



Rapport sur l'enquête sur les usages et les  
besoins pour la gestion des données de la  
recherche sur le site de l'Université  
Grenoble Alpes

10 mai 2022

## Préambule

Ce travail s'inscrit dans la convergence de grands enjeux liés, d'une part à l'augmentation exponentielle des données dans tous les domaines scientifiques, et d'autre part à l'amplification du mouvement d'ouverture de la science.

Ces changements modifient profondément les processus de recherche, et les méthodologies de travail des communautés scientifiques.

En particulier, la bonne gestion des données de la recherche devient une nécessité pour faire face à ces défis. Les bénéfices associés sont nombreux puisqu'elle facilite le travail de recherche, individuel et collectif, sécurise les fruits de la recherche, et simplifie leur diffusion.

Elle implique cependant une appropriation d'outils (par exemple des plateformes de stockage) et de concepts (comme les plans de gestion de données) dont la méconnaissance, ou la complexité peuvent être des freins à leur utilisation.

L'objectif de cette enquête est de réaliser un état des lieux des pratiques autour des données de la recherche auprès de toutes les communautés scientifiques du site de Grenoble de façon à adapter l'accompagnement dans cette dynamique de changement extrêmement rapide.

## Ce qu'il faut retenir

Les participants à cette enquête sont bien représentatifs de la population recherche du site, tant en terme de discipline scientifique, de fonction ou d'expérience professionnelle.

Cette photographie du paysage grenoblois sur les données de la recherche permet de mettre en lumière des différences disciplinaires.

D'une façon générale, les Plans de Gestion de Données, documents qui décrivent la façon dont les données seront produites ou obtenues, (ré)utilisées, traitées, organisées, stockées, sécurisées, préservées, documentées, partagées, au cours et à l'issue d'un projet de recherche, sont assez méconnus.

Compte tenu du fait qu'ils sont désormais obligatoires pour tout projet financé, il est clair que cette tendance va rapidement évoluer mais ce sont aussi des outils importants pour la bonne gestion des données que les communautés doivent s'approprier en tant que tel.

Par ailleurs, sur les habitudes de stockage, les volumes de données en jeu influent sur les pratiques, et certaines communautés (par exemple la physique des particules, l'astrophysique, les géosciences, l'environnement et l'écologie) sont mieux structurées que d'autres sur ce sujet. De même, le degré de complexité, tant règlementaire que technique, lié au travail sur des données sensibles, impacte plus certaines communautés (sciences humaines et sociales, santé) que d'autres.

De manière générale, la réutilisation de données existantes n'est pas simple, car bien souvent les principes FAIR c'est-à-dire trouvables, accessibles, interopérables et réutilisables, ne sont pas complètement respectés.

Enfin, l'action de diffuser ses données comme un produit de recherche à part entière n'est pas encore bien intégré dans le processus de recherche.

Ces constats encouragent la poursuite des efforts réalisés en ce qui concerne l'information autour des outils et services déjà existants. Ils confortent la nécessité de renforcer l'accompagnement qui doit être au coeur des dispositifs de soutien.

Les actions de formation doivent être consolidés et développés, à tous les niveaux.

Enfin, il est essentiel d'aller vers une simplification de tout le dispositif pour une meilleure compréhension et une bonne adhésion de toutes les communautés scientifiques, quel que soit leur niveau de maturité dans le domaine de la gestion des données de la recherche.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Profils des participants</b>	<b>6</b>
2.1	Pôles de recherche de l'Université Grenoble Alpes . . . . .	6
2.2	Analyse de la représentativité des participants . . . . .	7
<b>3</b>	<b>Description des données en jeu et problématique du stockage</b>	<b>9</b>
3.1	Types des données . . . . .	9
3.2	Volumes des données et stockage . . . . .	11
3.3	Synthèse . . . . .	21
<b>4</b>	<b>Sujet des données sensibles</b>	<b>23</b>
4.1	Analyse des réponses . . . . .	23
4.2	Synthèse . . . . .	25
<b>5</b>	<b>Pratiques collaboratives</b>	<b>27</b>
5.1	Habitudes de collaborations . . . . .	27
5.2	Outils pour le travail collaboratif . . . . .	29
5.3	Synthèse . . . . .	32
<b>6</b>	<b>Réutilisation et diffusion des données</b>	<b>34</b>
6.1	Réutilisation des données . . . . .	34
6.2	Difficultés à la réutilisation des données . . . . .	36
6.3	Pratiques de diffusion . . . . .	36
6.4	Profils des déposants . . . . .	38
6.5	Diffusion et volumétrie . . . . .	39
6.6	Mode de diffusion . . . . .	40
6.7	Aide à la diffusion . . . . .	41
6.8	Synthèse . . . . .	41
<b>7</b>	<b>Archivage</b>	<b>42</b>
7.1	Analyse des réponses . . . . .	42
7.2	Synthèse . . . . .	46
<b>8</b>	<b>Plans de Gestion des Données</b>	<b>48</b>
8.1	Niveau d'appropriation des Plans de Gestion de Données . . . . .	48
8.2	Formations, expérience de rédaction et difficultés rencontrées . . . . .	50
8.3	Synthèse . . . . .	51
<b>9</b>	<b>Personnes ressources pour l'aide à la gestion des données</b>	<b>52</b>
9.1	Analyse . . . . .	52
9.2	Synthèse . . . . .	53
<b>10</b>	<b>Besoins exprimés</b>	<b>55</b>
10.1	Besoins sur l'accompagnement et la formation . . . . .	55
10.2	Besoin sur les services numériques . . . . .	56
10.3	Besoins sur les entrepôts de données . . . . .	56
10.4	Synthèse . . . . .	57

<b>11 Focus doctorants</b>	<b>59</b>
11.1 Profils . . . . .	59
11.2 Type de données . . . . .	59
11.3 Pratiques durant la recherche . . . . .	61
11.4 Personnes ressources . . . . .	62
11.5 Stockage . . . . .	62
11.6 Pratiques collaboratives . . . . .	64
11.7 Besoins exprimés . . . . .	64
11.8 Diffusion des données . . . . .	65
11.9 Archivage . . . . .	65
11.10 Synthèse . . . . .	65
<b>12 Conclusions, perspectives et recommandations</b>	<b>66</b>
12.1 État des lieux . . . . .	66
12.2 Sensibilisations et information . . . . .	66
12.3 Formations et accompagnements à prévoir . . . . .	66
12.4 Nouveaux outils ou services à prévoir . . . . .	67
12.5 Mise en Perspective . . . . .	67
<b>A Questions de l'enquête</b>	<b>68</b>

# 1 Introduction

L'Université Grenoble Alpes (UGA) est une université pluridisciplinaire (Etablissement Public Expérimental) avec :

- 59 000 étudiants, dont 10 000 étudiants internationaux et 2 900 doctorants.
- 3 100 enseignants et chercheurs.
- 3 700 personnels administratifs et techniques.
- 4 établissements : l'UGA ; Grenoble INP, Institut d'ingénierie et de management - UGA, Sciences Po Grenoble - UGA et l'École nationale d'architecture de Grenoble (ENSAG) – UGA.
- 120 projets européens, 13 labex et 11 equipex. 30 brevets déposés par an en moyenne.
- Depuis 2007, elle héberge 124 lauréats ERC affiliés à l'UGA et aux organismes et institutions partenaires.
- 100 laboratoires organisés en 6 pôles de recherche.
- 13 Ecoles Doctorales.

L'UGA a des partenariats forts au niveau local avec le CHU Grenoble Alpes, au niveau national avec le CNRS, le CEA, l'INSERM, l'INRAE, INRIA et l'IRD, et au niveau européen avec les grands instruments présents sur le site : l'ESRF, l'ILL et l'EMBL.

Devant l'importante croissance des problématiques liées aux données de la recherche pour l'ensemble des communautés scientifiques du site, l'Université Grenoble Alpes a structuré depuis bientôt 3 ans son offre d'accompagnement grâce à la Cellule Data Grenoble Alpes (CDGA).

Ce questionnaire, proposé par la CDGA, vise à connaître les pratiques et besoins des chercheurs et ingénieurs de l'UGA en matière de gestion des données de la recherche.

Il a été diffusé à l'ensemble des communautés scientifiques sur une période allant du 2 novembre 2020 au 13 décembre 2020.

Ce questionnaire a été réalisé sur une instance locale de Limesurvey en collaboration avec des étudiants de M2 MDLE information et documentation que nous remercions chaleureusement, ainsi que leurs enseignants Mathieu Fusi et Aude Inaudi. L'ensemble des questions est accessible en annexe, page A.

Il a fait l'objet d'un total de 877 réponses, dont 421 réponses complètes et 456 réponses incomplètes. Parmi ces dernières, 157 sont exploitables, ce qui donne un total de 578 réponses analysées au moyen de scripts Python et R.

Le pré-traitement des données a consisté à supprimer les réponses vides ou dont seules les questions personnelles (questions Q1 à Q4) avaient été remplies et qui ne fournissaient donc pas d'information en matière de gestion des données. Le nom des questions a été également raccourci afin de faciliter la lecture des graphiques, le nom complet des questions figure en annexe. Le pôle "Autre" (question Q1) et le statut "Autre" (question Q3) ont été, autant que possible, reformulés dans une des catégories proposées en se basant sur les réponses personnelles (questions Q1 à Q4) afin de pouvoir exploiter ces réponses.

Les graphiques des questions et des croisements de questions ont été réalisés en se basant uniquement sur les réponses non vides des questions concernées. Les données sont également normalisées lors de la création des graphiques, cette normalisation est réalisée suivant la variable affichée en légende. Par exemple pour la figure 1, le statut des répondants est affiché suivant le pôle d'appartenance : les répondants ayant indiqués le statut "Chercheur" ont également

indiqué le pôle “CBS” dans 31% des cas, le pôle MSTIC dans 13% des cas, etc. De même, pour la figure 2, les répondants du pôle CBS ont entre 0 et 5 ans d’ancienneté pour 20% d’entre eux, entre 5 et 10 ans d’ancienneté pour 16% d’entre eux, et enfin plus de 10 d’ancienneté pour 63% d’entre eux. Certains pourcentages ne somment pas à 100% du fait des arrondis effectués lors de l’affichage des graphiques (pour des raisons de lisibilité, les pourcentages sont donnés sans chiffre après la virgule).

## 2 Profils des participants

L'objectif de cette enquête était de réaliser un état des lieux des usages et besoins des communautés scientifique sur la question de la gestion des données de recherche.

Il était donc essentiel de nous assurer que nous avions une représentation suffisante de l'ensemble des pôles de recherche, mais également de la diversité des métiers de la recherche présents dans les laboratoires.

### 2.1 Pôles de recherche de l'Université Grenoble Alpes

La recherche à l'Université Grenoble Alpes est organisée en six pôles regroupant les unités et de structures de recherche relevant de disciplines et de cultures proches, avec des objets et/ou des méthodes de recherche similaires.

**Pôle « Chimie, Biologie, Santé », CBS** Ce pôle de recherche regroupe les laboratoires qui mènent des recherches dans les domaines de la biologie, de la chimie orientée sciences du vivant, du médical et de la pharmacie. Ses recherches couvrent tout le spectre de la recherche fondamentale à la recherche appliquée et à la recherche clinique. Il s'appuie sur une collaboration forte avec le CHU Grenoble Alpes et l'Établissement français du sang.

**Pôle « Mathématiques, sciences et technologies de l'information et de la communication », MSTIC** Ce pôle concerne les activités de recherche en mathématiques et dans les sciences et technologies de l'information et de la communication (informatique, automatique, traitement du signal). Ses recherches vont du fondamental aux applications.

**Pôle « Physique, ingénierie, matériaux », PEM** Ce pôle mène des recherches dans les domaines de la physique et des nanosciences, de la mécanique, de l'électronique, du génie électrique, de la science de matériaux, du génie des procédés et des sciences de l'ingénieur et de la production. Ces travaux ont des implications dans des domaines à fort enjeu sociétal, comme la santé, la transition énergétique, la durabilité, le renouveau industriel et la société de l'information et de la communication.

**Pôle « Physique des particules, astrophysique, géosciences, environnement et écologie », PAGE** Ce pôle mène des recherches couvrant les domaines de la physique des particules et noyaux, de l'astrophysique, de la Terre, de l'environnement et de l'écologie. Elles contribuent au développement et à l'exploitation de grands instruments, des moyens d'observation sur le long terme et d'infrastructures de recherche, dans le cadre d'une programmation nationale et internationale.

**Pôle « Sciences humaines et sociales », SHS** Ce pôle a vocation à étudier l'Homme dans les contextes sociaux, sociétaux, culturels, interculturels et environnementaux. Les problématiques concernant le langage, la cognition, la création, l'éducation, les techniques en société, sous tous leurs aspects, sont centrales dans les recherches menées au sein de ce pôle.

**Pôle « Sciences sociales », PSS** Ce pôle mène des recherches couvrant tout le champ disciplinaire des sciences sociales, en mobilisant un large éventail de méthodes (observation, enquête, expérimentation, modélisation, analyse comparative, herméneutique, classification...) qui témoignent de la richesse et du dynamisme des sciences sociales.

Dans toute la suite de ce document, les pôles de recherche seront indiqués par leur acronyme pour des questions de lisibilité. Une description plus détaillée peut être trouvée sur le site de l'université.

## 2.2 Analyse de la représentativité des participants

La figure 1 montre que les répondants sont bien représentatifs de la population de recherche de l'UGA, avec une présence des doctorants plus importante dans le pôle SHS. On retrouve également une proportion plus notable d'enseignants-chercheurs que de chercheurs dans des pôles comme MSTIC, PSS et SHS. On peut également noter le faible nombre de post-doctorants parmi les participants. Ils sont majoritairement dans le pôle PAGE.

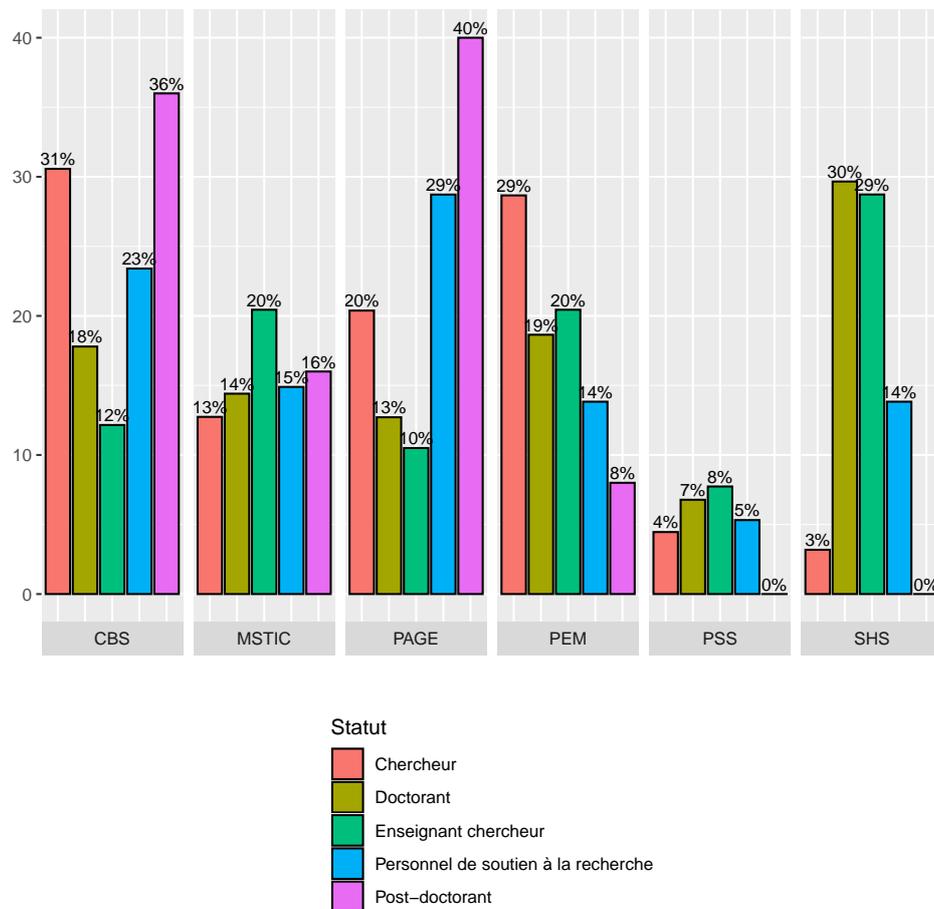


FIGURE 1 – Pôle de recherche par statut des répondants

Si on croise l'expérience professionnelle avec le pôle de recherche (figure 2), on constate une majorité de répondants expérimentés en recherche.

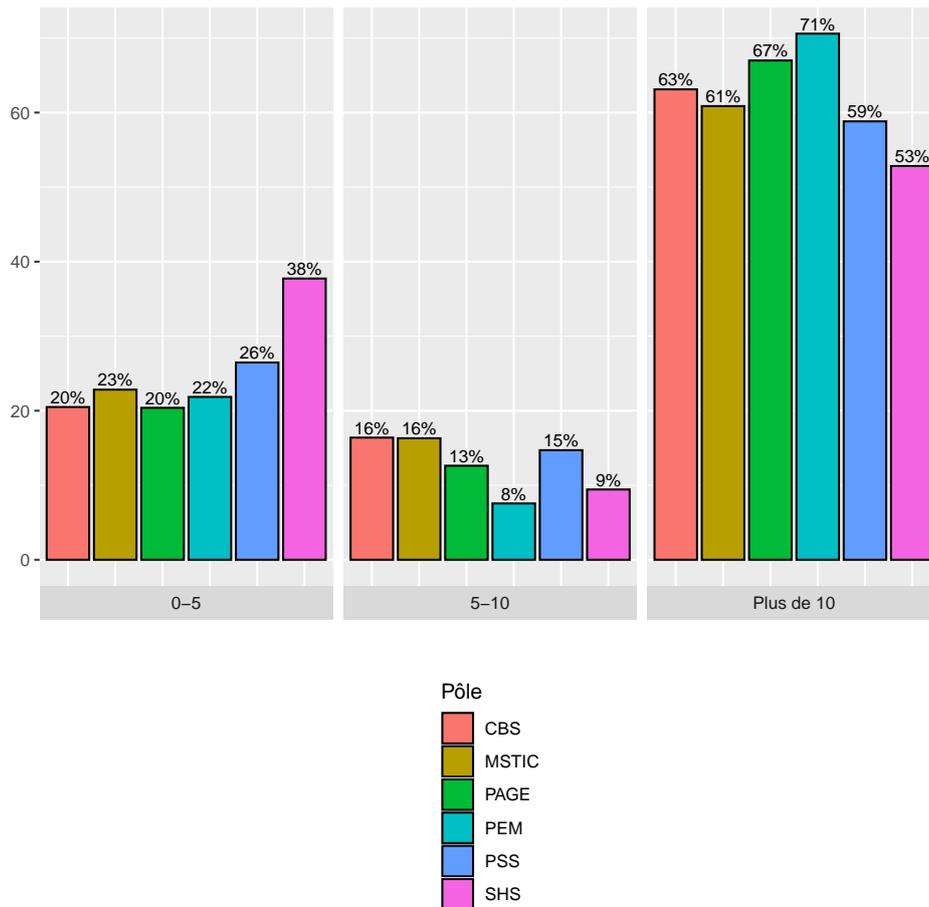


FIGURE 2 – Expérience en recherche par pôle de recherche

La population des participants à l'enquête couvre bien l'ensemble des disciplines présentes sur le site de l'UGA, ainsi qu'une diversité suffisante de profils différents en terme de fonction et d'expérience professionnelle qui vont permettre de faire des analyses croisées intéressantes pour comprendre les spécificités communautaires, et les usages et problématiques qui pourraient caractériser plus particulièrement les jeunes chercheurs.

## 3 Description des données en jeu et problématique du stockage

La finalité de cette partie est la bonne compréhension des données qui sont manipulées par les communautés scientifiques du site, ainsi que des pratiques en matière de stockage.

### 3.1 Types des données

Les réponses permettent d'avoir une cartographie assez complète des différents types de données manipulés par les communautés scientifiques du site de Grenoble Alpes, et des spécificités disciplinaires.

Il faut d'abord noter que les termes proposés ont peut-être une interprétation un peu différente d'une communauté à l'autre mais nous retrouvons, à travers l'analyse des figures 3 et 4, les grandes tendances attendues :

- l'importance des données textuelles pour le pôle SHS,
- les données des réseaux sociaux majoritairement dans le pôle PSS,
- la présence des codes sources dans les pôles MSTIC, PAGE et PEM,
- de façon cohérente, les données de simulation se retrouvent plutôt également dans les pôles PEM, MSTIC et PAGE,
- enfin, les workflows sont très présents dans les pôles CBS et PAGE.

