66 (60,5 %)	42 (38,5%)

	< 30 ans	**30-50 ans**	**> 50 ans**	**Total**
Non	1 = 1 % des non = 100 % de la tranche d'âge	56 = 58 % des non = 85 % de la tranche d'âge	40 = 41 % des non = 95 % de la tranche d'âge	**97 (89 %)**
Oui		10 = 83 % des oui = 15 % de la tranche d'âge	2 = 17 % des oui = 5 % de la tranche d'âge	**12 (11 %)**
Total	**1**	**66**	**42**	

Figure 56 : Connaissance du PGD selon l'âge

On a considéré ensuite le statut des répondants. Les catégories majoritaires étant par ordre descendant les MCF (56 %), les Professeurs (32 %) et les HDR (8 %), on observe dans le tableau ci-dessous que proportionnellement à leur nombre, les trois statuts les plus représentés ont répondu par la négative dans des proportions à peu près analogues (les réponses « non » correspondent respectivement à 90, 91 et 89 % des effectifs de chaque catégorie de statut), tout en totalisant 57 % des non pour les MCF, 33 % pour les Professeurs et 8 % pour les HDR. Le statut du chercheur ne semble donc pas peser ici de façon sensible, du moins dans les ordres de grandeur des échantillons recueillis.

Chercheur	**HDR**	**Ingénieur**	**MCF**	**Professeur**
2 (2 %)	9 (8 %)	2 (2 %)	61 (56 %)	35 (32 %)

	Chercheur	HDR	Ingénieur	MCF	Professeur	Total
Non	1 = 1 % des non = 50 % des répondants du statut	8 = 8 % des non = 89 % des répondants du statut	1 = 1 % des non = 50 % des répondants du statut	55 = 57 % des non = 90 % des répondants du statut	32 = 33 % des non = 91 % des répondants du statut	**97 (89 %)**
Oui	1 = 8 % des oui = 50 % des répondants du statut	1 = 8 % des oui = 11 % des répondants du statut	1 = 8 % des oui = 50 % des répondants du statut	6 = 50 % des oui = 10 % des répondants du statut	3 = 25 % des oui = 9 % des répondants du statut	**12 (11 %)**
Total	**2**	**9**	**2**	**61**	**35**	

Figure 57 : Connaissance du PGD selon le statut

Enfin, une analyse des réponses par laboratoire montre qu'une minorité de laboratoires (7 sur 19) a répondu positivement en indiquant une connaissance du PGD, et que parmi ceux-ci, les Sciences sociales totalisent le plus de réponses : 3 laboratoires de Sciences Sociales, ESO-Rennes, Tempora et LiRIS (6 réponses, dont 4 pour le laboratoire ESO), 2 laboratoires d'Arts, Lettres, Communication, APP et PREFics (3 réponses), 1 laboratoire de Langues, le LIDILE (2 réponses), enfin 1 laboratoire de Sciences Humaines, le CREAD (1 réponse).

	Non	Oui	Total
ACE	3		**3**
APP	10	1	**11**
CELLAM	5		**5**
LiRIS	7	1	**8**
COSTEL	4		**4**
CRBC	3		**3**
CREAD	3	1	**4**
LP3C	12		**12**
ERIMIT	7		**7**
ESO	11	4	**15**
HCA	3		**3**
IRMAR	1		**1**
LAHM	1		**1**
LIDILE	4	2	**6**
M2S	5		**5**
PREFics	3	2	**5**
Recherches en Psychopathologie	2		**2**
Tempora	8	1	**9**
VIPS²	5		**5**
Total	**97**	**12**	

Figure 58 : Connaissance du PGD selon les laboratoires

3.5.1.2 Quels opinions et souhaits en termes de stratégie des données de recherche à l'université ?

Nous avons souhaité ensuite connaître le positionnement personnel des chercheurs de Rennes 2 sur les dispositifs et la démarche institutionnelle qui seraient à leurs yeux souhaitables à l'université pour faciliter la gestion de leurs données de recherche, indiquant par là même à la fois leurs attentes et leurs représentations de ce que pourrait être une véritable stratégie dans ce domaine. C'est donc l'objet de la question 29, question à facettes, qui a sollicité leur avis sur les différents aspects de la gestion des données en proposant une gradation dans leurs réponses, depuis « tout à fait d'accord » jusqu'à « pas du tout d'accord ».

Les réponses ci-dessous sont classées selon le total des réponses « Tout à fait d'accord » et « Plutôt d'accord » :

	Tout à fait d'accord	Plutôt d'accord	*Total accord*	Sans opinion	Plutôt pas d'accord	Pas du tout d'accord	Nombre de réponses	Pourcentage de participation
Je souhaiterais pouvoir stocker mes données de recherche au-delà de la durée de vie d'un projet	46	43	***89***	7	2	1	**99**	69%
Je pense qu'une déontologie des données pourrait devenir un enjeu lorsque les données de recherches sont partagées avec les autres	41	42	***83***	14	0	1	**98**	69%
Je suis prêt à partager mes données de recherche avec les autres	15	52	***67***	8	20	7	**102**	71%
Je souhaiterais que l'université dispose d'un plan de gestion de données	28	28	***56***	28	1	1	**86**	60%
Je suis familier avec les enjeux de l'open access	9	39	***48***	12	23	14	**97**	68%
Je souhaiterais que l'université dispose d'un ensemble de métadonnées prescrit pour verser les données vers un entrepôt	19	15	***34***	42	3	3	**82**	57%
Je souhaiterais que l'université utilise/recommande un système de nommage standard de fichiers	12	20	***32***	40	6	8	**86**	60%
Total	**170**	**239**	***409***	**151**	**55**	**35**		

Figure 59 : Tableau de synthèse des positions personnelles et souhaits en matière de données de recherche

Cette question était optionnelle. Le taux de réponse moyen sur l'ensemble des items est de 64,85 %, mais l'aspect le plus renseigné se rapporte au partage des données : 71 % de réponses à la question sur le partage personnel des données de recherche, 69 % sur le principe d'une déontologie relative aux données de recherche dans le cadre d'un partage, 68 % sur la

connaissance des enjeux du libre accès : c'est un aspect qui semble focaliser les attentions. Le taux de réponse sur la question du stockage pérenne est aussi parmi les plus importants (69 %). Faisant peut-être moins débat, les aspects plus techniques sur le Plan de Gestion de Données, les métadonnées ou le nommage des fichiers, même s'ils obtiennent des taux de réponses non négligeables (entre 57 et 60 %) ont moins mobilisé les chercheurs.
Pour ce qui est des réponses en elles-mêmes, il est intéressant de voir si des points de consensus, des objets de débat ou des réticences fortes s'expriment à travers les réponses.

A/ Sur le stockage des données

Le constat le plus évident est celui d'un quasi-consensus dans les attentes en matière de stockage à long terme des données.

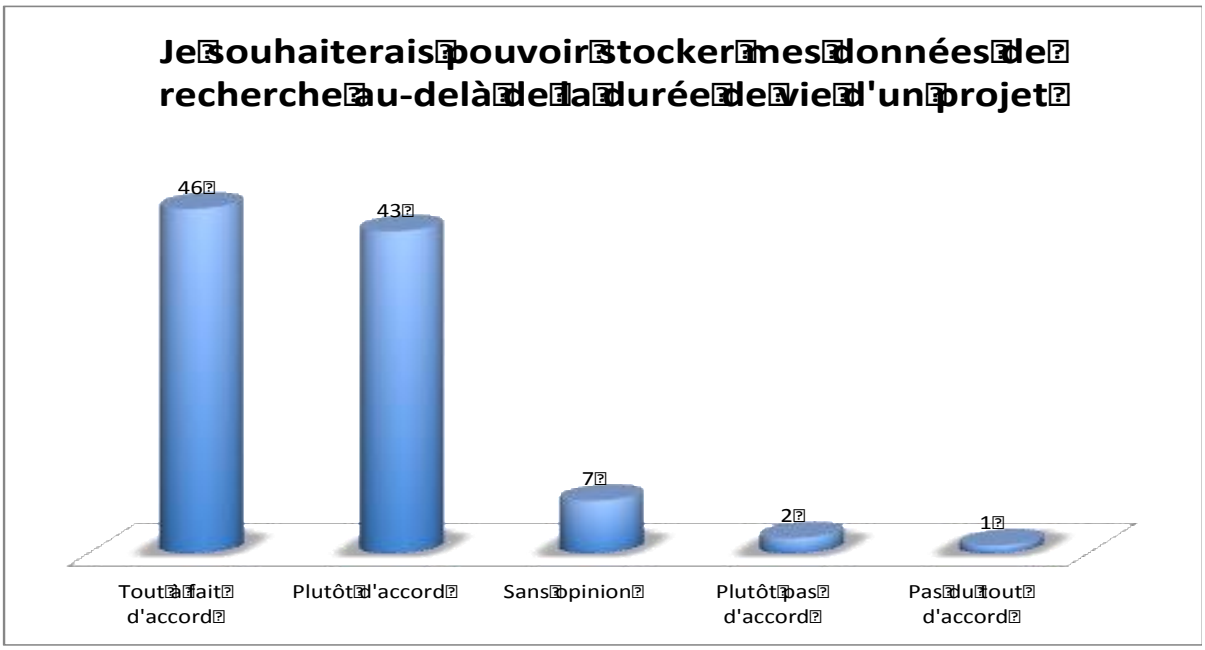


Figure 60 : Sur un stockage pérenne des données (N = 99)

En effet, 89 des 99 chercheurs ayant répondu sur cet item (soit près de 90 %) s'accordent sur l'intérêt de pouvoir stocker les données au-delà de la durée de vie du projet de recherche, 3 réponses seulement faisant état d'un avis négatif. La question ne permet pas de dire dans quelles conditions ce stockage pourrait s'effectuer, et ne préjuge en rien d'un besoin homogène en termes de capacité, de volume (voir les résultats sur les pratiques de stockage) ; il semble cependant patent que la constitution d'un corpus personnel de données, sinon permanent, du moins accessible sur le long terme, fait l'objet d'une communauté de vues. Ce point est à mettre en relation à la fois avec les éléments de réponse sur l'absence ou l'ignorance d'une politique d'archivage dans les laboratoires (questions 15 à 18), avec les constats exprimés dans les entretiens sur les problèmes d'archivage sécurisé des données et avec la question 30 sur les services et aides souhaités. Il est aussi à corréler avec les déclarations sur les possibilités de réutilisation des données.

C'est un souhait partagé dans les mêmes proportions, quel que soit l'âge des répondants, puisque 90 % des 30-50 ans (54 répondants sur 60) et 89 % des plus de 50 ans (34 répondants sur 38) ont choisi les éléments de réponse « plutôt d'accord » et « tout à fait d'accord » (voir

tableaux dans l'Annexe), représentant respectivement 61 % pour les 30-50 ans et 38 % pour les plus de 50 ans, des réponses globalement favorables au stockage des données sur le long terme. Le statut des répondants, en revanche, ne semble pas totalement indifférent et on note une position très forte des Professeurs en faveur d'un stockage long des données de recherche, dans la mesure où 97 % des professeurs (30 répondants) sont plutôt d'accord ou tout à fait d'accord, ce qui représente 34 % des réponses globalement favorables (par rapport au total des 2 items), contre 88 % des MCF (51 répondants), ce qui représente 57 % de ces mêmes réponses. Les HDR, 7 répondants, sont globalement favorables à hauteur de 71,5 % des répondants du statut et 6 % des réponses positives, enfin 2 chercheurs sur 2 sont tout à fait d'accord. Les 3 réponses négatives viennent de MCF et il serait intéressant de pouvoir en connaître les raisons, alors que les entretiens qualitatifs menés avec les chercheurs (voir infra) confortent ce souhait de stockage à long terme.
Lorsqu'on réunit les opinions tout à fait d'accord et plutôt d'accord, on voit aussi que cinq laboratoires sont particulièrement intéressés au stockage long de leurs données de recherche : ESO Rennes (14 réponses favorables sur 14), LP3C (11 réponses favorables sur 12), Tempora (8 réponses favorables sur 8), APP (7 réponses favorables sur 10), LIDILE (6 réponses favorables sur 6).

B/ Sur la gestion des données

La question 29 comportait trois éléments demandant aux chercheurs s'ils souhaiteraient qu'au niveau de l'université se mettent en place une démarche et un dispositif pour faciliter et homogénéiser la gestion des données de recherche : « Je souhaiterais que l'université dispose d'un Plan de Gestion de Données », « Je souhaiterais que l'université dispose d'un ensemble de métadonnées prescrit pour verser les données vers un entrepôt », et enfin « Je souhaiterais que l'université utilise / recommande un système de nommage standard de fichiers ».

- **Le Plan de Gestion de Données :**

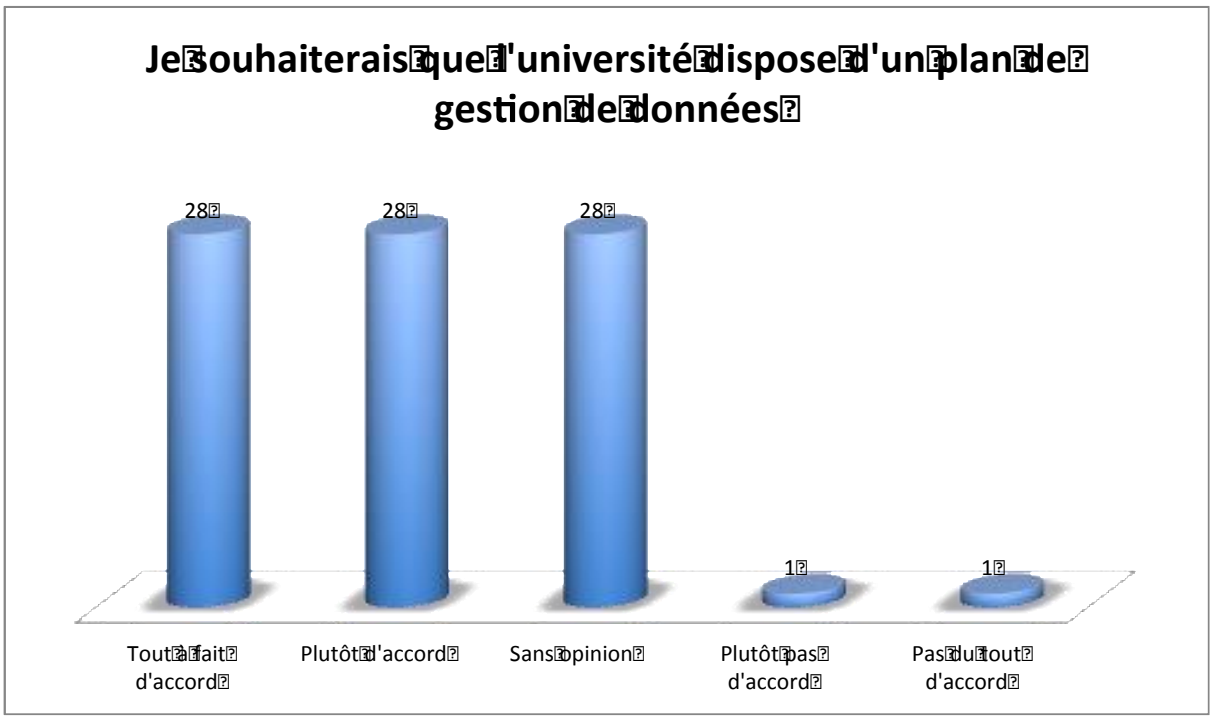


Figure 61 : Sur un Plan de Gestion de données commun à l'université (N = 86)

La question était intéressante à poser si l'on se rappelle que 97 répondants sur 109 avaient indiqué (à la question 28) ne pas savoir ce qu'était un Plan de Gestion de Données. Les réponses montrent qu'il n'y a pas d'opposition au développement d'un PGD au niveau de l'université : seuls 2 répondants ne sont pas d'accord, 56 répondants sur 86, c'est-à-dire environ 65 % des répondants, souhaitent un PGD (28 sont tout à fait d'accord et 28 plutôt d'accord).

Les croisements avec l'âge et le statut, rapportés au nombre de répondants dans chaque catégorie, ne donnent pas de différence significative ; étant entendu par ailleurs que 28 répondants, soit environ un tiers, sont sans opinion (16 MCF, 10 Professeurs, 2 HDR), ce qui semble naturel compte tenu de l'ignorance dans laquelle sont beaucoup de chercheurs par rapport au PGD. Il est d'ailleurs à noter qu'une partie des répondants connaissant mal le PGD se retrouve forcément dans les réponses qui s'accordent à le souhaiter. L'analyse des réponses par laboratoire montre que 17 laboratoires sur 19 ont répondu à cet item, et parmi eux, ceux qui ressentent le plus la nécessité d'un PGD au niveau de l'université sont ESO (9 réponses favorables sur 11), LIDILE (6 réponses favorables sur 6), APP (6 réponses favorables sur 10) et LIRIS (5 réponses favorables sur 8). Reste que la part des « sans opinion » est prégnante et que cette mention se retrouve, à deux exceptions près, dans les réponses de tous les laboratoires. Il semble qu'un travail de sensibilisation et avant tout d'information soit nécessaire avant d'engager toute démarche en ce sens.

