

Données de la recherche : Quelles pratiques ? Quels besoins ? Enquête à Aix-Marseille Université

Caroline Sophie Donati

▶ **To cite this version:**

Caroline Sophie Donati. Données de la recherche : Quelles pratiques ? Quels besoins ? Enquête à Aix-Marseille Université. [Rapport de recherche] Aix Marseille Université. 2019. ⟨hal-02493679⟩

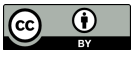
HAL Id: hal-02493679

https://amu.hal.science/hal-02493679v1

Submitted on 28 Feb 2020

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 CC BY 4.0 - Attribution - International License

DONNEES DE LA RECHERCHE :
QUELLES PRATIQUES ? QUELS BESOINS ?

Enquête à Aix-Marseille Université



https://pixabay.com (libre utilisation)

Caroline Sophie DONATI

SCD – Cellule Open Access

Septembre 2019

Aix*Marseille université
Initiative d'excellence

CC BY

Table des matières

Contexte ---- 2

Résumé ---- 2

Quelques données « socio-démographiques » ---- *4*

Présentation des résultats ---- 6

1. Enseignants chercheurs ---- *6*

1.1. Activité de recherche ---- 6

1.2. Typologie des données de la recherche ---- 8

1.3. Pratiques de stockage et d'archivage des données ---- 11

1.4. Pratiques de partage et de diffusion des données ---- 14

2. Personnels de l'IST ---- *19*

2.1. Gestion de projet ---- 20

2.2. Les données de la recherche ---- 24

2.3. Stockage et archivage des données ---- 25

3. Besoins et attentes ---- *27*

3.1. Les principaux freins rencontrés par les personnels IST ---- 27

3.2. Sur quels points souhaiteriez-vous bénéficier d'une aide ? ---- 28

Perspectives ---- 32

Quelle suite donner à cette enquête ? ---- *32*

Contexte

La réflexion autour de la gestion des DR s'inscrit dans une dynamique impulsée par l'État en 2016, concrétisée par la publication du Plan national pour la science ouverte en juillet 2018. La Science ouverte est un mouvement consistant à diffuser les matériaux et les résultats de la recherche en accès ouvert, sans obstacle technique, juridique, géographique ou commercial, et idéalement sans aucun délai.

Entrer dans ce mouvement, c'est répondre à de nouveaux ordres de grandeur concernant la capture, la recherche, le partage, le stockage, l'analyse et la présentation des données et donc se préoccuper de leur bonne gestion et de leur conformité au regard des principes FAIR qui président à ce nouvel écosystème.

Comme nous l'avons vu lors du premier comité de pilotage (le 6 mai 2019), AMU s'engage en faveur de la science ouverte et de l'ouverture des données de la recherche. Pour rappel, par donnée de la recherche nous pensons aux :

- données produites lors des activités des chercheurs donc réalisées dans le cadre de la mission de service public d'un établissement de recherche ;
- enregistrements factuels (chiffres, textes, images et sons…) utilisés comme sources principales pour la recherche scientifique ; et généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche ;
- données achevées.

Dans ce contexte il est important de faire un état des lieux des pratiques en termes de perception des données de la recherche, de compréhension du mouvement de la science ouverte et des pratiques de la gestion des données de la recherche.

Cette enquête est l'un des piliers de notre plan d'actions et doit valider ou invalider les choix que nous avons fait concernant cette première année structurante autour du projet « données de la recherche ».

Résumé

Disponible de mi-juin à mi-septembre 2019, cette enquête a été préparée par le comité de pilotage des données de la recherche. Elle est composée de deux formulaires à destination de la communauté scientifique d'Aix-Marseille Université :

- Le premier, 36 questions, pour les enseignant-chercheurs (EC) ;
- Le second à l'attention des personnels de l'information scientifique et technique (IST) en soutien à la recherche, 33 questions.

Elle a été diffusée via les listes suivantes, ainsi que par nos contacts et par la DRV, que nous remercions :

- espe-chercheurs@univ-amu.fr
- sciences-chercheurs@univ-amu.fr
- pharmacie-chercheurs@univ-amu.fr
- odontologie-chercheurs@univ-amu.fr
- medecine-chercheurs@univ-amu.fr
- fdsp-chercheurs@univ-amu.fr
- feg-chercheurs@univ-amu.fr
- iut-chercheurs@univ-amu.fr
- itrf@univ-amu.fr

Toutes les questions n'étant pas obligatoires, sur les 676 réponses (310 IST et 366 EC soit une participation d'environ 10 %) récoltées, 40,09 % (271 formulaires) sont inexploitables, nous verrons par la suite pourquoi, et 405 formulaires sont exploitables (59,91 %). Tous ont été dépouillés mais seuls **258** formulaires complets (38,17 %) ont été utilisés pour réaliser cette analyse. Nous disposons ainsi de **61** enquêtes IST et de **197** enquêtes EC (représentant 5.55 % de la communauté recensée au 31/12/2018[1]) **complètes** pour notre état des lieux.

Pourquoi tant de formulaires incomplets ?

Il est apparu que de nombreuses enquêtes ont été lancées en fin d'année universitaire, rendant moins visible notre démarche. De plus, la gestion des données de la recherche n'est pas (encore) considérée comme une priorité mais plutôt comme une contrainte supplémentaire dans l'activité du chercheur. Subie, imposée, elle n'attire pas l'attention et ne fait pas sens aux yeux de certains… Nombreux sont les EC qui ne pensent pas devoir rendre des comptes sur l'exploitation de leurs données, et encore moins les ouvrir (ce terme étant mal interprété par la communauté scientifique).

Ensuite, Lime Survey compte un « participant » à chaque clic vers l'enquête : du moment qu'elle est ouverte et qu'une réponse est cochée, Lime Survey enregistre une participation partielle. Par exemple, 151 répondants IST (sur les 310 ayant ouvert l'enquête) se sont arrêtés dès la première page du formulaire (question 2). D'autres ont ouvert l'enquête sans lire le mail (expliquant à qui elle était destinée) et se sont rendus compte qu'ils n'étaient pas concernés par la thématique.

Enfin, certains EC ont eu, de leur propre aveu, le sentiment d'être espionnés, oppressés, ne comprenant pas pourquoi toutes ces questions leur étaient posées. La plupart, ne sachant quoi répondre, ont préféré abandonner plutôt que d'indiquer « je ne sais pas » et aller au bout de la démarche. Une phase de rencontre sera indispensable pour comprendre ce blocage et essayer d'y remédier.

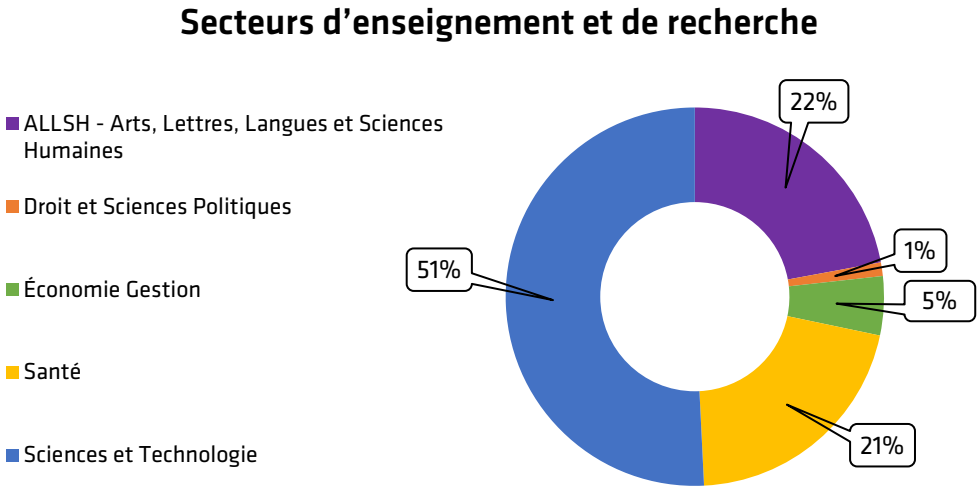
Malgré tout, cet échantillon de 258 réponses est suffisant pour avoir un premier aperçu des pratiques en termes de gestion des données de la recherche et des besoins existant à AMU. Il nous permettra, nous le verrons, de statuer sur la viabilité de notre plan d'actions et des démarches en cours.

49,22 % des répondants souhaitent recevoir des informations sur le sujet (127 personnes seront donc inscrites à la liste de diffusion). 67 répondants acceptent d'être contactés pour un entretien (25,97 %) ce qui ouvre d'autres perspectives car rien ne remplace un échange. Nous allons poursuivre les rencontres avec les volontaires en espérant impliquer plus largement la communauté chemin faisant.

[1] 2 817 EC au 31/12/2018.

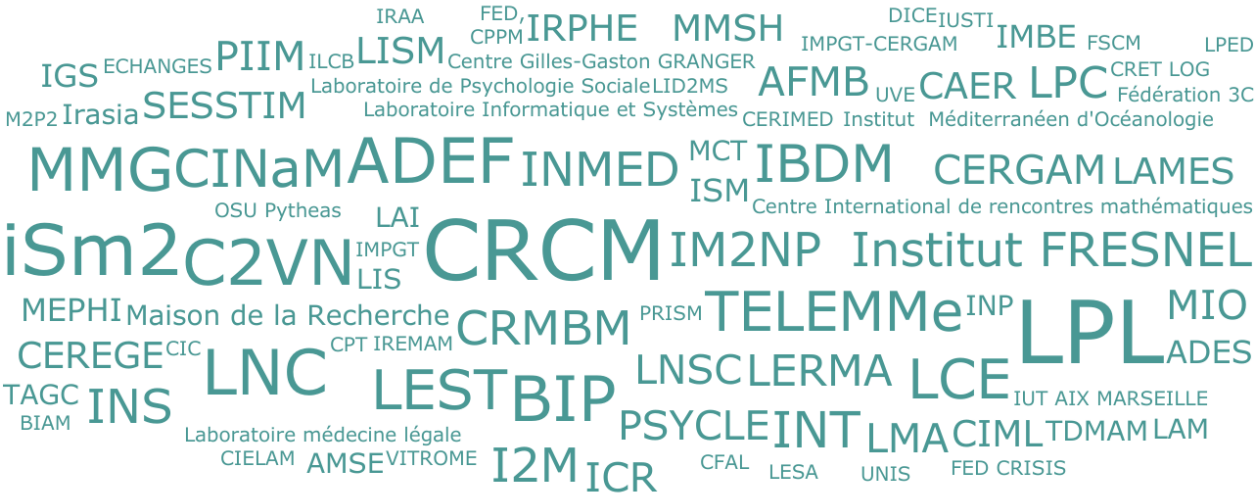
Quelques données « socio-démographiques »

Toutes les disciplines sont représentées. 50,78 % des participants proviennent des *Sciences et Technologie*, 22,34 % des *Arts, Lettres, Langues et Sciences Humaines (ALLSH)* suivis de près par le secteur *Santé* (20,93 %). L'*Économie Gestion* et le *Droit* réunissent respectivement 5,04 % et 1,16 % des répondants.



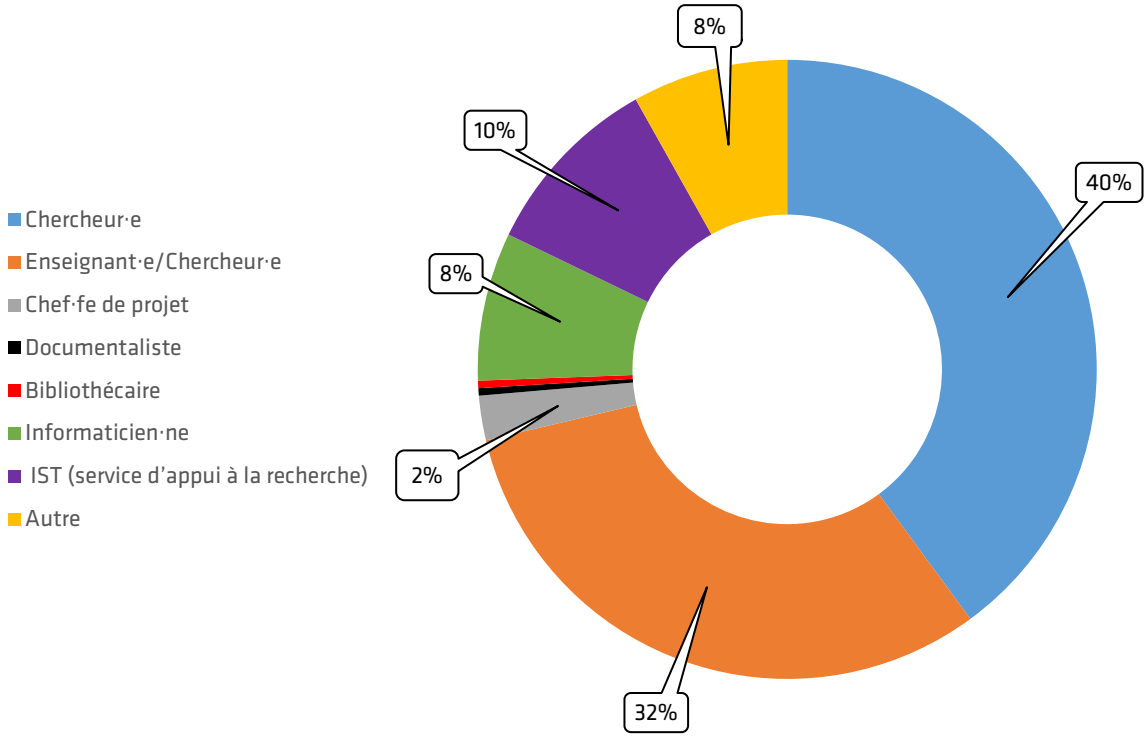
Il est intéressant et satisfaisant de constater que les unités qui se sont le plus manifestées lors de cette enquête sont celles que nous avons rencontrées en amont. La présentation du projet et surtout l'annonce de l'enquête ont été déterminantes dans la participation[2].