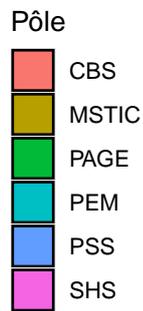
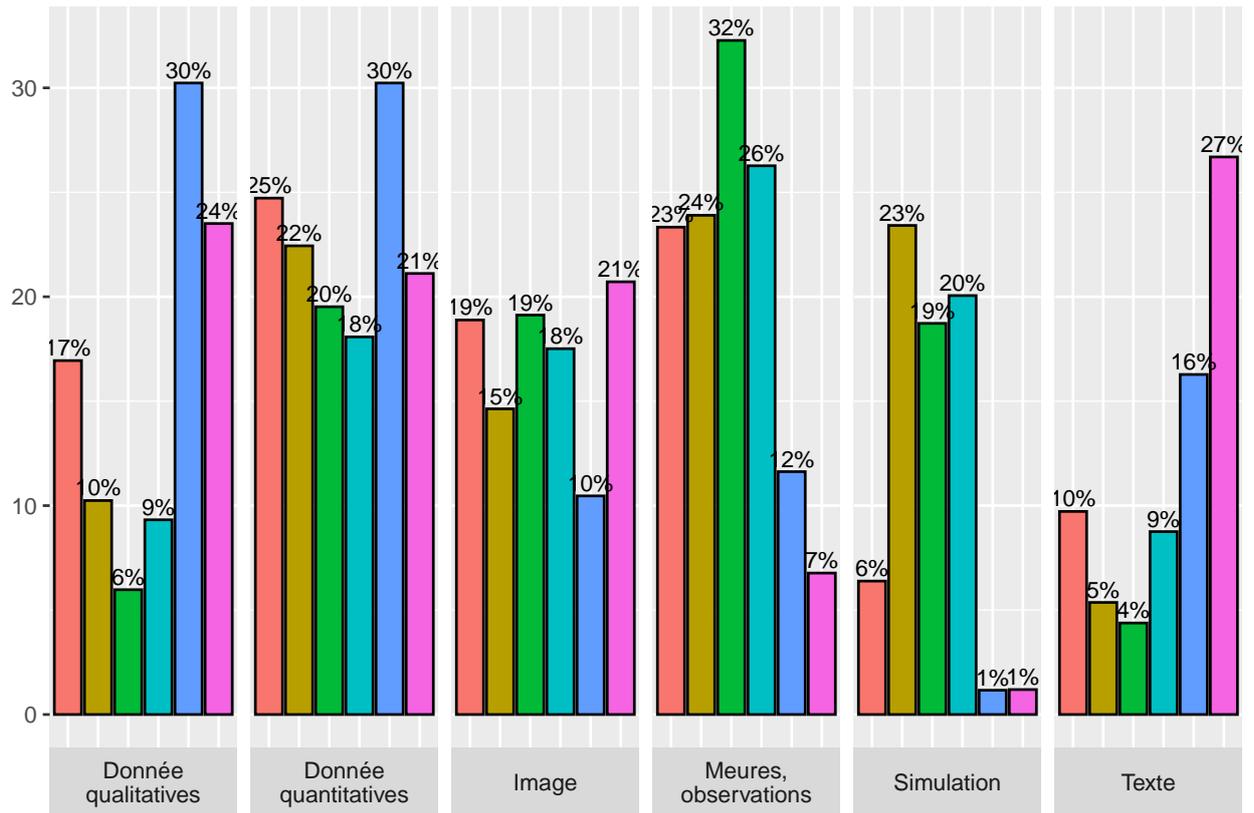


FIGURE 3 – Types de données par pôles de recherche (partie 1/2)

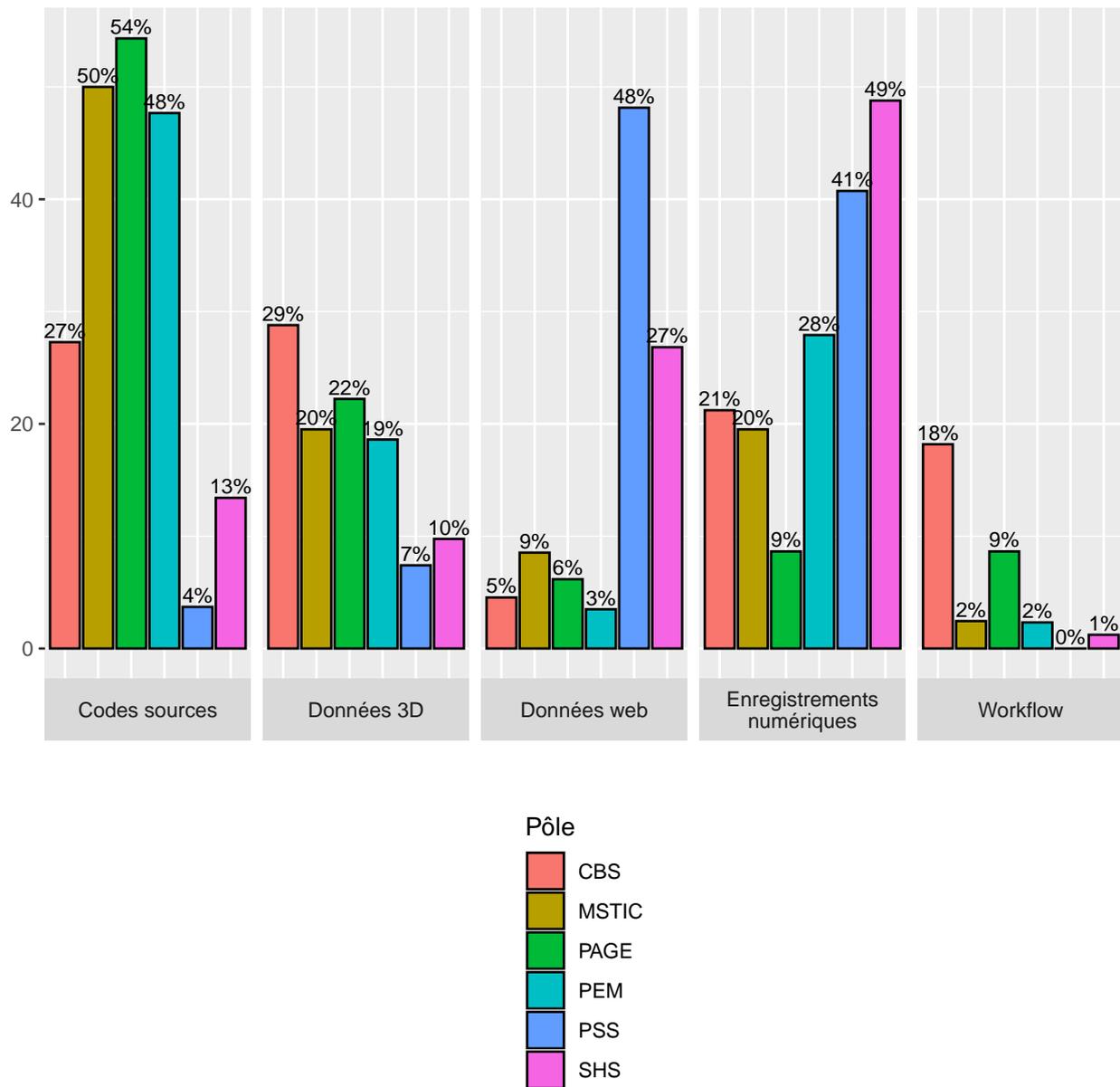


FIGURE 4 – Types de données par pôles de recherche (partie 2/2)

### 3.2 Volumes des données et stockage

Si on met en regard les volumes de données concernés avec l'appartenance aux pôles de recherche (figure 5), il existe des différences notables mais attendues entre les disciplines :

- On retrouve de gros volumes de données dans les pôles PAGE, CBS et dans une moindre mesure PEM.
- Il faut noter que les personnes qui n'ont pas réellement d'idée de la volumétrie de leurs données appartiennent majoritairement au pôle SHS.
- La grande majorité des répondants se positionnent sur des volumes de l'ordre de 1 à quelques centaines de Go.

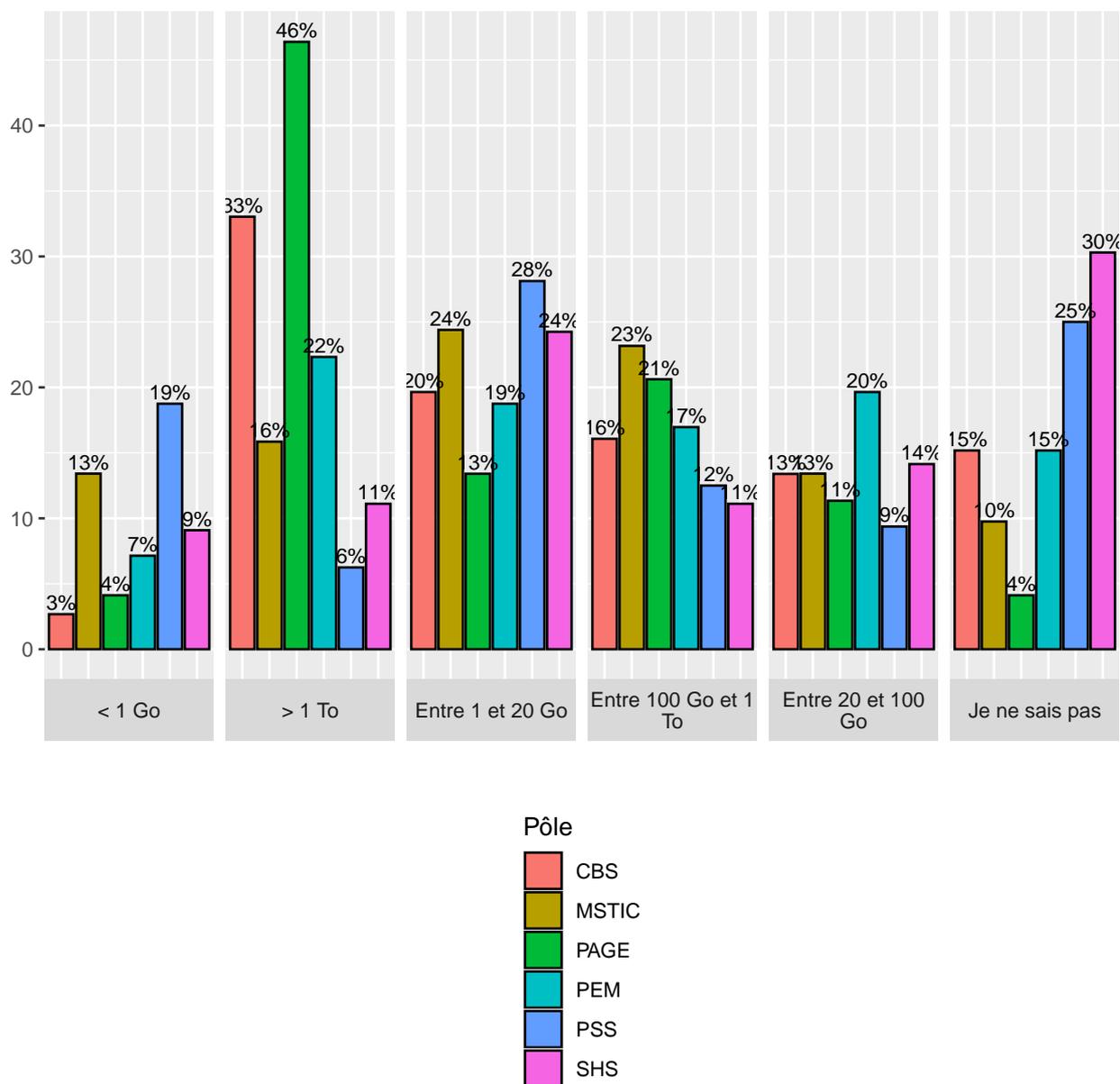


FIGURE 5 – Volumétries des données par pôles de recherche

Ces réponses interrogent nécessairement sur la façon dont sont stockées ces données. Si on croise le pôle d'appartenance avec les habitudes de stockage (figures 6 et 7), plusieurs éléments ressortent de l'analyse :

- L'utilisation de plateformes institutionnels, en particulier au plus près de là où sont produites les données se retrouvent beaucoup dans le pôle PAGE.
- L'utilisation de matériels personnels et de clouds privés est très majoritaire dans le pôle SHS, et dans le pôle PSS pour la partie clouds privés.
- L'utilisation de disques durs externes professionnelles semblent une pratique très ancrée dans certains pôles comme MSTIC ou PEM.
- L'utilisation de plateforme d'établissement ou de serveurs de laboratoire est assez développée dans les pôle CBS et PAGE.

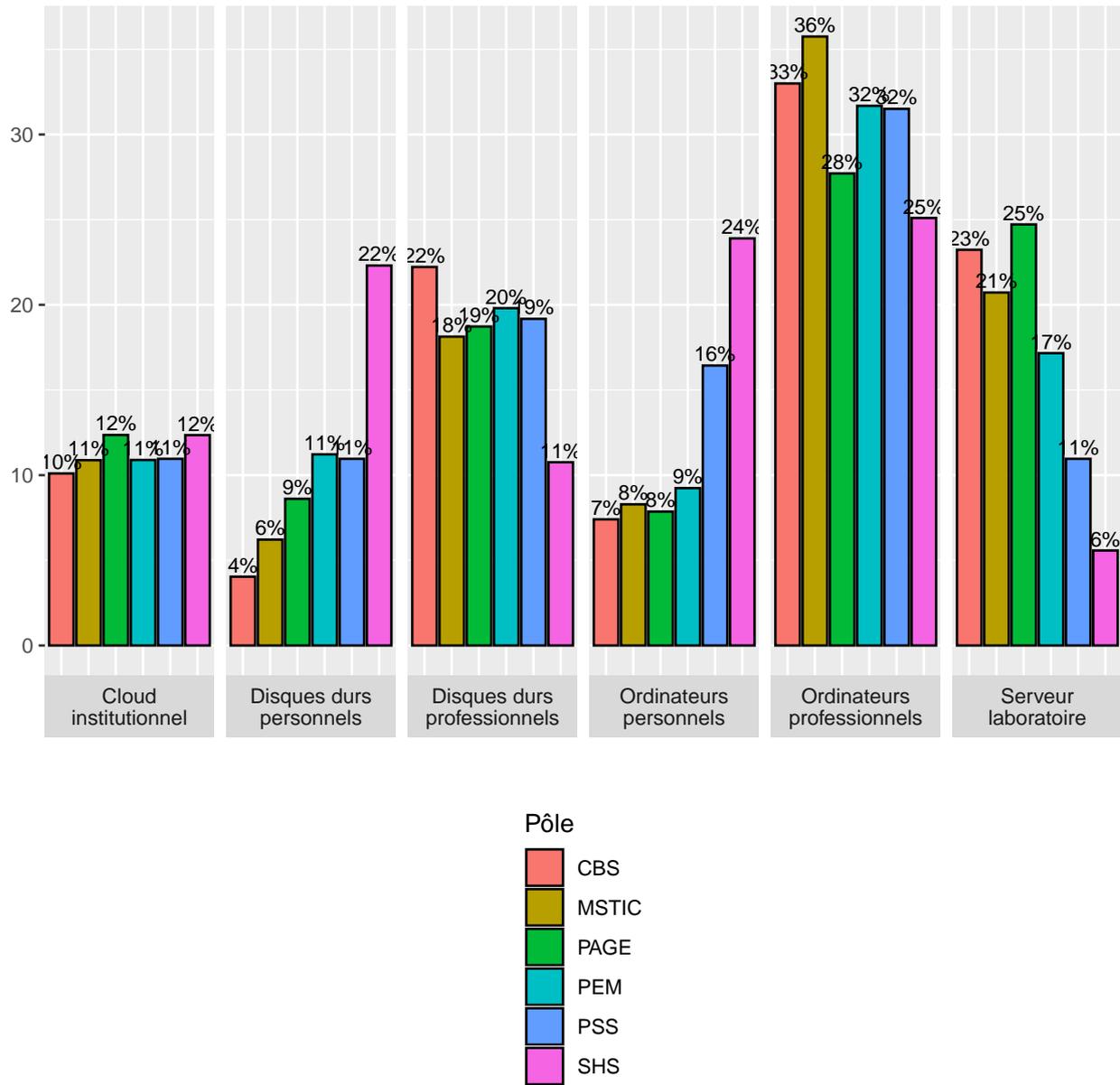


FIGURE 6 – Lieux de stockage des données par pôles de recherche (partie 1/2)

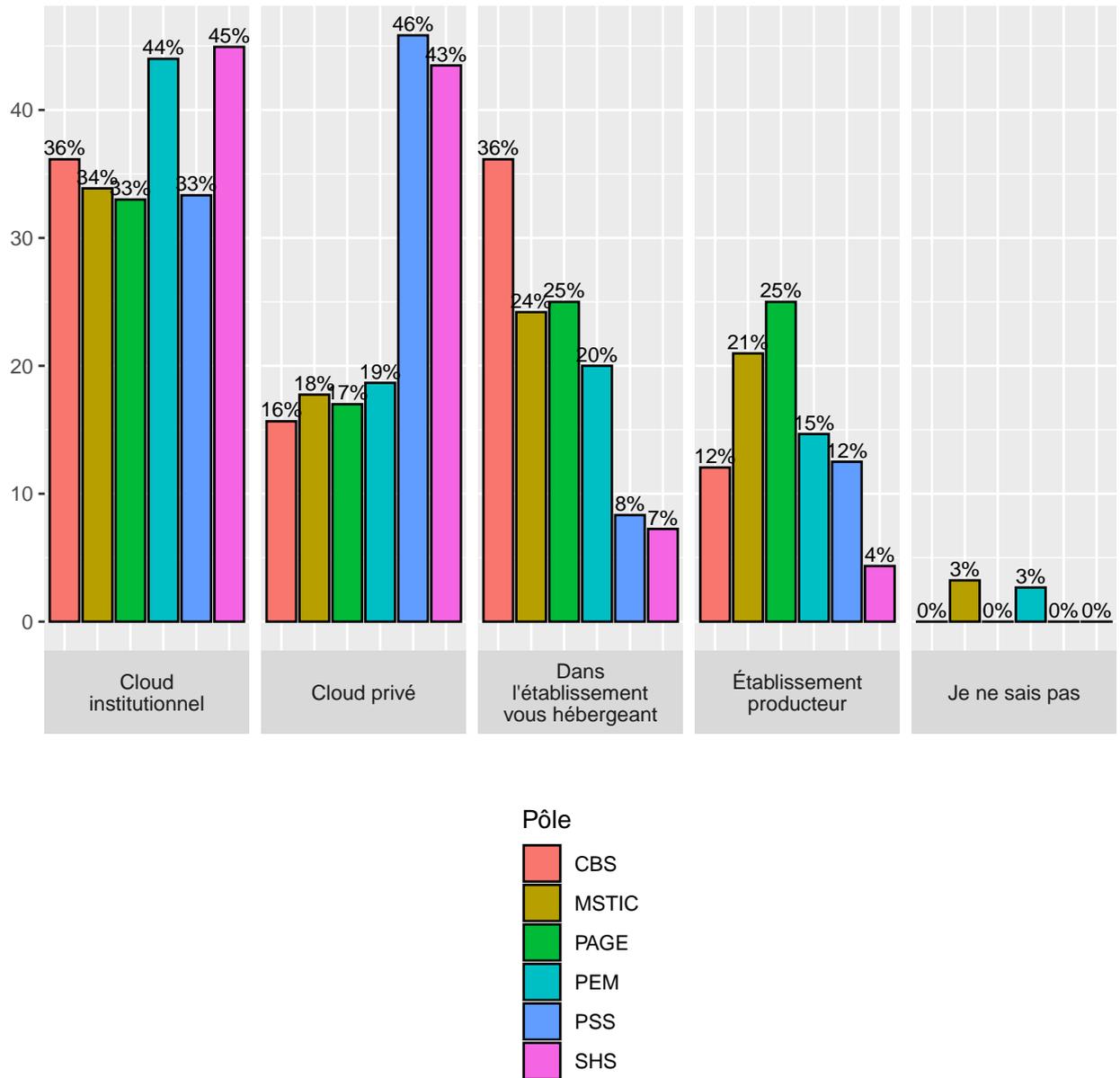


FIGURE 7 – Lieux de stockage des données par pôles de recherche (partie 2/2)

Il est aussi intéressant de croiser les choix pour le stockage avec le niveau d'expérience (figures 8 et 9).

En particulier, on constate que pour les jeunes scientifiques (en particulier les doctorants), l'utilisation de matériel personnel est majoritaire, de même que les outils de type clouds privés.

L'utilisation de serveurs de laboratoires ou de plateformes institutionnelles est plutôt du fait de personnes plus expérimentées, même si les plus jeunes utilisent aussi les serveurs de laboratoires.