- **Les métadonnées pour la description des données de recherche :**

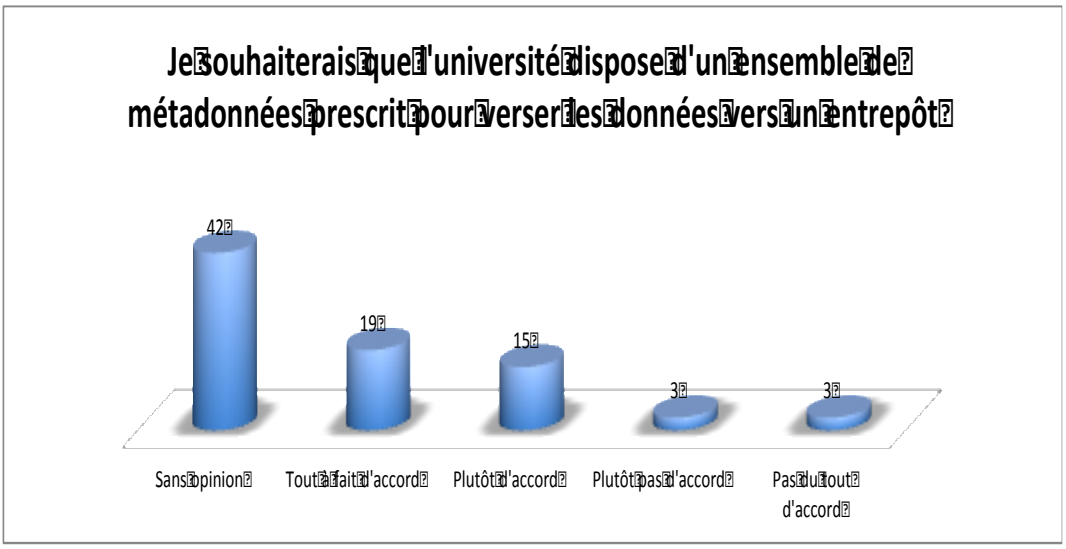


Figure 62 : Sur une politique d'établissement pour les métadonnées (N = 82)

On a vu à la section 3.3.1.3 sur les pratiques en matière de description des données que les chercheurs étaient très peu nombreux à utiliser des standards de métadonnées, et même visiblement qu'ils étaient très peu nombreux à les connaître. Les réponses obtenues à la question 29 sur le besoin d'une mise à disposition de métadonnées par l'université, témoignent d'une cohérence avec les usages relevés plus tôt. En effet, ce sont les « sans opinion » qui prédominent, soit 42 réponses sur 82, plus de la moitié des répondants ; il est probable que ces réponses traduisent chez beaucoup une vision peu claire de ce que recouvrent les métadonnées, et par conséquent de l'intérêt qu'il y aurait à en disposer.

On observe toutefois que 34 répondants, soit plus de 41 %, souhaiteraient un standard de métadonnées puisqu'ils ont coché les cases « Plutôt d'accord » ou « Tout à fait d'accord » ; et

si l'on regarde les résultats par tranche d'âge des répondants (cf. Annexe), il apparaît que 19 de ces 34 répondants ont entre 30-50 ans (soit 56 % des réponses favorables), 14 ont plus de 50 ans (soit 41 % des réponses favorables). On remarque aussi que, proportionnellement au nombre de répondants, ce besoin s'exprime plus fortement chez les plus de 50 ans, qui se sont déclarés tout à fait d'accord ou plutôt d'accord à 48 % de leur classe d'âge, tandis que 36 % seulement des 30-50 ans ont exprimé les même avis favorables sur une politique d'établissement pour les métadonnées.

Relativement aux statuts des répondants, les Professeurs ont été 14 sur 27, soit 52 % de leur classe d'âge à être plutôt d'accord ou tout à fait d'accord sur cette question, ce qui représente 41 % de l'ensemble des réponses favorables. Les MCF ont été également 14 à exprimer le même point de vue (41 % des réponses favorables), mais avec 30 % de la classe d'âge cette fois, ils ont été proportionnellement moins nombreux que les Professeurs. Le nombre de répondants HDR, chercheurs et ingénieurs n'étant pas suffisamment significatif, il n'a pas été pris en compte pour l'analyse de ces croisements. En tout cas, finalement, rares sont ceux qui refusent le principe d'une prescription de métadonnées par l'établissement (6 seulement). Ce qui cadre plutôt bien avec le fait que les chercheurs sont assez nombreux à informer leurs données, que ce soit par le biais du cahier de laboratoire ou même d'une manière personnelle (cf. questions 19 et 20). Reste que beaucoup n'expriment pas d'avis sur le sujet. La répartition des réponses par laboratoire met bien en évidence le poids des interrogations autour de ces standards de métadonnées, ainsi même lorsqu'un laboratoire comme ESO Rennes a un regard positif, il y 3 réponses « tout à fait d'accord », 3 réponses « plutôt d'accord », mais aussi 5 réponses «sans opinion ».

C/ Sur un système de nommage des fichiers

Sur le souhait de voir l'université préconiser un système de nommage standard de fichiers, on constate là encore une forte proportion de « sans opinion », soit que la nature technique de la question ait créé l'indécision, soit que le chercheur se soit fait son propre système de nommage auquel il reste attaché. Quoi qu'il en soit, près de la moitié des répondants (40 chercheurs sur 86) ne se sont pas prononcés sur cet item. Reste que, parmi ceux qui se sont prononcés, se dégage une majorité (32 personnes, soit 37 %) en faveur d'un standard de nommage.

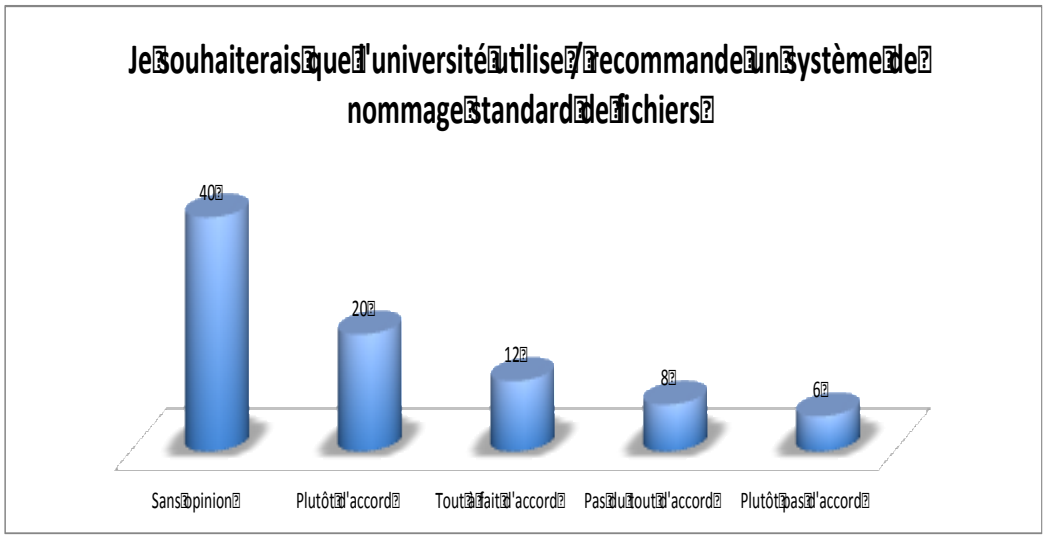


Figure 63 : Sur un système de nommage de fichiers (N = 86)

Sur ces 32 réponses positives, 59 % ont été exprimées par des répondants de 30 à 50 ans, mais ces 59 % représentent 35 % de la classe d'âge ; 37,5 % sont le fait de répondants de plus de 50 ans, soit 38,5 % de la classe d'âge. Il n'y a donc pas sur cet élément de réponse de véritable écart générationnel. Quant au statut, 50 % des réponses favorables (tout à fait d'accord et plutôt d'accord) viennent des MCF (37,5 % des Professeurs), cependant les MCF sont moins nombreux proportionnellement (32 %) à s'engager sur une norme universitaire de nommage des fichiers que les Professeurs (44 % de leur classe d'âge) ; 51 % des MCF sont sans opinion sur le sujet, et représentent 65 % des sans opinion, contre 44 % des Professeurs (30 % des sans opinion). La répartition par laboratoires montre une indécision assez marquée de la part de laboratoires qui par ailleurs ont pu s'engager assez fortement sur les questions précédentes, tels ESO (6 réponses favorables sur 12, 5 sans opinion, 1 réponse défavorable), ou Tempora (1 réponse favorable, 6 sans opinion, 1 réponse défavorable).

D/ Sur le partage des données et le libre accès

On a vu que 3 items de la question 29 concernent plus particulièrement le partage des données et le libre accès : la conscience ou non des enjeux du libre accès ; l'accord ou non du chercheur pour partager ses données ; la conviction ou non qu'une déontologie des données est importante dans le cadre de leur partage.
Les réponses à ces questions ont été traitées dans la partie 3.4 « Pratique de partage et de diffusion des données de recherche », en 3.4.1.2 « Quelles représentations, quelles opinions sur le partage des données de recherche ? ». Nous renvoyons donc aux analyses présentées dans cette section.

3.5.1.3 Quels besoins de services pour la gestion des données de recherche ?

On s'est intéressé ensuite plus particulièrement aux besoins concrets que les chercheurs pouvaient faire remonter en termes de services, infrastructures ou outils pour les aider dans la gestion de leurs données de recherche avec la question 30 : « Sur quels points souhaiteriez-vous bénéficier d'une aide (de votre unité de recherche, de la DSI, du SCD, de l'URFIST, de la MSHB…) ? ». Sur ces besoins potentiels de services, la question 30 s'inspire très largement de l'enquête de Lille, dont elle reprend textuellement 7 éléments sur 8 : un espace d'archivage sécurisé et fiable, des conseils pour la gestion des données, des conseils sur les questions techniques, des conseils juridiques, des conseils relatifs à la publication et à la citation des données, une assistance technique, une aide pour préparer un plan de gestion des données. Il a semblé pertinent d'intégrer en outre dans les services éventuels, une offre de formation sur la question des données de recherche. La question, en introduisant une notion de priorité dans les réponses possibles (en priorité, éventuellement, non) permettra de hiérarchiser le développement de services qui pour certains se mettent déjà progressivement en place localement (formations doctorales, stages de l'URFIST, services d'appui à la recherche du SCD).

Besoins exprimés	En priorité	Éventuel-lement	Non	Nombre de réponses	% de participation
Un espace d'archivage sécurisé et fiable pour mes données de recherche	54 *(57 %)*	31	9	94	66 %
Des conseils sur des questions techniques	40 *(41 %)*	49	8	97	68 %
Des conseils juridiques	35 *(38 %)*	48	9	92	64 %
Une assistance technique	34 *(38 %)*	44	12	90	63 %
Des conseils d'ordre général pour la gestion des données de la recherche	30 *(31 %)*	57	11	98	69 %
Une offre de formation sur la question des données de recherche	25 *(27 %)*	53	13	91	64 %
Des conseils relatifs à la publication des données de la recherche	22 *(23 %)*	52	20	94	66 %
Une aide pour préparer un plan de gestion de données	22 *(27 %)*	44	15	81	57 %
Total	**262**	**378**	**97**		

Figure 64 : Tableau des besoins par ordre décroissant de priorité

A/ Archivage sécurisé, conseils techniques et juridiques : trois besoins jugés prioritaires

Si l'on regarde l'expression des besoins prioritaires, le taux de réponses montre que c'est d'abord la mise à disposition d'un espace d'archivage sécurisé et fiable qui arrive en tête, avec 54 réponses sur 94, soit plus de 57 % des réponses. Ce qui semble cohérent avec le souhait de pouvoir stocker ses données sur le long terme (« je souhaiterais pouvoir stocker mes données de recherche au-delà de la durée de vie d'un projet ») exprimé largement à la question précédente et qui fait écho notamment aux problèmes de pertes de données et aux réticences quant au stockage sur le cloud, évoquées dans les entretiens. Il s'agit là d'un besoin d'infrastructures et d'outils, dont il faudra déterminer l'échelle de mise en place.
Viennent ensuite des demandes de conseils et d'assistance, techniques, juridiques ou de nature plus générale (respectivement 41,23 % des réponses pour les conseils sur les questions techniques, 38,04 % pour les conseils juridiques, 37,77 % pour l'assistance technique et 30,61 % pour les conseils plus généraux sur la gestion des données de recherche), exprimant là un besoin assez diversifié d'accompagnement. Une formation à proprement parler sur cette question des données de recherche n'arrive qu'en sixième position, avec 25 choix de cet item (27,47 %). Il semble donc que soit plutôt réclamé un accompagnement plus informel, ciblé et personnalisé.
Enfin les deux derniers items, concernant un accompagnement prioritaire autour de la publication des données de recherche et du Plan de Gestion de Données, totalisent chacun seulement 22 réponses (avec 23,40 et 27,16 % des réponses). Et ce sont aussi ces deux propositions qui comptent le plus de réponses « non » quant à leur utilité (respectivement 21 % et 19 %).
Tous ces résultats seraient sans doute à pondérer en fonction des pourcentages de participation, qui varient selon les items, mais dont cependant l'écart reste modéré (de 57 % à 66 %).

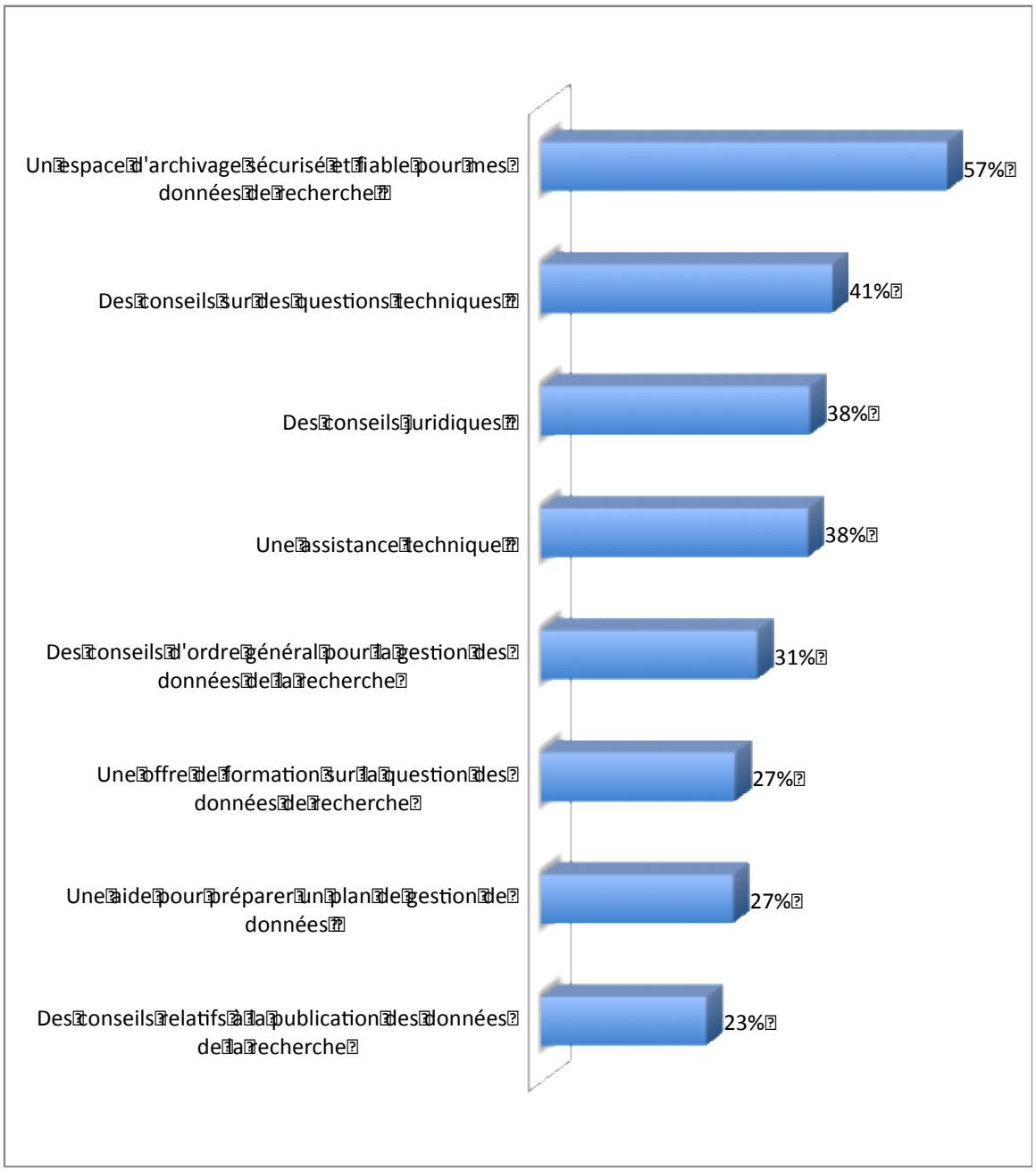


Figure 65 : Besoins par ordre décroissant de priorité

Certaines similitudes se dégagent avec l'enquête menée à Lille 3 : il est frappant de voir que le besoin d'un archivage sécurisé des données est une évidence dans les deux cas, demandé à Lille par 54 % des répondants, à Rennes par 57 %. On retrouve également un consensus sur les services de conseil : les conseils techniques et juridiques sont là aussi plébiscités avec 49 % et 48 % de taux de réponse (taux de réponses moindres à Rennes). En revanche, dans l'enquête lilloise, les conseils pour la gestion des données en général arrivent en tête et les conseils relatifs à la publication et la citation des données de recherche recueillent un assentiment nettement plus large (43 % contre 23 % à Rennes). L'assistance technique est souhaitée à peu près dans les mêmes proportions (40 contre 38 %), de même que l'aide au PGD, qui ne constitue pas une priorité et dont un nombre finalement relativement réduit de chercheurs éprouvent la nécessité.

B/ Une attente globale autour de la gestion des données de recherche

Si l'on agrège maintenant les réponses des besoins estimés comme prioritaires à celles des attentes secondaires, pour avoir une vision globale de ce qui est souhaité, quel qu'en soit le degré de nécessité ressenti, on obtient un ordre légèrement différent, bien qu'analogue dans les grands équilibres.