Aperçu des unités qui ont participé à l'enquête (EC et IST confondus)



[2] Vous trouverez en annexe la liste des unités.

Diverses catégories des personnels scientifiques sont représentées.



La catégorie « *Autre* » représente 21 personnes qui pour la plupart ont répondu au mauvais formulaire. Ainsi 13 réponses du formulaire EC affichent les réponses suivantes : IE (2), doctorant (4), IR (4), ITA, prof émérite, chercheure-associée (activité secondaire) et chargé d'affaires alors que dans le formulaire IST nous avons 5 réponses que nous n'avons pu classer parmi lesquelles : IGR, IGE, normalien élève, gestionnaire, secrétaire pédagogique.

Nous avons donc la confirmation que la science ouverte, à travers la gestion des données de la recherche, concerne une multitude de métiers et qu'il ne faut en aucun cas réduire sa dimension aux seuls producteurs de données. Plus large le profil professionnel est, plus riche la communauté sera.

Présentation des résultats

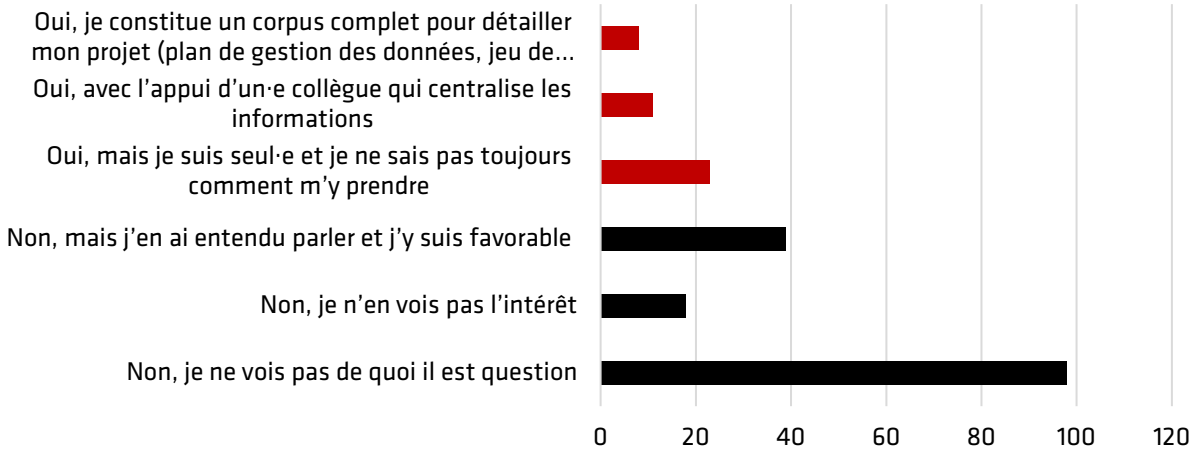
Cette partie va nous permettre d'analyser les réponses des EC et des personnels IST en suivant peu ou prou l'ordre de chaque questionnaire.

1. Enseignants chercheurs

1.1. Activité de recherche

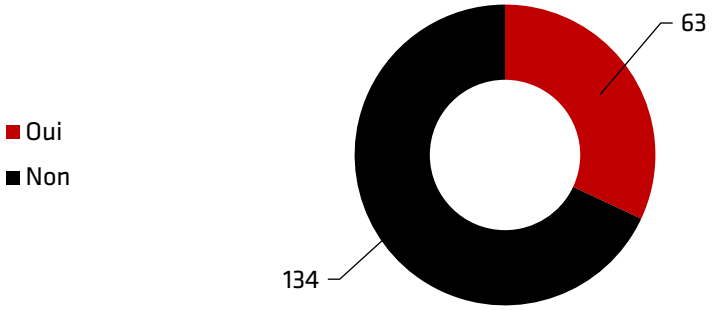
Fixer le cadre et savoir exactement où en sont les EC en 2019 était l'objectif de cette première question : *Dans le cadre de vos projets de recherches, avez-vous rédigé des documents détaillant les données produites et leur cycle de vie ?*
Depuis cette année, les financeurs (France, Europe) exigent la rédaction de plans de gestion des données de la part de tous les lauréats. A priori, il est aisé de s'inquiéter de la réponse puisque la moitié des répondants (98) ne sait pas de quoi il est question. 39 répondants (19,80 %) sont favorables à cette pratique, et 23 (11,68 %) ne savent pas comment faire mais le font.

Fig. 1. Dans le cadre de vos projets de recherches, avez-vous rédigé des documents détaillant les données produites et leur cycle de vie ?



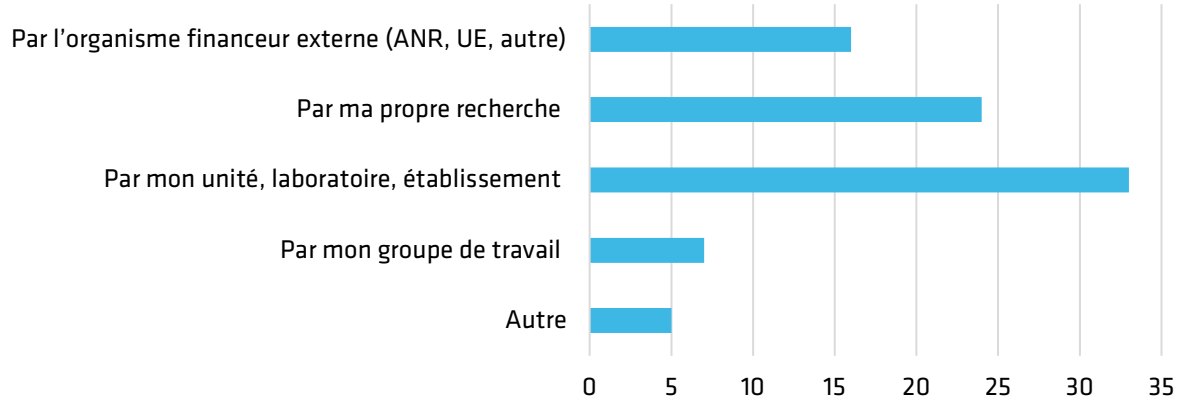
À partir de cette première question nous arrivons doucement au constat que la communauté scientifique à AMU est mal informée et que la thématique de la science ouverte n'est pas centrale dans la recherche, méconnue même (pour l'échantillon interrogé). Par exemple 68,02 % ne connaissent pas de directives ou de recommandations pour la gestion des DR (Fig. 2).

Fig. 2. Connaissez-vous des directives ou recommandations pour la gestion des données de recherche ?



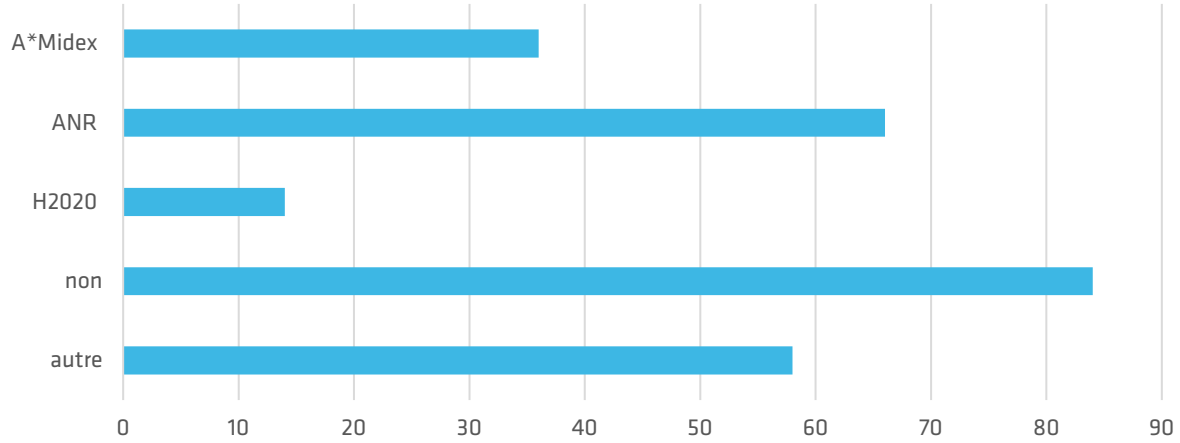
Si nous nous intéressons aux presque 32 % de « Oui », nous découvrons comment circule l'information. Il est étonnant de lire que l'organisme financeur arrive en troisième position dans la figure 3. En effet nous aurions pensé que les organismes financeurs seraient les premiers à diffuser de l'information au sujet des injonctions qu'ils mettent en place mais ce n'est pas le cas. Peut-être y aurait-il une réflexion à mener avec ces derniers pour une meilleure dissémination de l'information sur les injonctions qu'ils mettent en place. L'unité reste le point d'entrée principal pour faire circuler l'information, cela est rassurant dans la perspective des entretiens.

Fig.3. Comment en avez-vous été informé.e ?



Même si 84 des répondants EC ne bénéficient pas de financement actuellement, nombreux sont ceux qui ont des projets soutenus par des financeurs publics (Fig 4). Il est donc important de préparer des campagnes de communication en premier lieu pour les lauréats 2019, en second lieu pour tous les EC et de poursuivre les rencontres.

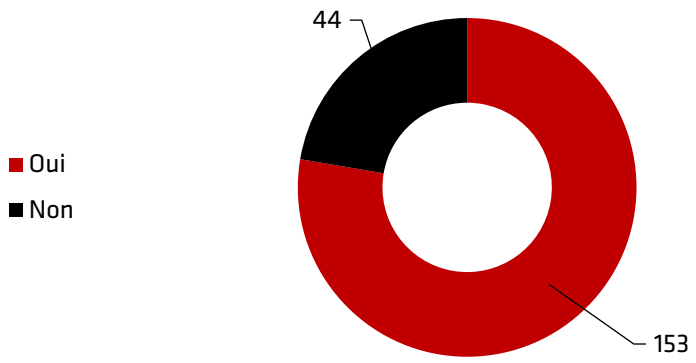
Fig. 4. Je bénéficie d'un financement pour mon projet de recherche



1.2. Typologie des données de la recherche

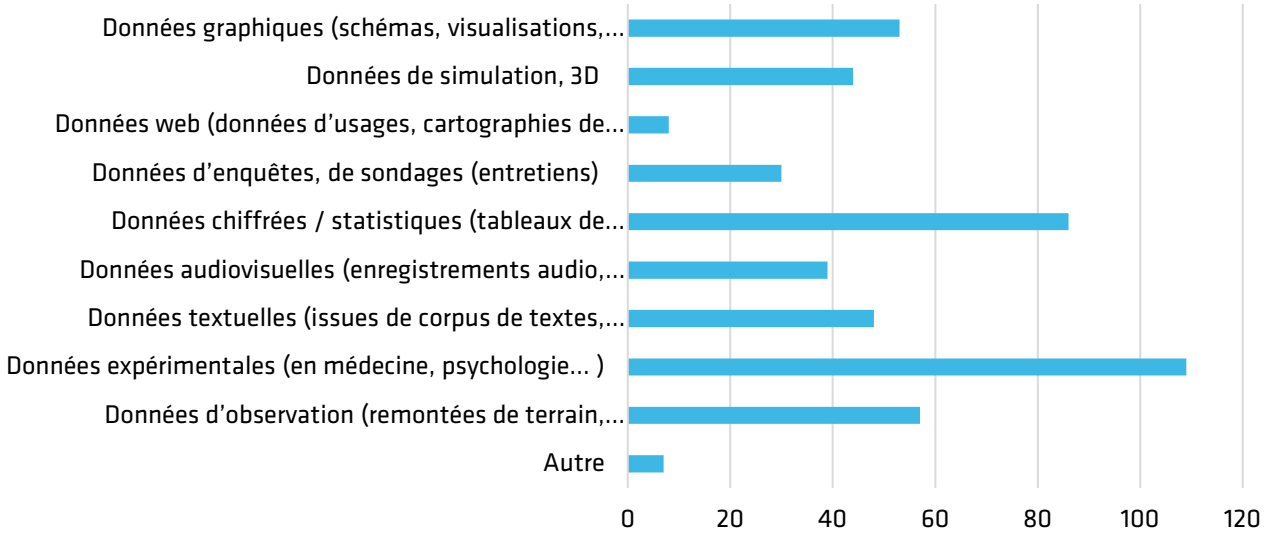
Nous avons choisi dans cette partie d'avoir une approche globale des données pour aller vers les questions essentielles sur l'utilisation de ces dernières. Pour introduire la thématique, nous avons ainsi posé la question basique mais ô combien complexe *Savez-vous ce qu'est une donnée de la recherche ?* La majorité des EC a répondu « Oui » (77,66 %), nous verrons dans la pratique ce que cela signifie car il est très difficile de définir cette notion et nous voyons, au quotidien, qu'une certitude peut devenir floue devant la quantité de données existante.

Fig. 5. Savez-vous ce qu'est une donnée de la recherche ?