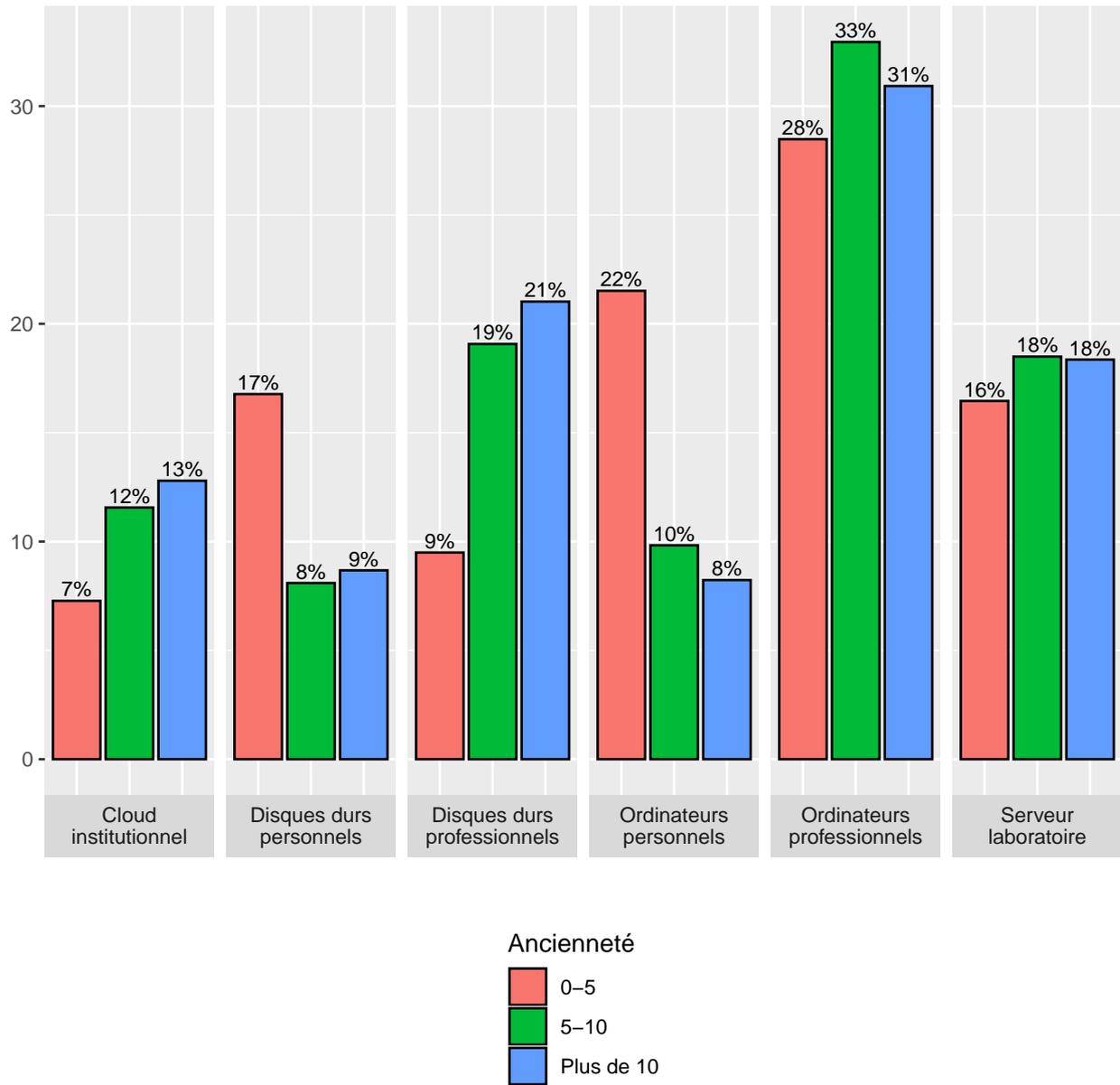


FIGURE 8 – Lieux de stockage des données par ancienneté des répondants (partie 1/2)

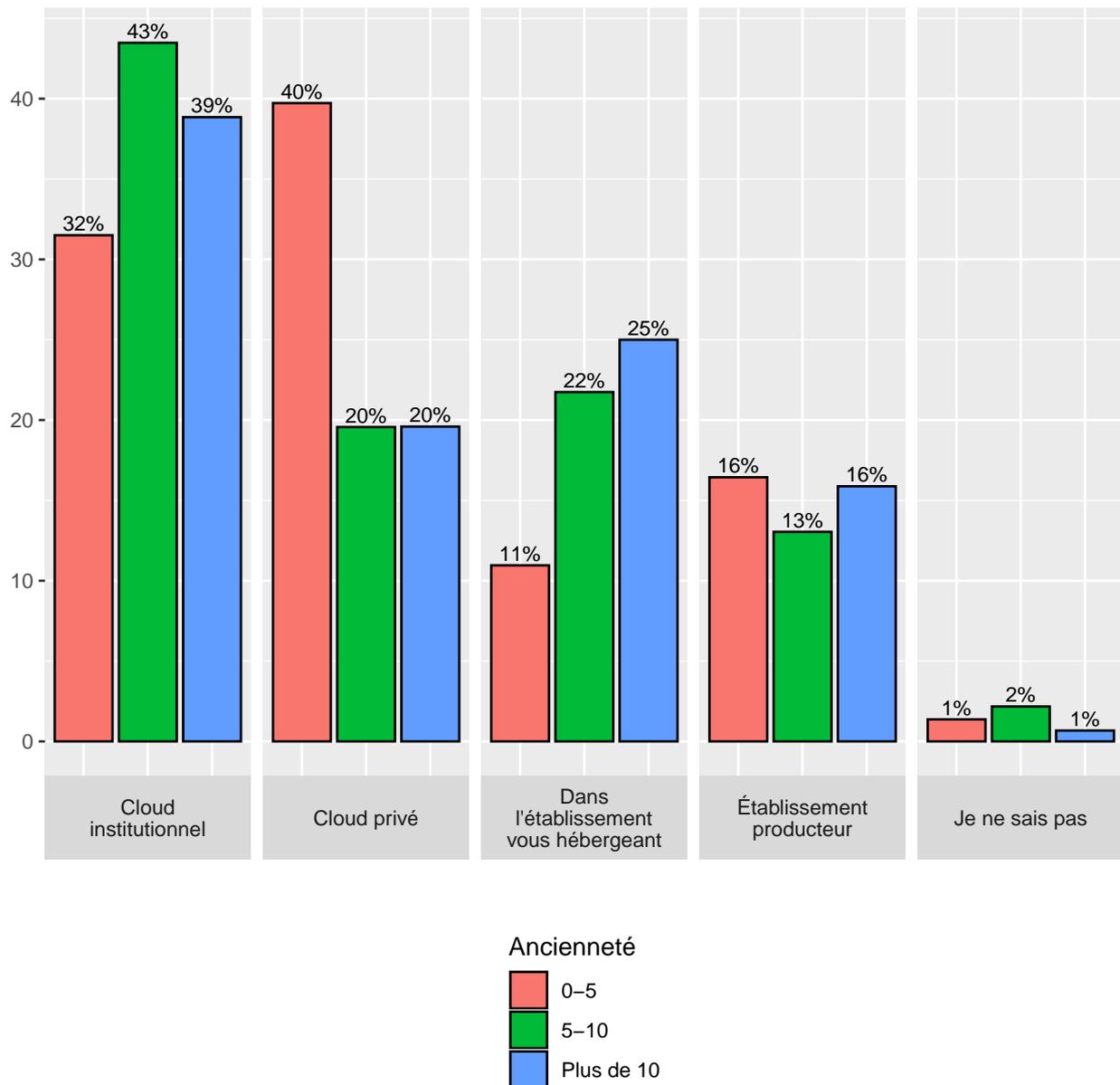
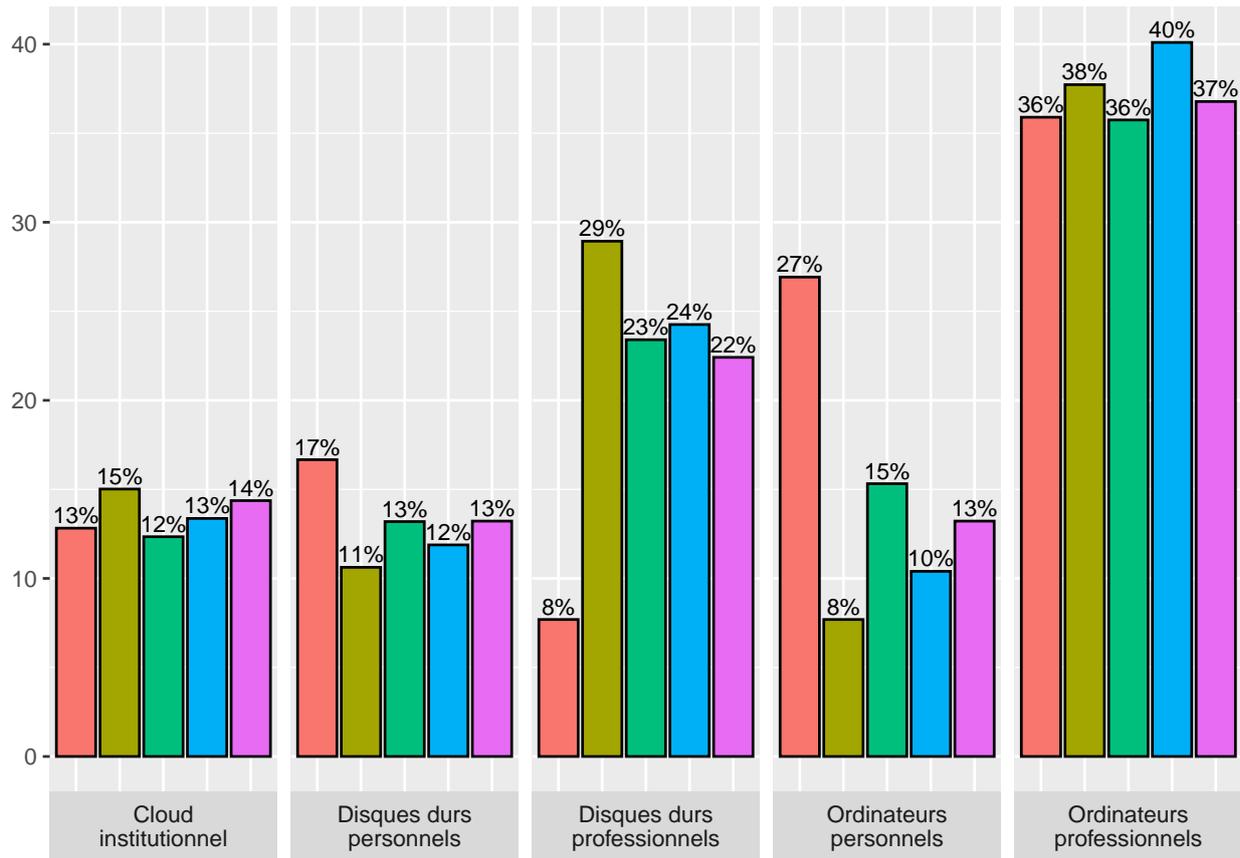


FIGURE 9 – Lieux de stockage des données par ancienneté des répondants (partie 2/2)

La question de savoir si le volume de données à stocker a un impact sur le choix du lieu et type de stockage est illustrée par les figures 10 et 11. Les éléments suivants peuvent être notés :

- Lorsque les volumes commencent à être importants, les données sont plutôt stockées où elles ont été produites ou au sein de plateformes de stockage institutionnelles.
- Les autres solutions de stockage sont utilisées de façon assez homogène.
- Sur de petites volumétries, le stockage sur des sites de cloud privé reste important.



Volumétrie des données



FIGURE 10 – Lieux de stockage des données par volumétries des données (partie 1/2)

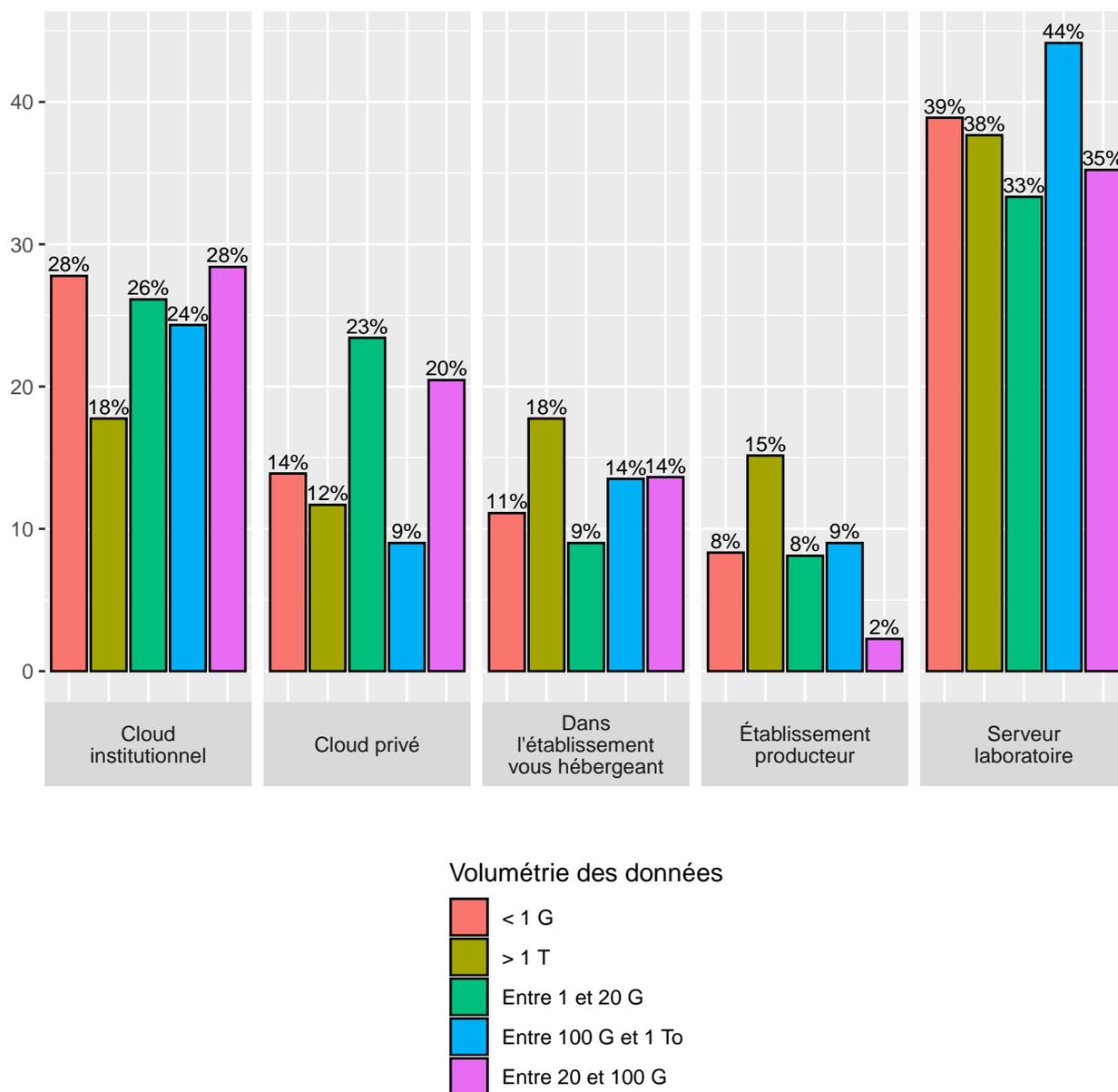


FIGURE 11 – Lieux de stockage des données par volumétries des données (partie 2/2)

Un des blocs de questions de l'enquête portait sur la problématique du stockage sécurisé. Au regard des réponses, ainsi que des commentaires associés, la compréhension de ce terme a été très diversement perçue.

Le stockage sécurisé était, dans l'esprit du questionnaire, un stockage proposant une résilience automatique suffisante pour éviter la perte de données liée à des problèmes matériels ou humains. Il est en ce sens en particulier une base pour l'organisation d'une architecture technique adaptée au stockage de données sensibles.

Les pratiques en terme de stockage sécurisé (figure 12), selon la compréhension qu'en ont eu les répondants, montrent des usages assez divers en fonction des communautés :

- Peu de stockage sécurisé pour MSTIC où les volumes de données sont assez faibles.
- Utilisation importante de plateformes d'établissement dans les pôles CBS et PEM.
- Utilisation forte de plateformes nationales pour PAGE, en adéquation avec les instruments

disponibles dans cette communauté, mais aussi en SHS, avec la mention d'infrastructures comme HumaNum.

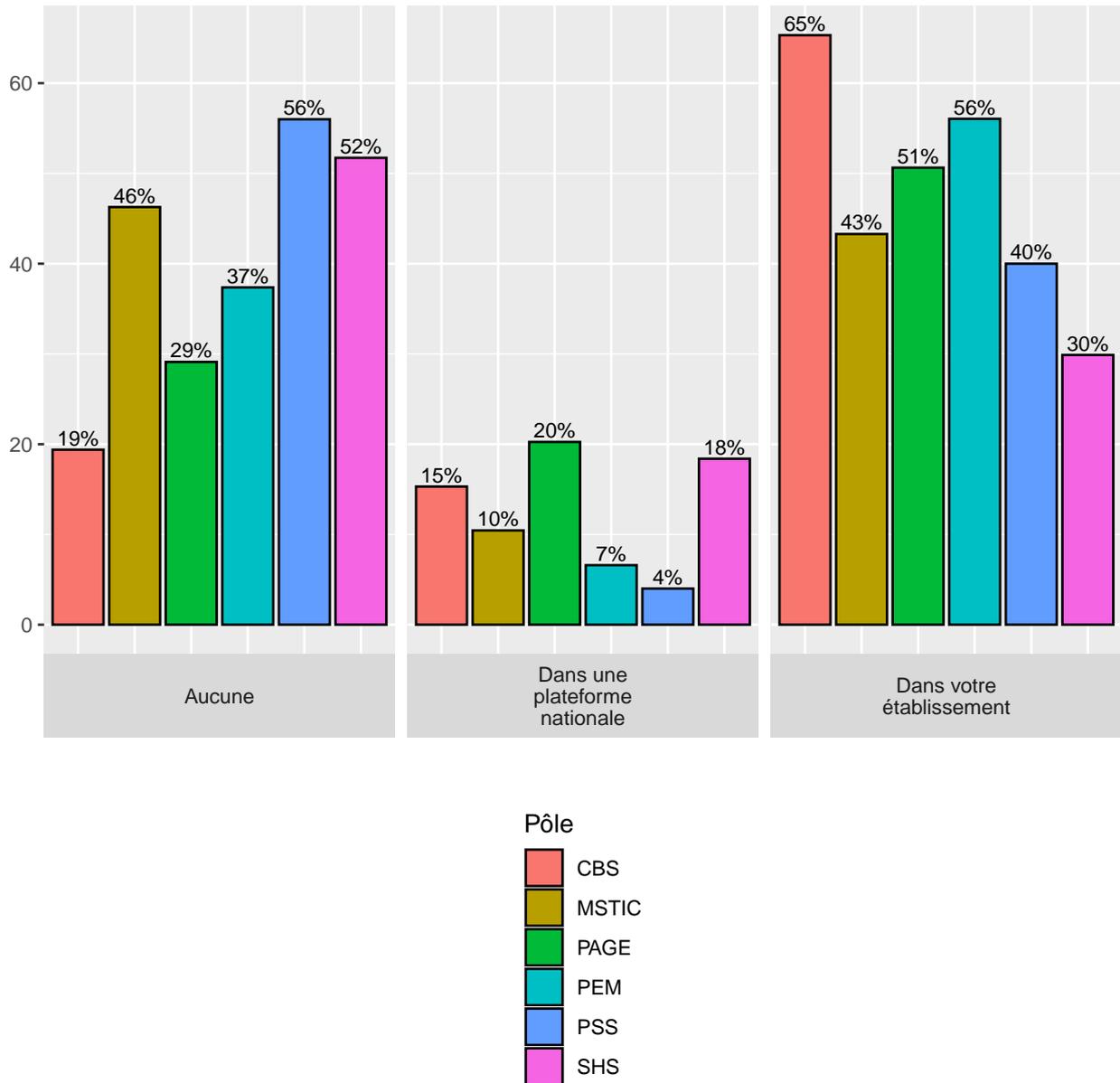


FIGURE 12 – Lieux de stockage sécurisés des données par pôle de recherche

Les lieux de stockage, quand ils ont été précisés, apportent les éléments suivants :

- Concernant les plateformes de stockage des établissements, on retrouve majoritairement les infrastructures institutionnelles : la plateforme SUMMER, les espaces de stockage de GRICAD, les plateformes du CEA, certains cloud institutionnels, mais aussi un certain nombre de réponses concernant les serveurs de laboratoires.
- La plateforme gitlab est également mentionnée plusieurs fois dans cette question « stockage sécurisé ».
- Au niveau national, les structures citées sont essentiellement le CINES, HumaNum, certaines autres Infrastructures de Recherche (Data Terra, FLI-IAM, Grid'5000 par exemple) ainsi que le cloud CNRS.

Si on analyse les besoins exprimés au niveau de l'accompagnement autour du stockage en général (figure 13), la moitié environ des répondants sont intéressés par cette aide, quelque soit le pôle de rattachement. De plus, on note, à travers l'analyse d'une autre question, une demande légèrement plus importante sur la partie stockage que dépôt des données.