Type de besoins ***P = prioritaire*** ***P + E : prioritaire + éventuellement***	**Globalement souhaités** ***prioritaires + éventuellement***
1. *P : Un espace d'archivage sécurisé et fiable pour mes données de recherche* P + E : Des conseils sur des questions techniques	91,75 %
2. *P : Des conseils sur des questions techniques* P + E : Un espace d'archivage sécurisé et fiable pour mes données de recherche	90,42 %
3. *P : Des conseils juridiques* P + E : Des conseils juridiques	90,21 %
4. *P : Une assistance technique* P + E : Des conseils d'ordre général pour la gestion des données de recherche	88,77 %
5. P : *Des conseils d'ordre général pour la gestion des données de recherche* P + E : Une assistance technique	86,66 %
6. *P : Une offre de formation sur la question des données de recherche* P + E : Une offre de formation sur la question des données de recherche	85,71 %
7. *P : Une aide pour préparer un plan de gestion de données* P + E : Une aide pour préparer un plan de gestion de données	81,48 %
8. *P : Des conseils relatifs à la publication des données de la recherche* P + E : Des conseils relatifs à la publication des données de la recherche	78,72 %

Figure 66 : Intégralité des souhaits exprimés (prioritaires et secondaires)

Trois besoins forts ressortent :

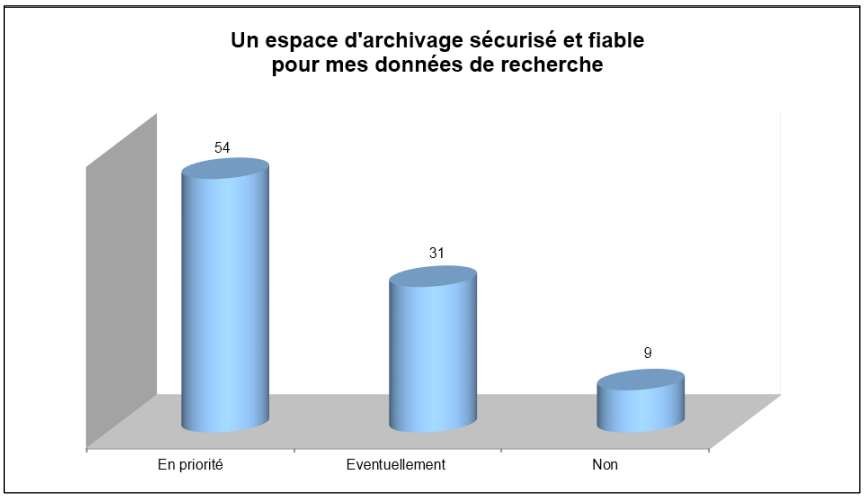


Figure 67 : Besoin d'un espace d'archivage des données (N = 94)

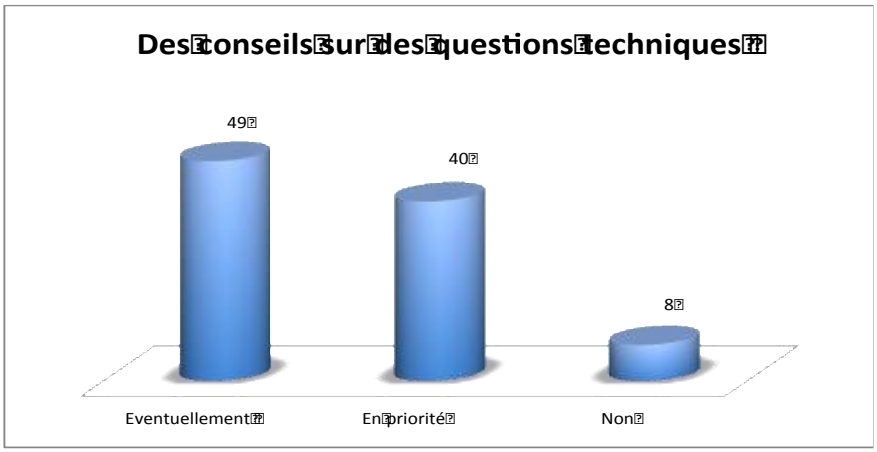


Figure 68 : Besoin de conseils techniques (N = 97)

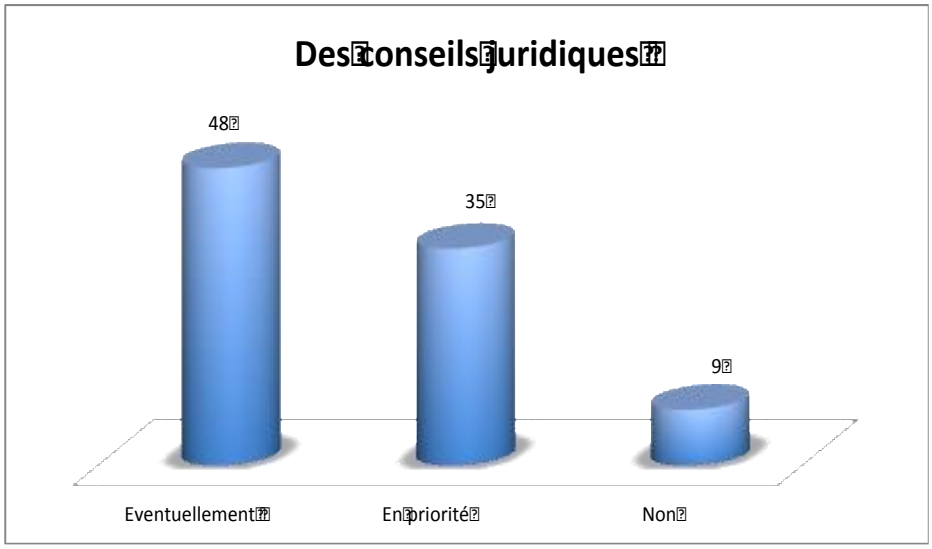


Figure 69 : Besoin de conseils juridiques (N = 92)

Ces trois souhaits prioritaires varient-ils en fonction de l'âge ou du statut des chercheurs ? L'espace d'archivage sécurisé et fiable est considéré comme prioritaire en premier lieu chez les 30-50 ans (pour 62 % de la tranche d'âge ; 48,5 % de la tranche d'âge des plus de 50 ans). Les conseils sur les questions techniques sont particulièrement privilégiés par les plus de 50 ans (ils sont prioritaires pour 51 % de la tranche d'âge), tandis que le chiffre passe à 36 % de la classe d'âge des 30-50 ans. Quant aux conseils juridiques, ils sont aussi majoritaires chez les plus de 50 ans, à hauteur de 41 % de la tranche d'âge, contre 37 % pour les 30-50 ans.

Sur les 5 items suivants, on a des valeurs plus élevées chez les chercheurs de plus de 50 ans dans 4 cas sur 5 : seule l'offre de formation sur la gestion des données de recherche est moins considérée comme prioritaire chez les plus de 50 ans que chez les 30-50 ans, mais peut-être est-ce la formation elle-même comme catégorie de service qui est moins sollicitée chez des chercheurs plus âgés ; en revanche, des services de type aide et conseils, sans doute plus facilement personnalisables et plus ponctuels, retiennent davantage leur attention.

Réponses « En priorité »	% de la tranche d'âge : 30-50 ans	% de la tranche d'âge : > 50 ans
Assistance technique	33 %	47 %
Conseils d'ordre général	28 %	36 %
Offre de formation sur les DR	32 %	19 %
Aide au PGD	22 %	37 %
Conseils relatifs à la publication des DR	20 %	29 %

Si l'on considère le statut des répondants, on voit que l'espace d'archivage vient en tête chez les MCF, qui sont 68,5 % de leur catégorie professionnelle à le privilégier, contre 37 % des professeurs. Les deux choix prioritaires suivants : les conseils techniques et les conseils juridiques sont en revanche mentionnés en priorité par les Professeurs, en particulier pour les conseils techniques qui ont été indiqués par 45 % des Professeurs et par 38,5 % des MCF ; les conseils juridiques, quant à eux, constituent une priorité pour 38 % des Professeurs et pour 36,5 % des MDC.

Les Professeurs sont encore les plus nombreux, à une exception près (l'aide à la préparation du PGD, qui peut apparaît comme un dispositif lourd et chronophage ; voir les entretiens) à souhaiter en priorité les autres services :

Réponses « En priorité »	% des répondants du statut MCF	% des répondants du statut Professeurs
Assistance technique	34 %	41 %
Conseils d'ordre général	21 %	47 %
Offre de formation sur les DR	24,5 %	31 %
Aide au PGD	27 %	18 %
Conseils relatifs à la publication des DR	13 %	38 %

Le détail des analyses croisées avec l'âge et le statut pour tous ces items est disponible dans l'Annexe des croisements.

Le croisement entre les items de la question 30 sur les services souhaités et l'appartenance à un laboratoire fait apparaître une différence assez marquée entre certains laboratoires qui ont des taux de réponse relativement élevés et d'autres unités qui ne répondent que peu, voire pas du tout sur les éléments de réponse évoqués. On étudiera ici les réponses des 5 laboratoires ayant en moyenne les plus forts taux de réponse, soit à partir de 6 réponses positives (« en priorité » et « éventuellement ») : ESO-Rennes, Tempora, LiRIS, LP3C et APP, soit 3 laboratoires de Sciences sociales, 1 laboratoire de Sciences humaines et un laboratoire d'Arts Lettres Communication. Les pourcentages correspondent au taux de réponses favorables de chaque laboratoire, différencié entre « en priorité » (dans la première colonne) et « éventuellement » (dans la deuxième colonne), par rapport à l'ensemble des réponses faites par ce laboratoire sur l'item considéré.

Le tableau ci-dessous permet de constater une approche finalement assez différenciée des besoins prioritaires selon les laboratoires :

- Si l'espace d'archivage sécurisé est largement souhaité par l'ensemble des 5 laboratoires avec des taux allant, pour la catégorie « en priorité », de 60 % à 75 % des réponses de chaque laboratoire, ce dernier pourcentage étant celui de Tempora (75 % de 8 réponses apportées), d'autres besoins peuvent être répartis de façon très différentes d'un laboratoire à l'autre. Il est à noter par ailleurs qu'ESO a toujours un nombre de réponses assez élevé et apparaît souvent à la fois largement sensibilisé aux problématiques des données de recherche et en même temps partant pour des services diversifiés en la matière ;
- Un apport de conseils sur des questions techniques est considéré comme une priorité par ESO, APP, Tempora ; en revanche il n'est souhaité qu'éventuellement pour LP3C et LiRIS ;
- Une différence d'appréciation se fait également sentir entre les laboratoires sur les conseils juridiques : alors qu'ils sont une priorité pour Tempora, avec un pourcentage majoritaire des réponses « en priorité », et dans une moindre mesure pour ESO et LiRIS (avec des taux inférieurs à ceux de la réponse « éventuellement », mais néanmoins non négligeables), ils ne semblent pas être une nécessité dans un premier temps pour APP et LP3C et sont plutôt vus comme un service secondaire ;
- L'assistance technique est plus particulièrement sollicitée comme une priorité par APP et LiRIS, tandis que les autres laboratoires la choisissent assez largement en service éventuel ;
- Il semble en tout cas que des aides et conseils ciblés soient plutôt mis en avant, de préférence à des conseils d'ordre général, qui ne sont pas vraiment prioritaires pour les 5 laboratoires, avec moins de mentions « en priorité » que de mentions « éventuellement » sur cet élément de réponse (seul APP a répondu « en priorité » et « éventuellement » dans les mêmes proportions sur cet item). De même la proposition d'une offre de formation n'a pas recueilli une majorité de réponses « en priorité » et elle apparaît surtout en service éventuel. Est-ce parce que les chercheurs, disposant de peu de temps, ne sont pas enclins à participer à des formations formalisées, ou parce que leur préférence va à des aides ciblées, personnalisées sur tel ou tel aspect de la gestion de leurs données de recherche ? Les entretiens menés avec les chercheurs semblent conforter le besoin de conseils et de soutiens sur la durée, notamment en termes de méthodologie et d'outils de gestion des données ;
- Enfin, on note que sur les conseils pour la publication et la citation des données, seul APP le voit comme une priorité ; en revanche, c'est une chose que les 5 laboratoires citent largement en service éventuel ;
- De même l'aide à l'élaboration d'un PGD est souhaitée plutôt éventuellement que prioritairement : APP et LiRIS la mentionnent cependant à égalité en service prioritaire et en service éventuel.

Items de la question 30	**Réponses en priorité / éventuellement**									
Nombre de réponses et pourcentages par rapport au nombre total de réponses de chaque laboratoire	**ESO**		**APP**		**LP3C**		**Tempora**		**LiRIS**	
Espace d'archivage sécurisé et fiable	**8** **61,5 %**	5 38,4 %	**6** **60 %**	3 30 %	**6** **60 %**	4 40 %	**6** **75 %**	2 25 %	**5** **62,5 %**	1 12,5 %

Nombre de réponses et pourcentages par rapport au nombre total de réponses de chaque laboratoire	**ESO**		**APP**		**LP3C**		**Tempora**		**LiRIS**	
Conseils techniques	**7 53,8 %**	5 38,4	**5 55,5 %**	3 33,3 %	**1 9%**	8 72,7 %	**6 66,6 %**	3 33,3 %	**3 37,5 %**	4 50%
Conseils juridiques	**5 41,6 %**	7 58,3 %	**1 11,1 %**	4 44,4 %	**2 18,1 %**	8 72,7 %	**6 66,6 %**	3 33,3 %	**3 37,5 %**	4 50 %
Assistance technique	**3 23 %**	8 61,5 %	**5 55,5 %**	3 33,3 %	**2 22,2 %**	6 66,6 %	**3 33,3 %**	4 44,4 %	**5 62,5 %**	2 25 %
Conseils d'ordre général sur les DR	**5 35,7 %**	9 64,2 %	**4 44,4 %**	4 44,4 %	**2 20%**	4 40 %	**1 11,1 %**	8 88,8 %	**2 25 %**	5 62,5 %
Offre de formation sur les DR	**4 28,5 %**	9 64,2 %	**2 22,2 %**	5 55,5 %	**2 18,1 %**	5 45,4 %	**2 22,2 %**	6 66,6 %	**3 42,8 %**	3 42,8 %
Conseils sur la publication et la citation des DR	**1 7,6 %**	9 69,2 %	**4 44,4 %**	3 33,3 %	**2 20 %**	5 50 %	**2 25 %**	5 62,5 %	**2 25 %**	5 62,5 %
Aide au PGD	**4 33,3 %**	5 41,6 %	**3 37,5 %**	3 37,5 %	**1 11,1 %**	7 77,7 %	2 25 %	5 62,5 %	3 42,8 %	3 42,8 %

Figure 70 : Croisement des besoins pour cinq laboratoires

Si les besoins les plus prégnants (pour plus de 90 % des répondants) se concentrent, comme on le voit, toujours sur la sécurisation et l'archivage des données en milieu universitaire, ainsi que sur des conseils, soit d'ordre technique, soit d'ordre juridique, l'ensemble de ces chiffres semble indiquer aussi une attente et une attitude globalement ouverte face à l'ensemble des propositions d'accompagnement et de services : d'une part peu de chercheurs estiment inutile une aide, quelle qu'en soit la nature, pour la gestion de leurs données, d'autre part, on pourrait évidemment supposer que des chercheurs jugent délicat de déclarer une absence de besoins ; cependant le taux de réponses, qui est important, montre que beaucoup de chercheurs se sont sentis concernés par les différentes propositions et parle en faveur d'une mise en place de services spécifiques sur cette question. Là encore, on retrouve une analogie avec l'enquête de Lille 3, dans laquelle s'est manifestée également une demande très majoritaire de services, puisque seulement 5 % des chercheurs y indiquaient ne pas avoir besoin des services proposés.

Quelques chercheurs (9 sur 143, 6,2 %) ont même souhaité utiliser la question de texte libre (question 31) destinée à recueillir des suggestions complémentaires : « Si vous souhaitez bénéficier d'une aide sur un point non listé ci-dessus, précisez lequel »[61]. Lorsqu'on essaie de catégoriser le contenu de leurs réponses, on remarque que certaines demandes rejoignent et

[61] Certaines réponses ont été revues, pour respecter l'anonymat.

soulignent des besoins déjà recensés, notamment un besoin d'espace et d'aide technique pour le stockage et l'archivage des données : « *C'est vrai que beaucoup de données sont perdues du fait de l'évolution des supports. Un service qui nous aiderait à les mettre en sûreté et à les rendre disponibles pour la seule recherche publique serait alors un progrès* » ; ensuite, un besoin de formation : « *Une formation expliquant les différentes possibilités de stockage et les conditions de partage pourrait m'être utile. Idéalement cette formation devrait également reposer sur la gestion de données et les questions techniques* » ; enfin, un besoin de conseils juridiques : « *Mieux connaître les aspects juridiques et de protection des données* ».

Catégorisations des besoins	**Nombre de réponses**
Stockage et archivage sécurisés des données	2
Travail collaboratif	2
Politique multi-sites	2
Formation	1
Conseils juridiques	1
Tous les services proposés	1

Figure 71 : Tableau des besoins exprimés en réponse libre

En revanche, deux éléments originaux particulièrement intéressants sont mis en avant par 4 des réponses. Le premier renvoie à une pratique collaborative de la recherche et au besoin d'une infrastructure universitaire pour un archivage collaboratif dans le cadre d'un projet de recherche commun :

« *L'un des problèmes est celui de l'archivage collaboratif pour les projets de recherche. Pour l'heure, il n'existe aucun outil universitaire qui permet de partager des données entre chercheurs impliqués sur un même projet. Dropbox est limité en taille et l'université ne peut pas financer une dropbox. Googledrive est lui aussi limité et interroge aussi sur la sécurisation des données.* »

« *Non listé dans les points précédents mais les systèmes de documents partagés sont très intéressants (shareLatex, svn) dans nos travaux de recherche. Cela optimise le travail sur des données communes et toujours actualisées.* »

Le second élément interroge le niveau d'échelle d'une politique en matière de données de recherche et évoque la nécessité de prendre en compte l'environnement institutionnel des équipes de recherche : les besoins en matière de gestion des données de recherche ne s'apprécient pas exclusivement dans le périmètre de l'université et il est pertinent d'examiner aussi l'intérêt et la possibilité d'une politique multi-sites et multi-établissements :

« *Notre UMR étant multi-sites et CNRS on ne peut pas limiter la réflexion pour l'UMR à Rennes 2.* »

La question est posée en particulier pour le Plan de Gestion de Données :

« *Pour une unité multisite et multitutelle comme la mienne, il serait plus pratique d'avoir un seul PGD, et non pas un sur chaque site.* »

On voit peut-être, à partir de ce cas précis, toute l'importance d'engager une réflexion non seulement au plan local, mais encore dans le cadre d'une concertation des différentes instances de recherche impliquées.