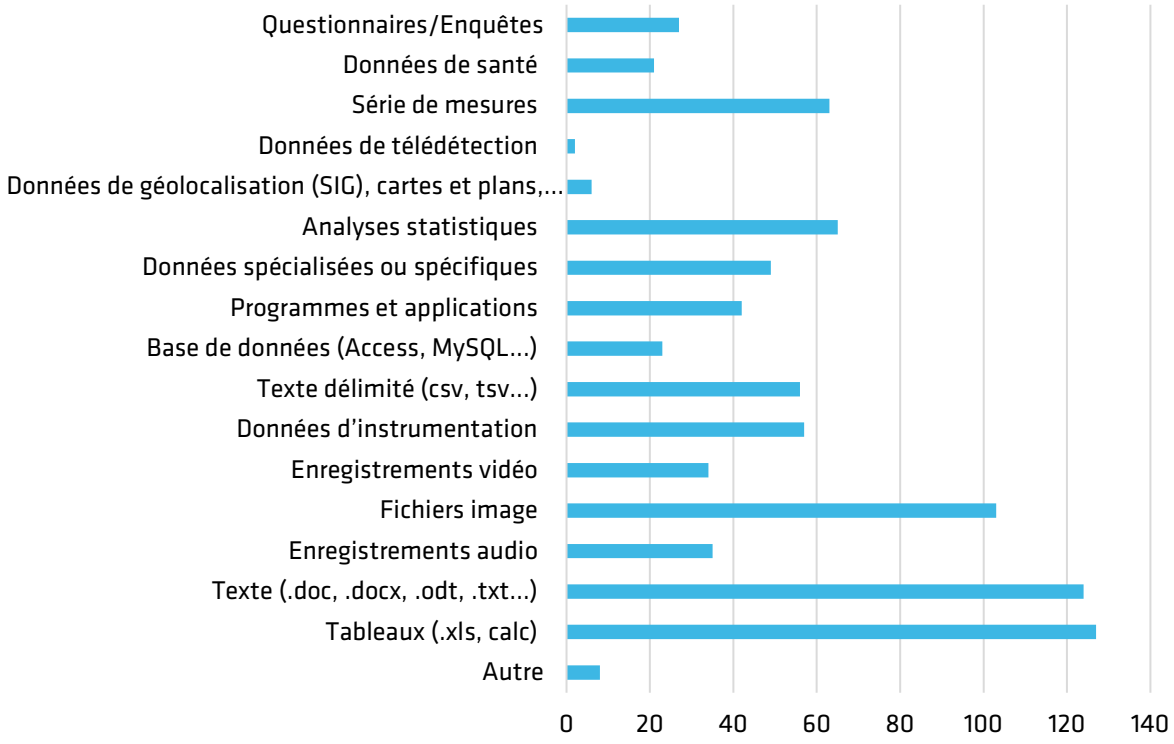
Nous avons repris les propositions des enquêtes de Berlin, Lille ou encore Rennes pour ensuite essayer de classer les données utilisées par nos répondants (**Fig. 6**). Le secteur science étant fortement représenté il est logique de retrouver en tête les données expérimentales (55,33 % des réponses) et les données chiffrées/statistiques (43,65 %). Les données d'observation et les données graphiques approchent les 30 %.
Malgré tout nous avons, sans surprise, un large éventail de catégories de données exploitées.

Fig. 6. Dans quelle.s catégorie.s classeriez-vous vos données ?



La **figure 7** montre les réponses à la question des types de documents produits par les EC. Les données textuelles et les tableaux sont massivement employés et représentent respectivement 62,94 % et 64,47 % des réponses, suivis de près par les fichiers de type image (52,28 %). La catégorie « autre » reprend finalement des fichiers qui auraient trouvé leur place dans les choix proposés et n'apportent pas d'indication supplémentaire sur les pratiques des EC : nous avons des fichiers bruts de données de spectrométrie de masse, des données de diffraction X, des bases de données, des fichiers de séquences, des fichiers issus de logiciels spécifiques à l'électrophysiologie...

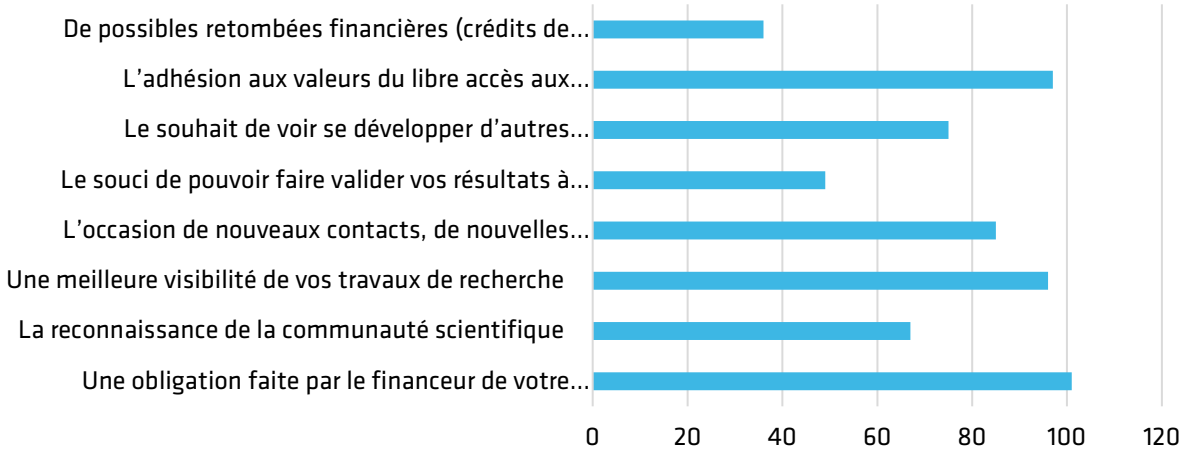
Fig. 7. De quel.s type.s de fichier.s s'agit-t-il ?



Ces deux questions démontrent la richesse et la variété des données produites, mais présagent aussi de la difficulté future à les trier… Il est souvent difficile de distinguer une donnée source et une donnée produite. Un commentaire de texte pourrait-il être considéré comme une source alors qu'il est lui-même issu d'une source ? Comment valoriser ces données dérivées ? Des photos d'œuvres d'art sont-elles des sources ou des données produites (par l'appareil photo) ? De fait, doit on sauvegarder et partager la capture **ET** le traitement, l'analyse qui en découlent ? En a-t-on seulement le *droit* ? Autant de questions sur lesquelles il est très compliqué de statuer pour l'instant faute de recul (et encore une fois nous ne parlons pas des droits sur ces dernières, juste de leur classification !). Malgré tout, ces éléments devaient nous permettre de trier de la manière la plus fiable possible les données produites à AMU.

Il est intéressant de connaitre les motivations des EC à ouvrir leurs données et entrer de fait dans le mouvement de la science ouverte. La question *Parmi les raisons suivantes, lesquelles vous inciteraient à rendre vos données de recherche accessibles en libre accès ?* a donné des résultats assez surprenants car forts équilibrés. Le diagramme suivant parle de lui-même.

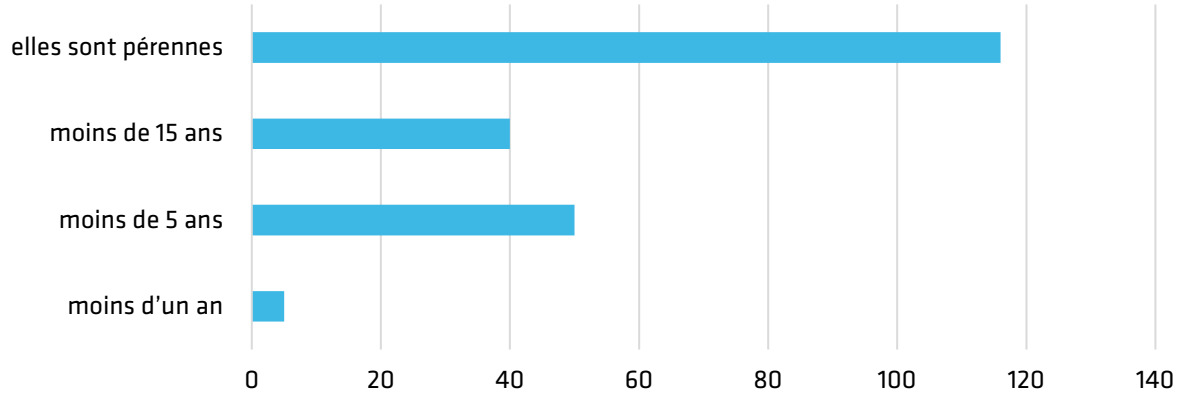
Fig. 8. Parmi les raisons suivantes, lesquelles vous inciteraient à rendre vos données de recherche accessibles en libre accès ?



Même si l'injonction faite par les financeurs est en tête avec 51,27 % des réponses, presque la moitié des répondants ont choisi *L'adhésion aux valeurs du libre accès aux résultats de la recherche scientifique* (49,24 %) et *Une meilleure visibilité de vos travaux de recherche* (48,73 %). Cela démontre une conscience de l'intérêt de ces démarches et une volonté d'entrer dans la science ouverte. C'est assez encourageant, d'autant que certains y voient l'opportunité de nouvelles collaborations (43,15 %), ce qui connote une vision positive de cette démarche.

La dernière question de cette partie est davantage technique. Elle est surtout un indicateur destiné à réfléchir sur une politique de stockage/archivage à cours, moyen et long termes. Ainsi la **figure 9** montre que la majorité des données sont jugées pérennes par les EC. Il faudra donc étudier finement les conditions d'archivage des entrepôts de données qui seront sélectionnés, et se rapprocher le cas échéant du CINES pour assurer la pérennisation de ces données. Cette question interpelle sur les conditions de stockage des données ayant une durée de vie à 1 et 5 ans. Qu'est-ce que cela signifie ? Seront-elles reprises et modifiées au-delà ? Seront-elles obsolètes et donc supprimées ? Cela change la stratégie de leur partage et donc leur avenir. Il sera intéressant d'approfondir ce point avec les EC et les services informatiques dédiés.

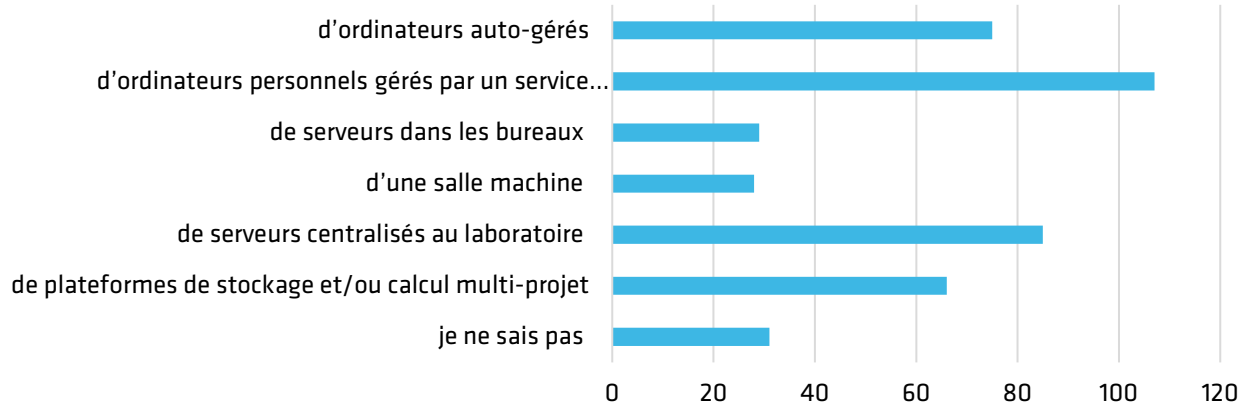
Fig. 9. Sur quelle période vos données restent-elles pertinentes ?



1.3. Pratiques de stockage et d'archivage des données

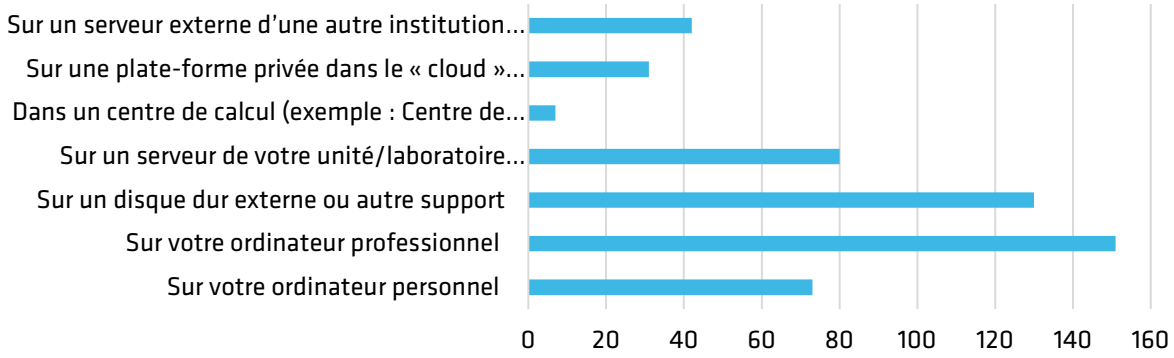
Nous essayons désormais d'en savoir davantage sur les pratiques des EC en matière de stockage et d'archivage. Pour commencer, nous nous intéressons aux moyens mis à disposition au sein de leur unité. Il est important de connaitre les besoins futurs des unités pour le stockage en fonction des pratiques actuelles. Ces informations sont données à titre indicatif, il faudra faire du cas par cas pour cette problématique.

Fig. 10. Votre laboratoire est-il doté



Plus intéressante est la question suivante (**Fig. 11**) *Où stockez-vous vos données de recherche ?* 76,65 % des répondant laissent leurs données sur leur ordinateur professionnel, et 65,99 % utilisent un support externe : la pratique majoritaire est donc de stocker en local. 31 EC disposent d'un cloud « privé », alors que 80 utilisent un serveur hébergé par AMU. Ces questions étant à choix multiple, nous pouvons imaginer qu'un EC utilise divers moyens de stockage, mais favorise la sauvegarde en local - parfois même sur un ordinateur personnel (37,06 %). Les entretiens nous permettront d'aller plus loin sur ce point.

Fig. 11. Où stockez-vous vos données de recherche ?