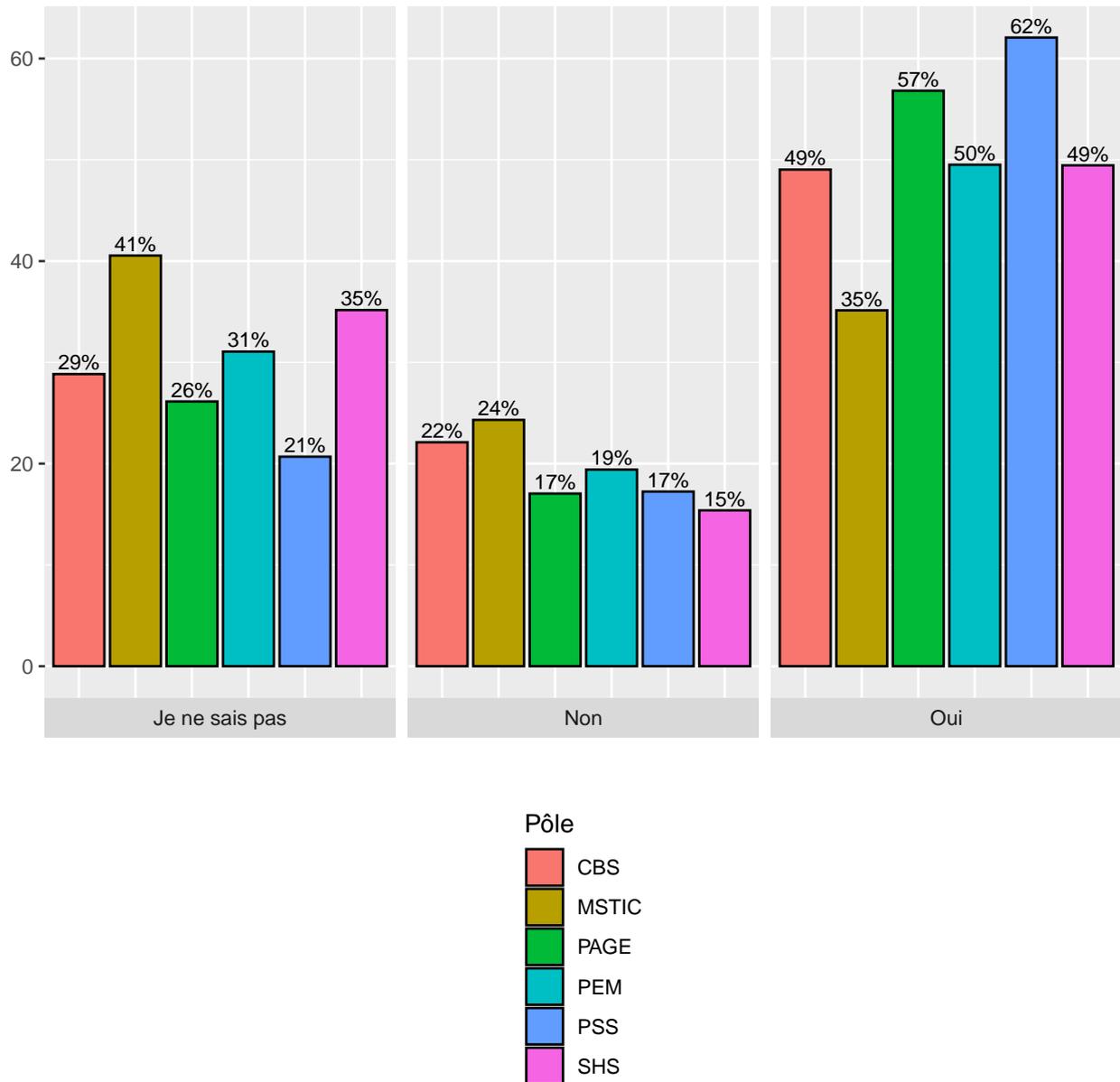


FIGURE 13 – Besoins d'accompagnement sur le stockage et le dépôt des données par pôle de recherche

Ce besoin est davantage mentionné par les chercheurs et enseignants chercheurs que par les personnels d'appui à la recherche (figure 14).

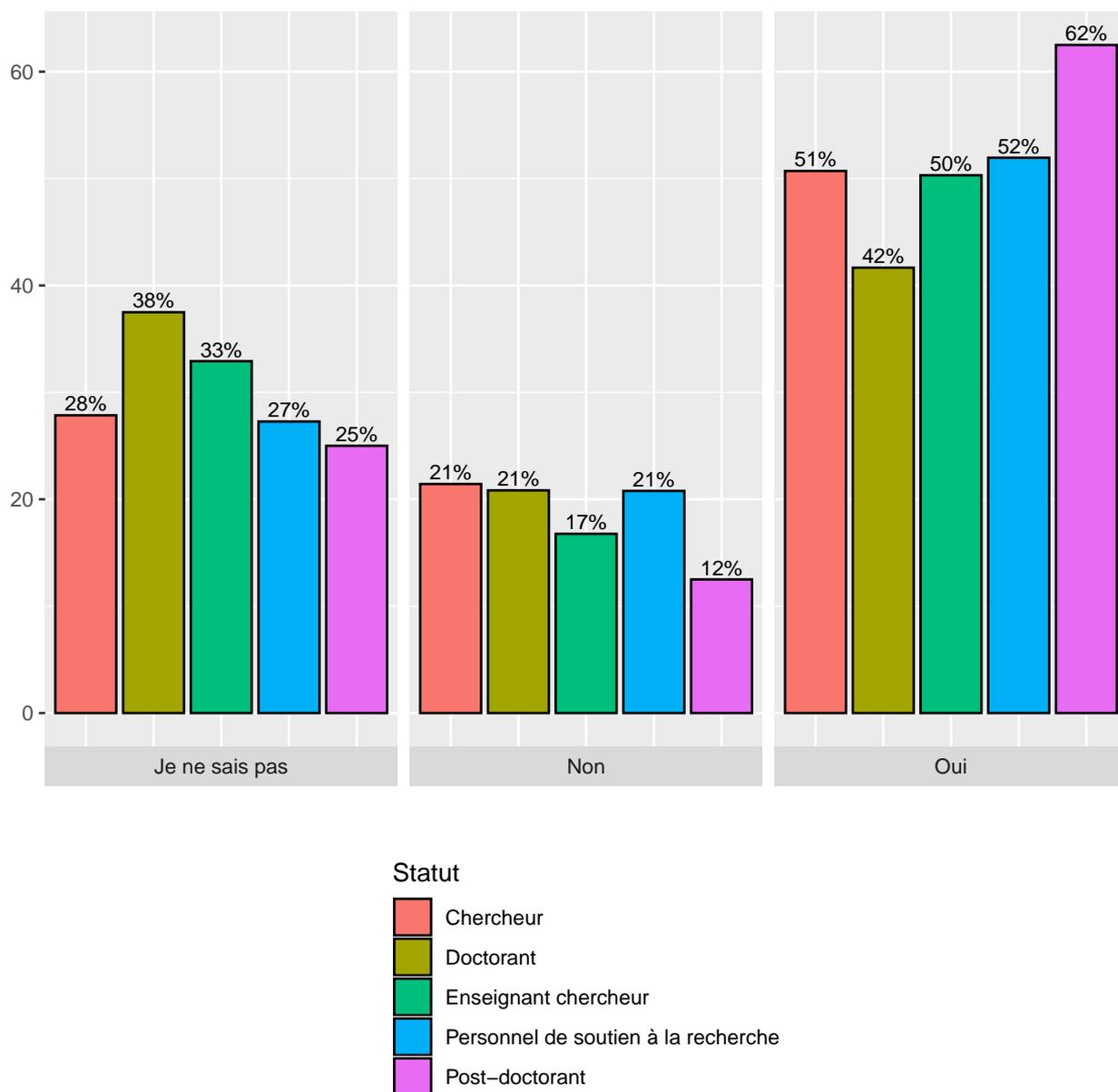


FIGURE 14 – Besoins d’accompagnement sur le stockage sécurisé par statut

### 3.3 Synthèse

Les pratiques de stockage sont assez fortement disciplinaires. Les grosses volumétries poussent certaines communautés à se tourner, pour certaines depuis longtemps, vers des plateformes institutionnelles ou dédiées. Certains pôles comme le pôle SHS sont a contrario peu utilisateurs d’espaces mutualisés portés par les établissements, même si la mention de la TGIR HumaNum montre son impact dans cette communauté. Les jeunes chercheurs s’appuient de façon assez importante sur du matériel personnel.

Il semble donc nécessaire de proposer un accompagnement spécifique pour certaines communautés, afin en particulier de les aider à surmonter la problématique technique associée à l’utilisation de plateformes de grande envergure.

De même, l’aide pour les jeunes chercheurs doit se développer afin d’aller au-delà de l’usage de matériel personnel peu sécurisé. Cette démarche doit s’inscrire dans un accroissement de la formation, de la communication, à la fois vers les doctorants mais aussi vers les directions de

thèse, ainsi que la mise en place d'un accompagnement spécifique.

Par ailleurs, il est fondamental de poursuivre l'amélioration des services académiques afin de détourner les scientifiques de l'usage de clouds privés qui peuvent se révéler très problématique pour la souveraineté des données et le cas des données sous RGPD.

Enfin, il faut également développer l'accompagnement des chercheurs et enseignants chercheurs, et donc s'assurer de la mise en place de formations adaptées.

## 4 Sujet des données sensibles

L'objectif de cette partie est de préciser les habitudes de gestion des données lorsqu'il s'agit de données sensibles, en particulier de santé ou à caractère personnel.

### 4.1 Analyse des réponses

La recherche avec des données sensibles concernent 17% des participants pour des données personnelles, 15% pour des données de santé, et 3% pour d'autres formes de données sensibles.

Le croisement entre les pôles d'appartenance et la manipulation de données sensibles (figure 15) est conforme à ce qui était attendu à savoir :

- Les pôles PSS et SHS manipulent en partie des données personnelles, et également de façon moins importantes des données de santé.
- L'utilisation de données de santé se retrouve dans le pôle CBS.
- Le pôle MSTIC est concerné également à la fois par des données de santé et des données personnelles ce qui illustre les liens interdisciplinaires existants entre ce pôle et les autres.

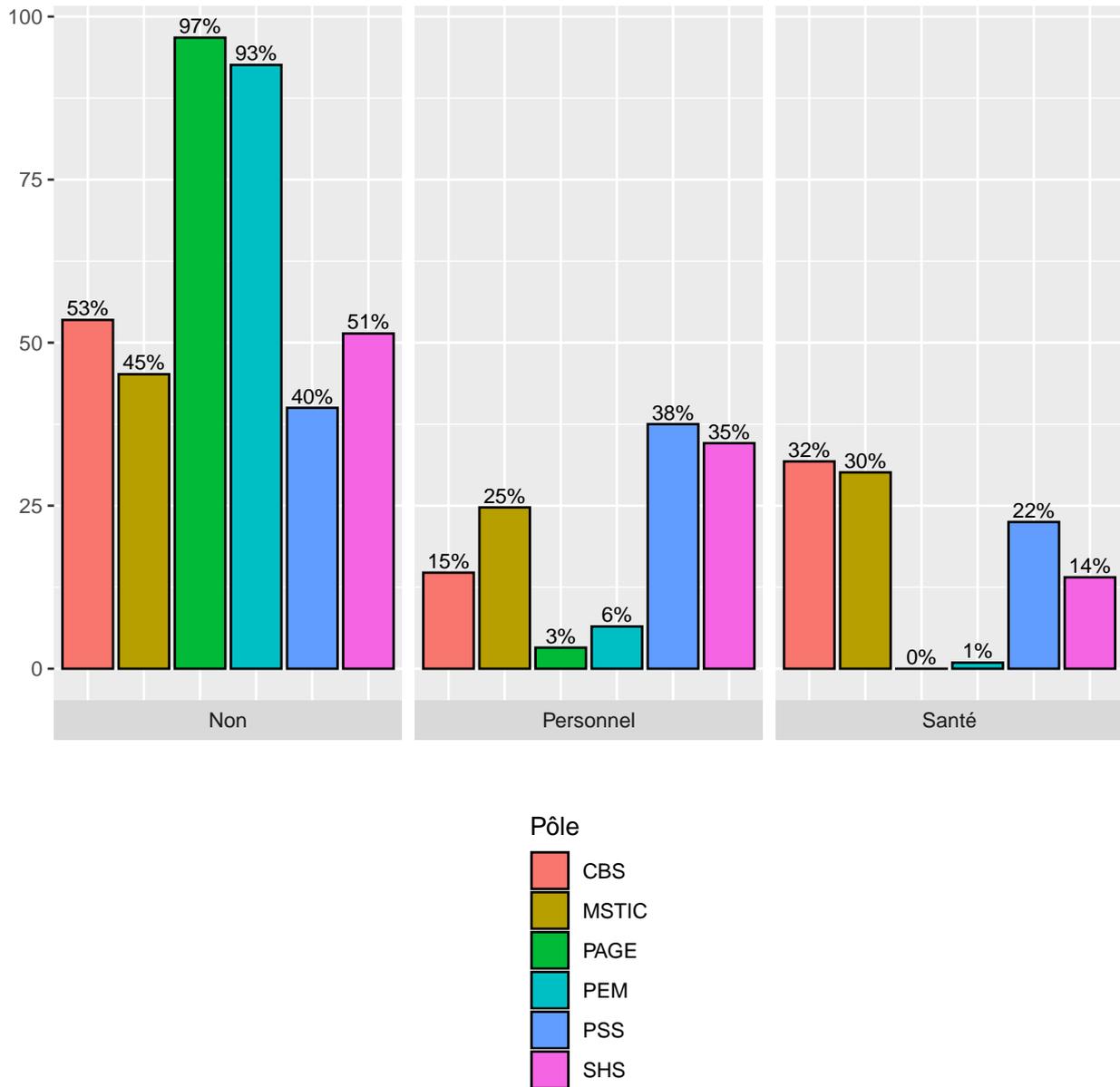


FIGURE 15 – Type de données sensibles par pôle de recherche

Il faut également noter que la problématique des données confidentielles liées à des contrats avec des entreprises a été mentionnée à plusieurs reprises par les répondants.

Concernant la diffusion des données, le fait de travailler avec des données sensibles implique que les personnes concernées adaptent les jeux de données afin de pouvoir les diffuser (figure 16).

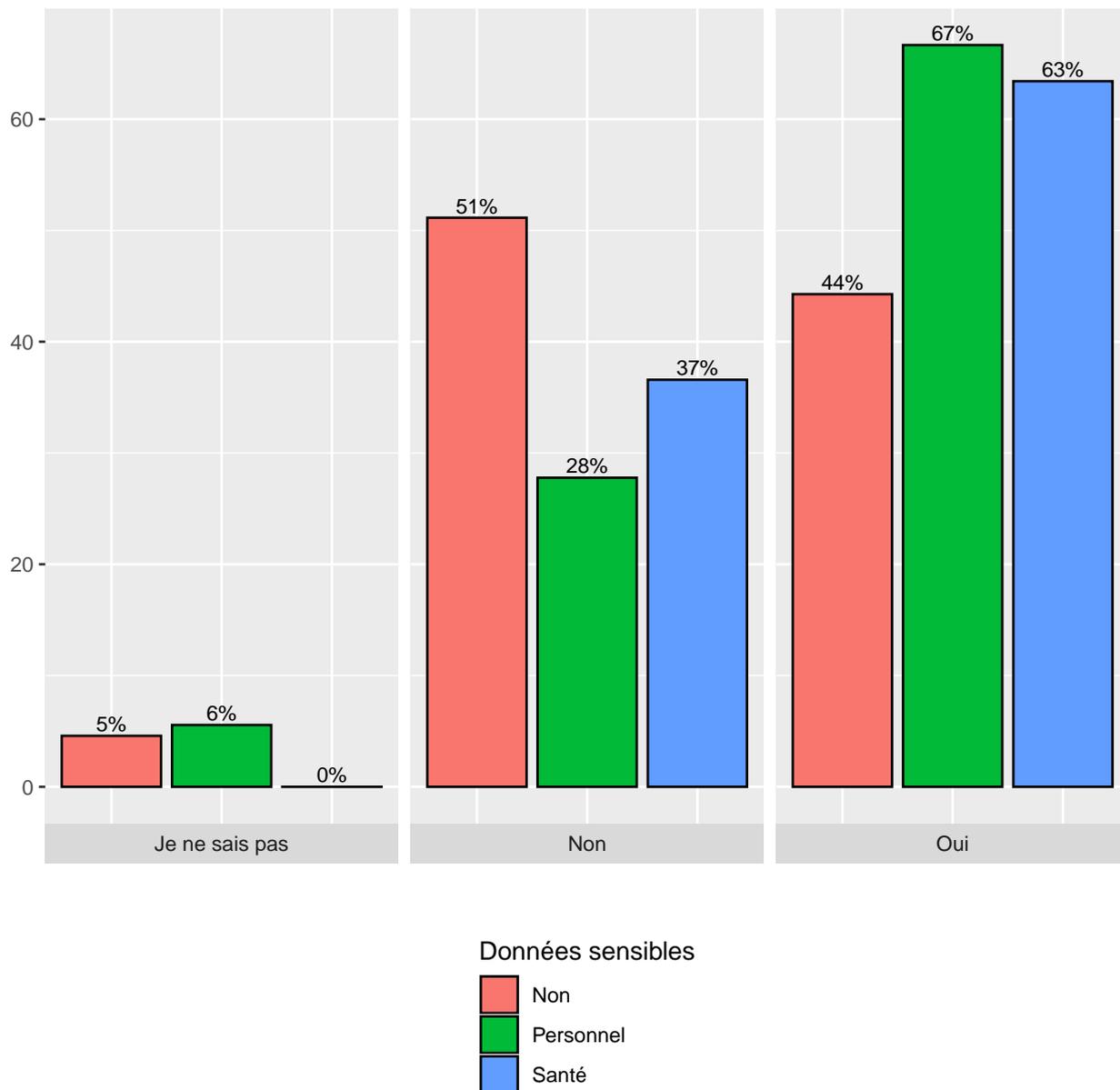


FIGURE 16 – Adaptation des jeux de données en vue de leur diffusion par type de données sensibles

## 4.2 Synthèse

La problématique des données sensibles concernent directement plusieurs pôles de recherche. Les contraintes techniques et juridiques associées sont importantes. L'aide juridique fait d'ailleurs partie des besoins exprimés les plus importants (voir partie 10). Par ailleurs, il est également important d'accompagner sur l'enjeu du stockage qui nécessite la mise en place d'architecture adaptée et sécurisée, éventuellement dans des environnements dédiés si nécessaire. Enfin, la diffusion de ce type de données ne peut se faire qu'à travers un processus d'anonymisation robuste et approprié.

Il est ainsi indispensable de pouvoir identifier sur le site les expertises permettant de répondre à ces questionnements sur les plans juridique et technique et de construire les services d'accompagnement les plus pertinents : aide à la rédaction des Plans de Gestion de Données sur ces chapitres réglementaires, accompagnement pour la mise en place des architectures tech-

niques, conseil pour la mise en œuvre de processus d'anonymisation.

## 5 Pratiques collaboratives

Cette partie a pour but de comprendre les habitudes de travail collaboratif, en particulier lorsque des données partagées sont en jeu.

### 5.1 Habitudes de collaborations

Si on croise les réponses sur ce sujet avec l'appartenance aux pôles de recherche (figure 17), il est assez clair que la plupart des répondants travaillent dans le cadre de projets collaboratif de façon plus ou moins régulière, avec une différence notable pour le pôle SHS (et dans une moindre mesure PSS) dont les travaux sont plutôt individuels, et le pôle PAGE (et dans une moindre mesure CBS) dont les travaux sont très majoritairement dans le cadre de collaborations.