3.5.2 Les résultats des entretiens

Les besoins et les attentes qui ont été exprimés lors des entretiens reflètent la grande diversité des situations de recherche (volume de données, sensibilisation antérieure, travail collaboratif ou solitaire, etc.). Il est possible toutefois de dégager des thématiques qui touchent de manière transversale une assez large partie des chercheurs. Par commodité, nous les regrouperons en deux grands ensembles : l'un touchant davantage les besoins techniques, l'autre portant sur les services, sachant que, dans la pratique, ils sont fortement articulés.

3.5.2.1 Les besoins techniques

A/ La nécessité d'un archivage

L'un des besoins les plus couramment abordés durant ces échanges est la nécessité d'un archivage. Les enseignants-chercheurs constatent que les solutions jusqu'à présent sont réalisées avec les moyens du bord, en fonction des contraintes et des possibilités du moment : « *Sur l'archivage, aucune politique n'est mise en place […] c'est que du bricolage* ». Le volume de stockage est évoqué à plusieurs reprises : « *Il faudrait qu'on ait des espaces de stockage pour pouvoir sauver les données* ». Et si les besoins en la matière sont étroitement liés à la nature des données gérées, les possibilités actuelles ne semblent pas satisfaire toutes les situations : « *au labo c'est vrai qu'on atteint les limites […] s'il y avait quelque chose à Rennes 2, on serait content d'avoir ça* ».
Concernant le traitement des données, beaucoup observent une rigueur insuffisante au sein de leur laboratoire : « *je ne pense pas qu'on doit être très rigoureux chez nous* ». Parmi les actions susceptibles de les impliquer davantage, les chercheurs souhaitent qu'on mette à leur disposition des procédures très simplifiées, résumées avec humour par l'un d'entre eux : « *on a une solution clé en main pour vous, vous avez juste à cliquer sur ce bouton et hop, c'est fait !* ». Cela traduit plus profondément un manque symptomatique de temps dû à une « bureaucratisation » de leur quotidien qui les conduirait naturellement à confier cette tâche spécialisée à un tiers : « *franchement, je n'ai pas le temps ! J'ai appris beaucoup de choses, ça je ne veux pas l'apprendre. Je crois vraiment que l'on est maintenant dans une période où on a besoin de professionnels de l'information qui soient dédiés* ».

Simultanément, plusieurs expriment une frustration que les données, une fois la recherche effectuée, demeurent difficilement accessibles, voire inaccessibles si l'on souhaite de nouveau les reprendre. Ils ont conscience que cela empêche le partage et a fortiori une mise en valeur : « *ça me peine un peu parce qu'en fait, on n'est pas dans une logique de valorisation des travaux antérieurs* ». Selon eux, remédier à cette situation supposerait, dans un premier temps, un lieu de stockage pérenne et sécurisé des données : « *C'est important qu'on ait un lieu unique avec un rangement, dans lequel on sait qu'on peut trouver les choses* » et, dans un second temps, d'améliorer leur accessibilité notamment grâce à un travail sur les métadonnées, une « *mise en valeur, en mots clés, en description […] qu'il y ait une organisation* ».
Une solution institutionnelle serait naturellement privilégiée si elle offrait les mêmes services appréciés chez certains prestataires payants : « *avec les mêmes potentialités, avec la même souplesse d'utilisation, un accès depuis n'importe quel poste* ». La démarche doit être simplifiée, permettant au chercheur d'acquérir un degré d'autonomie satisfaisant : « *il faut aussi qu'il y ait une forme de souplesse, d'ergonomie […] il faut pouvoir avoir la main dessus* ». L'idée exprimée étant de ne pas dépendre complètement d'un soutien technique qui manque de continuité à leur goût, ce qui a tendance à bloquer les procédures lorsque le personnel n'est pas ou plus disponible. C'est pourquoi la possibilité de solliciter un personnel travaillant

précisément sur ces questions et s'inscrivant dans la durée remporte une large adhésion : « *on a besoin de professionnels de l'informatique qui soient dédiés […] de façon pérenne, qu'on ne change pas d'ingénieur tous les ans* ».
Plusieurs fois mentionnés, un cadre méthodologique, voire une politique d'archivage contribueraient de manière décisive à installer des pratiques normées et collectives : « *je pense que ça mériterait d'avoir une ligne, une ligne collective* ». À ce titre, le plan de gestion de données (PGD), une fois sa finalité précisée, pourrait en constituer une pièce centrale : « *ça va devenir un truc incontournable, on ne va pas pouvoir faire comme si ça n'existait pas, parce que on y est confrontés concrètement dans le cadre de projets.* ».

B/ Un espace de travail collaboratif sécurisé

Une autre demande forte est celle d'un espace de travail collaboratif : « *on a besoin d'outils collaboratifs pour arriver à travailler en projet avec des données multiples qui viennent s'agréger* ». Certains établissements sont pris en exemple pour les solutions satisfaisantes qu'ils proposent (ex. CNRS, INRIA), plus rarement des infrastructures comme Huma-Num : « *quand on bosse avec des gens du CNRS, on a une plateforme, c'est sécurisé […] on n'a pas une solution ici pour travailler ensemble* ». Mais le plus souvent, quand la pratique est mise en place, c'est au moyen d'outils propriétaires (type Google Drive, Dropbox, etc.), car ils offrent des solutions faciles et rapides à installer notamment pour des collaborations hors établissement : « *on le fait par solution de facilité, on a tous un compte Google Drive* ». L'absence d'outil institutionnel équivalent motive la plupart du temps ce recours : « *j'étais bien en peine pour trouver un espace de partage, un espace collaboratif qui nous convienne, c'est pour ça qu'on est parti chez Drive* ». Cependant, ils émettent de sérieuses réserves sur ce type de service et déplorent ainsi que l'offre devienne payante si l'on souhaite bénéficier de capacités de stockage étendues : « *si je veux augmenter ma taille de stockage, je suis obligé de payer de ma poche puisque l'université ne veut pas prendre en charge, ce que je peux comprendre […] je n'ai pas le choix, je suis obligé de collaborer avec des collègues en Suisses, ailleurs* ». De plus, la politique d'utilisation de ces outils en termes de sécurité est clairement une préoccupation exprimée par les chercheurs : « *je crois que c'est assez critiquable sur les normes de sécurité, c'est pas du tout sécurisé* » de même que se pose la question de la conservation pérenne des données car la « *viabilité à long terme* » de ces outils n'est pas garantie. Un service aux fonctionnalités équivalentes avec les garanties que pourrait offrir une structure institutionnelle serait favorablement accueilli mais à condition de bien mesurer son champ d'action : « *c'est difficile d'enfermer ce type de service à l'échelle de l'université […] si on ne peut pas partager ses données avec les collègues extérieurs, eh bien ça ne sert à rien* ».

3.5.2.2 Les besoins de services

A/ Une sensibilisation préalable

Conformément à la difficulté déjà observée de circonscrire la notion de données et de ses nombreuses implications, les répondants expriment le besoin initial d'avoir une vision plus claire du paysage numérique auxquelles elles appartiennent : « *le premier besoin […] c'est la clarification, il faudrait mieux pouvoir s'orienter dans toutes les plateformes qui existent, dans tous les types de programmes* ». Pour répondre à ce besoin d'information ressenti dans leur équipe, certains directeurs d'unités préconisent « *un gros travail de pédagogie à faire* » ou, à tout le moins, une « *sensibilisation auprès des collègues* ». Au cours des entretiens émergent des demandes répétées d'éclaircissement sur tels ou tels aspects de la gestion des données, par exemple, au sujet de l'archivage, une enseignante-chercheure observe : « *moi je me rends*

compte que je suis nulle en archivage, il me faudrait un cours de rangement », un autre constate par rapport à la notion méconnue de plan de gestion de données : « *il faudrait nous éclairer un peu* ».

Le sentiment largement partagé est celui de manquer d'information face à un paysage numérique en pleine mutation, notamment en ce qui concerne le cadre réglementaire : « *le fait est que, en tout cas à l'université on n'est pas tellement au courant de cela, loin de là, et des évolutions très rapides* ».

B/ Des besoins de formations et de conseils spécialisés

La thématique de la formation revient très régulièrement dans les entretiens. La problématique des données s'inscrivant elle-même dans un environnement dont les évolutions sont régulières, certains enseignants-chercheurs réclament : « *une formation continue en termes juridiques et techniques* ». Les contenus cités des formations portent généralement sur la méthodologie ou les outils : « *avant même l'exploitation des données, je pense que pour beaucoup d'entre nous en SHS, c'est la formation aux méthodes, aux outils* ». Il faut souligner la demande récurrente concernant le cadre juridique appliqué aux données de la recherche qui soulève des problématiques complexes et inédites. Les questionnements peuvent être liés à la collecte : « *sur des données communes, qui a la propriété des données ? Et là, je ne parle pas des résultats scientifiques, je parle des données brutes* ». L'absence de réponse pouvant aboutir à une situation de blocage très délicate : « *on s'est retrouvé bloqués pour les utiliser parce qu'on n'avait pas de réponse sur ce qu'on avait le droit de faire avec ces données* ». Cet exemple illustre la frontière parfois floue entre un besoin d'information pouvant être satisfait par une formation classique et le besoin de recourir à un véritable conseil juridique que pourrait délivrer un service dédié, l'un n'excluant pas l'autre. Il peut tout autant s'agir de questions liées à la diffusion des données de recherche : « *qu'est-ce qu'on a le droit de mettre en ligne ?* ». Enfin, ce soutien pourrait être sollicité pour établir un plan de gestion des données, document dont la constitution et l'application reste très largement méconnue, mais dont l'enjeu est souvent perçu : « *ça doit être une politique d'accompagnement de tous les laboratoires* ».

L'une des principales attentes des chercheurs concerne la possibilité de solliciter un personnel qualifié, un « référent », un « expert », qui soit en mesure de suggérer les « bonnes pratiques » liées à la gestion des données : « *si on peut être accompagné en interne, pas seulement sur un plan technique, aussi un accompagnement avec un personnel compétent sur ce dossier* ».

C/ Une collaboration accrue

À côté de cette définition plus traditionnelle de la formation, une demande s'exprime pour un lien beaucoup plus étroit entre le technicien et le chercheur, à l'image de ce qui existe dans certaines structures et dont les résultats sont jugés très positifs : « *quelqu'un qui soit vraiment partie prenante de la recherche et pas juste un simple technicien auquel on dit « tu fais ça, tu me rentres ça* ». Certes, l'expertise technique d'un ingénieur (notamment en humanités numériques) demeure fondamentale mais l'efficacité de son action sera fonction de son implication personnelle : « *on est très dépendant malgré tout de la dimension technique et donc aussi de la personnalité de l'ingénieur et de sa volonté de participer pleinement à l'élaboration de la recherche* ». Le profil recherché devrait donc posséder à la fois de solides compétences techniques et une bonne connaissance des objets de recherche qui peut s'acquérir au contact prolongé des chercheurs et de leurs problématiques : « *quelqu'un qui, même s'il n'est pas spécialiste [...], au bout d'un moment comprend le pourquoi, peut proposer du nouveau*

matériel, des nouvelles méthodes ». C'est l'une des plus-values attendues de ce genre de collaboration, s'appuyant sur les compétences élargies de l'ingénieur, les possibilités du chercheur s'en trouvent accrues : « *on a besoin de gens qui ont des compétences méthodologiques très particulières qui vont nous aider et nous pousser sur de nouvelles méthodes* ».

L'importance des données dans l'élaboration de la recherche, l'évolution récente de leur gestion, tels qu'ils sont mis en évidence dans les entretiens, réclament deux grands types de besoins selon les répondants. L'un plus traditionnel, sous la forme d'un soutien technique et informationnel mais qui demande à être renforcé et pérennisé. L'autre tient compte d'une évolution souhaitée des rapports entre les enseignants-chercheurs et le personnel de l'information scientifique et technique dans le sens d'une plus grande coopération : « *il faut vraiment une collaboration, une pluridisciplinarité importante* ».

3.5.3 Synthèse sur les besoins et attentes

Les résultats de l'enquête statistique et des entretiens qualitatifs laissent émerger une image un peu plus précise et nuancée des positionnements, des pratiques, mais aussi des attentes autour de la question des données de recherche. C'était l'un des enjeux de cette enquête que de s'appuyer sur les retours obtenus pour conduire une réflexion sur les services les plus adéquats qui pourraient être mis en œuvre en vue d'aider les chercheurs et de faciliter leur gestion des données de recherche.

3.5.3.1 La sécurisation et la préservation des données : un point d'entrée et un impératif à court terme

Il ressort des réponses au questionnaire comme des échanges lors des entretiens que si les différentes questions se posant à propos de la gestion des données de recherche ne sont pas toutes appréhendées, ni peut-être clairement, ni forcément de la même manière, en revanche l'enjeu d'un stockage sécurisé des données accompagnant la recherche est parfaitement bien saisi et fait écho, aussi bien dans les réponses au questionnaire que dans les entretiens, à l'inquiétude devant le risque de perte de données. Pour autant les pratiques, hétérogènes, souvent artisanales et parfois imprudentes, en matière de stockage et d'archivage ne viennent pas appuyer cette conscience, comme on a pu le voir plus haut (partie 3.3). Et c'est peut-être par là qu'il faudrait commencer lorsque l'on s'interroge sur les besoins des chercheurs. Car outre que c'est un sujet qui parle directement au chercheur, rendu parfois plus présent par des expériences malheureuses (voir les entretiens), c'est aussi un problème qui est rencontré chaque jour et se pose de manière très concrète. Et de fait, dans les besoins exprimés, la demande la plus forte et la plus consensuelle va à la mise à disposition d'un espace d'archivage sécurisé.
Elle est d'ailleurs renforcée par le souhait très majoritaire de pouvoir stocker ses données sur le long terme, au-delà de la durée d'un projet de recherche, quelles que soient par ailleurs la nature, l'homogénéité et la volumétrie des données à stocker. Toutes choses qui devraient faire l'objet d'une analyse approfondie, avec une prise en compte de la complexité et de la diversité des configurations, avant de pouvoir envisager un dispositif universitaire allant dans ce sens. Quoi qu'il en soit, l'accessibilité des données sur le long terme et leur organisation au sein d'une structure institutionnelle adaptée est une nécessité qui a été mise en avant par bien des chercheurs, dans le questionnaire comme dans les entretiens.

Aussi, les perspectives directement liées à cette question pourraient prendre trois formes principales : une action de l'ordre de la sensibilisation, visant à mettre en cohérence le souci de préserver ses données et les pratiques réelles, une action relevant plus concrètement de l'appui méthodologique, et la mise en place d'une structure universitaire permettant le dépôt de données dans un espace sécurisé et pérenne.

3.5.3.2 Une dimension supplémentaire : le collaboratif

Si le questionnaire n'y faisait pas directement allusion, la question finale en texte libre et les entretiens ont permis de dégager un souhait assez partagé allant plus loin que le seul stockage : c'est le stockage partagé des données, permettant par exemple aux membres d'un même projet de recherche ou d'un même laboratoire de travailler sur ces données au sein d'un espace collaboratif au moyen d'outils partagés. Les besoins en la matière conduisent généralement les chercheurs à utiliser des solutions propriétaires extérieures (Google, Dropbox…), mais ils le font plutôt en l'absence d'outils institutionnels appropriés et en s'interrogeant sur les conditions de sécurité des données. La question qui reste posée est celle de l'échelle : face à des projets qui mettent en jeu des structures et des établissements différents, l'université est-elle l'échelon à privilégier ? Faut-il concevoir différents types de dispositifs ?

3.5.3.3 Information et clarification : des préalables

Les questions posées sur les besoins et attentes, en brassant certaines notions comme le Plan de Gestion de données, les métadonnées ou l'open access, ont mis en évidence chez une partie des chercheurs interrogés une méconnaissance de certains concepts, outils, méthodes liés à la gestion des données. Cela est particulièrement vrai du PGD. Les entretiens, en revenant sur ces notions pour les expliciter à la demande de chercheurs, ont d'ailleurs eu une fonction de clarification par rapport au questionnaire, de sensibilisation aussi et finalement de conviction, même si la logique de l'enquête, loin de toute injonction, privilégiait l'observation et la neutralité. Il est devenu en tous cas manifeste que ce travail d'explicitation et de clarification est à engager, en préalable, en parallèle ou dans le cadre de possibles actions de sensibilisation, selon des modalités qui peuvent être diverses.