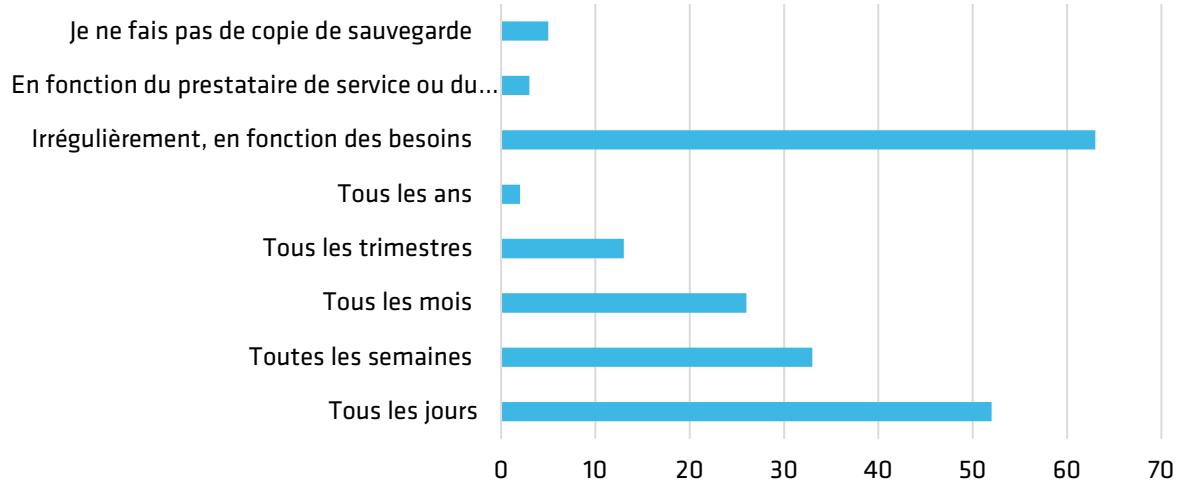
La question complémentaire, *Stockez-vous vos données sur un autre support ?*, confirme l'intérêt pour les supports externes (clé USB, disque dur, CD/DVD) et montre que certains n'ont qu'un seul lieu de stockage (24,87 %). En revanche, les répondants ayant coché « autre » ont ajouté dans le champ libre :

- Copie partielle sur ordinateur de bureau ;
- Boîte mail ;
- NAS ;
- Cahier ;
- Serveur du laboratoire NON hébergé au datacenter de AMU ;
- Cloud ;
- AMUbox ;
- Sauvegarde chiffrée.

Les réponses grisées auraient pu être cochées dans les propositions initiales. Peut-être les éléments de réponse n'étaient-ils pas assez clairs, ou bien les EC ne maitrisent pas totalement le vocabulaire. Nous verrons cela lors d'entretiens plus poussés. Le papier n'a pas disparu et doit être pris en considération comme « mode de stockage ».

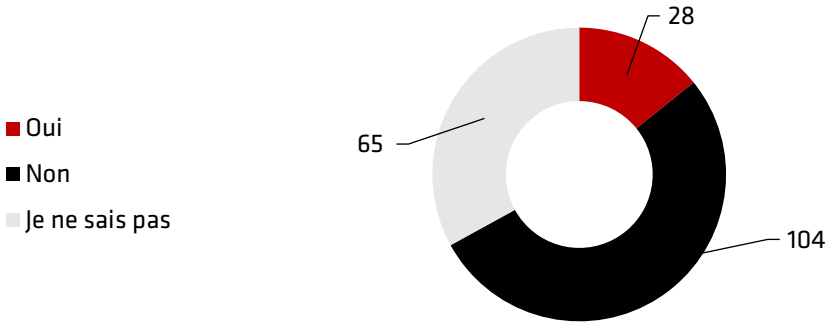
À quelle fréquence les EC sauvegardent-ils les données ? 26,40 % déclarent lancer une sauvegarde tous les jours, et 31,98 % le font irrégulièrement en fonction des besoins (**Fig. 12**). Cela témoigne de l'absence de règle et de charte de bonnes pratiques à ce sujet. Toutefois ce n'est pas parce qu'une unité fait des préconisations qu'elles sont suivies. Cette problématique est de la responsabilité de l'EC du moment où ses données sont sur son ordinateur ou son disque dur. La solution d'avoir des serveurs qui gèrent les sauvegardes est intéressante, encore faut-il que les EC y déposent les données… Une réflexion doit avoir lieu car elle pose directement la question de la politique d'archivage dans les unités (**Fig. 13**) et de la responsabilité de la sauvegarde des données (**Tableau A**).

Fig. 12. À quelle fréquence sauvegardez-vous vos données ?



En effet, alors 52,79 % des EC répondent que leur laboratoire n'a pas de politique d'archivage des données, 32,99 % ne peuvent répondre à cette question. Cela suppose que, sur les 197 répondants, 169 ne sont pas informés et/ou n'ont pas de politique d'archivage dans leur unité.

Fig. 13. Y a-t-il une politique d'archivage des données de recherche dans votre laboratoire (critères, modalités, nommage des fichiers...) ?



Les 28 répondants « oui » ont eu à répondre à la question de la responsabilité de la sauvegarde des données. Le choix multiple, pour représenter tous les cas, révèle que la moitié confie cette tâche au service informatique (50 %), les autres procèdent eux-mêmes ou organisent la sauvegarde au sein de leur groupe de travail. Dans le tableau des résultats (**Tableau A**), on constate que les personnels de bibliothèque ne sont pas sollicités, certainement parce que les données se gèrent en local, dans l'unité, l'ouverture n'étant pas encore ancrée comme pratique. L'on pourrait imaginer, dans quelques temps, solliciter les compétences au sein des bibliothèques pour la documentation des données et ainsi leur en confier le dépôt sur des espaces dédiés.

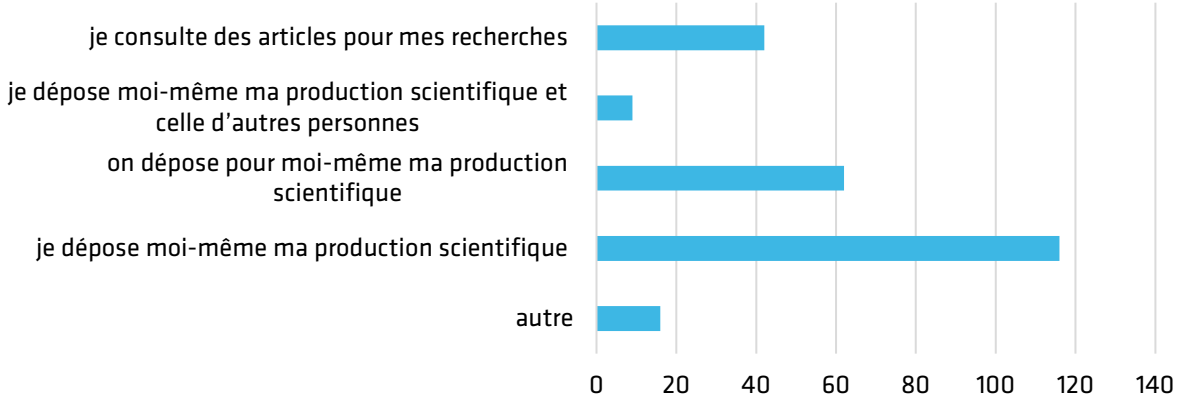
Tableau A. Qui est responsable du stockage, de la sauvegarde et de l'archivage des données ?		
Moi-même	9	32,14 %
Un·e doctorant·e	1	3,57 %
Un·e chef·fe de projet	1	3,57 %
Une personne dédiée au sein du groupe de travail	6	21,43 %
Le personnel du service informatique	14	50,00 %
Le personnel de la bibliothèque	0	0,00 %
Un·e prestataire de service externe	0	0,00 %
Je ne sais pas	0	0,00 %
Mon assistant·e	0	0,00 %
Autre un IR les membres de l'équipe	2	7,14 %

Dans l'ensemble, les pratiques de sauvegarde dépendent de l'unité dans laquelle l'EC se trouve. Les unités dotées de moyens disposent souvent d'un pôle informatique qui peut prendre en charge cette activité, ou du moins mettre en place des serveurs et des outils pour la communauté. D'autres unités se contentent de pratiques plus personnelles et locales, chacun gérant ses données sans vraiment se demander si c'est de « la gestion » d'ailleurs. Il est important de travailler auprès des unités pour aider à penser la gestion des données (y compris la sauvegarde), et donc avoir une politique de gestion. Cela ne pourra se faire que si l'établissement propose des outils et des solutions techniques adaptées et accessibles aux publics, et qu'il est à l'initiative de ces démarches.

1.4. Pratiques de partage et de diffusion des données

Une question sur un outil « connu » pour commencer le dossier épineux du partage des données : **HAL**. La majorité des EC déposent eux-mêmes leurs publications (58,88 %) et 31,47 % confient cette tâche à un collègue. HAL est, pour 42 EC, l'occasion de consulter la production de leurs collègues. En revanche, les commentaires au sujet de l'archive ouverte prouvent qu'il y a encore de la sensibilisation à faire…

Fig. 14. Concernant l'utilisation de HAL,

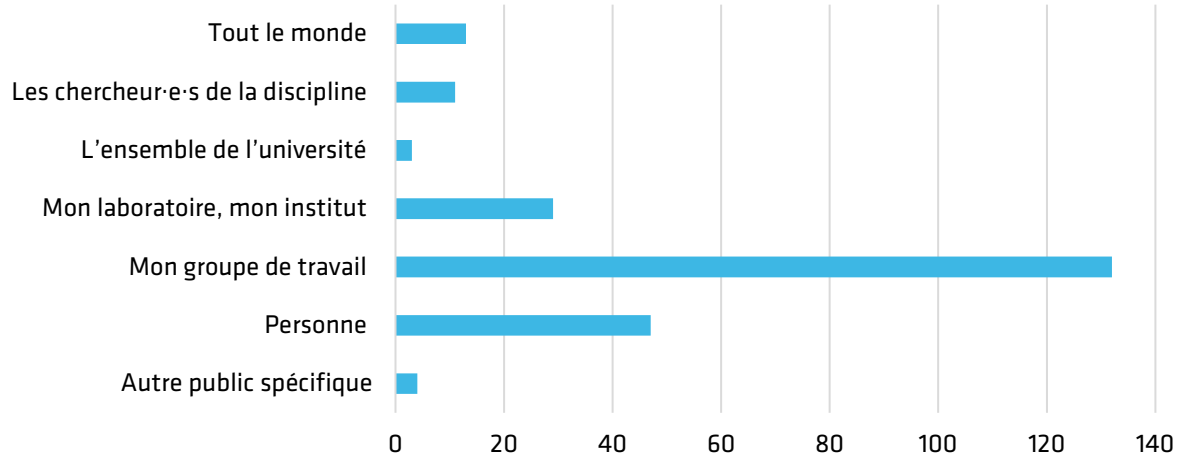


Voici les commentaires laissés pour la case « autre » :

non utilisé
n'utilise pas encore HAL
je n'arrive pas à déposer ma production scientifique
parfois
Ne sais pas
HAL n'est pas pertinent, c'est une agrégation de dépôts d'dans des bases de donnes standardisées pour la discipline. EN tant que tel, HAL est une contrainte, pas une aide à la recherche.
Pas d'usage
pas d'utilisation de HAL
je n'utilise pas (encore) HAL
je ne l'utilise pas encore
HAL est une usine à gaz
Jamais utilisé
hal ?
Jamais utilisé HAL
ni dépôt ni consultation
je ne dépose pas
Surtout pas HAL.... grosse merde
je dois déposer moi-même ma production scientifique mais j'estime que ce n'est pas mon rôle, nous chercheurs avons un grand besoin de soutien à ce sujet

À la question, *En dehors de vous, qui peut accéder à vos données ?*, 239 réponses ont été récoltées. 47 répondants ne partagent pas leurs données, 161 ne partagent qu'au sein de leur unité ou de leur équipe. La pratique du partage des données est donc très localisée et fermée. Seulement 30 répondants partagent de manière élargie à la communauté ou à l'université (12,55 % des réponses).

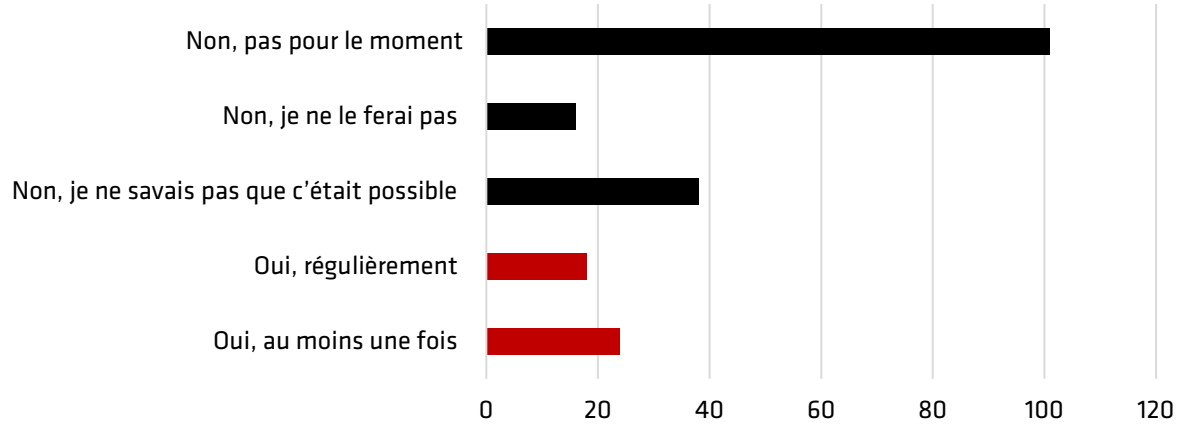
Fig. 15. En dehors de vous, qui peut accéder à vos données ?