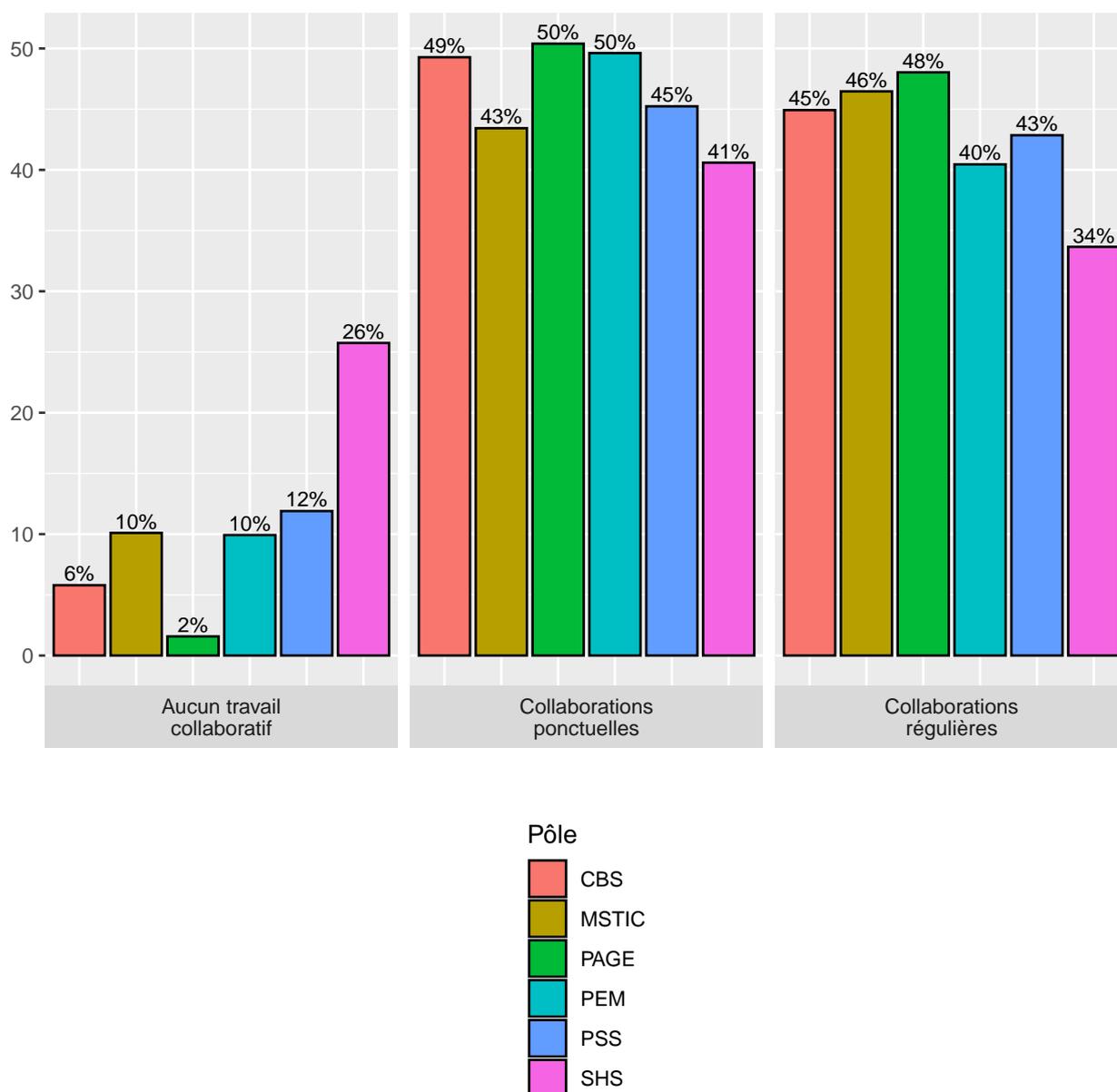


FIGURE 17 – Pratiques collaboratives par pôle de recherche

Il est intéressant de croiser également ces pratiques de travail avec l'expérience en recherche (figure 18), ainsi que la fonction (figure 19). Dans les répondants travaillant plutôt seuls, les

doctorants et les post-doctorants sont très représentés (avec 29 et 30% respectivement) contrairement aux chercheurs (3%) et aux personnels en soutien à la recherche (4%). Les enseignants-chercheurs sont un peu plus partagés (9%).

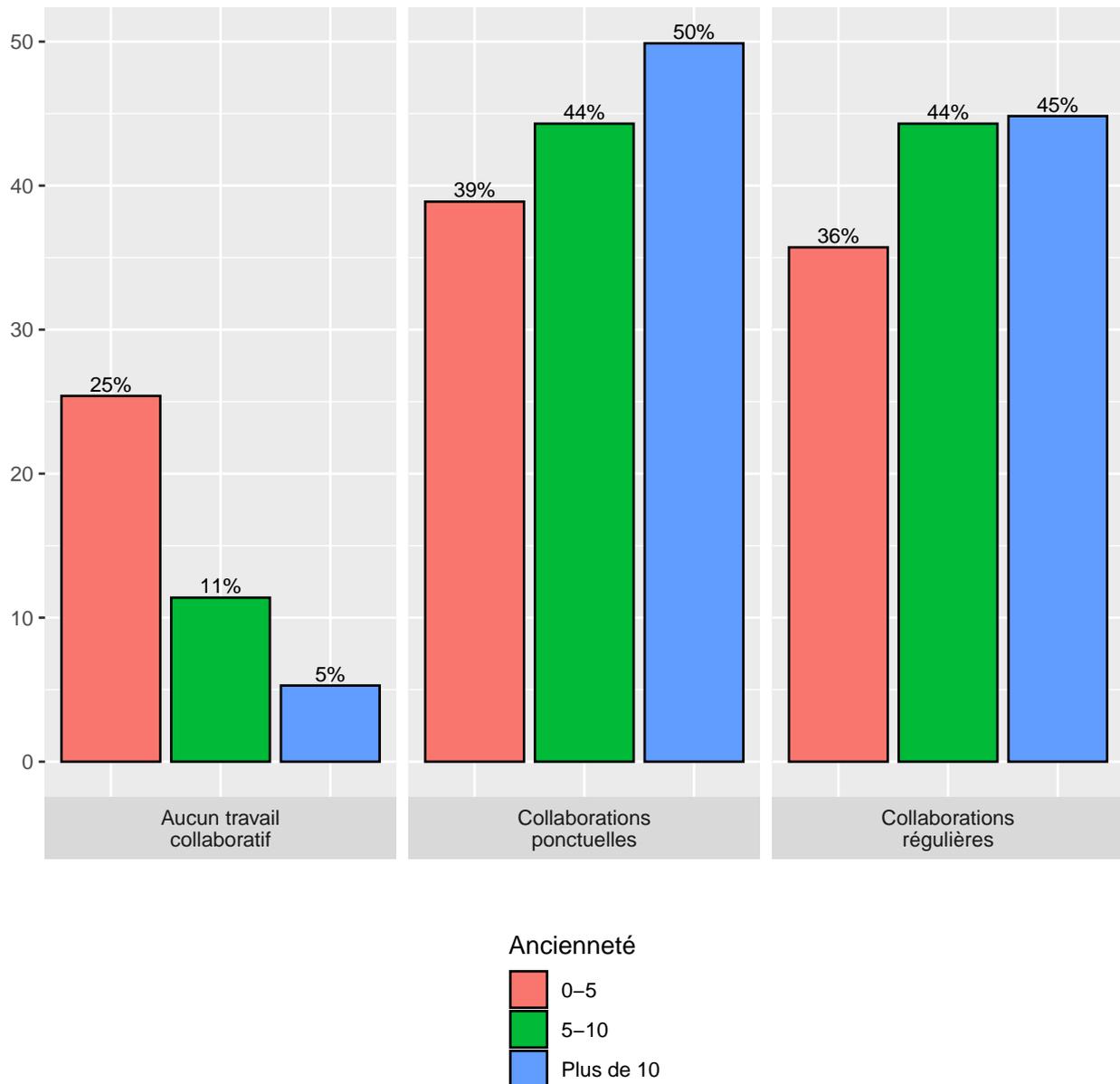


FIGURE 18 – Pratiques collaboratives par expérience en recherche

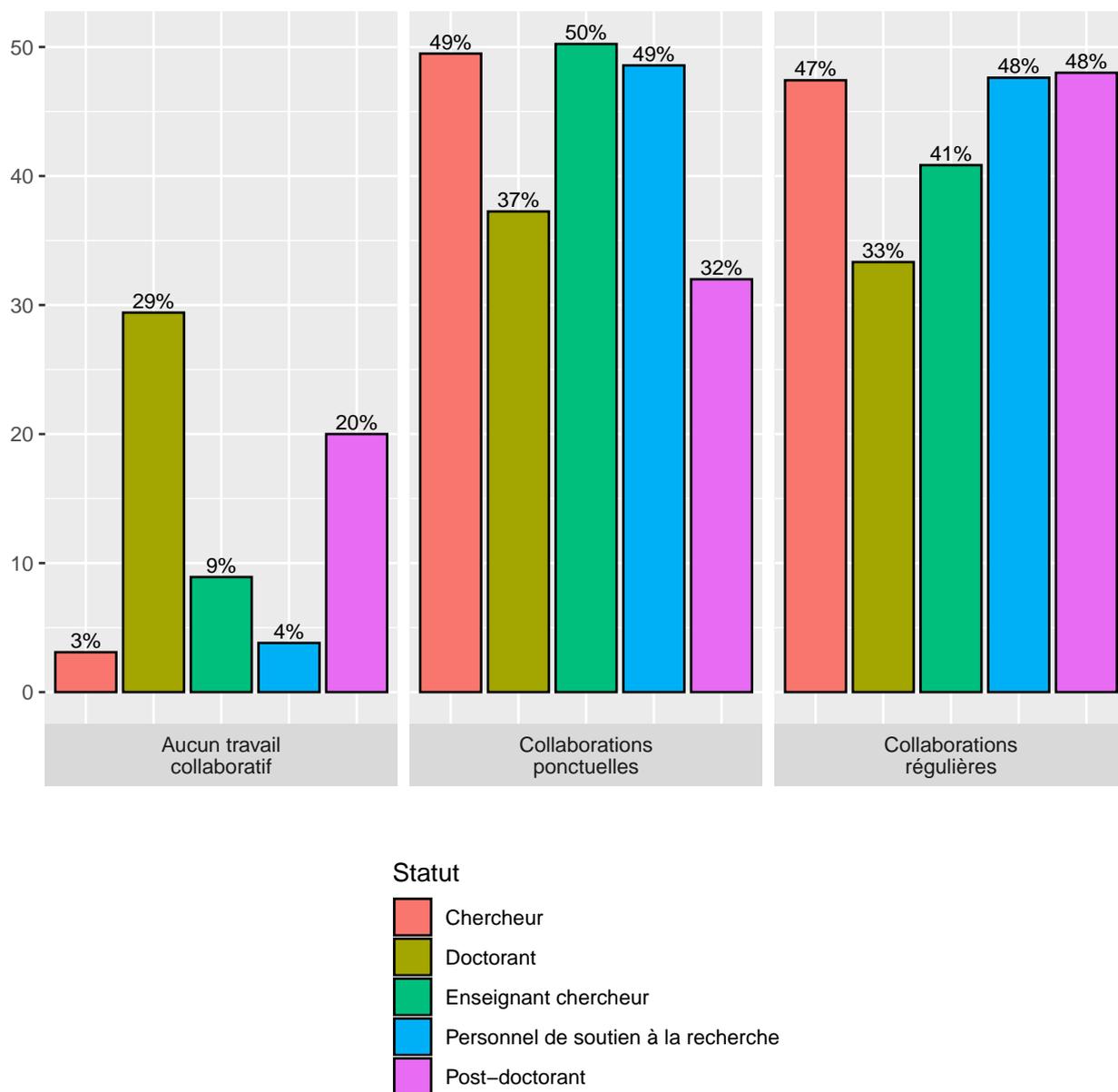


FIGURE 19 – Pratiques collaboratives par statut

## 5.2 Outils pour le travail collaboratif

Si on s'intéresse aux outils utilisés pour ces collaborations (figure 20), on peut noter que :

- le pôle PAGE dont le travail en collaboration est important, utilise aussi bien des outils institutionnels que commerciaux, de même que CBS et PEM,
- l'usage d'outils collaboratifs est moins répandu pour le pôle PEM,
- le pôle MSTIC a des réponses assez proches mais on peut noter une utilisation un peu moins importantes des outils commerciaux,
- les pôles SHS et PSS utilisent majoritairement des outils commerciaux.

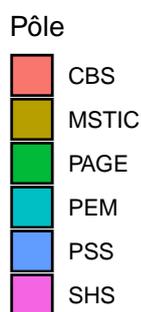
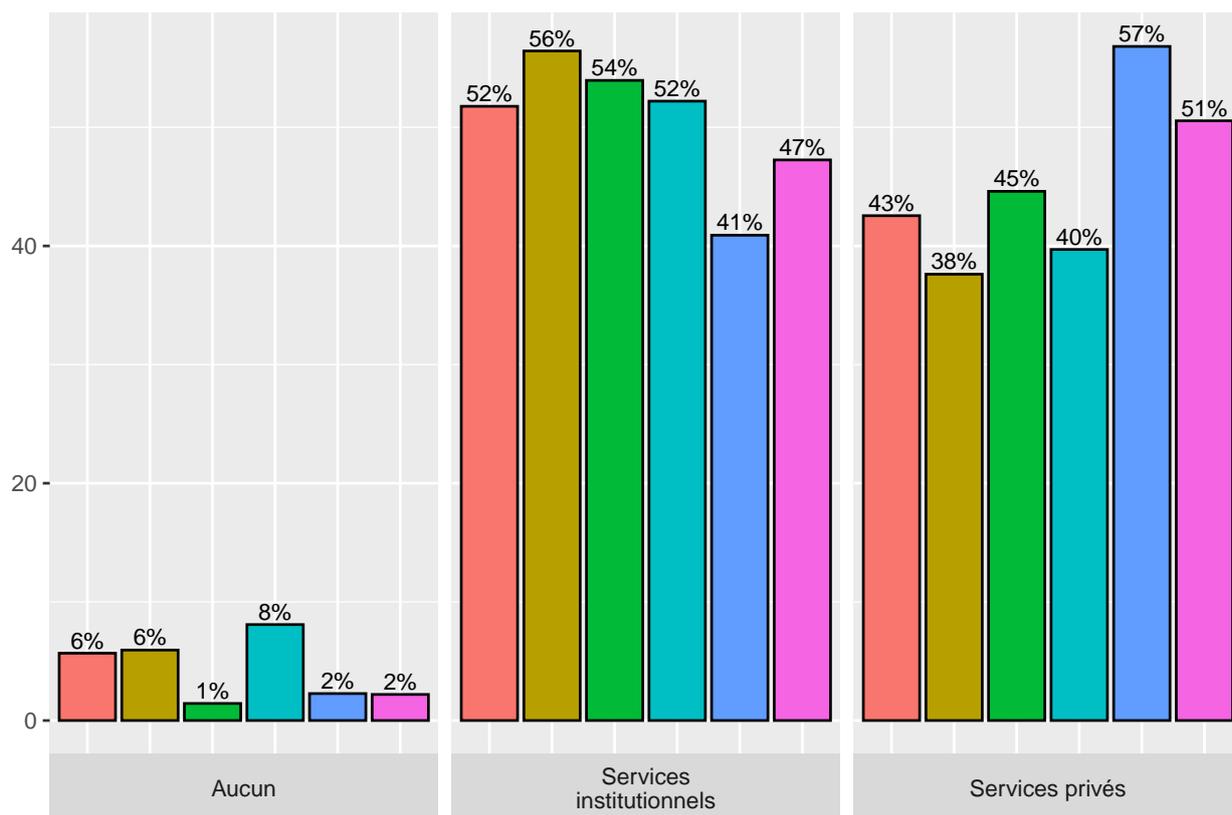


FIGURE 20 – Outils collaboratifs utilisés par pôle de recherche

Enfin, en mettant en regard les profils des répondants avec les outils utilisés (figure 21), on peut constater que :

- les doctorants, qui déclarent travailler assez souvent de façon individuelle, utilisent peu d'outils collaboratifs, et surtout des outils commerciaux,
- les chercheurs et enseignants-chercheurs sont dans le cas inverse, avec une utilisation à peu près équivalente entre outils institutionnels et outils commerciaux,
- les personnels d'appui à la recherche utilisent majoritairement des outils institutionnels.

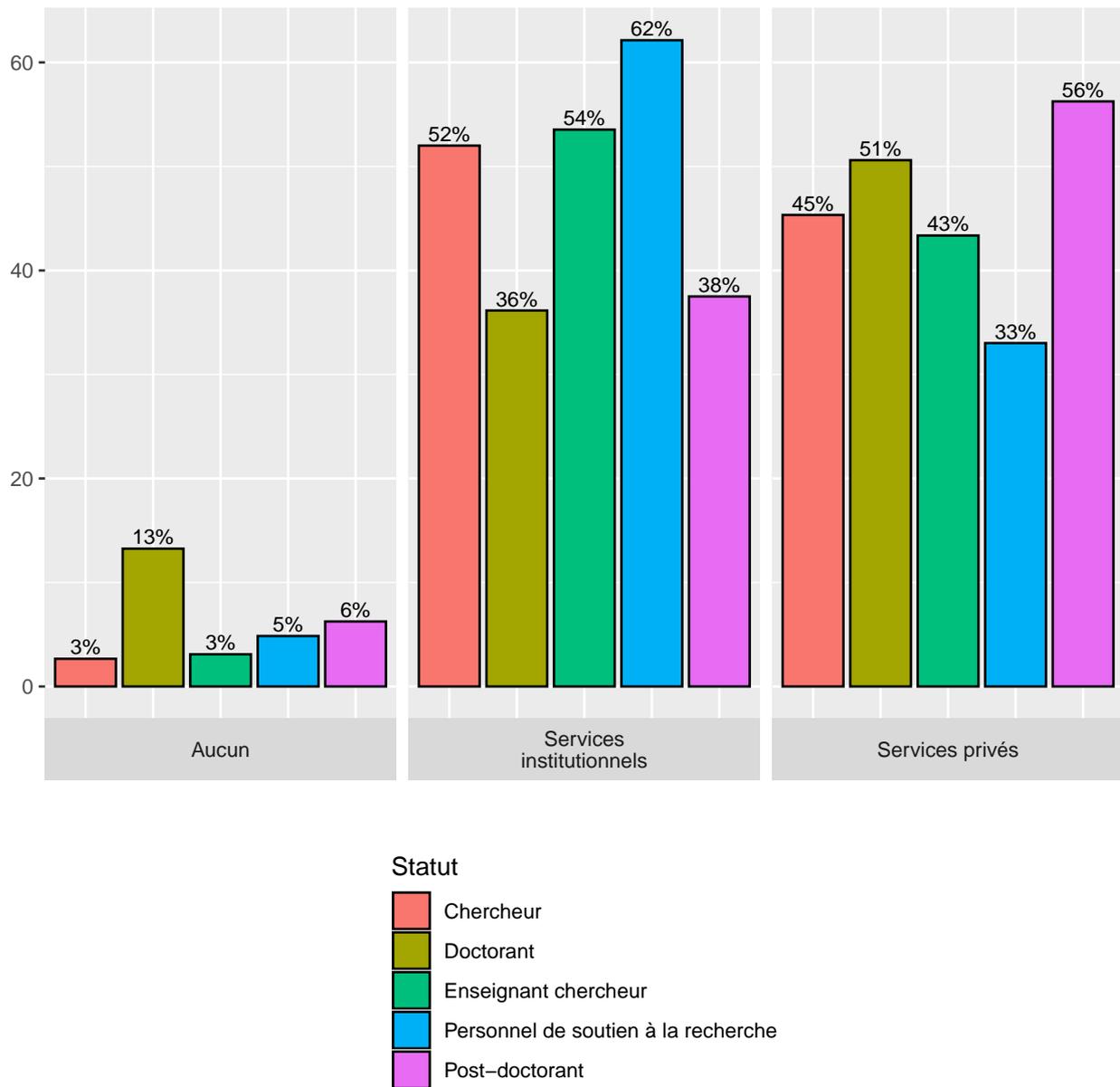


FIGURE 21 – Outils collaboratifs utilisés par statut

L'expérience professionnelle joue un rôle important dans le travail collaboratif comme l'illustre la figure 22.

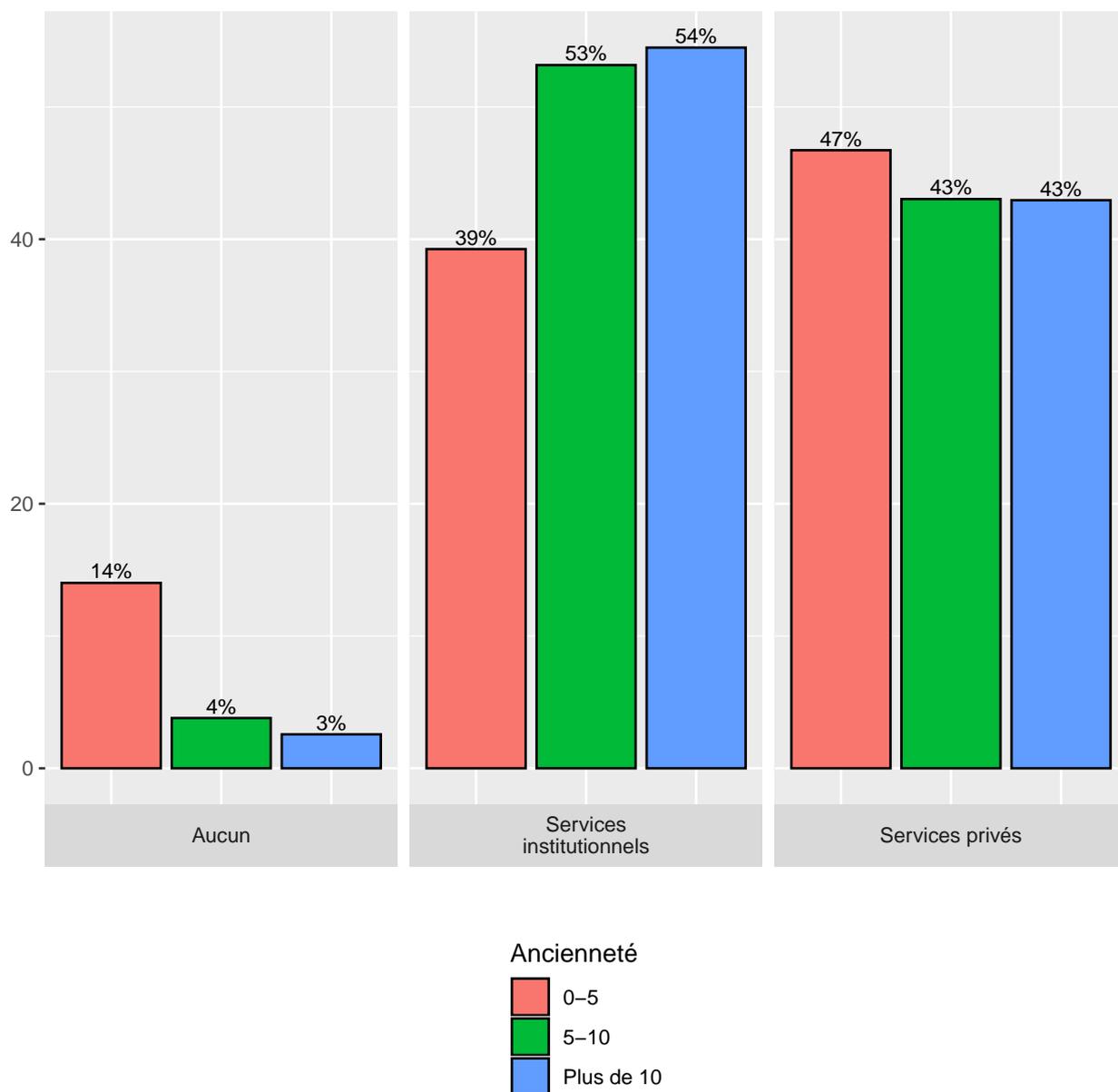


FIGURE 22 – Pratiques collaboratives par expérience en recherche

### 5.3 Synthèse

La majorité de la recherche se fait dans un cadre collaboratif avec quelques disparités liés à la discipline, comme pour le pôle SHS, et à la fonction, comme pour les doctorants, exemples pour lesquels on note une tendance vers le travail individuel.