3.5.3.4 Un appui au long cours

Le questionnaire sollicitait les chercheurs d'une part sur leur appui potentiel à une stratégie d'établissement en matière de données de recherche, notamment avec les questions sur la mise à disposition d'un plan de gestion de données, d'un standard de métadonnées, et d'un système de nommage de fichiers ; d'autre part, il proposait une gamme de services possibles pour accompagner le chercheur dans la gestion de ses données.
Le premier constat qui peut être fait c'est que les chercheurs, dans leur majorité, sont plutôt ouverts à cette idée de stratégie institutionnelle, même si pour certains le bon échelon par rapport à des projets de recherche intégrant différents établissements n'est pas nécessairement l'université ; par ailleurs, que la part des sans opinion est très importante, traduisant là sans doute le peu de proximité avec le processus de gestion des données, évoqué plus haut ; c'est patent sur les items évoquant la mise en place d'un PGD, et plus encore sur ceux évoquant la recommandation d'un ensemble de métadonnées ou d'un système de nommage de fichiers.

De la même manière, les services proposés recueillent un assentiment assez général ; certes, certains focalisent davantage l'attention et même si l'archivage sécurisé, des conseils sur des questions techniques et des conseils juridiques sont les trois besoins qui apparaissent comme les plus prioritaires, les pourcentages de réponses obtenus par les autres services potentiels

témoignent d'une attente globale en termes d'assistance à la gestion des données de recherche. Les entretiens viennent compléter cette impression, tout en insistant sur l'importance cruciale, outre de disposer d'un espace sécurisé pour ses données, d'être accompagné dans la durée par des conseils et une assistance, aussi bien technique que juridique, d'être formé aux méthodes et aux outils, aux différentes étapes de la gestion des données de recherche. Parallèlement à des modes classiques de formation, le besoin d'un « référent » qui accompagne le chercheur de manière ciblée et personnalisée, en interne, pour diffuser les bonnes pratiques, a été explicitement évoqué en entretien.

Il importera de prendre en compte les particularités disciplinaires, de même que les différences de contexte d'activité des laboratoires, apparues au fil des entretiens avec les chercheurs et s'avérant dans certains cas des facteurs discriminants au regard de l'action à engager. Mais c'est donc tout un dispositif, à la fois souple et pensé en amont, qu'il faudra concevoir avec des moyens adaptés, et à des niveaux qui peuvent être à géométrie variable, pour répondre aux attentes exprimées et développer des conditions satisfaisantes de gestion des données de recherche.

4. Synthèse des leçons de l'enquête

Quelles sont les principales leçons à tirer de cette double enquête ?

A/ Sur les données de recherche

- d'abord, la confirmation d'un constat général : **celui de la complexité et de la spécificité des données de recherche en général et dans les SHS en particulier**. Qu'il s'agisse de leurs définitions, difficiles et multiples, de leur distinction pas toujours évidente d'avec les publications, de la difficulté à séparer parfois données collectées et données produites, de l'entremêlement des phases du cycle des données, de leur immense variété... : les données de recherche en SHS ne se laissent pas aisément définir et saisir ;

- cette première complexité est redoublée d'une deuxième, que l'enquête qualitative a bien mis en lumière : **l'importance des aspects épistémologiques** (données *vs* obtenues) **et scientifiques** (la question de la contextualisation), dans plusieurs disciplines de sciences humaines et sociales ;

- la troisième leçon porte sur la **très grande diversité disciplinaire et des domaines de recherche, l'importance du niveau micro**, lié aux communautés de recherche et aux méthodes utilisées, et aussi la diversité des méthodes de recueil et de production des données.
  Autrement dit, les données de recherche en SHS ne sont pas faciles à définir et à cerner, leur statut épistémologique est sujet à débat et elles recouvrent une infinie diversité de réalités.

- une quatrième leçon émerge naturellement des trois précédentes : **l'impossibilité d'une politique générale, unique et identique pour tous**. Une politique des données de recherche se devra d'abord de partir de cette complexité, de cette diversité, en tâchant d'en épouser les multiples variations.

B/ Sur le stockage et l'archivage des données

- la confirmation du **constat des pratiques « artisanales » des enseignants-chercheurs en matière de stockage et d'archivage des données** : confirmation, car ce constat est récurrent à travers toutes les enquêtes, et la nôtre n'apporte pas d'éléments fondamentalement nouveaux sur ce point. Mais elle donne à voir, par les entretiens, la réalité humaine, concrète, de ces pratiques ;

- **l'importance de la sécurisation des données** : conséquence du premier point, le manque de sécurité des données est bien montré dans l'enquête, aussi bien statistique que qualitative. C'est sans nul doute l'un des points les plus cruciaux mis en exergue : il y a réellement urgence à donner une véritable sécurité au stockage et à l'archivage de nombreux jeux de données en SHS. Cette sécurisation des données doit être envisagée sur deux plans : celui de la sécurisation technique, matérielle, du stockage, et celui de la sécurisation des accès, notamment pour les données sensibles (en médecine, etc.) ;

- **la diversité des niveaux d'échelle, des pratiques, des besoins** : produit direct de la diversité des données elles-mêmes, évoquée ci-dessus, la diversité des situations et des

besoins, en matière de stockage et d'archivage, est également une grande leçon de notre enquête, et constituera sans doute l'un des principaux défis à relever pour une politique des données de recherche. Il faudrait adapter différentes solutions de stockage et d'archivage, à la fois selon les types de données, les volumes, le nombre d'acteurs concernés, la situation institutionnelle des équipes de recherche (mono-site, multi-sites, etc.) ;

- ces trois premières leçons peuvent se résumer en une quatrième : au-delà de la diversité des situations, des problèmes et des solutions, **la question du stockage, de l'archivage, de la description et de la sécurisation des données est la plus transversale** posée par les données de recherche, car elle concerne absolument toutes les disciplines et tous les acteurs, qu'il s'agisse d'un enseignant-chercheur seul, faisant sa recherche individuellement, ou bien de grandes équipes de recherche mobilisées sur des projets internationaux, qu'il s'agisse aussi de la recherche en train de se faire ou de projets à leur terme ;

- **la situation du stockage des données de recherche est perçue comme très insatisfaisante** par les chercheurs, et deux besoins s'expriment assez fortement : d'une part, des besoins d'espaces de stockage plus importants, mieux sécurisés, plus souples et surtout mieux adaptés aux besoins, d'autre part, des besoins de stockage collaboratif, des possibilités de partage interne des données ;

- enfin une sixième leçon ressort également de l'enquête : **la distinction nécessaire à établir entre stockage des « données chaudes » et archivage des « données froides »[62],** en distinguant également les niveaux de solution ; et en renforçant aussi sur ce point l'accompagnement des chercheurs, notamment dans la connaissance des solutions d'archivage pérenne sur les entrepôts de données.

C/ Sur le partage et la diffusion des données de recherche en libre accès

- la première leçon porte sur **l'écart entre les déclarations, les représentations et les pratiques** : si des différences ou des divergences ont été également mises à jour concernant le stockage et l'archivage des données, c'est sur la question du partage qu'apparaissent les écarts les plus grands, voire les contradictions les plus flagrantes. Sur la question du partage des données, beaucoup d'enseignants-chercheurs sont dans l'incertitude, l'interrogation, voire un certain flou, tous états propices aux contradictions entre le dire et le faire, entre les proclamations et les pratiques. Et cette situation d'interrogation, parfois de trouble, est assez explicable, compte tenu de la complexité de la question ;

- une deuxième a trait à ces nombreuses interrogations, aux inquiétudes devant le partage des données, à la peur du plagiat..., en bref **au poids des « facteurs internes » dans les réticences des enseignants-chercheurs à diffuser leurs données** en libre accès. Et ces inquiétudes, pour excessives qu'elles puissent parfois paraître aux yeux des spécialistes de l'IST, doivent être entendues ;

- **l'importance du facteur disciplinaire**, là encore, **dans la problématique du partage**. Outre la grande différence entre sciences sociales et arts-lettres-langues, de fortes nuances, sinon d'évidentes disparités, séparent les enseignants-chercheurs de psychologie

[62] Par données « chaudes », nous entendons les données en cours de traitement, et par données « froides » les données finalisées et archivées.

expérimentale, ou de biomécanique et ceux d'histoire, d'esthétique ou de sociolinguistique. La diversité des données, attachées aux champs disciplinaires, va fortement conditionner les possibilités et les pratiques de partage ;

- **l'hétérogénéité des pratiques collaboratives et de partage**, depuis l'individualisme de la recherche dans plusieurs disciplines, ou chez plusieurs chercheurs, jusqu'au travail collaboratif quotidien. La propension au partage des données est étroitement corrélée à ces pratiques collaboratives, plus ou moins développées ;

- la cinquième leçon redouble les deux premières sur les données de recherche et porte sur la **complexité objective de la question du partage des données en SHS.** L'enquête n'a rien révélé de vraiment nouveau sur ce point, mais elle permet de mettre l'accent sur cette complexité réelle, de l'incarner à travers des exemples et des témoignages. Questions juridiques liées aux données (avec la protection des données personnelles, la difficulté de l'anonymat, la reconnaissance de la propriété intellectuelle ou de la paternité des données, etc.), ou questions scientifiques (liées à la nécessaire contextualisation dans plusieurs disciplines) pèsent lourdement, voire entravent les possibilités d'une diffusion en libre accès ;

- **le poids des « écosystèmes » et des pratiques de recherche des chercheurs dans plusieurs disciplines** : place prépondérante des revues traditionnelles, individualisme, concurrence, difficultés à se constituer un terrain de recherche personnel, faible culture du Libre Accès, et parfois du numérique... ; de nombreuses raisons, autres que purement scientifiques, juridiques ou « internes », viennent également expliquer certaines réticences à diffuser les données de recherche ;

- la septième et dernière leçon, étroitement corrélée aux précédentes, porte **sur les discours de sensibilisation à tenir à l'égard de la communauté des enseignants-chercheurs** et concerne notamment les professionnels de l'IST, les personnels de soutien à la recherche et les décideurs : non seulement il faut **rejeter tout discours d'injonction**, de pression moralisatice ou culpabilisante sur la question du partage des données, et opter résolument pour la sensibilisation progressive, mais il faut de plus adapter les discours de sensibilisation à la complexité du paysage des données, tenir compte à la fois de la valeur affective attachée aux données de recherche par leurs producteurs, des nombreuses craintes et interrogations manifestées (sur le plagiat, etc.), des réticences, souvent légitimes, à un partage trop systématique, et enfin il conviendrait de cibler les priorités sur les besoins les plus urgents.

D/ Sur les besoins

Deux leçons peuvent être tirées sur les besoins :

- **la diversité des besoins exprimés, mais aussi non exprimés** : plusieurs besoins ont été particulièrement bien formulés dans les entretiens (sur la nécessité de conseils juridiques, sur la formation, sur des aides techniques, etc.), de même que dans l'enquête statistique. Mais l'enquête laisse aussi entrevoir en creux de nombreux besoins non exprimés : en matière de culture de l'IST, notamment des métadonnées ou du Libre Accès, parfois de culture numérique, les réponses révèlent souvent une certaine méconnaissance des nombreuses dimensions des données de recherche ; et, partant, des besoins d'information ou de formation d'autant plus forts ;

- **l'importance des besoins sur les aspects techniques et juridiques** : ces deux types de besoins apparaissent prioritaires pour les enseignants-chercheurs ; besoins en équipements, en espaces de stockage, en outils collaboratifs, en solutions d'archivage, mais besoin aussi de l'appui de personnes compétentes et disponibles ; besoins de conseils juridiques, pour mieux s'orienter dans le maquis du droit des données.

E/ Leçons méthodologiques

Enfin sur les aspects méthodologiques de l'enquête, quelles leçons pouvons-nous tirer de cette riche expérience ? Nous nous situons ici à une sorte de niveau « méta », qui intéressera surtout les professionnels et les chercheurs de l'IST. Nous avons identifié au moins trois leçons méthodologiques :

- d'abord, **l'importance des tests du questionnaire** : avant de lancer l'enquête statistique, il nous a paru essentiel de pouvoir tester le questionnaire auprès d'un petit panel d'enseignants-chercheurs ; et ces tests ont permis d'améliorer le questionnaire et d'éprouver sa pertinence ;

- **le nécessaire questionnement sur ses propres présupposés** : aussi bien dans l'orientation et la formulation des questions, dans le choix des thèmes, il est très important de se questionner, voire de remettre en question ses propres présupposés : est-ce que telle question sera bien comprise des collègues (nous en avons eu la preuve contraire par exemple sur l'item des « valeurs du Libre accès », que nous n'avions pas définies), est-ce que nos catégories « parlent » aux enseignants-chercheurs, etc ? On retrouve ici à la fois la question générale, et sans doute classique dans les enquêtes sociologiques, de la distance entre les représentations des questionneurs et des questionnés, et la question plus spécifique de l'écart entre le petit monde de l'IST et celui des enseignants-chercheurs. Peut-être faudrait-il partir davantage des perceptions spontanées du public visé, des formulations courantes ? À tout le moins, il convient d'expliciter davantage les questions trop spécifiques, trop techniques ;

- **la richesse du couplage quantitatif / qualitatif** : si notre enquête n'avait été que quantitative, elle aurait certainement perdu une grande part de son intérêt. Les entretiens nous ont apporté une vision beaucoup plus fine, des nuances impossibles à faire émerger dans un questionnaire statistique. Certes, une enquête qualitative représente une somme de travail non négligeable (que nous avions d'ailleurs sous-estimée), elle n'est pas toujours possible à réaliser et demande des conditions particulières. Mais l'observation des pratiques informationnelles demande de descendre à un niveau de granularité beaucoup plus fin que de simples questionnaires statistiques, forcément réducteurs.

Conclusion : recommandations et propositions

Cette enquête n'avait pas seulement une visée de connaissance générale ; parmi nos objectifs de départ, figurait également une dimension « politique », de recommandations et de propositions.

Des recommandations pour qui ?

Au final, quelles sont les principales recommandations possibles, qui peuvent être formulées à l'issue de cette enquête ? Trois ensembles de recommandations nous paraissent utiles, différenciées selon leurs destinataires :

A/ Pour les responsables politiques de la recherche

Une politique des données de recherche à Rennes 2 pourrait être articulée autour de quatre axes complémentaires :
- la sensibilisation : l'objectif serait de développer, sous toutes ses formes, la sensibilisation de la communauté de recherche de Rennes 2 (chercheurs, enseignants-chercheurs, ingénieurs de recherche et d'étude, personnels des bibliothèques, personnels de la DRV, différents services...) autour de la question des données de recherche. Et notamment la sensibilisation (et non l'injonction) au partage des données et au Libre Accès, car il s'agirait de convaincre les enseignants-chercheurs encore réticents des apports du libre accès aux résultats de la recherche ;
- le conseil et la formation juridique : aucune pratique de diffusion et de partage des données ne saurait se développer sans un minimum de garanties juridiques ;
- le développement de solutions de stockage et de partage interne des données : la question du stockage et de la sécurisation des données, mais aussi de leur partage collaboratif, nécessite d'être réellement prise en compte au niveau de l'établissement ;
- la facilitation du dépôt des jeux de données : pour diffuser et partager ses données, encore faut-il bien connaître les supports techniques, les infrastructures de dépôt et éventuellement de partage. Une politique active d'information systématique de la communauté et des équipes de recherche de Rennes 2 sur les entrepôts de données en SHS pourrait être menée, notamment sur la TGIR Huma-Num et sur Nakala.

B/ Pour les responsables des équipes de recherche

Si l'équipe de recherche n'est pas obligatoirement et dans tous les cas le seul niveau pertinent de décision pour une politique en matière de données de recherche, elle joue un rôle essentiel et irremplaçable, sur plusieurs points :
- relayer au sein du laboratoire les actions de sensibilisation des chercheurs, mises en place au niveau de l'établissement ;
- faire de la question des données de recherche un thème de débat au sein du laboratoire ;
- mettre en place des actions de sensibilisation et de formation systématique des doctorants et des nouveaux chercheurs, sur la gestion des données de recherche ;
- étudier la pertinence, l'intérêt et l'utilité d'élaborer un modèle de Plan de Gestion de Données propre au laboratoire ;
- encourager et aider les chercheurs dans l'élaboration de PGD, pour les projets de recherche collectifs importants (ANR, européens, régionaux…) ;

- étudier, en lien avec la DSI, la pertinence, l'intérêt et la faisabilité d'un espace collaboratif de stockage et de partage des données de recherche entre membres du laboratoire ou d'une équipe de recherche ;
- favoriser la culture collaborative chez les chercheurs, en développant l'usage d'outils collaboratifs de partage.

C/ Pour les acteurs du Libre accès et du partage des données, les chercheurs en IST

- développer des approches différenciées selon les situations, les disciplines et sous-disciplines ;
- tenir compte des réticences des enseignants-chercheurs sur le partage des données ;
- donner la priorité à la sensibilisation aux garanties juridiques sur la propriété intellectuelle des données, le plagiat, etc. ;
- approfondir la réflexion sur les aspects méthodologiques des enquêtes sur les pratiques informationnelles des chercheurs.

Six propositions pour l'Université Rennes 2

Nous soumettons à la réflexion et à la discussion de la direction, des différents acteurs de la recherche et de l'ensemble de la communauté de recherche de l'Université Rennes 2 les six propositions suivantes, présentées de manière distincte en Annexes.