Cette pratique est totalement confirmée par la question suivante sur l'utilisation des données d'autres chercheurs. Sur les 197 répondants, 155 soit 78,68 % ne téléchargent pas de données provenant d'autres chercheurs pour des raisons diverses, voir **figure 16.** Cependant l'espoir vient des raisons

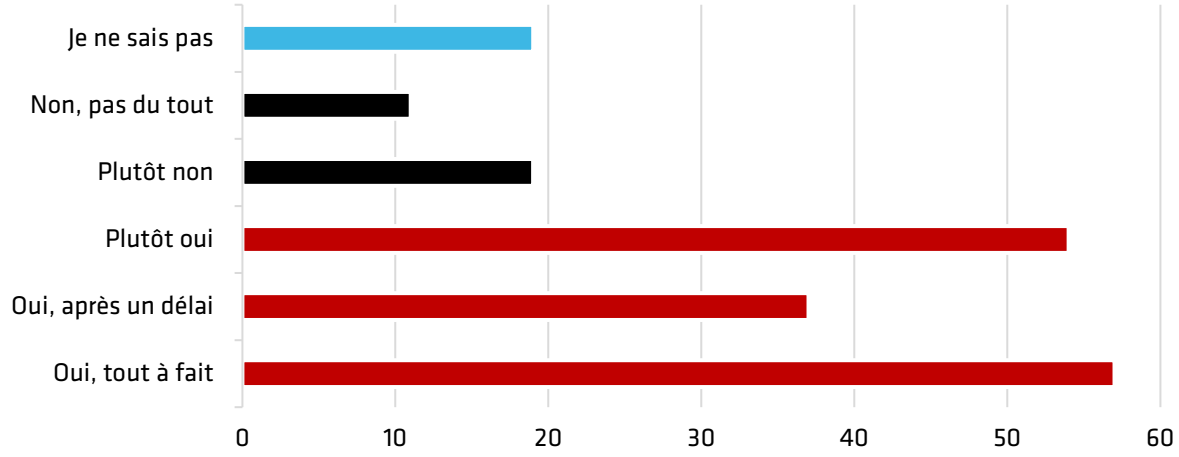
exprimées pour ce « non ». En effet, 139 EC **ne le font pas pour le moment** ou **ne savent pas que c'est possible.** Rien n'est donc définitif ![3].

Fig. 16. Avez-vous déjà téléchargé les données d'autres chercheur·e·s ?



Paradoxalement (**Fig. 17**), la majorité des EC serait plutôt d'accord pour partager les données (75,13 %). Cela reflète bien l'ignorance plutôt que le refus autour de cette pratique. Si un EC envisage de partager ses données, on peut penser qu'il finira par consulter celles des collègues du moment qu'il maitrise les outils et les services. Notons tout de même que 21,32 % utilisent les données de tiers.

Fig. 17. Seriez-vous d'accord pour diffuser vos données de recherche en libre accès (à condition qu'elles soient diffusables et protégées juridiquement) ?

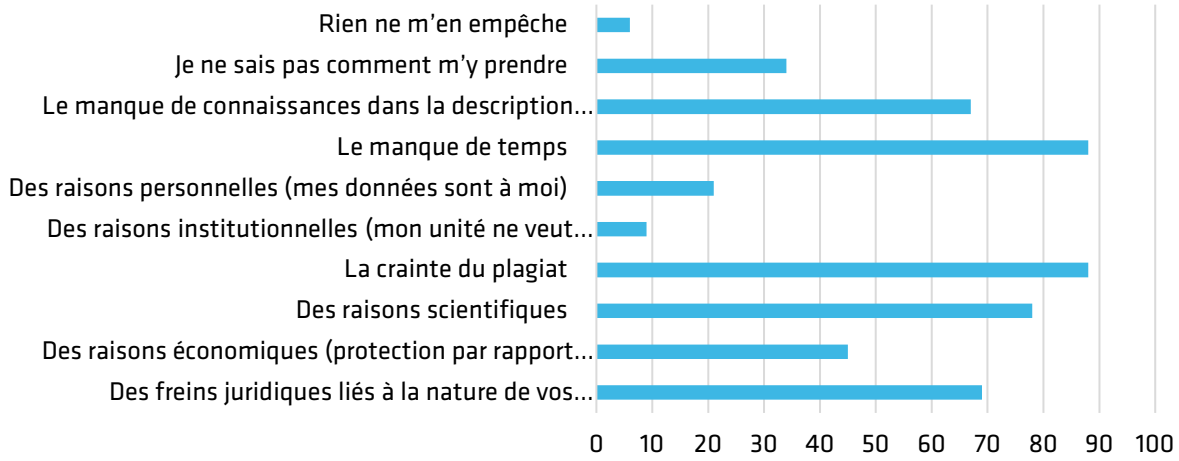


De fait, *Quels seraient les principaux freins à la diffusion ?* Puisque nous savons désormais que les EC sont plutôt réceptifs à l'idée de partager les données (sous conditions de protection juridique et technique), il faut comprendre quelles sont leurs craintes afin de proposer la bonne information, la bonne méthode et les rassurer.
La figure 18 nous permet d'observer que le manque de temps (44,67 %) est un facteur bloquant important, qui ne pourra être amélioré qu'en revisitant le cycle de travail autour de la donnée, les

[3] Les commentaires sont en annexes.

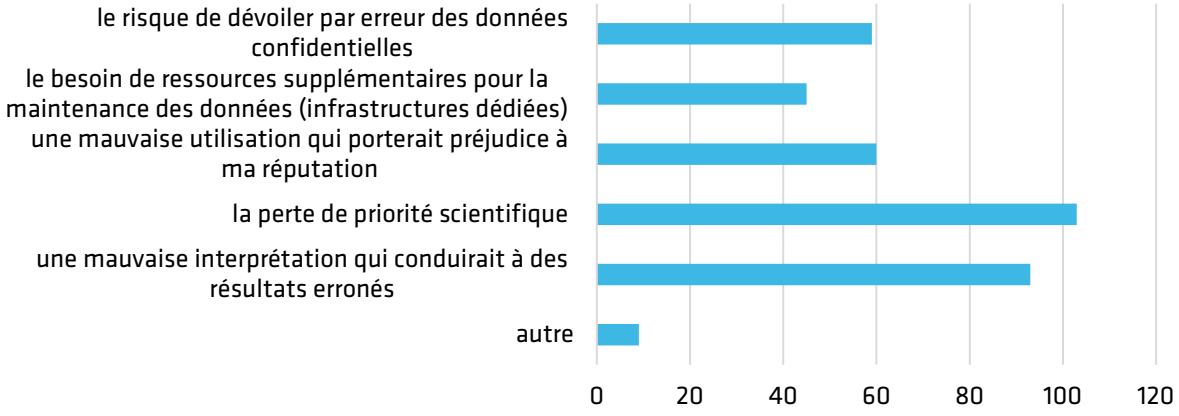
interactions, interventions et responsabilités étant à définir pour les étapes du projet. Comment constituer des équipes d'accompagnants ? Quels personnels ? Comment les intégrer au processus de recherche ? …

Fig 18. Quels seraient les principaux freins à la diffusion ?



De même la figure 20 présente les risques liés à l'ouverture des données, question qui a inspiré les répondants puisque 369 réponses ont été récoltées[4]. Plus de la moitié des réponses porte sur la perte de priorité scientifique et la mauvaise interprétation des données. Ces craintes sont légitimes, et il est important de rassurer la communauté en expliquant les conditions de partage, les méthodologies pour partager (description des jeux de données, documentation des données, …) pour éviter toute méprise. Cette question sert d'indicateur pour les thématiques de formation et/ou de communication.

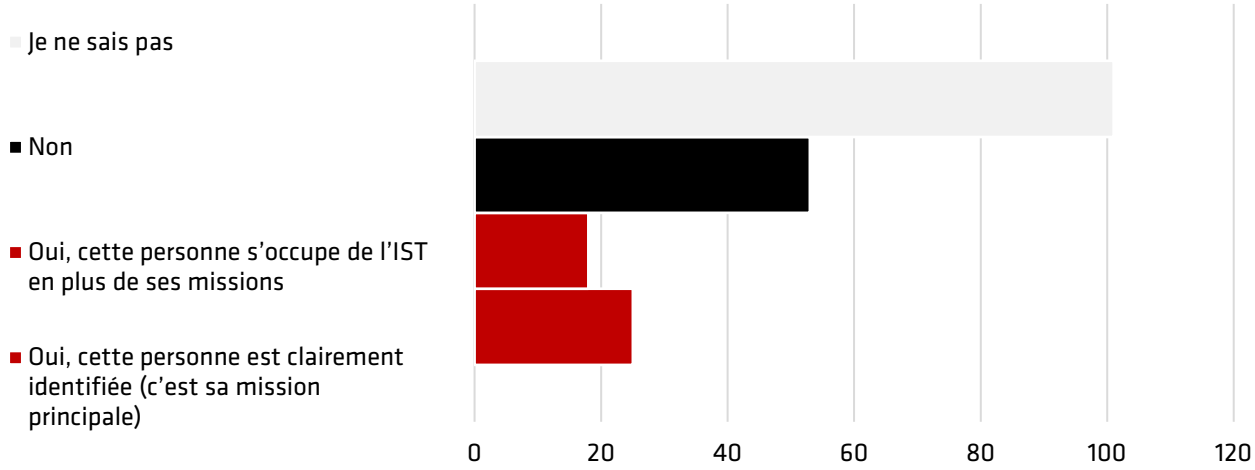
Fig. 19. Pour vous, quels sont les risques liés à la mise à disposition de vos données de recherche ?



Dès lors, nous nous demandons si nos 197 répondants ont un référent ou un collègue spécialiste de l'IST dans leur unité. Là encore, la moitié (51,27 %) l'ignore, mais il est encourageant de voir que 43 EC ont un référent IST. Reste à savoir si ce dernier est sollicité pour travailler sur les projets de recherche, à quel moment et dans quelles conditions (nous verrons ce point avec l'enquête IST[5])

[4] Les commentaires sont disponibles en annexe.
[5] Voir *infra* p 19.

Fig. 20. Avez-vous un référent IST dans votre unité ?



Dans l'optique de développer un accompagnement au sein des unités, il est indispensable de connaitre les ressources disponibles et/ou opérationnelles. Nous ne pouvons envisager de proposer une charte de bonnes pratiques à AMU si nous n'avons pas des relais techniques sur place. C'est pour cela qu'il était essentiel de solliciter la communauté des personnels IST qui sont confrontés à ces problématiques et qui rencontrent également des difficultés pour répondre aux nouvelles injonctions.

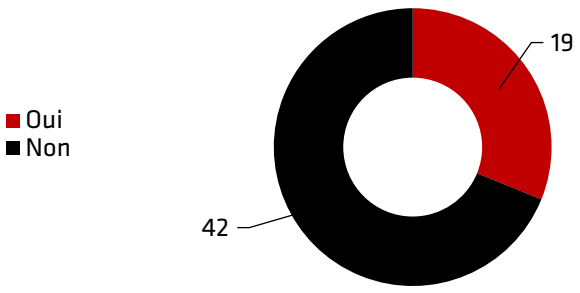
2. Personnels de l'IST

« L'information scientifique et technique (IST) regroupe l'ensemble des informations produites par la recherche et nécessaires à l'activité scientifique comme à l'industrie. De par sa nature, l'IST couvre tous les secteurs scientifiques et techniques et se présente sous de multiples formes : articles, revues et ouvrages scientifiques, spécifications techniques décrivant des processus de fabrication, documentation technique accompagnant les produits, ..., bases de données bibliographiques, banques de données brutes, archives ouvertes et entrepôts de données accessibles sur Internet, portails, etc ... »[6]

61 agents de l'IST ont répondu à l'enquête jusqu'au bout. L'objectif était d'en apprendre davantage sur leurs pratiques et leur implication dans les projets de recherche. Sont-ils concernés ? Sollicités ? Quand et par qui… ? Parviennent-ils à porter assistance aux EC, comment ?
Toutes ces réponses seront précieuses pour élaborer une offre de service leur permettant d'être d'une part notre relais au sein des unités, et d'autre part autonomes dans l'accompagnement des EC et doctorants.

En premier lieu, il était intéressant de savoir si les personnels IST faisaient déjà partie d'un un réseau, puisque notre objectif à terme est de constituer une communauté autour des questions de science ouverte.
Seulement 19 répondants (31,15 %) ont répondu « oui ».

Fig. 21. Êtes-vous inscrit·e dans un réseau professionnel ?