Ces méthodes de travail s'accompagnent de l'usage d'outils collaboratifs. L'information la plus importante est sans aucun doute l'utilisation plus ou moins importante d'outils commerciaux.

Cela sous-entend que les outils institutionnels sont insuffisamment connus et peut être mal adaptés à l'ensemble des usages. Cela pose clairement des questions autour des enjeux de souveraineté et de protection des données et du patrimoine scientifique.

Il est donc nécessaire, d'une part de faire connaître les outils existants, d'aider à leur usage, mais aussi d'identifier les freins à leur utilisation. Cela ne pourra se faire que par le biais d'un accompagnement de proximité et de la mise en place d'échanges simples et faciles avec les

utilisateurs.

## 6 Réutilisation et diffusion des données

L'objectif est d'appréhender les pratiques concernant la diffusion des données et d'identifier les besoins d'accompagnement pour adapter les dispositifs existants.

### 6.1 Réutilisation des données

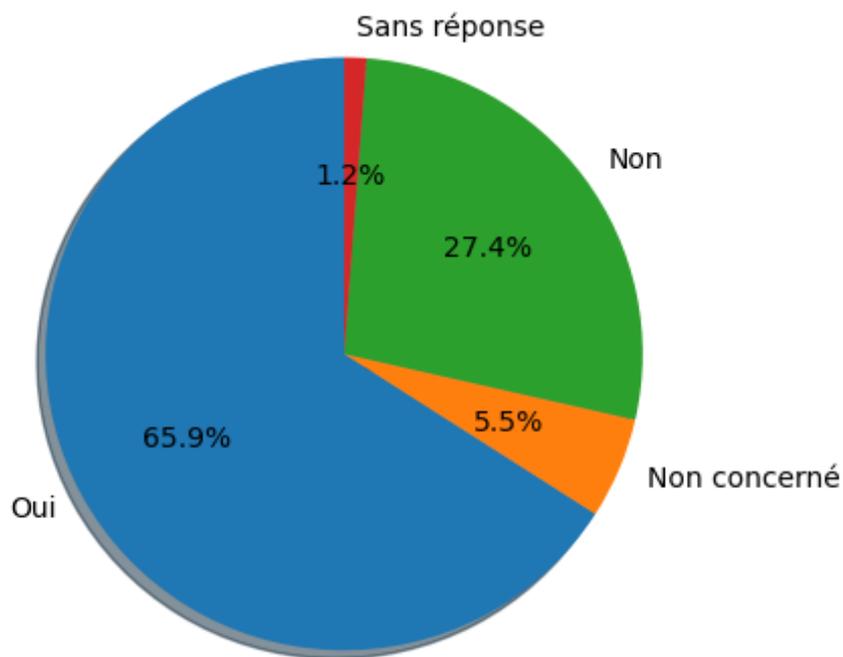


FIGURE 23 – Réutilisation de données publiées

D'après la figure 23, une très large proportion (près de 66%) des répondants a déjà réutilisé des données publiées par d'autres chercheurs. Les répondants relevant des pôles PEM, PAGE et CBS sont parmi les plus gros réutilisateurs (voir figure 24. On peut mettre en regard ce chiffre avec les pratiques de diffusion indiquées par les répondants pour leurs propres données : seuls 41% des répondants ont déjà diffusé leurs données.

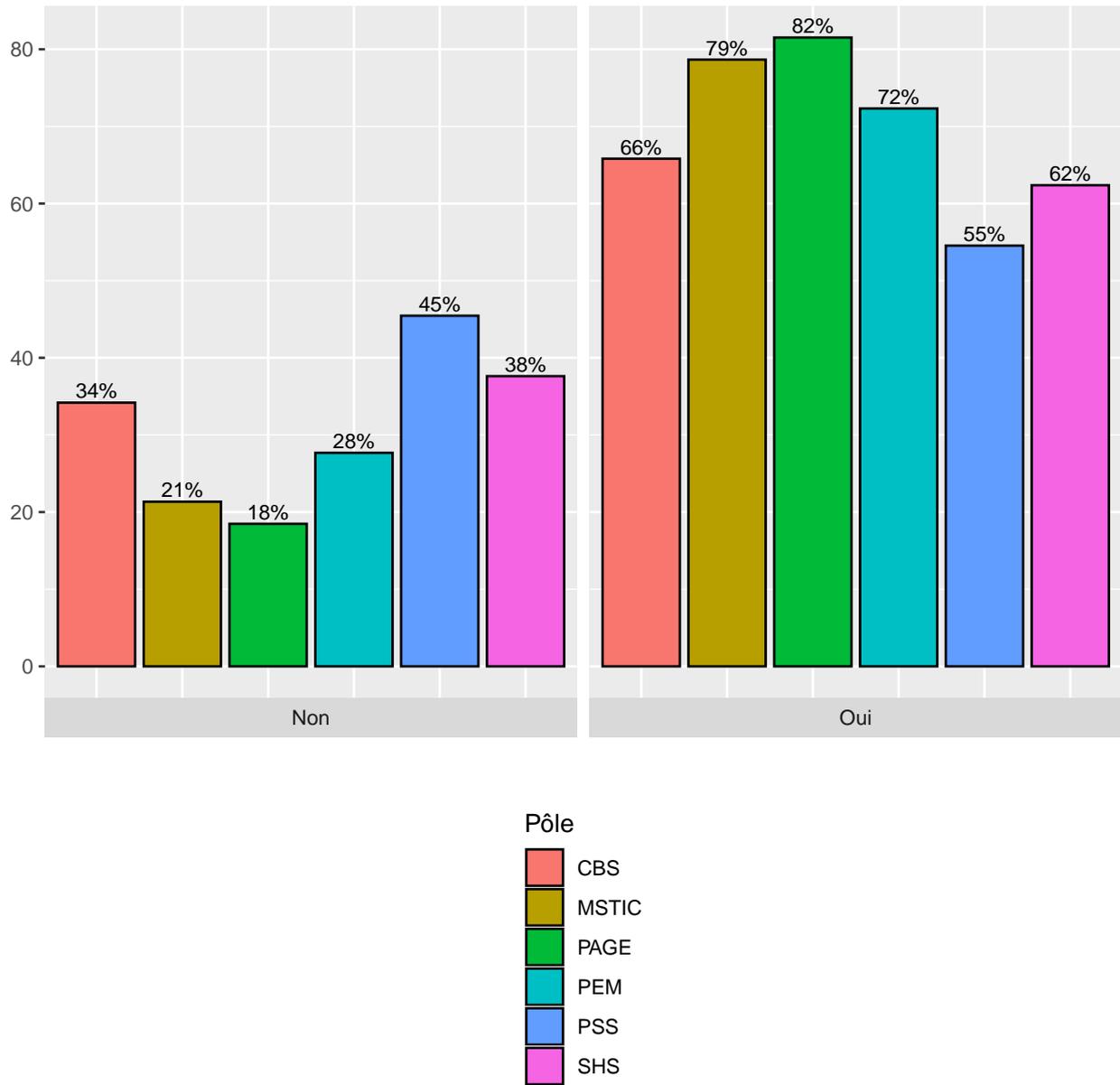


FIGURE 24 – Réutilisation de données par pôles de recherche

## 6.2 Difficultés à la réutilisation des données

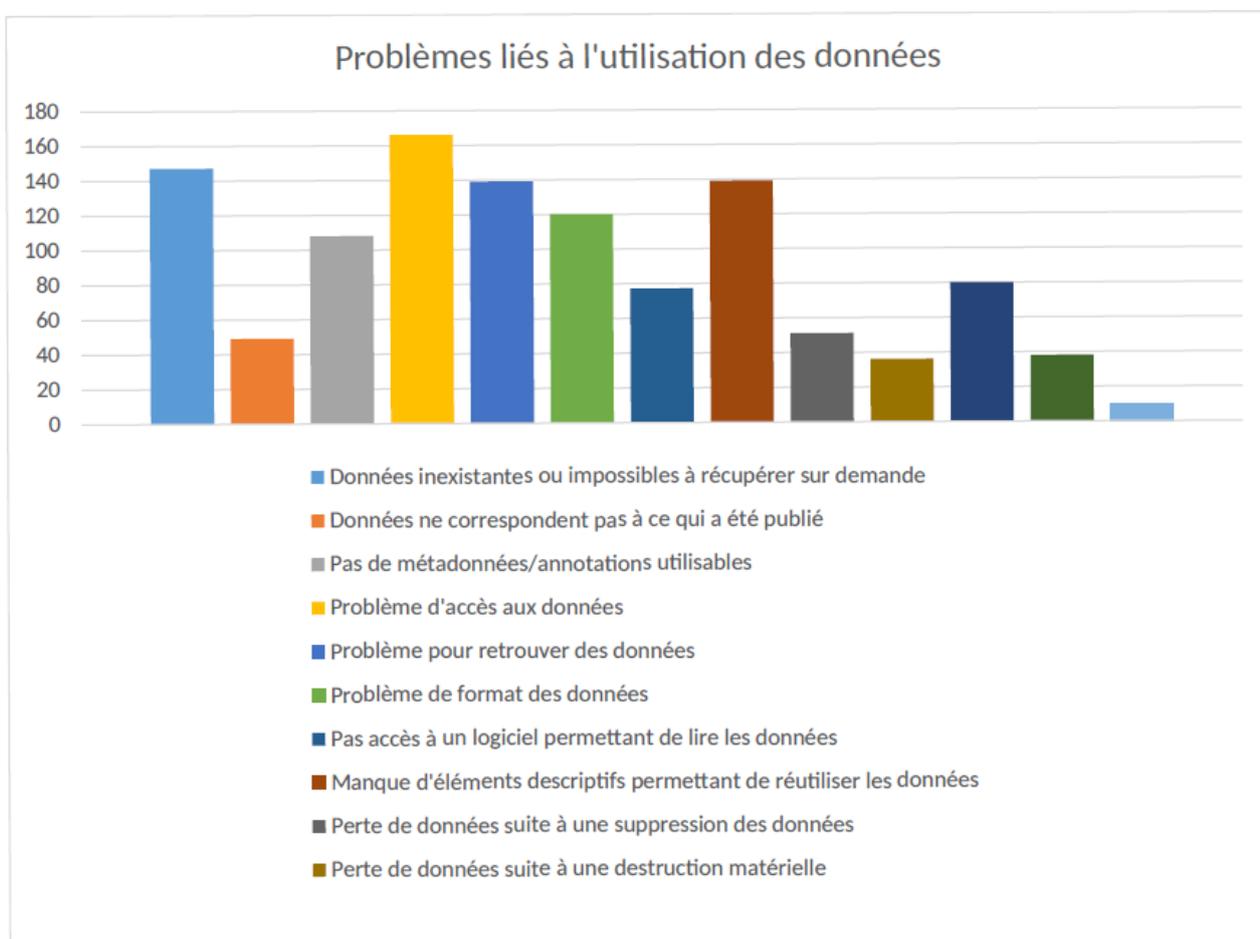


FIGURE 25 – Problèmes liés à l'utilisation des données

80% des répondants indiquent avoir rencontré des difficultés dans l'utilisation des données, ce qui est une proportion considérable.

Selon la figure 25, les participants sont nombreux à citer des problèmes d'accès aux données (38%), s'y ajoute l'impossibilité à récupérer des données sur demande, voire des données inexistantes (33%).

Dans les réponses libres, certains précisent que ces difficultés peuvent relever de questions juridiques (liées au copyright par exemple). Autre écueil relevé par 30% des répondants : le manque de documentation des données qui nuit à leur ré-usage ou à leur interprétation. Ainsi, 22% relèvent l'absence de métadonnées ou d'annotations utilisables. Dans la même proportion, les participants signalent des difficultés à retrouver des données.

Les problèmes également évoqués concernent les questions de format (27%) ou de manque de logiciel (18%).

## 6.3 Pratiques de diffusion

54% des répondants déclarent ne pas avoir diffusé leurs données, ce qui représente une proportion très importante. La diffusion des données n'est donc pas une pratique très répandue, malgré les recommandations des tutelles et les directives des financeurs.

Comme l'illustre la figure 26, le pôle PAGE diffuse majoritairement ses données (66%). Cela correspond à une pratique disciplinaire de partage des publications et des données assez répandues dans les sciences de l'environnement et en astronomie-astrophysique. Suivent les

pôles MSTIC et CBS, (respectivement 50 et 49%). On peut constater que les pôles PSS et PEM diffusent très peu leurs données (respectivement 28 et 26%).

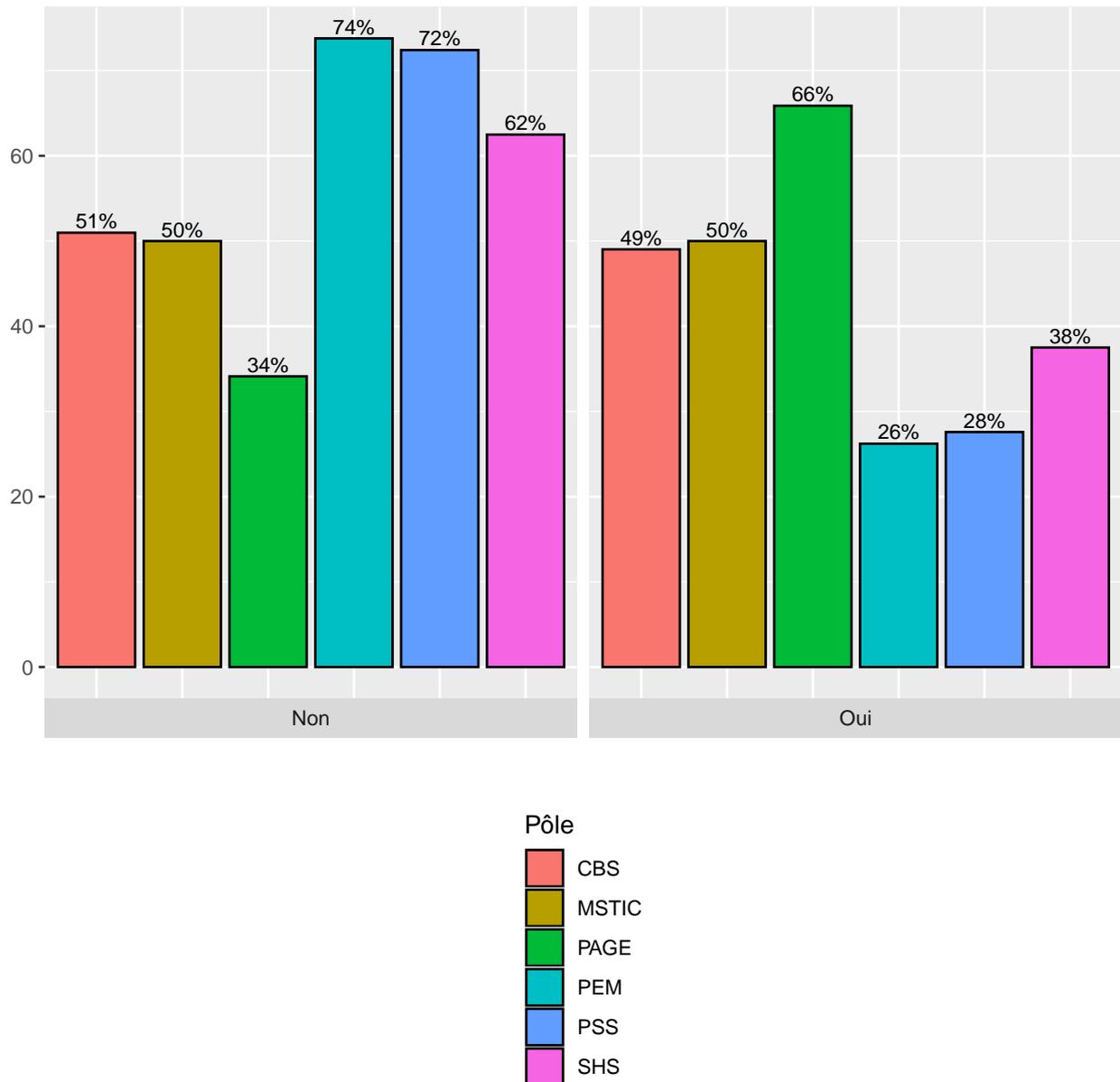


FIGURE 26 – Pratique de diffusion par pôle de recherche

Selon la figure 27, ce sont les répondants avec le plus d'expérience qui diffusent leur recherche.

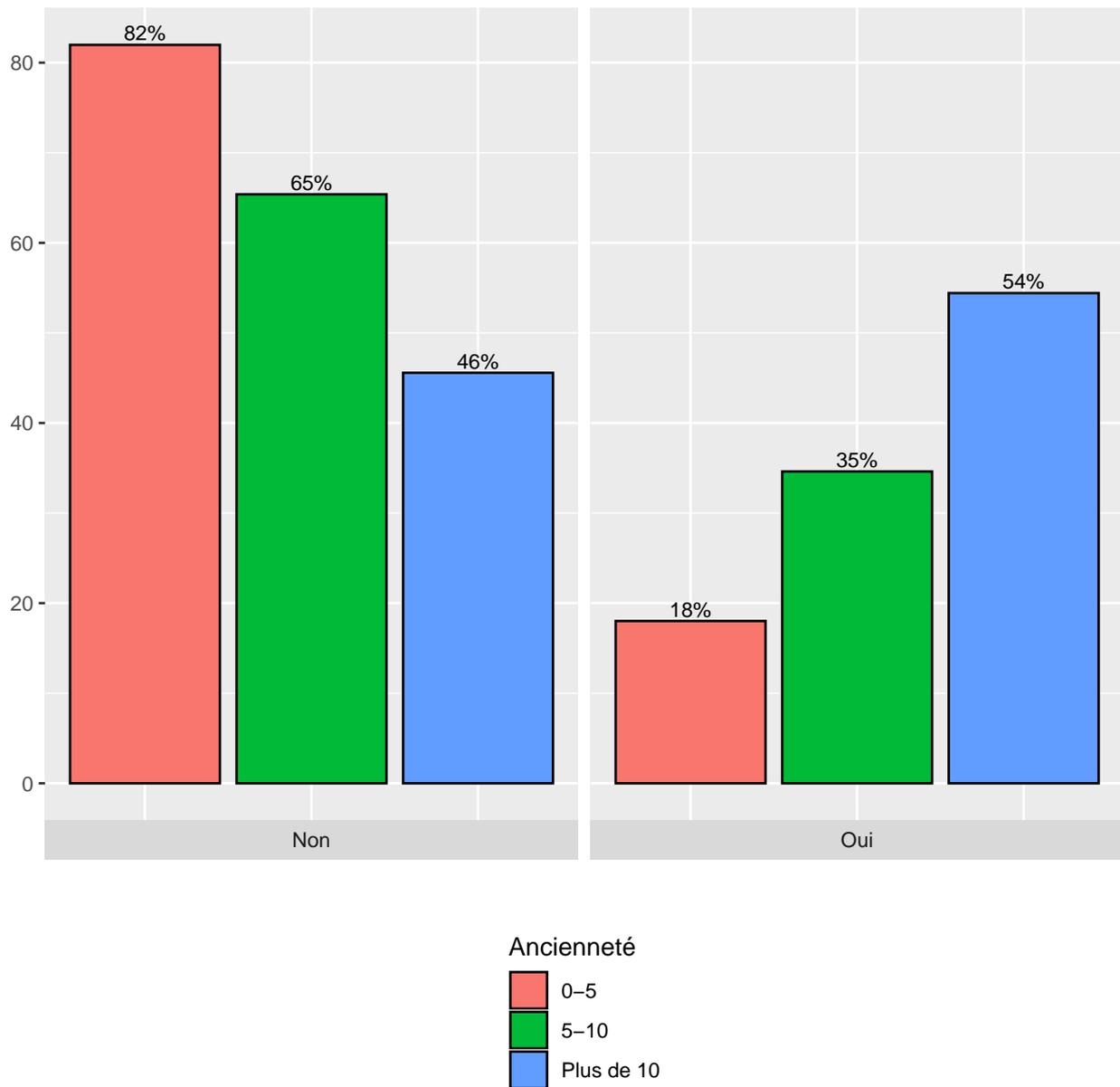


FIGURE 27 – Pratique de diffusion par expérience en recherche

## 6.4 Profils des déposants

À une écrasante majorité (voir figure 28), ce sont les chercheurs, et enseignants-chercheurs producteurs de données qui se chargent de leur diffusion. Les doctorants sont peu nombreux à assumer ce rôle.