A/ Un groupe de travail pour l'élaboration d'une politique d'établissement des données de recherche

Les données de recherche sont généralement au cœur du travail de recherche des chercheurs. À ce titre, elles relèvent d'un régime d'une extraordinaire hétérogénéité : de formes, de supports, de types, d'approches méthodologiques, de procédures de collecte, d'outils et de méthodes de traitement, d'aspects juridiques, de conditions de diffusion, etc. En conséquence, vouloir définir une seule politique, une seule approche en matière de gestion et de partage des données de recherche, qui s'appliquerait à toutes les disciplines, toutes les équipes et tous les chercheurs, serait illusoire et constituerait même une erreur stratégique pour un projet voué à l'échec. Pour autant, il est tout à fait possible d'élaborer une politique d'établissement globale et cohérente. La réalité des données de recherche comme les résultats de notre enquête nous conduisent donc à proposer une politique à géométrie variable, respectueuse de cette diversité foncière des données de recherche, et fondée sur quelques principes généraux communs et partagés, sur la mobilisation de multiples acteurs et sur l'articulation fine de différents niveaux de décision et d'action. Pour élaborer cette politique, il s'agirait, dans un premier temps, de procéder, si possible de manière exhaustive, à deux types de recensements :

- d'une part, le recensement complet de tous les acteurs impliqués dans cette question des données de recherche, en soulignant les responsabilités de chaque acteur (qu'il s'agisse des enseignants-chercheurs, des équipes de recherche, des services de l'université, des types de personnels, etc.) ;
- d'autre part, essayer de déterminer finement ce qui relève de chaque niveau de décision et d'action : depuis le niveau individuel de l'enseignant-chercheur jusqu'au niveau national, en passant par celui des unités de recherche, des Cellules Recherche, de l'établissement, des différents établissements de tutelle, etc.

Pour ce faire, nous proposons la mise en place d'un groupe de travail à l'Université Rennes 2, qui serait chargé de la réflexion sur une politique des données de recherche. Ce groupe de travail

devrait être fortement soutenu par la direction de l'université, et composé de plusieurs types de profils et d'acteurs volontaires (enseignants-chercheurs, bibliothécaires, ingénieurs de recherche, informaticiens, responsables d'unités de recherche, organismes de formation, archivistes…). Son objet serait d'échanger sur les besoins, d'alimenter la réflexion et de faire des propositions de politiques de données de recherche, en définissant les bons niveaux de décision et d'action possibles.

B/ Sensibiliser, accompagner, former sur les données de recherche

Cette enquête approfondie sur les représentations et les pratiques des enseignants-chercheurs a permis de faire apparaître de nombreux besoins, explicites ou implicites : les données de recherche constituent souvent un « impensé » de la recherche en SHS, comme l'ont montré plusieurs entretiens ; leur définition même, leur statut, sont loin d'être clairement appréhendés, les enjeux liés à la description des jeux de données ne sont pas clairement perçus, ceux qui portent sur le stockage, la sécurisation et l'archivage sont certes mieux identifiés mais il demeure une certaine méconnaissance des solutions possibles, les mutations du contexte juridique restent bien souvent méconnues, etc. En bref, les besoins d'accompagnement, de conseils et de formation sont, à Rennes 2 comme ailleurs, particulièrement importants, pour faire des données de recherche une thématique parfaitement maîtrisée par le plus grand nombre de chercheurs et de personnels.
Notre deuxième proposition porte donc sur le développement et la mutualisation d'un ensemble de services pour la sensibilisation, l'accompagnement et la formation des chercheurs, enseignants-chercheurs, doctorants et personnels de recherche : l'élaboration de documents d'accompagnement des chercheurs (ou l'intégration dans le Vade-Mecum du chercheur d'une partie sur les données de recherche), l'information sur les ressources disponibles, la diffusion d'une veille sélective sur les données de recherche, l'organisation de formations régulières, etc. Le Groupe de travail, qui serait créé sur les données de recherche, aurait à se pencher sur la mise en place d'une véritable politique de sensibilisation et de formation.

C/ L'importance des aspects juridiques

Les données de recherche posent de redoutables questions juridiques, parfois d'une grande complexité, à la fois au niveau de leur typologie (quel droit spécifique à chaque type de donnée), de leur collecte (par exemple pour la collecte de données personnelles, les entretiens, etc.), de leur traitement (droit de la « fouille de texte »), de leur archivage (quelle durée de conservation, quelle accessibilité...), et plus encore de leur diffusion en libre accès et de leur réutilisation (sous quel régime de licence). Par ailleurs, le changement récent de contexte juridique, avec la Loi Lemaire du 7 octobre 2016 pour une République numérique et notamment son article 30, a changé profondément la donne, en faisant de l'ouverture des données publiques le principe général. En effet, désormais, les établissements d'enseignement et de recherche sont soumis à l'obligation de rendre accessibles et librement réutilisables leurs données. Autrement dit, l'ouverture des données devient la règle et leur fermeture l'exception, ce qui va transformer progressivement les pratiques habituelles. De même, l'entrée en application du règlement européen sur la protection des données personnelles, prévue en mai 2018, aura également des effets sur le cadre juridique de la collecte et de la diffusion des données à caractère personnel, utilisées dans la recherche.
Or, notre enquête montre que les chercheurs et enseignants-chercheurs ne connaissent pas bien les règles de la propriété intellectuelle et sont peu préparés devant cette complexité du droit des données de recherche. Les besoins de conseil, de formation, d'aide technique sont donc très importants sur ces aspects juridiques, même s'ils ne sont pas toujours clairement formulés.

Ce constat nous conduit à faire une troisième proposition : la création d'un service juridique, spécialisé sur les questions de propriété intellectuelle, de libre accès des publications et des données de recherche, qui serait à la disposition de toutes les équipes et de tous les personnels de recherche et de soutien à la recherche. Il conviendrait évidemment d'étudier les conditions de mise en place et de fonctionnement d'un tel service, notamment son périmètre le plus pertinent : soit au niveau de l'Université Rennes 2, soit au niveau de Rennes, en partenariat avec les autres établissements du site rennais. Nous sommes convaincus qu'un tel service juridique rendrait de grands services à tous les acteurs de la recherche.

D/ Améliorer les solutions internes de stockage et de partage des données de recherche, et sensibiliser aux solutions d'archivage

L'enquête a montré à la fois l'importance et la diversité des besoins des enseignants-chercheurs en matière de stockage et de partage interne des données. Ces besoins sont plus ou moins aigus selon les disciplines, les unités de recherche, voire les individus eux-mêmes, et ils concernent les aspects suivants :

- l'adaptation des espaces personnels de stockage, offerts par l'université, aux besoins individuels des chercheurs : certains chercheurs ont des besoins de stockage plus importants que d'autres, et un recensement plus fin pourrait être entrepris sur ce point ;
- la mise à disposition, pour les équipes de recherche qui en feraient la demande, d'espaces collaboratifs de stockage et de partage, sécurisés et capables d'offrir aux chercheurs une alternative crédible aux solutions du secteur privé (Google Drive, DropBox), souvent utilisées par les chercheurs ;
- l'archivage pérenne des jeux de données, soit en interne, soit sur des plateformes externes.

Le Groupe de travail, qui pourrait être créé, aurait notamment en charge le recensement précis des besoins des équipes de recherche, en matière de stockage et de partage interne des données de recherche. Mais d'ores et déjà, il est possible et souhaitable de proposer une politique de stockage et de partage des données, adaptée aux besoins spécifiques de chaque unité de recherche, qui aurait besoin d'une solution locale. Sur l'archivage pérenne, les solutions existent déjà et il s'agit surtout de renforcer l'information et la sensibilisation des enseignants-chercheurs.
D'où nos deux propositions suivantes :

- étudier la faisabilité technique d'une mise à disposition de chaque unité de recherche de Rennes 2, qui en ferait la demande, d'un espace collaboratif protégé, pour le stockage et le partage des données au sein de chaque unité.
- sensibiliser et informer systématiquement l'ensemble de la communauté de recherche de Rennes 2 sur les solutions d'archivage des jeux de données : le dispositif Huma-NumBox, proposé par la MSHB, ou les infrastructures nationales, comme la plateforme Nakala, proposée par Huma-Num.

E/ Sur la description des jeux de données

Désormais obligatoire dans les projets de recherche européens (Horizon 2020), encouragé fortement dans les projets ANR (et peut-être à terme également obligatoire), réclamé dans certains projets internationaux, ou par quelques revues dans certaines disciplines, le Plan de Gestion de Données est un instrument appelé à prendre de plus en plus d'importance dans les

années à venir, pour un nombre croissant de chercheurs. Or notre enquête a clairement montré l'ignorance d'une grande majorité de collègues de ce type de document, permettant de décrire l'intégralité des jeux de données produits dans une recherche.
Aussi faisons-nous une sixième proposition :

- sensibiliser les enseignants-chercheurs et les chercheurs à l'importance de la description des données, encourager et aider les unités de recherche de Rennes 2 à l'élaboration de Plans de Gestion de Données, pour les projets de recherche collectifs importants (ANR, européens, régionaux…).

Cette proposition pourrait se traduire par la mise en place d'actions de sensibilisation des chercheurs, directement dans les laboratoires, par une formation systématique au Plan de Gestion des Données des personnels de soutien à la recherche (ingénieurs de recherche, Cellules Recherche des UFR, bibliothécaires spécialisés dans le soutien aux chercheurs…), qui pourraient ensuite accompagner les équipes souhaitant élaborer un PGD pour un projet de recherche.

Dans son Projet d'établissement 2017-2021, l'Université Rennes 2 s'engage, à plusieurs reprises dans le projet, dans une politique en faveur du libre accès aux publications et aux résultats de la recherche : « *une politique globale sera donc menée pour mettre à la disposition de la communauté de Rennes 2 des outils et des compétences au service de la diffusion des connaissances, et pour accroître la « visibilité » des publications, non seulement celles de ses enseignant.e.s-chercheur.e.s et de ses laboratoires, mais aussi celles des doctorant.e.s et des étudiant.e.s de master rédigeant des mémoires*. »[63]. De même, dans les actions menées en faveur des Humanités Numériques, les données de recherche sont également citées comme un axe de développement : « *Afin de favoriser les applications concrètes de ces recherches aux besoins de la pratique scientifique quotidienne (production, conservation, diffusion des données, visibilité et traçabilité des sources et des auteurs des données, accès aux données, métadonnées, etc.), il convient d'envisager de multiples coordinations au sein de l'établissement et avec les partenaires : la MSHB et ses plateformes, les TGiR Huma-Num et Progedo, la DSI, le SCD, l'URFIST, les PUR, le futur Institut de l'Innovation et de la Société Numérique, L'aire d'u, la plateforme Loustic, Immermove, etc. Les objectifs de ces coordinations sont multiples : identifier des sources, corpus et fonds à numériser, sauvegarder des sources et fonds nativement numériques, envisager les moyens nécessaires à la numérisation et à la conservation, préciser les conditions d'accès et d'exploitation selon la sensibilité des données, identifier et valoriser des plateformes ou des dispositifs déjà existants et les inscrire dans le réseau de plateformes nationales*… »[64].
Gageons que cette enquête, qui était clairement annoncée dans le projet d'établissement et signalée dans la Fiche Action 72, permettra d'aller encore plus loin en jetant les bases d'une politique d'établissement, à géométrie variable, en matière de données de recherche.

[63] Université Rennes 2. Projet d'établissement 2017-2021. Août 2017. p. 45
[64] *Ibid*.

Table des annexes

Table des figures

Les six propositions

La liste des unités de recherche de Rennes 2

Le questionnaire de l'enquête statistique

Voir aussi les annexes séparées :

Annexe 1 : Résultats de l'enquête statistique

Annexe 2 : Résultats des croisements de l'enquête statistique

Annexe 3 : Extraits des entretiens

Table des figures

Figure 1 : Structuration du questionnaire 12
Figure 2 : Taux de participation à l'enquête des enseignants-chercheurs de Rennes 2 (N=496) 13
Figure 3 : Canevas des questions pour les différents entretiens 16
Figure 4 : Affichage des entretiens et de leurs thématiques sur Sonal 17
Figure 5 : Répartition des répondants par tranche d'âge (N=140) 22
Figure 6 : Taux de participation des répondants par statut 22
Figure 7 : Répartition des répondants par types de responsabilités exercées (N=139) 23
Figure 8 : Croisement des responsabilités et des tranches d'âge (N = 139) 24
Figure 9 : Nombre de responsabilités et tranche d'âge 24
Figure 10 : Nombre de répondants par unité de recherche (N=134) 25
Figure 11 : Répartition des répondants par statut et par unité de recherche 26
Figure 12 : Participation selon les disciplines CNU (N=138) 27
Figure 13 : Catégories des données sources (N = 127) 29
Figure 14 : Nombre de réponses par répondant sur les catégories de données sources 30
Figure 15 : Support matériel des données sources (N = 127) 32
Figure 16 : Support matériel des données sources par laboratoire (N = 127) 33
Figure 17 : Catégories de données produites (N = 127) 35
Figure 18 : Nombre de catégories de données produites indiquées par les répondants 36
Figure 19 : Nature des données produites (N = 125) 37
Figure 20 : Nombre de réponses par répondant sur la nature des données produites 38
Figure 21 : Différentes natures de données produites 38
Figure 22 : Espace de stockage requis par les chercheurs (N = 95) 48
Figure 23 : Espace de stockage requis par les chercheurs selon leur laboratoire (N = 95) 49
Figure 24 : Lieux de stockage (N = 122) 50
Figure 25 : Types de lieux de stockage par répondant (N = 122) 51
Figure 26 : Fréquence de sauvegarde des données (N = 121) 52
Figure 27 : Élaboration d'une politique d'archivage dans le laboratoire 53
Figure 28 : Perceptions des politiques d'archivage dans les laboratoires (N = 122) 54
Figure 29 : Personne responsable des données dans une politique d'archivage du laboratoire (N = 9) 54
Figure 30 : Élaboration d'une politique personnelle en cas d'absence de politique d'archivage du laboratoire (N = 64) 54
Figure 31 : Gestion et conservation des données produites par des chercheurs non-permanents du laboratoire (doctorants, chercheurs associés, partenaires…) (N = 54) 55
Figure 32 : Types de métadonnées attribuées par les chercheurs (N = 115) 55
Figure 33 : Méthodes d'attribution d'informations supplémentaires aux données (N = 55) ... 56
Figure 34 : Accessibilité des données des chercheurs (N = 116) 64
Figure 35 : Téléchargement des données d'autres chercheurs (N = 116) 65
Figure 36 : Avez-vous déjà stocké vos données en libre accès sur le web ? (N = 116) 66
Figure 37 : Facteurs d'incitation au libre accès de ses données de recherche (N = 107) 67
Figure 38 : Facteurs de frein au libre accès de ses données de recherche (N = 108) 68
Figure 39 : Tableau de synthèse sur les facteurs de frein au libre accès de ses données de recherche 70
Figure 40 : Position sur la diffusion en libre accès de ses données de recherche (N = 115) ... 71
Figure 41 : Sur le partage de ses données de recherche (N = 102) 71
Figure 42 : Sur l'enjeu déontologique dans le partage des données (N = 98) 72
Figure 43 : Sur la connaissance des enjeux de l'Open Access (N = 97) 72

Figure 44 : Support à privilégier pour un dépôt en libre accès (N = 96) 73
Figure 45 : Besoin de conseils sur la publication des données de recherche (N = 94) 74
Figure 46 : Synthèse de la variable de l'âge dans le partage des données 75
Figure 47 : Variable de l'âge dans les facteurs d'incitation au partage des données 77
Figure 48 : Variable de l'âge dans les facteurs de frein au partage des données 78
Figure 49 : Synthèse de la variable du statut dans le partage des données 80
Figure 50 : Synthèse de la variable du laboratoire dans le partage des données..... 82
Figure 51 : Synthèse de la variable de la section CNU dans le partage des données 83
Figure 52 : Synthèse de la variable de la nature des données collectées dans le partage des données..... 84
Figure 53 : Synthèse de la variable du support matériel des données collectées dans le partage des données 85
Figure 54 : Synthèse de la variable de la nature des données produites dans le partage des données..... 86
Figure 55 : Connaissance du Plan de Gestion des Données (N = 96)..... 107
Figure 56 : Connaissance du PGD selon l'âge..... 108
Figure 57 : Connaissance du PGD selon le statut 109
Figure 58 : Connaissance du PGD selon les laboratoires..... 109
Figure 59 : Tableau de synthèse des positions personnelles et souhaits en matière de données de recherche..... 110
Figure 60 : Sur un stockage pérenne des données (N = 99)..... 111
Figure 61 : Sur un Plan de Gestion de données commun à l'université (N = 86)..... 112
Figure 62 : Sur une politique d'établissement pour les métadonnées (N = 82) 113
Figure 63 : Sur un système de nommage de fichiers (N = 86)..... 114
Figure 64 : Tableau des besoins par ordre décroissant de priorité 116
Figure 65 : Besoins par ordre décroissant de priorité..... 117
Figure 66 : Intégralité des souhaits exprimés (prioritaires et secondaires) 118
Figure 67 : Besoin d'un espace d'archivage des données (N = 94) 118
Figure 68 : Besoin de conseils techniques (N = 97)..... 119
Figure 69 : Besoin de conseils juridiques (N = 92)..... 119
Figure 70 : Croisement des besoins pour cinq laboratoires 122
Figure 71 : Tableau des besoins exprimés en réponse libre 123

Propositions pour l'Université Rennes 2

SIX PROPOSITIONS SUR LES DONNEES DE RECHERCHE
POUR L'UNIVERSITE RENNES 2
Novembre 2017

PROPOSITION N° 1	
▶ **Intitulé :**	**Créer un groupe de travail chargé de la réflexion sur une politique des données de recherche**
▶ **Contexte :**	- Émergence de la thématique des données au niveau européen et international - « Impensé » des données de recherche et méconnaissance actuelle des enjeux et des solutions - Nécessité d'élaborer une politique d'établissement - Complexité des niveaux de décision