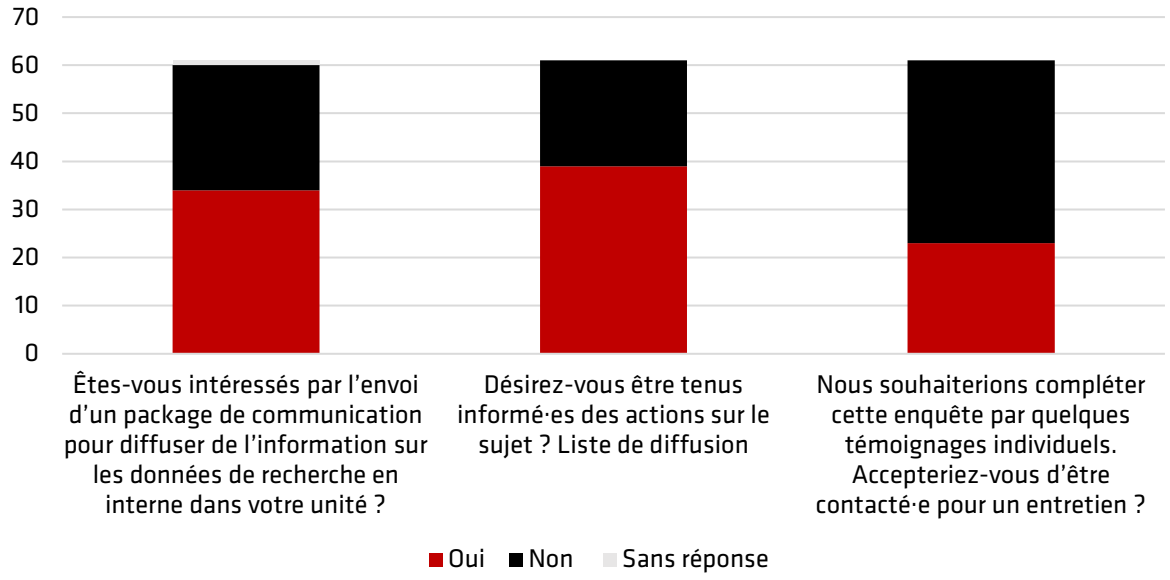


Cette pratique n'est donc pas commune (seuls les entretiens nous permettront d'en connaitre la raison), malgré cela 39 répondants sont intéressés par la liste de diffusion, et 34 par l'envoi d'un package de communication à diffuser dans leur unité. 23 personnes ont accepté d'être contactées pour l'entretien (**Fig. 22**). Il est plutôt positif de pouvoir toucher plus d'un répondant sur deux. En allant à la rencontre des personnels IST, on peut espérer toucher quelques EC. En rencontrant les EC, l'on peut

6 http://www.enseignementsup-recherche.gouv.fr/cid20438/les-missions-de-l-information-scientifique-et-technique.html

demander si un correspondant IST est présent dans leur unité, et si on peut entrer en contact avec cette personne ; cela permet d'élargir le réseau.

Fig. 22. Communication / Liste de diffusion / Entretien individuels



2.1. Gestion de projet

Nous devons désormais comprendre comment fonctionnent les unités et les interactions entre EC et personnels IST.
La question (à choix multiples) *Au sein de votre unité vous êtes* (**Fig. 23**), a recueilli 138 réponses. Les 10 réponses « autre » ne sont pas toute exploitables, certains répondants ayant décrit leur fonction[7]. L'objectif de cette question était de savoir comment se positionnent les répondants et comment ils sont identifiés sur ces questions. Ce diagramme est révélateur : il montre que d'une part, les agents sont sollicités, de par leur fonction et/ou leurs compétences, pour intervenir sur les projets (45,65 %) (rouge) mais que, d'autre part, tous ne sont pas officiellement reconnus et missionnés pour le faire. Certains interviennent à la demande (39,86 %), on peut imaginer en plus de leur travail quotidien, selon un planning qu'ils ne maîtrisent pas forcément (vert). Si l'on va plus loin, seulement 14,29 % des personnels IST reconnus pour leurs compétences en science ouverte sont officiellement positionnés sur ces questions.

Dès lors, il est important de savoir à quel moment du projet ces derniers sont sollicités. Nous comprenons toute la difficulté de se « greffer » à un projet ou seulement sur une question, sans en connaître le contexte ou en maîtriser ses enjeux, souvent entre deux portes dans un couloir. Nous voyons **figure 25** que la réponse **Je sais que je vais être sollicité·e mais je ne sais pas comment ni quand** remporte le plus de suffrage (27 occurrences). Il est encourageant de constater que certains répondants sont sollicités tôt sur les projets, même si 16 réponses indiquent que cela est fait **Dans l'urgence, quand il a fallu produire des livrables**.

[7] Les commentaires sont en annexe.

Fig. 23. Au sein de votre unité vous êtes :

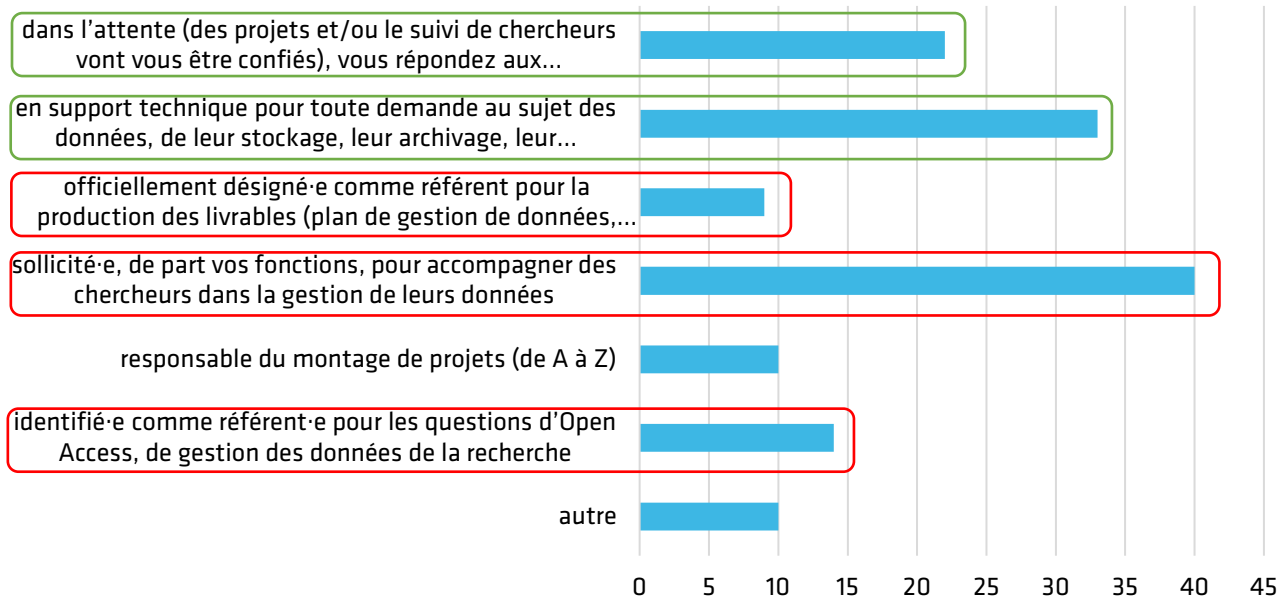
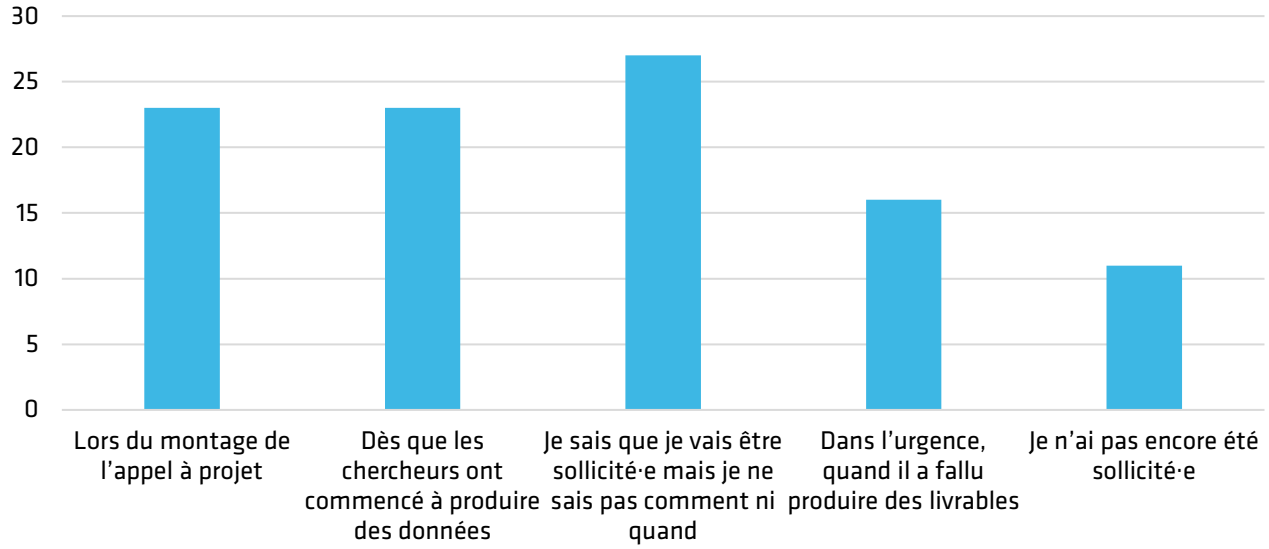


Fig. 24. À quel moment du projet êtes-vous généralement sollicité·e ?



À la question *Par qui ?* (**Fig. 25**), la réponse **Par ma direction** n'a que 19 entrées, ce qui nous semble révélateur des conditions dans lesquelles les professionnels de l'IST sont sollicités mais aussi de la manière dont le projet est organisé, pensé par les EC. La majorité des occurrences (64) montre que le travail collaboratif est le fait de demandes entre collègues, donc vraisemblablement hors cadre hiérarchique.

Fig. 25. Par qui ?

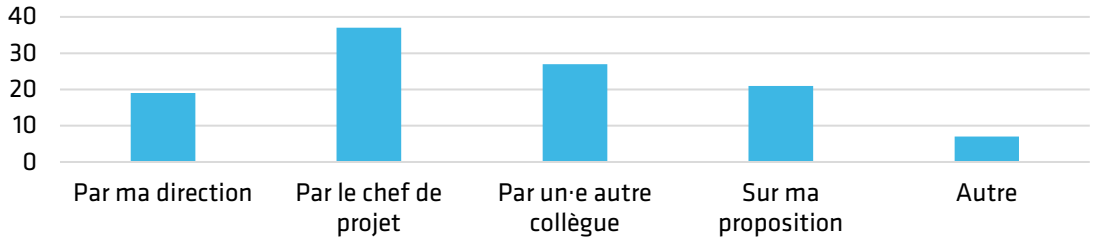
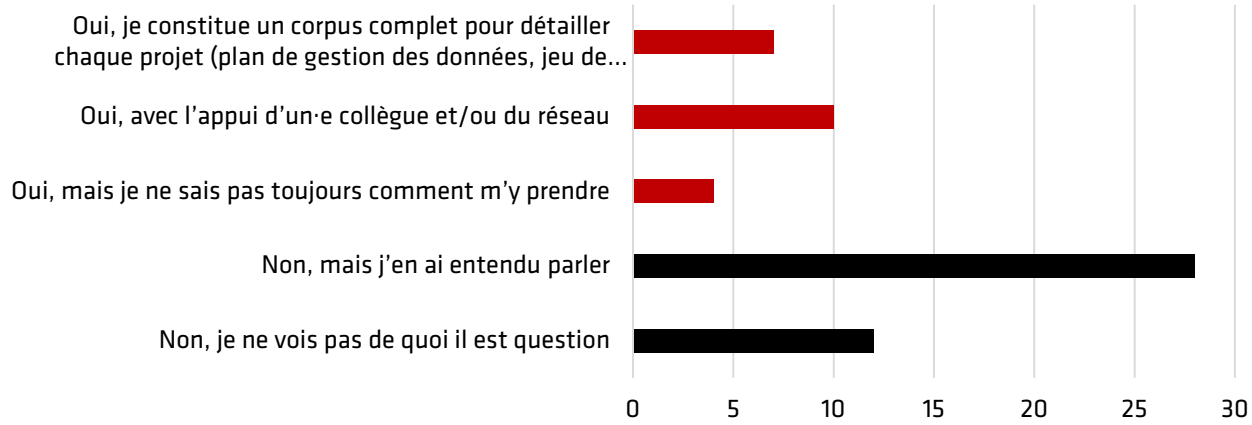


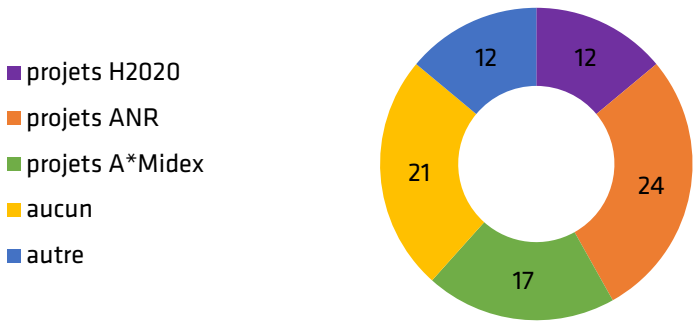
Fig. 26. Avez-vous déjà rédigé ou participé à la rédaction de documents détaillant les données produites et leur cycle de vie ?



Cette question permet d'avoir une idée des pratiques autour des livrables exigés par les financeurs. Presque 20 % des répondants ne connaissent pas ces documents, 45,90 % n'en rédigent pas mais en ont entendu parler. Force est de constater que rares sont ceux qui rédigent des plans de gestion des données ou constituent des « work package ». Ce qui semble cohérent avec les questions précédentes. Les agents de l'IST ne sont souvent sollicités que sur un point précis, et, de fait, seulement 34,43 % d'entre eux rédigent des livrables, non sans difficultés.

Enfin une question à titre indicatif sur les types de financements sur lesquels les personnels IST sont sollicités montre la place prépondérante des projets ANR et AMIDEX.