Les ingénieurs d'appui à la recherche ou les documentalistes sont cités, en association ou non avec un chercheur. Dans de rares cas, les chefs d'équipe ou de projet sont en charge de la diffusion des données.

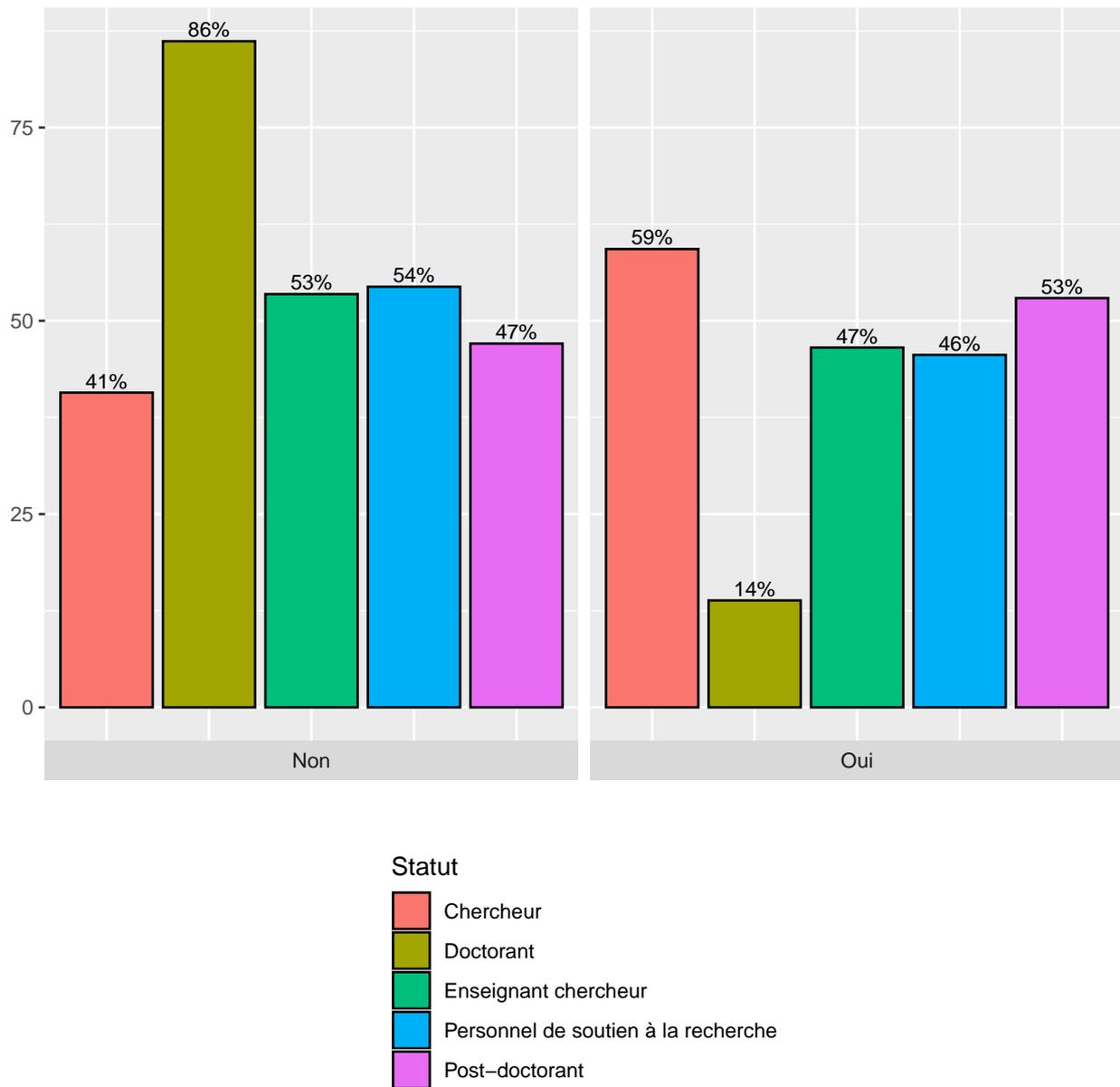


FIGURE 28 – Statut des personnes qui déposent des données

## 6.5 Diffusion et volumétrie

La figure 29 corrèle les réponses sur la diffusion des données aux volumétries déclarées. On constate que le volume des données n'est pas un frein pour la diffusion des données : il semble même que les volumes de données déclarés supérieurs à 100 Go et notamment supérieurs à 1 To soient diffusés plus fréquemment (respectivement 44 et 55%).

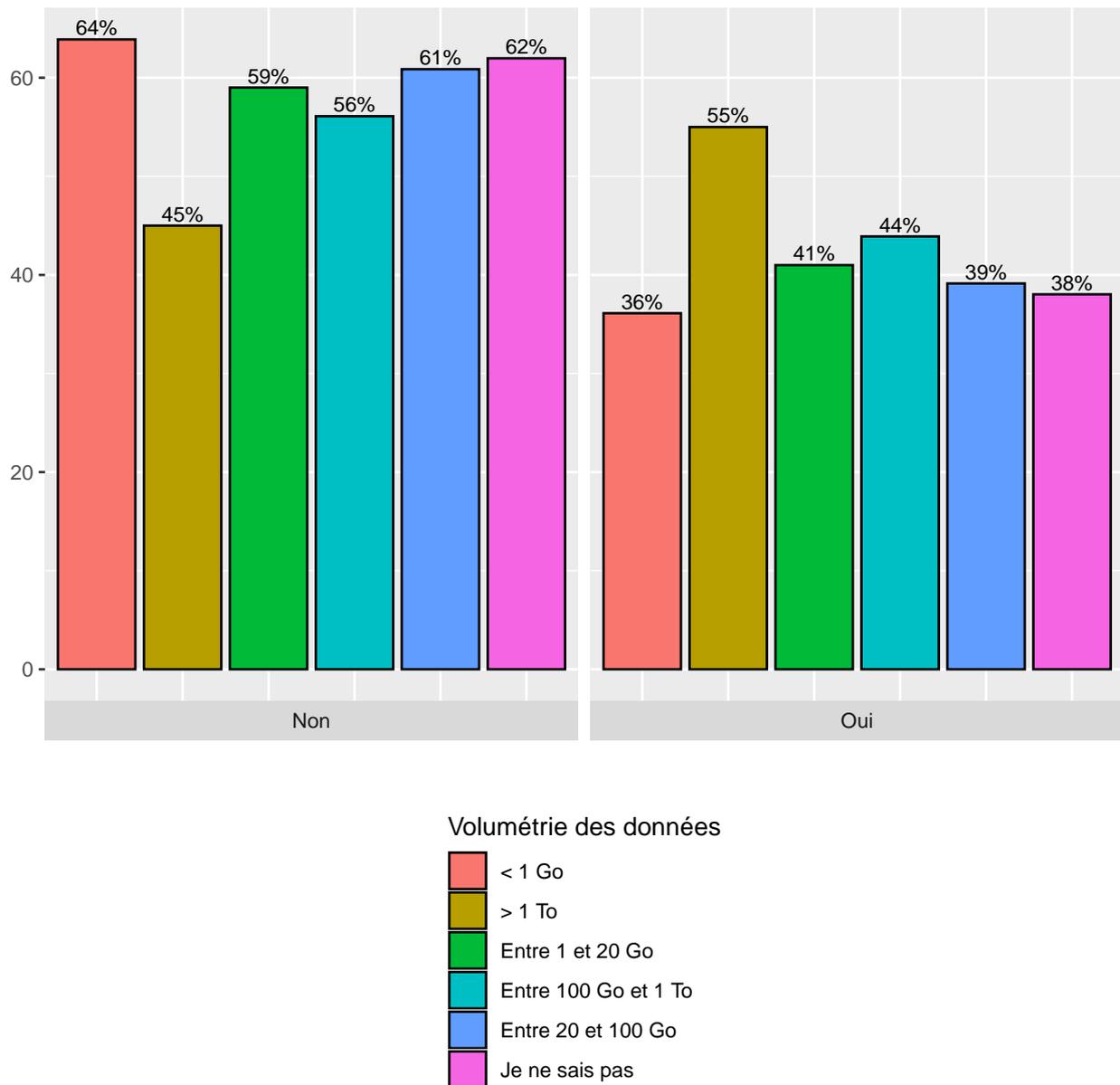


FIGURE 29 – Diffusion des données par volumétries manipulées

## 6.6 Mode de diffusion

Il est notable de constater que 19% des répondants n'arrivent pas à qualifier leur mode de diffusion des données : il leur est peut-être difficile d'identifier s'il s'agit d'entrepôts généralistes, thématiques, etc. Une connaissance plus précise des différentes modalités de diffusion serait sans doute à développer.

Le dépôt dans des entrepôts (thématiques ou généralistes) ou la rédaction de datapapers semble peu répandue.

La proportion est exactement la même entre dépôt dans un entrepôt généraliste et un entrepôt thématique (17%). Dans les commentaires, Zenodo se détache nettement dans les entrepôts généralistes. Sans surprise le dépôt dans des entrepôts thématiques bien identifiés semble plus fréquent dans certaines disciplines notamment sur les études génétiques (Proteomics Identifications Database, Protein Data Bank, European Genome-phenome Archive), en astronomie et astrophysique (CDS) et en sciences de l'environnement (RESIF, PANGEA).

On constate qu’une très large majorité des répondants (45%) ont utilisé d’autres modes de diffusion que les entrepôts ou les datapapers.

Les commentaires permettent de préciser : beaucoup associent leurs données à une publication par le biais des supplementary materials. Cette pratique montre l’attachement à la validation par les pairs, à l’intégrité scientifique et correspond aussi aux exigences des éditeurs qui sont de plus en plus nombreux à demander l’accès aux données liées à un article.

Dans les commentaires, les échanges directs vers collègues sont considérés également comme un mode de diffusion des données ; un autre mode de diffusion est assez fréquemment cité : gitlab/github, ce qui correspond aux pratiques disciplinaires.

Certains répondants choisissent d’utiliser plusieurs modalités de diffusion en parallèle : git + site web, publication et gitlab, gitlab et hal, gitlab et zenodo, etc. HAL est également mentionné ainsi que les sites web.

## 6.7 Aide à la diffusion

Une forte proportion (près de 40%) des répondants déclarent être intéressés par une aide à la diffusion des données, ce qui est très proche du pourcentage des répondants ayant déjà diffusé leurs données. Le pourcentage de réponse des « je ne sais pas » (39%) est sans doute à mettre en regard également avec la forte proportion de répondants n’ayant jamais diffusé des données.

L’aide au choix d’un entrepôt intéresse 64% des répondants et 54% souhaitent une aide directe au dépôt, ce qui montre une répartition assez équilibrée. Il était possible de préciser d’autres types d’aide : aide juridique (RGPD, propriété intellectuelle) ; aide sur l’anonymisation et la sécurisation des données ; aide à la structuration des données ; aide sur le lien entre codes et données pour assurer une certaine reproductibilité.

## 6.8 Synthèse

La réutilisation de données est une pratique assez courante, en particulier dans certaines disciplines mais elle s’accompagne de difficultés liées à différents facteurs : des données difficiles à trouver, non accessibles, peu interopérables et donc en définitive malaisément ré-utilisables. On retrouve ainsi l’importance des principes FAIR.

Par ailleurs, la diffusion des données n’est pas encore une activité intégrée dans le processus de recherche, en particulier pour certains pôles de recherche, mais également chez les jeunes chercheurs.

Ces constats montrent qu’il est donc essentiel de proposer une aide et des outils adaptés (entrepôt de données, conseils sur les standards de métadonnées, sur les formats etc.) pour permettre à toutes les communautés de diffuser leurs données selon les principes FAIR, mais aussi de renforcer les formations sur ces questions, en particulier au niveau des écoles doctorales.

## 7 Archivage

La notion d'archivage est complexe et elle a vraisemblablement été peu comprise dans l'enquête. Les termes stockage, stockage sécurisé et archivage ont donné lieu à des réponses qui laissent à penser que ces concepts sont loin d'être clairs pour la majorité des communautés scientifiques.

### 7.1 Analyse des réponses

Si on regarde les pratiques d'archivage par pôle de recherche (figure 30), on s'aperçoit que la majorité des répondants, quelque soit leur pôle de rattachement, n'ont aucune pratique d'archivage. Cependant, pour certaines communautés, cet archivage se fait sur les plateformes établissement (en particulier pour CBS et PEM), et minoritairement sur les plateformes nationales (en particulier PAGE et SHS).

L'étude des réponses ouvertes concernant le nom des plateformes utilisées montre une certaine confusion entre stockage et archivage :

- les plateformes établissement citées sont SUMMER, la plateforme du CEA, les serveurs de laboratoire, le cloud UGA, la plateforme gitlab, des disques durs externes, des entrepôts de données, HAL,
- les plateformes nationales citées sont des entrepôts comme RESIF, le CDS, THEIA ... Peu de réponses laisse penser que la question a été bien comprise : quelques mentions quand même du CINES, d'HumaNum par exemple.

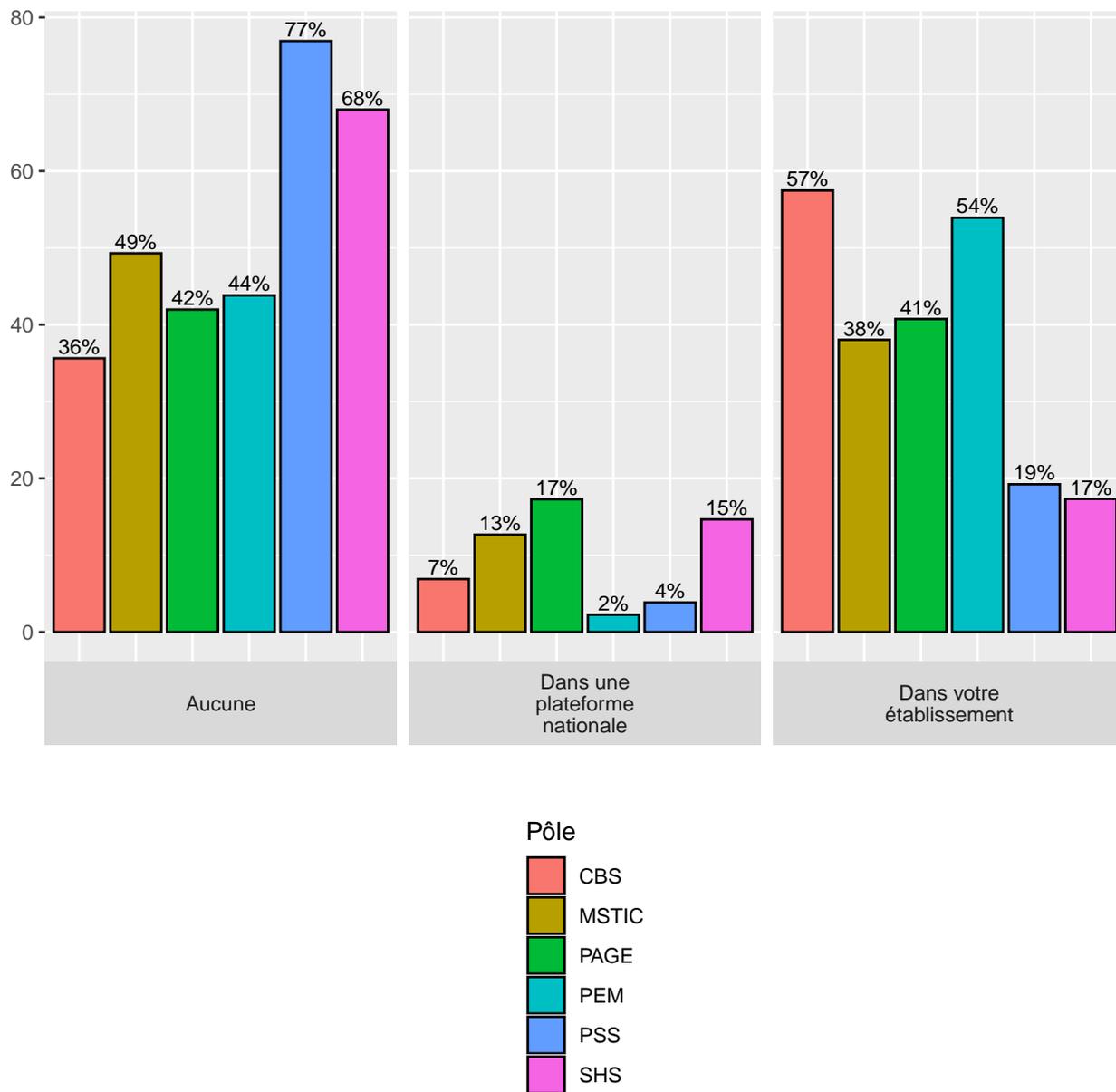


FIGURE 30 – Pratique d’archivage des données par pôle de recherche

Les réponses sur l’archivage corrélées à l’expérience en recherche (figure 31) permettent de constater que les personnes plus expérimentées ont une pratique plus importante de l’« archi- vage », notamment sur les plateformes nationales.

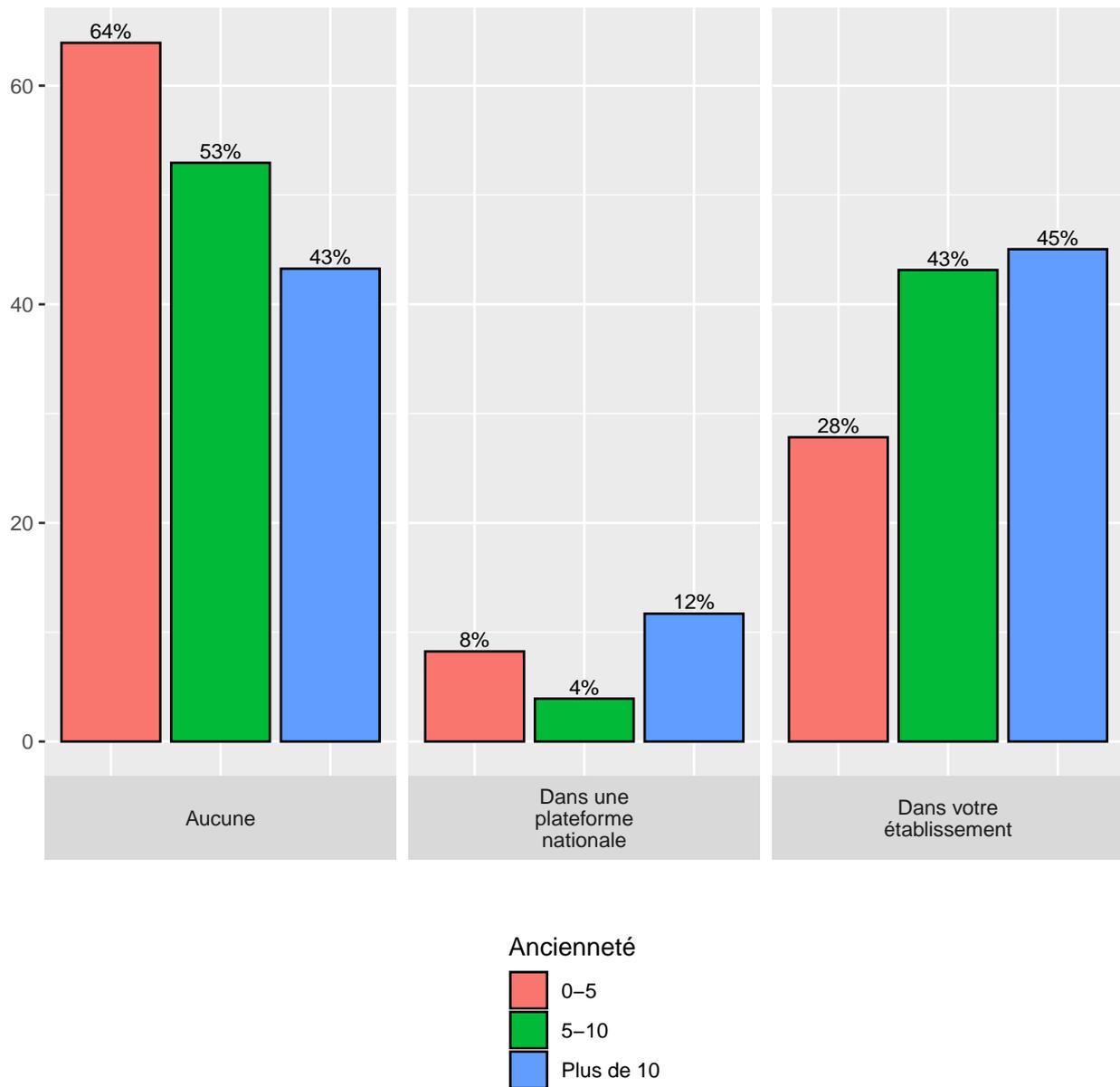


FIGURE 31 – Pratique d’archivage des données par expérience en recherche

Il y a une attente importante en terme d’accompagnement sur ce sujet, quelque soit le pôle de recherche. La présence non négligeable de réponse « Je ne sais pas » renforce l’idée que cette notion d’archivage n’est pas comprise.