DESCRIPTION	
▶ **Objectifs :**	- Faire émerger une politique d'établissement - Recenser tous les acteurs impliqués dans la question des données de recherche, en identifiant leurs responsabilités - Déterminer finement ce qui relève de chaque niveau de décision et d'action (depuis le niveau individuel jusqu'au niveau national, en passant par celui des unités de recherche, des cellules Recherche, de la DRV, des différents établissements de tutelle, etc.) - Améliorer la perception des enjeux liés aux données - Définir et faire connaître des solutions aux besoins identifiés (stockage, partage, formation …)
▶ **Activités proposées :**	- Identification des besoins - Définition des bons niveaux de décision et d'action possibles - Propositions de politiques de données de recherche et de priorités d'action - Élaboration d'une méthodologie…
▶ **Bénéficiaires :**	- Laboratoires et équipes de recherche en ALL et SHS - Ecoles Doctorales de l'UBL pour les disciplines concernées : ALL, ELIC, STT - Enseignants-chercheurs, chercheurs, doctorants, à titre individuel
▶ **Acteurs concernés :**	- Membres institutionnels : VP Recherche et Documentation, DRV, Cellules Recherche, un responsable d'unité de recherche, DSI, SCD, URFIST, MSHB ; - Membres volontaires (enseignants-chercheurs, bibliothécaires, ingénieurs de recherche, informaticiens, responsables d'unités de recherche, organismes de formation, archivistes…) -

PROPOSITION N° 2	
▶ **Intitulé :**	**Développer des services mutualisés pour la sensibilisation et la formation**
▶ **Contexte :**	- Manifestation de besoins implicites ou explicites avec les résultats de l'enquête - Besoins importants d'accompagnement et de formation sur les données de recherche - Coordination des services de soutien à la recherche à améliorer

DESCRIPTION	
▶ **Objectifs :**	- Mutualiser et coordonner différents services pour la sensibilisation, l'accompagnement et la formation des différents acteurs concernés par les données de recherche -
▶ **Activités proposées :**	- Élaborer des documents d'accompagnement (DRV) : par exemple, intégrer au Vade-Mecum du chercheur une partie sur les données de recherche - Mieux informer sur les ressources disponibles et diffuser une veille sélective (URFIST, MSHB, SCD) - Organiser des formations régulières sur les données de recherche : Recommander la formation systématique des personnels des Cellules Recherche, de la DRV, des directeurs d'unités... (DRV) Développer la formation des doctorants (SCD, URFIST) Organiser des formations techniques régulières sur les bonnes pratiques de nommage, sur les formats et les pratiques de stockage et d'archivage des jeux de données, sur les modalités de diffusion (URFIST, SCD, MSHB, DSI) Organiser des formations sur les enjeux et les évolutions de la thématique (URFIST, MSHB)
▶ **Bénéficiaires**	- Laboratoires et équipes de recherche en ALL et SHS - Enseignants-chercheurs, chercheurs, doctorants, à titre individuel - Personnels de la recherche
▶ **Acteurs concernés**	- *Groupe de travail « Données de recherche » (cf. Proposition n° 1)* - DRV - Cellules Recherche des UFR - Responsables des unités de recherche - Responsables des Ecoles Doctorales - Cellule juridique de Rennes 2 - Département Recherche du SCD - DSI - MSHB - URFIST de Rennes
▶ **Partenaires externes**	- SCD Rennes 1, Bibliothèque INSA - Ecoles Doctorales de l'UBL pour les disciplines concernées : ALL, ELIC, STT

PROPOSITION N° 3	
▶ **Intitulé :**	**Offrir un service juridique spécialisé**
▶ **Contexte :**	- Grande complexité juridique des données de recherche : dans leur typologie, leur collecte, leur diffusion… - Changement du contexte juridique avec la Loi Lemaire 2017 pour une République numérique, avec l'application future du Règlement européen sur la protection des données personnelles - Besoins importants des enseignants-chercheurs, des ingénieurs, des personnels de recherche… en matière juridique

DESCRIPTION	
▶ **Objectifs :**	- Offrir un service juridique spécialisé sur les questions de propriété intellectuelle, de libre accès des publications, et sur les données de recherche
▶ **Activités proposées :**	- Fournir des conseils, des accompagnements, des expertises… sur toutes les questions juridiques liées à la propriété intellectuelle - Faire et diffuser une veille juridique sur la propriété intellectuelle
▶ **Bénéficiaires :**	- Laboratoires et équipes de recherche en ALL et SHS - Enseignants-chercheurs, chercheurs, doctorants, à titre individuel - Personnels de la recherche (DRV, ingénieurs, cellules Recherche…) - Bibliothécaires spécialisés dans l'aide aux chercheurs
▶ **Acteurs concernés :**	- *Groupe de travail « Données de recherche » (cf Proposition n° 1)* - DRV - Cellule juridique de Rennes 2
▶ **Modalités :**	- Étudier le périmètre le plus pertinent d'un tel service : soit au niveau de l'Université Rennes 2 : soit au niveau de Rennes, en partenariat avec les autres établissements du site rennais. - Etudier les modalités les plus pertinentes : mutualisation de compétences existantes, création de poste...
▶ **Partenaires externes :**	- Services juridiques de Rennes 1 et des établissements de Rennes

PROPOSITION N° 4	
▶ **Intitulé :**	**Mieux adapter les espaces internes de stockage des données de recherche aux besoins individuels et collectifs des chercheurs**
▶ **Contexte :**	- Inadéquation des solutions actuelles de stockage et de partage des données à l'Université Rennes 2 - Pratiques individuelles peu sûres en matière de sauvegarde des données - Diversité des besoins individuels des enseignants-chercheurs en matière de stockage et de sécurisation - Importance et diversité des besoins des équipes de recherche pour stocker et partager les données, pendant les projets de recherche - Recherche d'alternatives crédibles aux solutions du secteur privé, souvent utilisées par les chercheurs -

DESCRIPTION	
▶ **Objectifs :**	- Adapter les espaces personnels de stockage, offerts par l'université, aux besoins individuels des chercheurs - Mieux répondre aux besoins de stockage, de sécurisation et de partage interne des données de travail des projets en cours, en mettant à disposition des équipes de recherche, qui en feraient la demande, des espaces collaboratifs de stockage et de partage, sécurisés -
▶ **Activités proposées :**	- Recenser finement les besoins, individuels et collectifs, de stockage des données dans les équipes de recherche - Etablir un cahier des charges pour le stockage et le partage internes des données - Etudier la faisabilité technique des solutions de stockage
▶ **Bénéficiaires :**	- Laboratoires et équipes de recherche en ALL et SHS - Enseignants-chercheurs, chercheurs, doctorants, à titre individuel
▶ **Acteurs concernés :**	- *Groupe de travail « Données de recherche » (cf Proposition n° 1)* - DSI - DRV - Responsables des équipes de recherche volontaires - Ingénieurs de recherche

PROPOSITION N° 5	
▶ **Intitulé :**	**Sensibiliser à la description des jeux de données et à la rédaction des Plans de Gestion de Données**
▶ **Contexte :**	- Importance croissante des PGD dans les projets de recherche : obligatoires dans les projets de recherche européens (Horizon 2020), encouragés dans les projets ANR, réclamés dans certains projets internationaux, ou par quelques revues dans certaines disciplines - Résultats de l'enquête montrant la méconnaissance du PGD par une grande majorité de collègues

DESCRIPTION	
▶ **Objectifs :**	- Sensibiliser les enseignants-chercheurs et les chercheurs de Rennes 2 à l'importance de la description des données - Encourager et aider les unités de recherche de Rennes 2 à l'élaboration de Plans de Gestion de Données, pour les projets de recherche collectifs importants (ANR, européens, régionaux…)
▶ **Activités proposées :**	- Actions de sensibilisation des enseignants-chercheurs et des chercheurs au PGD - Formation systématique des personnels de soutien à la recherche à la rédaction des PGD - Aide et soutien technique pour la rédaction de PGD
▶ **Bénéficiaires :**	- Laboratoires et équipes de recherche en ALL et SHS - Enseignants-chercheurs, chercheurs, doctorants, à titre individuel
▶ **Acteurs concernés :**	- *Groupe de travail « Données de recherche » (cf. Proposition n° 1)* - Responsables des équipes de recherche - DRV - Cellules Recherche - Ingénieurs de recherche - Département Recherche du SCD - MSHB - URFIST

PROPOSITION N° 6	
▶ **Intitulé :**	**Sensibiliser aux solutions institutionnelles de partage et d'archivage des jeux de données**
▶ **Contexte :**	- Importance des besoins d'archivage des données de recherche collectées et produites, après un projet de recherche - Existence d'infrastructures de recherche en SHS au niveau national, européen et international, encore peu connues des chercheurs - Enjeux de la diffusion des données en libre accès

DESCRIPTION	
▶ **Objectifs :**	- Sensibiliser et informer systématiquement toutes les équipes de recherche sur le dispositif Huma-Num Box, proposé par la MSHB - Développer l'archivage des données sur les infrastructures nationales, comme la plateforme Nakala, proposée par Huma-Num - Inciter au partage des données sur les infrastructures de recherche
▶ **Activités proposées :**	- Actions de sensibilisation des équipes de recherche - Formations sur les infrastructures de recherche en ALL-SHS
▶ **Bénéficiaires**	- Laboratoires et équipes de recherche en ALL et SHS - Enseignants-chercheurs, chercheurs, doctorants, à titre individuel
▶ **Acteurs concernés**	- *Groupe de travail « Données de recherche » (cf Proposition n° 1)* - DRV - Responsables des équipes de recherche - Ingénieurs de recherche - SCD Rennes 2 - MSHB - URFIST de Rennes
▶ **Partenaires externes**	- Infrastructure de Huma-Num

Liste des unités de recherche de l'Université Rennes 2

L'Université Rennes 2 comporte, en novembre 2017, 20 unités de recherche[65], dans son Contrat quinquennal 2017-20266. Au moment de l'enquête (janvier-mars 2017), elle n'en comptait que 19 ; l'unité de recherche PTAC (Pratique et Théorie de l'Art Contemporain) a été créée depuis, issue de APP, et ne figure donc pas dans les résultats. Nous donnons ici la liste de ces 19 unités, par ordre alphabétique, avec leur sigle figurant dans tous les tableaux et graphiques :

- **ACE** : Anglophonie : Communautés et Écritures
- **APP :** Arts : pratiques et poétiques (y compris PTAC : Pratique et Théorie de l'Art Contemporain)
- **CELLAM** : Centre d'Études des Langues et Littératures Anciennes et Modernes
- **COSTEL** : Climat et Occupation du Sol par Télédétection (composante Rennes 2 de LETG - CNRS UMR 6554)
- **CREAD** : Centre de Recherches sur l'Education, les Apprentissages et la Didactique
- **CRBC-Rennes** : Centre de Recherche Bretonne et Celtique (Rennes)
- **ERIMIT** : Équipe de Recherche Interlangues : Mémoire, Identité, Territoires
- **ESO-Rennes** : Espaces et Sociétés (composante Rennes 2 de ESO - CNRS UMR 6590)
- **HCA** : Histoire et Critique des Arts
- **IRMAR-Rennes 2** : Institut de Recherches Mathématiques de Rennes (composante Rennes 2 de l'IRMAR - CNRS UMR 6625)
- **LAHM** : Laboratoire Archéologie et Histoire Merlat (composante Rennes 2 du CREaHH - CNRS UMR 6566)
- **LIDILE** : Linguistique Ingénierie et Didactique des Langues
- **LiRIS** : Laboratoire Interdisciplinaire de Recherche en Innovations Sociétales (ex-CIAPHS)
- **LP3C** : Laboratoire de Psychologie : Cognition, Comportement, Communication (ex-CRPCC)
- **M2S** : Mouvement, Sport, Santé
- **PREFics** : Plurilinguismes, Représentations, Expressions Francophones, information, communication, sociolinguistique
- **Recherches en Psychopathologie** : nouveaux symptômes et lien social
- Tempora (ex-CERHIO)
- **VIPS²** : Violences, Identités, Politiques, Socialisations et Sports

[65] Voir sur https://www.univ-rennes2.fr/direction-recherche-valorisation/unites-recherche

Le questionnaire de l'enquête statistique

Enquête sur les pratiques des enseignants-chercheurs de l'université Rennes 2 en matière de données de recherche

Bonjour,

Merci de bien vouloir remplir ce questionnaire, qui vous demandera 10 à 15 minutes.

Cette enquête vise trois objectifs :
- établir un état des lieux le plus complet possible de la situation des données de recherche à Rennes 2 ;
- réfléchir à une offre de services, pouvant répondre aux besoins exprimés dans l'enquête, en termes d'accompagnement, de formation, de services, de ressources...
- alimenter la recherche et la réflexion sur les pratiques informationnelles des chercheurs et sur la place des données dans le travail de recherche.

Enquête organisée par l'URFIST de Rennes, la MSHB et le SCD Rennes 2, avec le soutien de l'Université Rennes 2.
There are 32 questions in this survey

Vous-même

1 Merci de cocher votre tranche d'âge *[67]
Veuillez sélectionner SEULEMENT UNE réponse
- < 30 ans
- 30-50 ans
- > 50 ans

2 Vous êtes ? *
Choisissez TOUTES les réponses qui conviennent :
- Chercheur-se
- Professeur-e
- HDR
- Maître-sse de conférences
- Post-doctorant-e
- Ingénieur-e

3 Exercez-vous actuellement des responsabilités ? *
Choisissez TOUTES les réponses qui conviennent :
- éducatives (direction de filière, de département, d'UFR...)
- scientifiques (direction ou co-direction de laboratoires, d'équipes de recherche, de projets ANR...)
- d'encadrement de thèses
- politiques (élu(e) à Rennes 2, à l'UBL...)
- éditoriales (direction de revue, membre de comité de rédaction...)
- administratives (responsabilité d'un service)
- aucune des responsabilités mentionnées

[67] Les questions suivies d'un astérisque étaient obligatoires

4 Quel est votre laboratoire ? *
Veuillez sélectionner SEULEMENT UNE réponse
- ACE
- Arts : pratiques et poétiques
- CELLAM
- CERHIO
- CIAPHS
- COSTEL
- CRBC
- CREAD
- CRPCC
- ERIMIT
- ESO
- Histoire et Critique des Arts
- IRMAR
- LAHM
- LIDILE
- M2S
- PREFics
- Recherches en Psychopathologie
- VIP&S
- Tempora

5 Quelle est votre section du CNU ? *
Choisissez TOUTES les réponses qui conviennent :
- 1. Droit privé et sciences criminelles
- 2. Droit public
- 4. Science politique
- 5. Sciences économiques
- 6. Sciences de gestion
- 7. Sciences du langage : linguistique et phonétique générales
- 8. Langues et littératures anciennes
- 9. Langue et littérature françaises
- 10. Littératures comparées
- 11. Langues et littératures anglaises et anglo-saxonnes
- 12. Langues et littératures germaniques et scandinaves
- 13. Langues et littératures slaves
- 14. Langues et littératures romanes : espagnol, italien, portugais, autres langues romanes
- 15. Langues et littératures arabes, chinoises, japonaises, hébraïque, d'autres domaines linguistiques
- 16. Psychologie, psychologie clinique, psychologie sociale
- 18. Architecture (ses théories et ses pratiques), arts appliqués, arts plastiques, arts du spectacle, épistémologie des enseignements artistiques, esthétique, musicologie, musique, sciences de l'art
- 19. Sociologie, démographie
- 20. Ethnologie, préhistoire, anthropologie biologique
- 21. Histoire, civilisations, archéologie et art des mondes anciens et médiévaux
- 22. Histoire et civilisations : histoire des mondes modernes, histoire du monde contemporain ; de l'art ; de la musique
- 23. Géographie physique, humaine, économique et régionale

24. Aménagement de l'espace, urbanisme
26. Mathématiques appliquées et applications des mathématiques
27. Informatique
70. Sciences de l'éducation
71. Sciences de l'information et de la communication
73. Cultures et langues régionales
74. Sciences et techniques des activités physiques et sportives
Autre:
Veuillez préciser, si besoin, votre discipline dans la dernière case.

Typologie des données (collectées et produites)

6 Dans quelles(s) catégorie(s) suivantes classeriez-vous vos "données sources" (i.e. vos données collectées) ?
Choisissez TOUTES les réponses qui conviennent :
des données d'observation (remontées de terrain, données de capteurs...)
des données expérimentales (en médecine, psychologie...)
des données textuelles (issues de corpus de textes, d'archives...)
des données audiovisuelles (enregistrements audio, séquences filmiques...)
des images fixes d'objets, de paysages, d'architectures, scans de textes
des données chiffrées (tableaux de données statistiques...)
des données d'enquêtes, de sondages (entretiens)
des données web (données d'usages, cartographies de sites, données de réseaux sociaux...)
des données 3D, de simulation
Autre :

7 Concernant leur support matériel, vos données sources (ou données collectées) sont-elles majoritairement : *
Veuillez sélectionner SEULEMENT UNE réponse
non numériques (par exemple : des sources archivistiques imprimées, des entretiens non enregistrés en numérique...)
numérisées (par exemple : des corpus numériques, des données statistiques...)
nativement numériques (par exemple : les données web)

8 Si vos données sources sont non numériques, est-ce que vous les numérisez en général (saisie d'entretiens, numérisations de textes...) ?
[Répondez seulement à cette question si vous avez répondu 'non numériques (par exemple : des sources archivistiques imprimées, des entretiens non enregistrés en numérique...) à la question 7]
Veuillez sélectionner SEULEMENT UNE réponse
Oui
Non

9 Dans quelle(s) grande(s) catégorie(s) classeriez-vous vos données produites, après exploitation, i.e. vos "données résultat" : *
Choisissez TOUTES les réponses qui conviennent :
des données statistiques (données chiffrées)
des données textuelles (extraits, champs textuels, extraits d'entretiens...)
des données graphiques (schémas, visualisations, cartographies, tableaux, scans de textes...)
des données audiovisuelles (film...)

☐ des données 3D, de simulation
Autre:
Plusieurs réponses possibles.