Fig. 27. Participation à des projets de recherche



Voici les projets renseignés par le champ « autre » :

Appel à projet ANRU
INCA
Tous types de financements
Financement régional ou privé
BPI France et DGA

Projets internationaux
FIR COLLOQUE
Projets et collaboration industriels
Projet GOLD GenOmic variability in heaLth & Disease
Projets financés par des organismes privés
Projet recherche avec Luxembourg et Belgique

2.2. Les données de la recherche

Si la grande majorité des personnels IST (81,97 %) savent ce qu'est une donnée de la recherche (**Fig. 28**), l'équilibre est parfait entre ceux qui connaissent des directives ou recommandations pour la gestion des données de la recherche et ceux qui n'en connaissent pas (**Fig. 29**). Il y a donc un travail de sensibilisation à faire pour impliquer la communauté IST, d'autant plus que 66,67 % d'entre eux indiquent s'informer de leur propre initiative (**Fig. 30**). L'intérêt est donc bien présent.
L'information circule donc peu, à l'instar de ce que nous avons constaté plus haut pour les EC. Peut-être faudrait-il élargir la communication, et en faire bénéficier toute la communauté scientifique (EC+IST) ? Il nous semble essentiel de toucher tous les acteurs possibles dans les projets de recherche, pour que le sujet des données devienne un réflexe, un acquis.

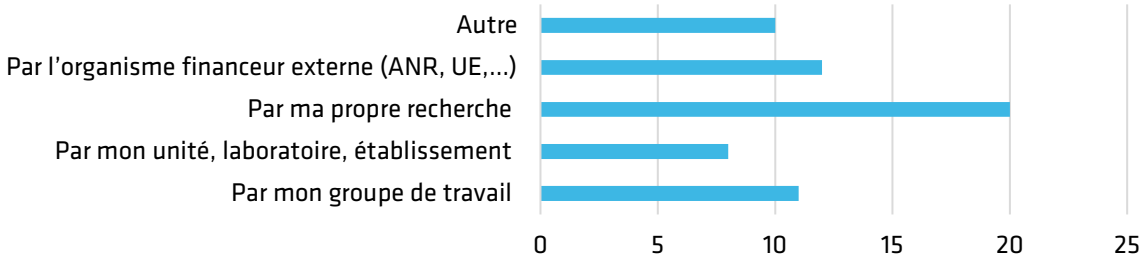
Fig. 28. Savez-vous ce qu'est une donnée de la recherche ?



Fig.29. Connaissez-vous des directives ou recommandations pour la gestion des données de la recherche ?



Fig. 30. Comment en avez-vous été informé·e ?



2.3. Stockage et archivage des données

59,02 % des répondants connaissent des entrepôts d'archives ou de données (**Fig. 31**). HAL est le plus connu, loin devant le Data Center AMU, Nakala ou encore ArXiv (**Fig. 32**). Connaître n'est pas utiliser, mais c'est cependant un bon début !

Fig. 31. Connaissez-vous des entrepôts d'archives et/ou de données ?

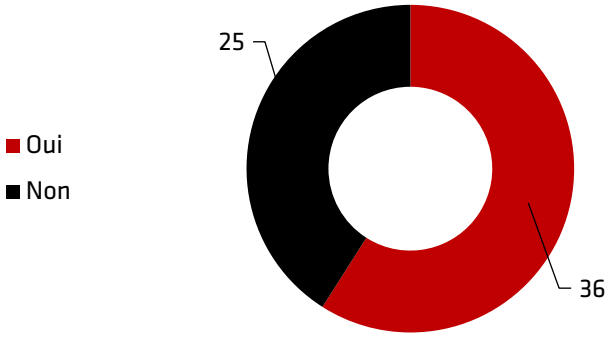
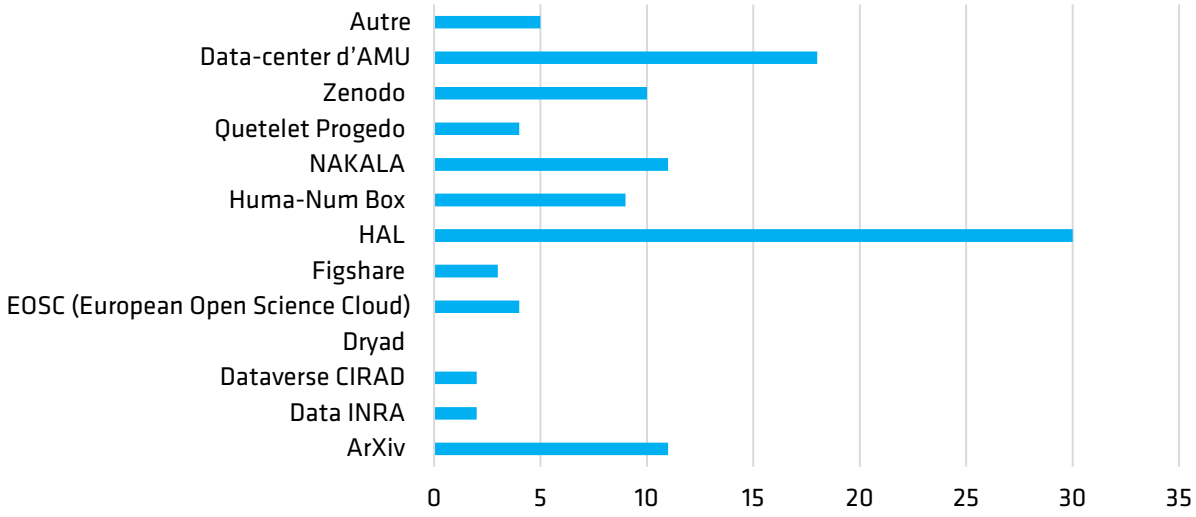


Fig. 32. Lesquels ?



Parmi les autres entrepôts librement cités se trouvent : European Pollen Database soutenue par l'OSU Pytheas, OpenNeuro, Dropbox, CINES, ESA. L'on voit avec « Dropbox » qu'il y a encore de la sensibilisation à faire…

La question suivante concernait la politique d'archivage des données. Comme pour les EC[8], la majorité des répondants indique « non » ou « je ne sais pas ». Les 18 « oui » ont pu répondre à la question suivante qui demandait *qui est responsable du stockage, de la sauvegarde et de l'archivage des données ?* (**Fig. 34**) L'un des répondant commente :

[8] Voir *supra*, p 13.

« La politique n'est hélas pas pleinement mise en œuvre, faute de moyens humains et matériels. »

Cela signifie que des initiatives existent, mais qu'elles ne peuvent s'appliquer. Les pratiques des personnels IST sont similaires à celles des EC, les services informatiques étant davantage sollicités par les EC. La gestion des données relève d'une pratique interne, propre à chacun en fonction des contraintes.

Fig. 33. Y a-t-il une politique d'archivage des données de recherche dans votre laboratoire ?

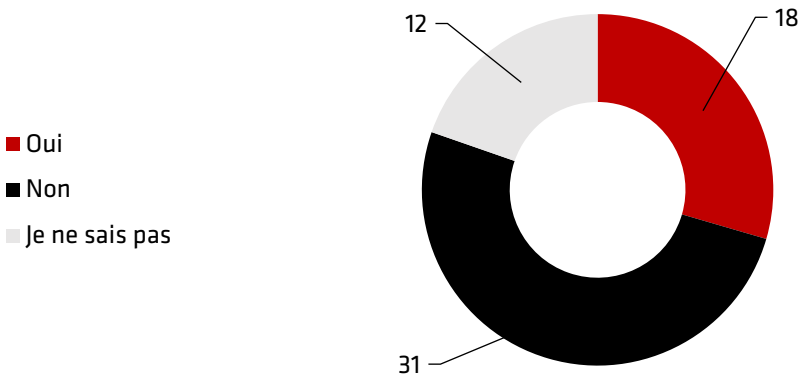
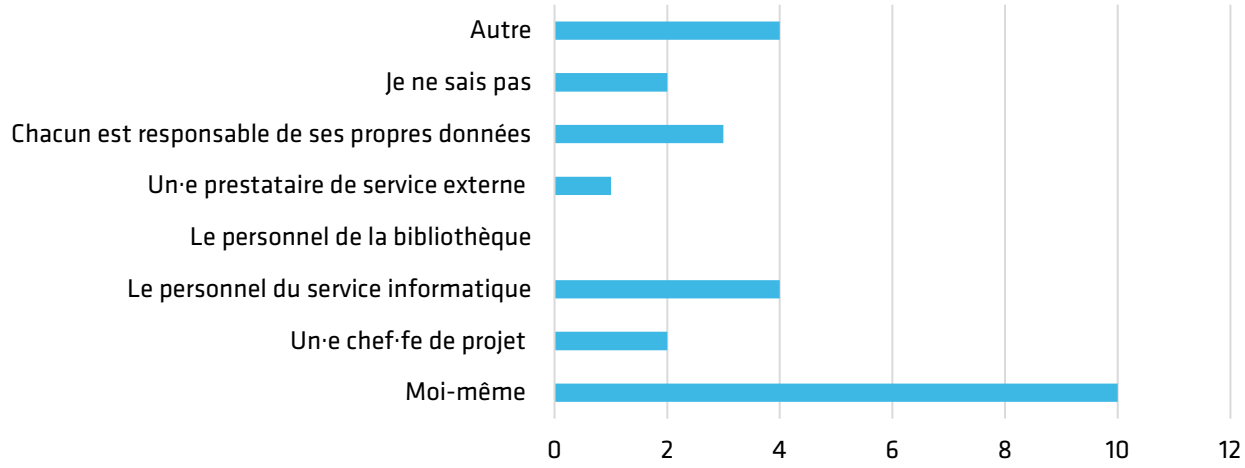


Fig. 34. Si oui, qui est responsable du stockage, de la sauvegarde et de l'archivage des données ?



3. Besoins et attentes

Cette dernière partie est essentielle car elle va nous permettre de faire des recommandations pour répondre aux besoins de la communauté scientifique d'AMU (EC et IST). Nous pourrons ainsi prioriser les actions et vérifier si notre plan d'actions, jusqu'à présent, est en phase avec les attentes.

Il est intéressant d'analyser les réponses des deux formulaires en même temps. Nous verrons que les attentes des uns reflètent les besoins des autres… Nous pouvons considérer cette partie comme un miroir EC/IST.
Les personnels IST ont cependant eu à répondre à une question supplémentaire, pour cibler les obstacles rencontrés dans leur pratique quotidienne.

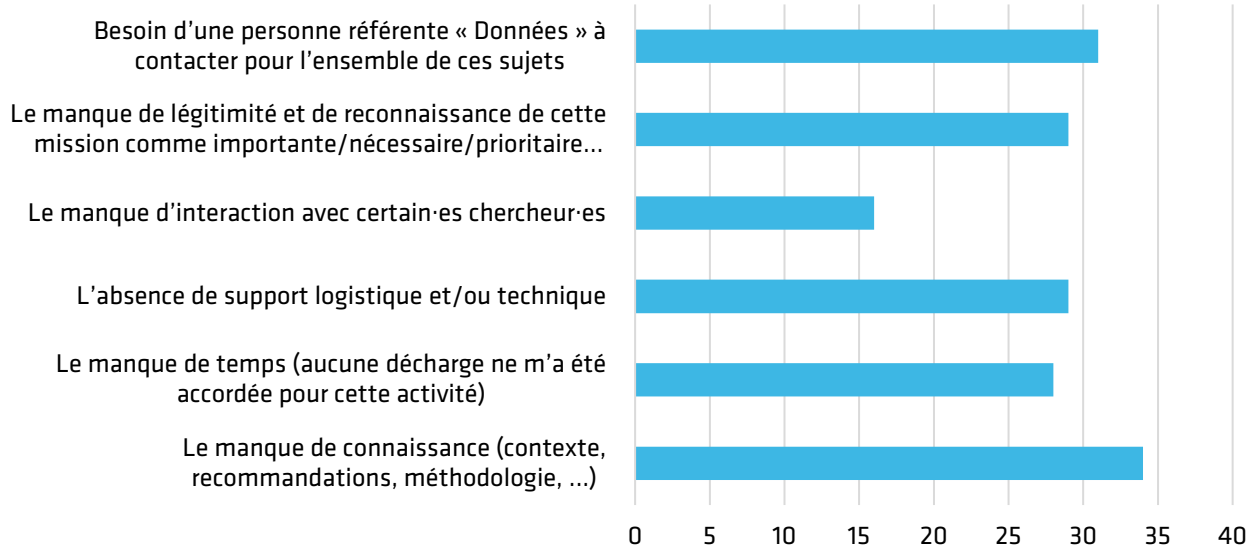
3.1. Les principaux freins rencontrés par les personnels IST

L'objectif de cette question était de comprendre les contraintes des personnels IST devant cette nouvelle activité. Comment les impliquer dans la science ouverte, comment faire entrer ces nouvelles tâches dans leur quotidien ?
Les réponses ne facilitent pas vraiment la prise de décision puisque qu'elles sont très équilibrées. Le manque de communication avec les chercheurs n'est pas autant bloquant que nous l'aurions imaginé. En revanche, toutes les entrées proposées ont chacune un fort taux de réponse. Cette question a particulièrement intéressé les participants puisque 167 réponses ont été recueillies. Parmi celles-ci, les freins les plus importants sont le manque de connaissances (20,36 %) et le besoin d'une personne référente pour cette problématique (18,56 %). Ces deux items sont au cœur de nos réflexions et des pistes vont prochainement être proposées.

En revanche, trois réponses sont à prendre en compte dans l'optique de développer une culture d'établissement. Trois points essentiels qu'il convient d'étudier et d'approfondir car ils représentent à eux trois 51,50 % des réponses (Fig. 36) :

1. Le manque de temps ;
2. L'absence de support logistique et/ou technique ;
3. Le manque de légitimité et de reconnaissance de cette mission comme importante, nécessaire, prioritaire au sein des laboratoires.

Fig. 35. Quels sont les principaux freins rencontrés dans votre quotidien en ce qui concerne les données de recherche ?



Le projet « données de la recherche » est un projet commun, inter-services, inter-métiers… et il semble compliqué de le conduire sans la combinaison de toutes les compétences.

Si AMU souhaite développer de bonnes pratiques et une culture de la science ouverte, il convient de réfléchir à comment impliquer et intégrer tous les acteurs au cycle de la recherche. Les EC dans leur métier de chercheur, les personnels IST dans leur métier de support à la recherche, en se donnant les moyens d'aller au bout du processus.

La question que doit désormais se poser l'établissement est « qui » pour « faire quoi » et « comment ». Le défi est de taille, car les moyens et les ressources ne sont pas extensibles. Il convient donc de réfléchir à comment répondre aux exigences de la science ouverte en positionnant les personnels compétents sur des missions d'accompagnement, de conseil, de support de la communauté des enseignants-chercheurs, afin de créer un cycle de travail efficace autour de la recherche et des données qui y sont produites.

3.2. Sur quels points souhaiteriez-vous bénéficier d'une aide ?

(de votre unité de recherche, de la DOSI, du SCD, de l'URFIST…)

À cette question, nous proposions les six réponses suivantes[9] :

1. **Conseils d'ordre général**
    a. pour les projets de recherche qui demandent la production de livrables (personnels IST)
    b. pour les données de la recherche (DR, RPGD…) (EC)
2. **Conseils d'ordre juridique :** restrictions d'accès, traitement des données sensibles, utilisation de licences…
3. **Formation sur la question des données de recherche :** type, gestion, diffusion…

[9] Variante selon les formulaires : **a** > pour les IST ; **b** > pour les EC.

4. **Assistance sur des questions techniques :** métadonnées, normes, archivage à long terme, publication des données…
5. **Préparation d'un plan de gestion de données et/ou d'un jeu de données**
6. **Utilisation d'un espace d'archivage sécurisé et fiable**
    a. pour les données de recherche dans les unités (personnels IST)
    b. pour les données de recherche des EC

Nous mettons en miroir les réponses des deux enquêtes pour cerner les priorités (**Fig. 36 / 37**).

Fig. 36. Réponses des IST

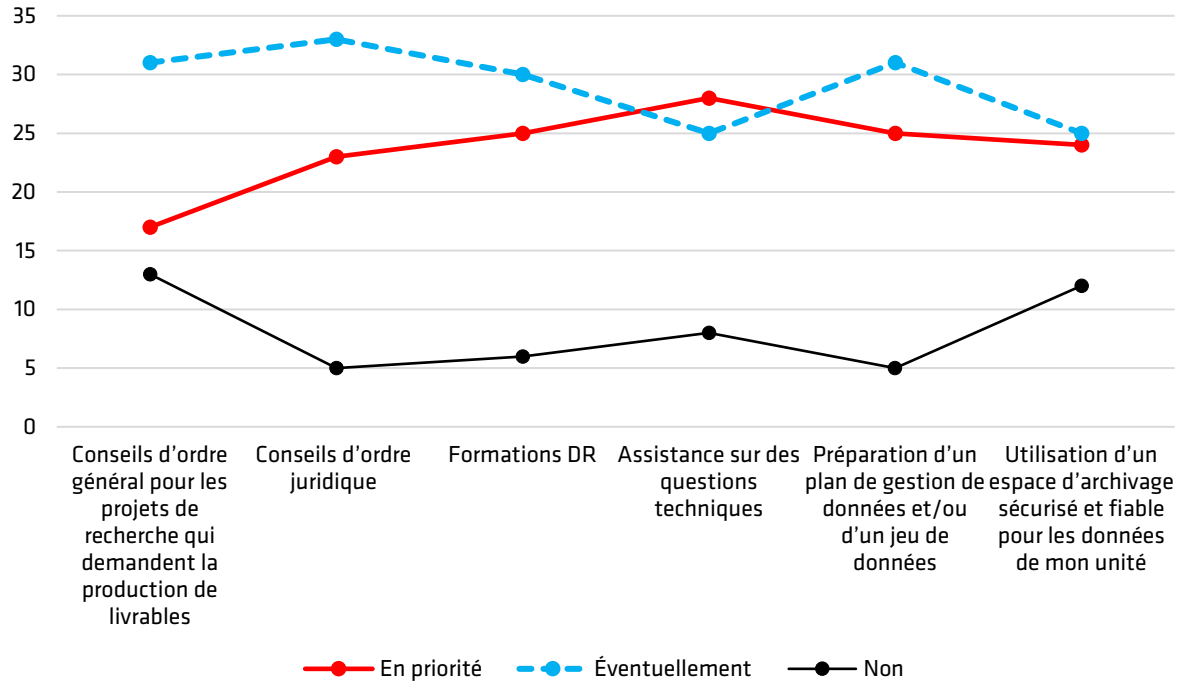
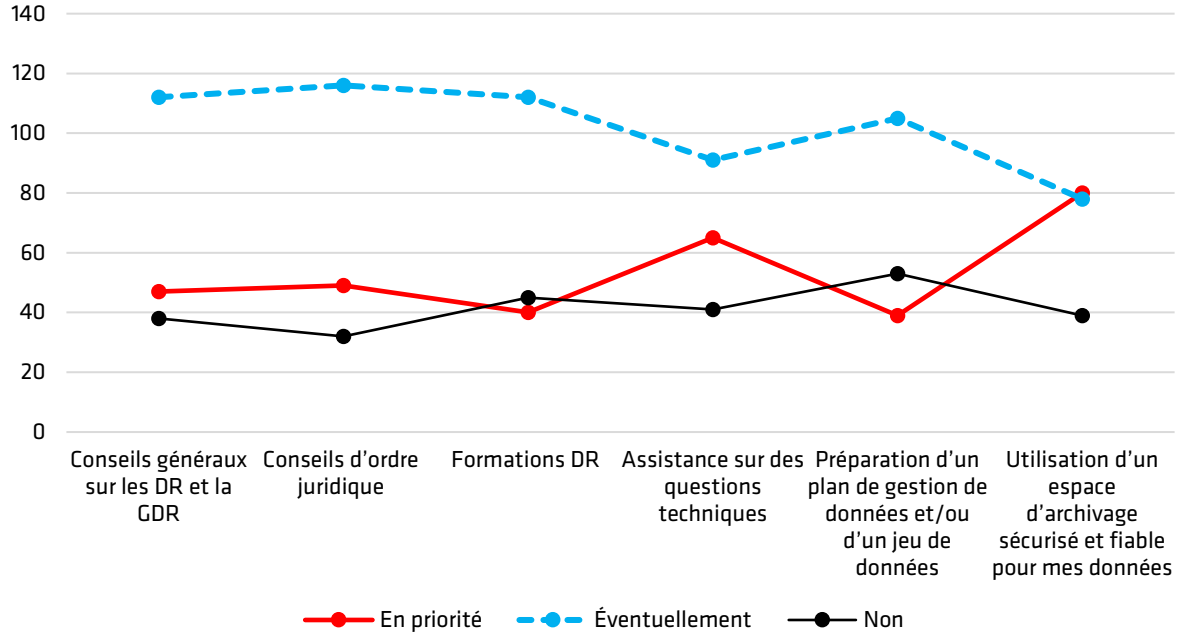


Fig. 37. Réponses des EC



Les répondants du questionnaire IST sont majoritairement intéressés par de l'assistance technique (45,90 %). Ce qui semble cohérent avec le fait qu'ils sont souvent sollicités sur des questions techniques dans les projets (cf. figure 23). De même, 40,98 % d'entre eux indiquent avoir besoin d'aide sur l'utilisation d'entrepôts, les réponses correspondent donc à un besoin concret d'assistance. Ainsi ils sont nombreux à être éventuellement intéressés - comprendre par-là « selon les attentes/demandes de ma communauté » -, par tous les axes d'aide proposés, avec une moyenne de 40 % à 50 % de réponses.

Les commentaires laissés suite à cette question confirment ce besoin d'accompagnement et de référents pour se mettre au point méthodologiquement et techniquement[10] :

« Nous aurions besoin de spécialistes dans la gestion du cycle de vie de la donnée qui viennent dans nos laboratoires pour travailler avec nous sur la mise au point de méthodologies adaptées à nos problématiques. »

~

« Data Center organisé sur le campus de Luminy pour rationaliser les investissements. »

~

« Besoin d'un contact avec le CINES et/ou d'un référent pour de l'archivage pérenne. »

~

« Avec les données de neurosciences, nous sommes confrontés plus spécifiquement aux problématiques sur les données sensibles (données de santé : IRM, génétique) en plus des données personnelles. Jusqu'à présent tous les interlocuteurs que j'ai eu ont traité les données personnelles, mais pas les données de santé. »

Les EC quant à eux sont soucieux de l'archivage de leurs données. 40,61 % des répondants souhaitent en priorité être aidés dans l'utilisation d'entrepôts. Cependant, l'écart entre les réponses « en priorité » et « éventuellement » est bien plus important que pour les personnels IST, et indique un intérêt moindre pour de la formation, peut-être parce qu'ils attendent davantage de soutien RH de la part de l'institution, plutôt qu'une formation leur permettant d'être autonomes. C'est en tout cas ce qui semble se dégager des réponses ouvertes :

« Il faut mesurer le temps que prend l'action de réellement partager des données. Il est considérable (anonymisation, vérification, nettoyage, formatage), c'est le frein considérable des personnes produisant des données pour le recherche et souhaitant les partager. Ce n'est pas avec un ingénieur « données » par labo que les données vont affluer (malgré toutes les bonnes volontés du monde). »

~

« La principale question est celle des moyens alloués à cette nouvelle tâche. S'il s'agit de passer de nombreuses heures à préparer un partage de données, de

[10] L'ensemble des commentaires est en annexe.

nombreuses heures à le maintenir, de nombreuses heures à répondre aux questions des éventuels utilisateurs… c'est catastrophique en termes de temps passé. »

~

« Qu'il y ait une infrastructure dédiée pour cela. Nous avons déjà de nombreuses tâches de secrétariat ou de gestion qui s'ajoutent à notre travail « normal », qui fagocitent le temps disponible pour la recherche et l'enseignement, et qui pèsent très négativement sur notre santé, sans parler du fait que nous devons financer sur fonds propres nos déplacements et nos outils de travail. En ajouter ne me paraît ni souhaitable, ni pertinent, bien au contraire. Quel que soit l'intérêt que cela puisse présenter. »

~

« Je pense que ça serait bien si AMU/CNRS fournissaient une infrastructure informatique pour le stockage des données qui soutiennent les article publiés (comme HAL mais pour les données). »

~

« Sans espaces de stockages massif mis à disposition rien de cela ne sera possible. »

~

« Ce serait bien qu'en plus de conseils et de formations collectives permettant officiellement de tout faire nous-même, on ait au niveau d'AMU une personne ressource qui passe dans les laboratoires nous aider personnellement et matériellement à le faire, au moins pour les premières fois, sur des données issues de nos propres recherches (qui ont souvent des caractéristiques propres spécifiques. »

~

« Je pense qu'il faut déjà développer la « culture de la donnée », nous sommes nombreux à ne pas avoir été formé, ce qu'on peut voir déjà dans nos façons d'archiver nos données. À mon avis, la réflexion sur un DMP permet de se poser de bonnes questions sur les métadonnées et l'utilité du partage. Mais l'étape de la mise en œuvre est encore difficile après, surtout la définition du périmètre des métadonnées. »

Perspectives

Quelle suite donner à cette enquête ?

Première étude du genre à Aix-Marseille Université, cette enquête entre dans le mouvement global d'ouverture dans lequel s'est engagé l'établissement (ouverture des publications avec HAL AMU et maintenant des données). Elle lance officiellement une série de démarches qui, nous l'espérons, contribuera à l'émergence d'une culture des données à l'université, avec :

- une structuration des actions ;
- des acteurs repérés autour du cycle de vie de la recherche ;
- l'adhésion au mouvement et aux principes de la science ouverte.

Cette démarche projet s'inscrit dans la durée, même si les injonctions sont là et que les délais pour y répondre sont très serrés. Nous devons coordonner plusieurs services et interlocuteurs, développer une offre de services, répondre aux urgences et prioriser un programme global auprès des doctorants d'une part, et des EC et des professionnels de l'IST d'autre part.

L'enquête est un outil précieux de communication. Elle a permis d'informer la communauté AMU de l'existence du projet, et de la démarche de formalisation de l'accompagnement et de la formation aux questions liées aux données de recherche.

L'enquête, comme nous nous en doutions, dévoile un panorama riche et très variable en terme de gestion des données. Il n'existe pas UNE bonne méthode qui conviendrait à toutes les situations, mais nous pouvons aider à construire, à AMU, une série de bonnes pratiques, et à mettre en place des pratiques FAIR, gages de succès auprès des financeurs et des instances nationales et européennes.
Cette « boite à outils de la science ouverte », nous devons la penser en fonction des verrous que nous aurons à percer (motivation, compétences, ressources informatique – humaine - financière), mais aussi – et surtout ! – en nous appuyant sur les bonnes volontés et les initiatives existantes.

Comme indiqué au début de ce document, presque un répondant sur deux (49,22 %) accepte de s'inscrire à la liste de diffusion. Malgré l'anonymat de l'enquête, 67 répondants ont donné leurs coordonnées pour participer à un entretien (25,97 %). Ces rencontres complémentaires sont primordiales pour impulser des collaborations, échanges, changements…
Nous sommes pour l'instant dans une première phase qui doit inciter la communauté à entrer dans le mouvement. Les EC sont certes obligés de s'y mettre (injonctions des financeurs), mais l'émulation peut entrainer des volontés, convaincre les plus réticents et faire avancer le projet.

Pour conclure, cette enquête consacre notre approche et valide notre feuille de route, ainsi que les actions que nous menons depuis plusieurs mois. Les deux formulaires confirment que nous sommes en accord avec les attentes de la communauté et que nous devons continuer notre état des lieux, notre réflexion prospective auprès de la communauté scientifique d'AMU.

CC BY