10 Pouvez-vous préciser davantage la nature des données produites à l'issue de vos recherches ?
Choisissez TOUTES les réponses qui conviennent :
☐ photos
☐ visualisations ou modèles multidimensionnels
☐ enregistrements audio
☐ enregistrements video
☐ textes
☐ tableaux
☐ bases de données
☐ programmes et applications
☐ Autres, données spécialisées ou spécifiques

11 Veuillez détailler :
[Répondez seulement à cette question si vous avez répondu 'Autres, données spécialisées ou spécifiques' à la question 10]
Choisissez TOUTES les réponses qui conviennent :
☐ analyses statistiques
☐ corpus de texte / annotations
☐ transcriptions d'entretiens
☐ données de géolocalisation (SIG)
☐ cartes et plans
☐ images satellite
☐ données de télédétection
☐ spectres
☐ séries de mesures
☐ modélisations du climat
☐ données de santé
☐ questionnaires
☐ cartographies de réseaux ou de corpus
Autre:

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

12 À combien estimez-vous l'espace nécessaire au stockage de vos données ?
Veuillez sélectionner SEULEMENT UNE réponse
☐ < 1 Go
☐ entre 1 et 20 Go
☐ entre 20 et 100 Go
☐ entre 100 Go et 1 To (1 tera-octet = 1000 Go ou 1000 milliards d'octets)
☐ >1 To
À titre indicatif : clé USB (de 1à 20Go), disque dur (autour de 1To), serveur ou cloud (au-delà du To)

13 Où stockez-vous les données de recherche ? *
Choisissez TOUTES les réponses qui conviennent :

☐ sur votre ordinateur personnel
☐ sur votre ordinateur professionnel
☐ sur un disque dur externe ou autre support
☐ sur un serveur de l'université
☐ sur une plate-forme privée dans le "cloud" (Google Drive, Dropbox, etc.)
☐ sur un serveur externe d'une autre institution
☐ sur le serveur d'une TGIR (Très Grande Infrastructure de Recherche)

14 À quelle fréquence faites-vous ces copies de sauvegarde de vos données ?
Veuillez sélectionner SEULEMENT UNE réponse
☐ tous les jours
☐ toutes les semaines
☐ tous les mois
☐ tous les trimestres
☐ tous les ans
☐ irrégulièrement, en fonction des besoins
☐ en fonction du prestataire de service ou du gestionnaire de site
☐ je ne fais pas de copie de sauvegarde

15 Y a-t-il une politique d'archivage des données de recherche dans votre laboratoire (critères, modalités, nommage des fichiers...): *
Veuillez sélectionner SEULEMENT UNE réponse
☐ Oui
☐ Non
☐ Je ne sais pas

16 Si oui, qui est responsable de l'archivage des données?
[Répondez seulement à cette question si vous avez répondu 'Oui' à la question 15] Choisissez TOUTES les réponses qui conviennent :
☐ un membre de votre équipe de recherche
☐ un membre d'un service de soutien à la recherche (bibliothèque, DSI...)
☐ un prestataire de service externe
☐ il n'y a pas de responsable identifié
Autre :

17 Si non, avez-vous mis au point vous-même une politique d'archivage de vos données ?
[Répondez seulement à cette question si vous avez répondu 'Non' à la question 15] Veuillez sélectionner SEULEMENT UNE réponse
☐ Oui
☐ Non

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

19 Attribuez-vous habituellement des informations supplémentaires (métadonnées) à vos données de recherches ?
Choisissez 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.)
☐ je n'attribue pas d'informations supplémentaires

20 Si vous attribuez des informations supplémentaires, comment procédez-vous ?
[Répondez seulement à cette question si vous avez répondu « des informations administratives (auteur, date de création, etc.) » ou « des informations techniques (format des fichiers, taille, logiciel, etc.) » ou « des informations scientifiques (origine des données, projet de recherche, mots-clés, etc.) » à la question 19]
Choisissez la réponse appropriée pour chaque élément :

	Oui	Non
vous élaborez une documentation qui vous est propre	☐	☐
vous utilisez un standard de métadonnées	☐	☐
vous utilisez un cahier de laboratoire (ou de recherche)	☐	☐

Pratiques de partage et de diffusion des données

21 En dehors de vous, qui peut accéder directement à vos données ?
Choisissez TOUTES les réponses qui conviennent :
☐ personne
☐ mon groupe de travail
☐ mon laboratoire, mon institut
☐ l'ensemble de l'université
☐ les chercheurs de la discipline
☐ tout le monde
☐ autre public spécifique

22 Avez-vous déjà téléchargé les données d'autres chercheurs ? *
Veuillez sélectionner SEULEMENT UNE réponse
☐ oui, au moins une fois
☐ oui, régulièrement
☐ non, car je ne savais pas que c'était possible
☐ non, et je ne le ferai pas
☐ non, pas pour le moment

23 Vous-même, avez-vous déjà stocké vos données en libre accès sur le web ? *
Veuillez sélectionner SEULEMENT UNE réponse
☐ oui, au moins une fois
☐ oui, régulièrement
☐ non, car je ne savais pas que c'était possible

☐ non, je n'y avais pas pensé
☐ non, et je ne le ferai pas
☐ non, car je ne sais pas le faire
☐ non, car je n'ai pas le temps
☐ non, car je pense que cette tâche nécessite le soutien d'un personnel qualifié

24 Seriez-vous d'accord pour diffuser vos données de recherche en libre accès (à condition qu'elles soient diffusables et protégées juridiquement) ? *
Veuillez sélectionner SEULEMENT UNE réponse
☐ oui, tout à fait
☐ oui, après un délai
☐ plutôt oui
☐ plutôt non
☐ non, pas du tout
☐ je ne sais pas

25 Parmi les raisons suivantes, qu'est-ce qui vous inciterait à rendre vos données de recherche accessibles en libre accès ?
Choisissez TOUTES les réponses qui conviennent :
☐ une obligation faite par le financeur de votre recherche (par ex. dans le cadre de Horizon 2020)
☐ la reconnaissance de la communauté scientifique
☐ une meilleure visibilité de vos travaux de recherche
☐ l'occasion de nouveaux contacts, de nouvelles coopérations scientifiques
☐ le souci de pouvoir faire valider vos résultats à partir des jeux de données
☐ le souhait de voir se développer d'autres recherches à partir de vos jeux de données
☐ l'adhésion aux valeurs du Libre Accès aux résultats de la recherche scientifique
☐ de possibles retombées financières (crédits de recherche...)

26 Quelles seraient les principales raisons qui vous empêcheraient ou vous freineraient pour rendre vos données de recherche accessibles en libre accès ?
Choisissez TOUTES les réponses qui conviennent :
☐ des freins juridiques liés à la nature de vos données (Exemple : données à caractère personnel, données sous droit (*text mining* de bases commerciales, données obtenues sous dérogation), données dites sensibles (liées à la Défense, à l'environnement, etc.)
☐ des raisons économiques (protection par rapport à la concurrence)
☐ des raisons scientifiques
☐ la crainte du plagiat
☐ des raisons institutionnelles (mon labo ne veut pas diffuser les données…)
☐ des raisons personnelles (je ne veux pas montrer mon "arrière-cuisine", mes données sont à moi)
☐ le manque de temps
☐ le manque de connaissances dans la description des jeux de données, les modalités de dépôt, etc.
☐ le manque de compétences
Autre :

27 Si vous souhaitez diffuser vos données en libre accès, quel est (ou serait) votre support principal ?
Choisissez TOUTES les réponses qui conviennent :

☐ une archive ouverte (de type HAL...)
☐ une plateforme spécialisée dans l'archivage et la diffusion des données de recherche :
☐ un réseau social académique (ResearchGate, Academia…)
☐ votre blog ou votre site personnel
☐ le site ou le blog du laboratoire
☐ un serveur de l'université
☐ un data paper (une publication décrivant un ou plusieurs jeux de données scientifiques)
Autre:

Vos besoins, vos attentes

28 Savez-vous ce qu'est un Plan de Gestion de Données (PGD), ou Data Management Plan (DMP) ? *
Veuillez sélectionner SEULEMENT UNE réponse
☐ Oui
☐ Non

29 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
Je suis familier(e) avec les enjeux de l'open access (accès ouvert)	☐	☐	☐	☐	☐
Je suis prêt à partager mes données de recherche avec les autres	☐	☐	☐	☐	☐
Je souhaiterais pouvoir stocker mes données de recherche au-delà de la durée de vie d'un projet	☐	☐	☐	☐	☐
Je souhaiterais que l'université dispose d'un plan de gestion de données	☐	☐	☐	☐	☐
Je souhaiterais que l'université dispose d'un ensemble de métadonnées prescrit pour verser les données vers un entrepôt	☐	☐	☐	☐	☐
Je souhaiterais que l'université utilise/recommande un système de nommage standard de fichiers	☐	☐	☐	☐	☐
Je pense qu'une déontologie des données pourrait devenir un enjeu lorsque les données de	☐	☐	☐	☐	☐

	Tout à fait d'accord	Plutôt d'accord	Sans opinion	Plutôt pas d'accord	Pas du tout d'accord
recherche sont partagées avec les autres					

30 Sur quels points souhaiteriez-vous bénéficier d'une aide (de votre unité de recherche, de la DSI, du SCD, de l'URFIST, de la MSHB...) ?
Choisissez la réponse appropriée pour chaque élément :

	En priorité	Eventuellement	Non
Des conseils relatifs à la publication et la citation des données de la recherche			
des conseils d'ordre général pour la gestion des données de la recherche			
des conseils sur des questions techniques (par exemple, métadonnées, normes, archivage à long terme)			
des conseils juridiques (par exemple, les restrictions d'accès,le traitement des données sensibles , l'utilisation de licences)			
une offre de formations sur la question des données de recherche (gestion, diffusion, aspects juridiques, etc.)			
une assistance technique (par exemple, pour la publication des données dans une revue)			
une aide pour préparer un plan de gestion de données			
un espace d'archivage sécurisé et fiable pour mes données de recherche			

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

32 Nous souhaiterions compléter cette enquête par quelques témoignages individuels. Accepteriez-vous d'être contacté pour un entretien ? Si oui, indiquez-nous votre adresse mail :

Envoyer votre questionnaire.
Merci d'avoir complété ce questionnaire.

TABLE DES MATIERES

REMERCIEMENTS **4**

1. INTRODUCTION **5**

2. METHODOLOGIE ET MODALITES DE L'ENQUETE **11**

2.1 L'ENQUETE STATISTIQUE 11
2.1.1 Le questionnaire *11*
2.1.2 L'échantillon *13*
2.1.3 Le nombre de réponses *13*
2.2 L'ENQUETE QUALITATIVE 14
2.2.1 Le canevas des entretiens *14*
2.2.2 La constitution du panel *16*
2.2.3 Le nombre et le déroulement des entretiens *16*
2.2.4 L'outil de transcription Sonal *16*
2.3 SYNTHESE SUR LA METHODOLOGIE DE L'ENQUETE 17
2.3.1 L'importance des tests en amont *17*
2.3.2 La perception du questionnaire par les chercheurs *18*
2.3.2.1 La compréhension des questions 18
2.3.2.2 Les limites d'un questionnaire directif 19
2.3.3 L'intérêt du couplage quantitatif / qualitatif *20*

3. PRESENTATION DES RESULTATS **21**

3.1 OBSERVATIONS SUR LES DONNEES SOCIO-DEMOGRAPHIQUES 21
3.1.1 Les données socio-démographiques de l'enquête statistique *21*
3.1.1.1 L'âge des répondants 21
3.1.1.2 Les statuts 22
3.1.1.3 Les responsabilités 23
3.1.1.4 Les laboratoires d'appartenance 24
3.1.1.5 Les sections CNU 26
3.1.2 Données sur les entretiens *28*
3.2 QUELLES DONNEES DE RECHERCHE ? 29
3.2.1 Les résultats de l'enquête statistique *29*
3.2.1.1 Les données sources 29
3.2.1.2 Support matériel 31
3.2.1.3 Les données « résultat », produites 35
3.2.2 Les résultats des entretiens *39*
3.2.2.1 Typologie et nature des données : une grande diversité de données collectées et produites 39
3.2.2.2 La donnée, une notion difficile à cerner dans le continuum de la recherche 40
3.2.2.3 La donnée comme construction 41
3.2.2.4 Valeur de la donnée : une reconnaissance scientifique et affective 42
3.2.2.5 La perte, révélateur de la valeur 44
3.2.2.6 Evolutions des pratiques liées aux données 44
3.2.2.7 Les unités et la thématique des données 46
3.2.3 Synthèse sur les données collectées et produites *46*
3.2.3.1 Complexité du paysage 46
3.2.3.2 Comment adopter un discours de sensibilisation efficace sur la question des données ? 47
3.2.3.3 Un paysage en évolution rapide 47
3.3 QUELLES PRATIQUES DE STOCKAGE ET D'ARCHIVAGE ? 48
3.3.1 Les résultats de l'enquête statistique *48*
3.3.1.1 Le stockage des données 48
3.3.1.2 Les politiques d'archivage 52
3.3.1.3 La description des données 55
3.3.2 Les résultats des entretiens *56*
3.3.2.1 Le stockage des données 56
3.3.2.2 L'archivage des données 59
3.3.3 Synthèse sur les pratiques de stockage et d'archivage *61*

3.4 PRATIQUES DE PARTAGE ET DE DIFFUSION DES DONNEES DE RECHERCHE 64
3.4.1 Résultats de l'enquête statistique 64
3.4.1.1 Quelles pratiques du partage des données ? 64
3.4.1.2 Quelles représentations, quelles opinions sur le partage des données de recherche ? 66
3.4.1.3 Quels besoins, quels souhaits en matière de partage des données ? 72
3.4.2 Quelles variables sur les pratiques de partage des données ? 74
3.4.2.1 La variable de l'âge 75
3.4.2.2 Les variables du statut et des responsabilités sur les pratiques de partage 79
3.4.2.3 Les variables scientifiques sur les pratiques de partage 81
3.4.2.4 Les variables liées à la nature des données sur les pratiques de partage 83
3.4.3 Résultats des entretiens sur le partage des données 87
3.4.3.1 Quelle position générale sur le principe du partage des données de recherche ? 87
3.4.3.2 Les facteurs d'incitation au partage des données 89
3.4.3.3 Les facteurs de frein au partage des données 92
3.4.3.3.1 Facteurs « externes » 93
3.4.3.3.2 Facteurs « internes » 98
3.4.3.4 Témoignages, interrogations, souhaits 101
3.4.4 Synthèse sur les pratiques de partage des données 103
3.4.4.1 Quelles leçons ? 103
3.5 QUELS BESOINS, QUELLES ATTENTES ? 107
3.5.1 Les résultats de l'enquête statistique 107
3.5.1.1 Quelle connaissance du Plan de Gestion de Données ? 107
3.5.1.2 Quels opinions et souhaits en termes de stratégie des données de recherche à l'université ? 110
3.5.1.3 Quels besoins de services pour la gestion des données de recherche ? 115
3.5.2 Les résultats des entretiens 124
3.5.2.1 Les besoins techniques 124
3.5.2.2 Les besoins de services 125
3.5.3 Synthèse sur les besoins et attentes 127
3.5.3.1 La sécurisation et la préservation des données : un point d'entrée et un impératif à court terme 127
3.5.3.2 Une dimension supplémentaire : le collaboratif 128
3.5.3.3 Information et clarification : des préalables 128
3.5.3.4 Un appui au long cours 128

4. SYNTHESE DES LEÇONS DE L'ENQUETE 130

CONCLUSION : RECOMMANDATIONS ET PROPOSITIONS 134

DES RECOMMANDATIONS POUR QUI ? 134
SIX PROPOSITIONS POUR L'UNIVERSITE RENNES 2 135

TABLE DES ANNEXES 139

TABLE DES FIGURES 140

PROPOSITIONS POUR L'UNIVERSITE RENNES 2 142

LISTE DES UNITES DE RECHERCHE DE L'UNIVERSITE RENNES 2 148

LE QUESTIONNAIRE DE L'ENQUETE STATISTIQUE 149

TABLE DES MATIERES 158

Résumé en français :

Quels sont les types de données de recherche collectées, traitées et produites dans une université de lettres et sciences humaines et sociales ? Quelles sont les pratiques des chercheurs en SHS en matière de stockage, d'archivage, de diffusion, de partage de leurs données de recherche ? Quelles sont leurs représentations et leurs définitions des données de recherche, leur position par rapport au libre accès ? Quels sont leurs besoins prioritaires en matière de gestion ou de partage des données de recherche ? Comment perçoivent-ils le bon niveau d'une politique des données ? C'est pour répondre à toutes ces questions qu'une double enquête, statistique et qualitative, a été menée à l'Université Rennes 2 au printemps 2017, enquête portée par l'URFIST (Unité Régionale de Formation à l'Information Scientifique et Technique) de Rennes, la Maison des Sciences de l'Homme en Bretagne et le Service Commun de Documentation de l'Université Rennes 2, avec le soutien des instances de l'université. Le rapport et ses annexes en présentent ici tous les résultats, avec un certain nombre de propositions pour une politique des données de recherche.

Résumé en anglais :

What types of research data do researchers collect, manage and create in a humanities university ? What are their habits as regards storing, archiving, dissemination and sharing of their research data ? How do they regard and define research data, as well as their own attitudes about open access? What are their main needs when it comes to research management or sharing? What do they consider to be the right level for a data policy? In order to answer all these questions, a statistical and qualitative survey was carried out at Rennes 2 University during the spring of 2017. The survey was led by the Rennes URFIST (regional STI training unit), the Maison des Sciences de l'Homme en Bretagne (Human sciences institute in Brittany) and the Rennes 2 University Library, with the support of university governance.
The report and its appendices lay out all the results, along with several proposals for the development of a research data policy.

Mots-clés en français :
données de recherche, SHS, enquête, pratiques informationnelles, chercheurs, Université Rennes 2, partage des données, pratiques de gestion des données,

Mots-clés en anglais :
research data, humanities, survey, information practices, researchers, Rennes 2 university, data sharing , data management