L’aide attendue concerne aussi bien le choix de la plateforme (majoritairement) que l’accompagnement à l’archivage de données lui-même (voir figures 32 et 33).

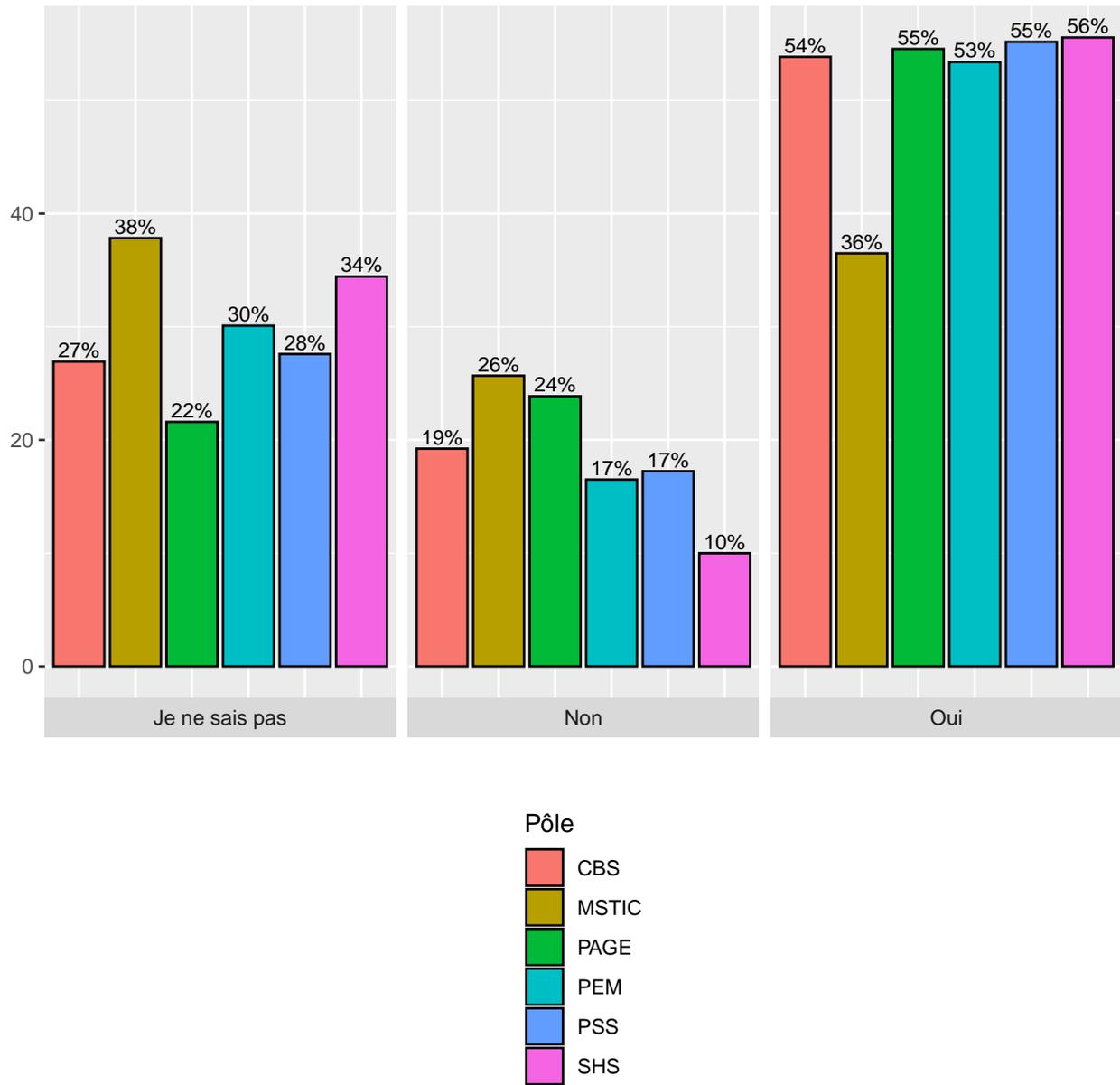


FIGURE 32 – Besoins d’accompagnement sur l’archivage des données par pôle de recherche

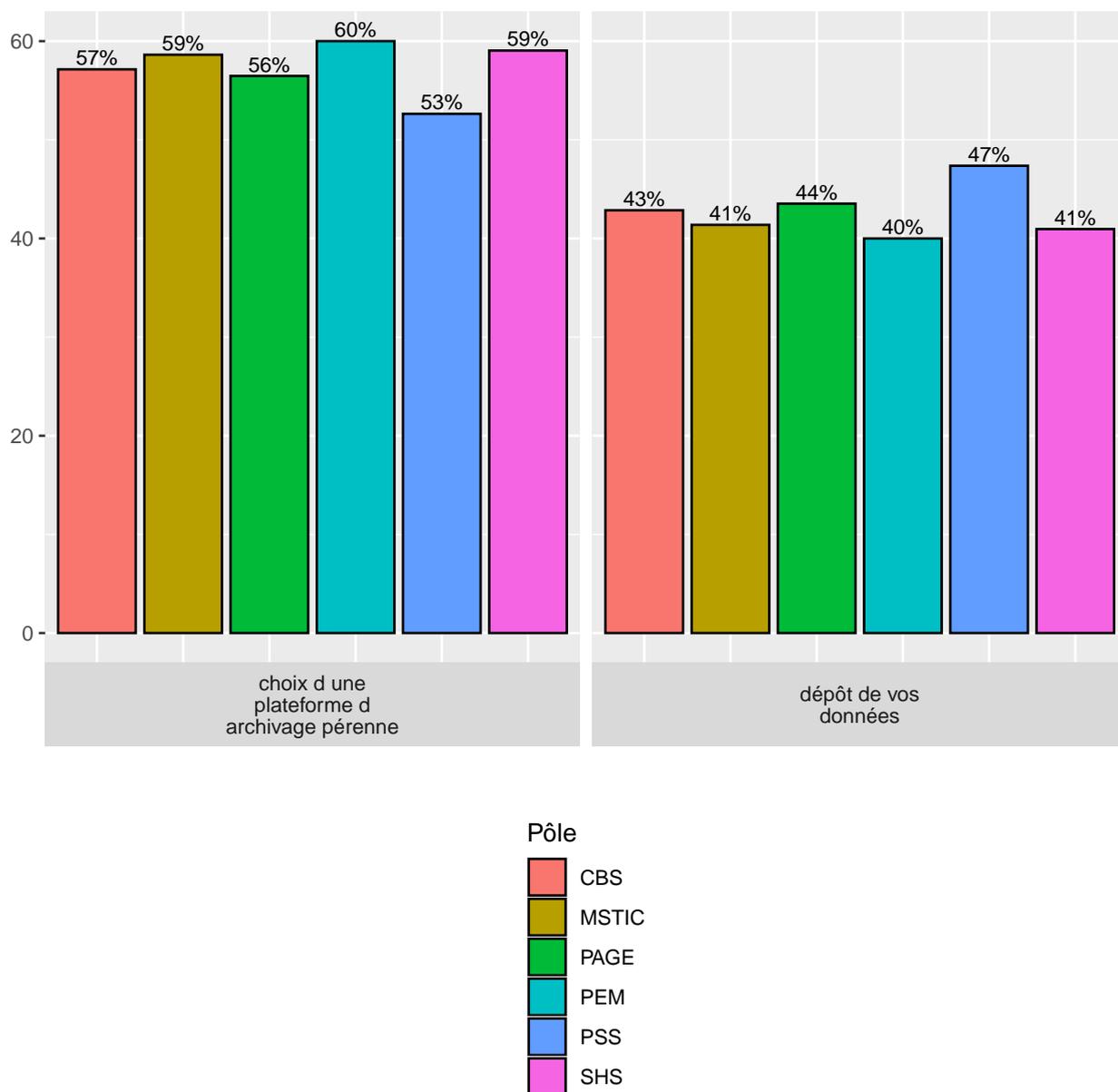


FIGURE 33 – Besoins d’accompagnement sur le choix de plateformes d’archivage et de dépôt des données par pôle de recherche

## 7.2 Synthèse

Les réponses aux questions concernant l’archivage, et en particulier les réponses libres, nous permettent de mesurer un manque d’information très majoritairement partagé sur ces concepts.

Il est donc indispensable de mettre en place des actions de sensibilisation et d’information. Cependant, il est aussi important de noter qu’il existe peu de solutions actuellement aux besoins d’archivage de données, que ces solutions ne sont pas réellement accessibles au chercheur dans son laboratoire, et pas ou peu d’expertise d’archiviste pour accompagner les communautés de recherche.

Il faut donc initier un travail sur le besoin réel, identifier les compétences sur le site et au niveau national, référencer les possibilités et mettre en place au niveau de l’établissement une réelle politique pour l’archivage des données de recherche.

Pour éclairer sur la notion d’archivage, le CINES propose des informations particulièrement

pertinentes : <https://www.cines.fr/archivage/un-concept-des-problematiques/>.

## 8 Plans de Gestion des Données

L'objectif est d'identifier le niveau de connaissances sur les plans de gestion des données (PGD) et les besoins exprimés associés.

### 8.1 Niveau d'appropriation des Plans de Gestion de Données

Un peu plus de la moitié des répondants (53%) ont déjà entendu parler des PGD mais seul 15% parmi eux savent précisément ce que cela recouvre.

Le croisement de ces réponses par pôle de recherche (figure 34) apporte des éléments intéressants sur les différences entre disciplines :

- Deux pôles, PAGE et PSS, se distinguent par leur connaissance du PGD, ainsi que le pôle CBS dans une mesure un peu moins importante.
- Pour les pôles MSTIC, PEM et SHS, les personnes connaissant les PGD sont minoritaires.

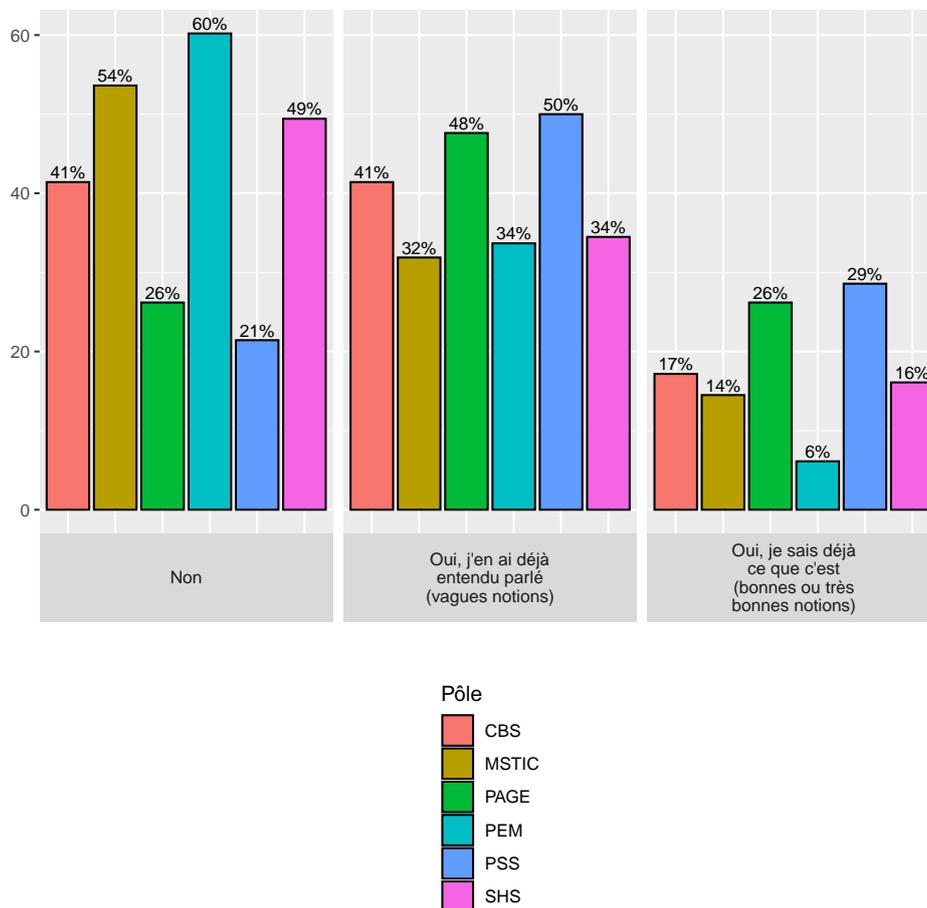


FIGURE 34 – Connaissance des Plans de Gestion de Données par pôle de recherche

Si on considère la fonction des répondants (figure 35), le personnel de soutien à la recherche a un bon niveau de connaissances sur le PGD : 75% savent ce qu'est un PGD avec 39% des répondants connaissant bien les PGD, tandis que les répondants dépendant d'autres statuts sont en dessous de 15% pour cette dernière catégorie. Les chercheurs et post-doctorants savent à 63% et 60% ce qu'est un PGD, tandis que les enseignants chercheurs et les doctorants n'arrivent qu'à 47% et 42%.

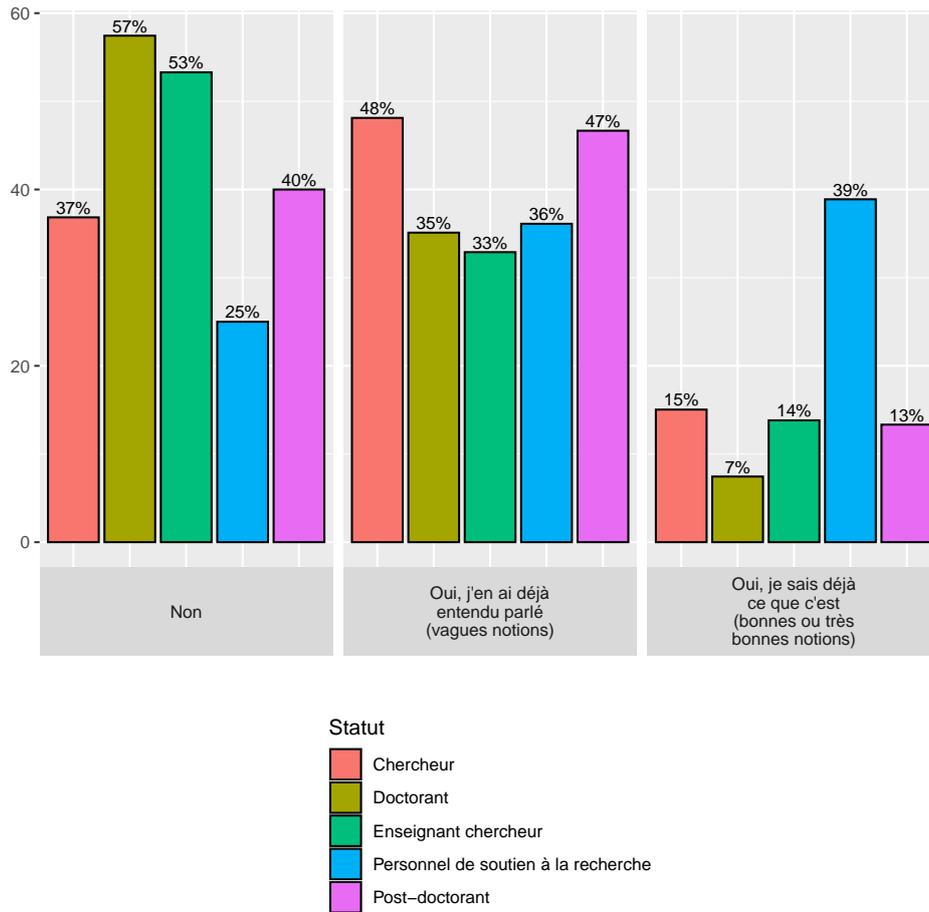


FIGURE 35 – Connaissance des Plans de Gestion de Données par statut

Enfin, la corrélation avec l'ancienneté de l'expérience en recherche (figure 36) montre que les moins expérimentés connaissent moins bien les PGD, et que ce sont ceux avec une ancienneté de 5 à 10 ans qui le maîtrisent le mieux.

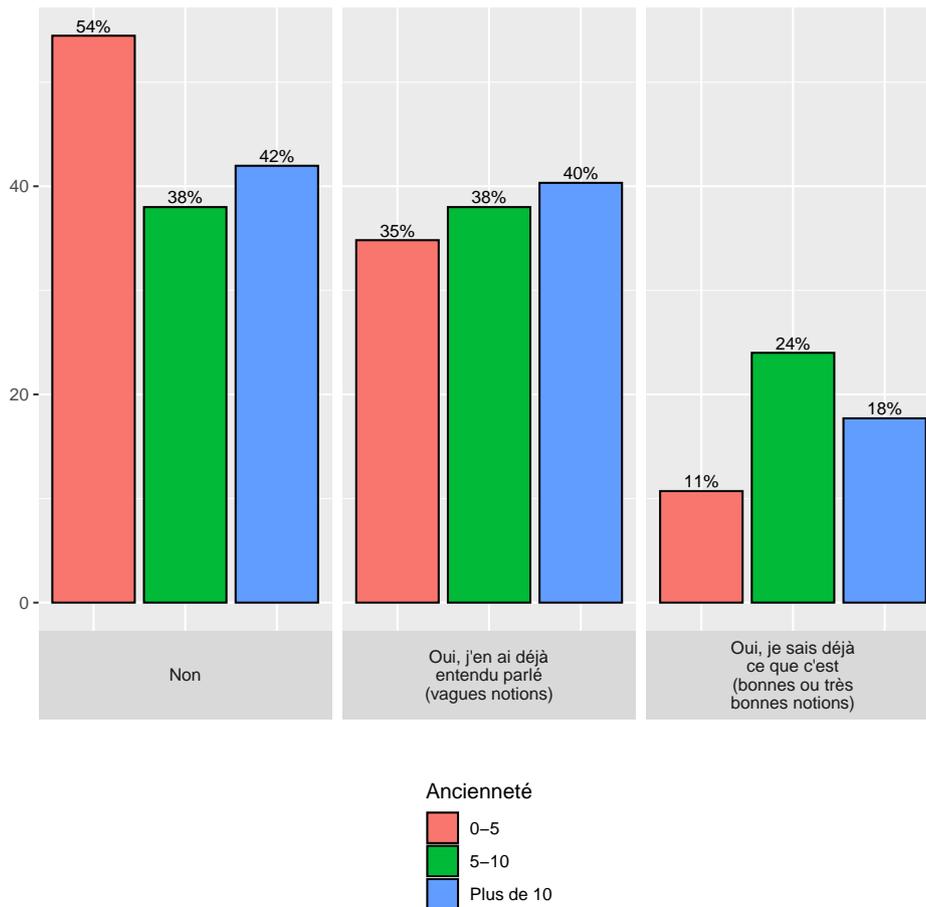


FIGURE 36 – Connaissance des Plans de Gestion de Données par ancienneté

## 8.2 Formations, expérience de rédaction et difficultés rencontrées

La majorité des répondants (71%) n'ont pas suivi de formation sur les PGD. De même 77% des participants à l'enquête n'ont jamais participé à un projet nécessitant la rédaction d'un PGD. Une majorité (55%) répondent ne pas savoir si les futurs projets de recherche dans lesquels ils participeront nécessiteront un PGD. Cela renforce le constat du manque de connaissances et d'informations sur ce sujet puisque la plupart des financeurs exigent désormais un PGD.

Si on s'intéresse aux réponses des participants ayant déjà été confrontés à la rédaction d'un PGD, soit 14% des répondants, 36% disent ne pas avoir rencontré de problème particulier, et 30% mentionnent quelques difficultés :

- un certain désarroi face à quelque chose qui n'est pas compris,
- le manque d'exemples disponibles,
- la problématique des questions éthiques et juridiques pour lesquelles il faut des expertises spécifiques,
- un manque d'expérience sur le sujet etc.

### 8.3 Synthèse

Les plans de gestion de données sont encore largement méconnus de la majorité des répondants, et très peu ont été réellement confrontés à leur rédaction.

Cette tendance est en train de s'inverser très rapidement avec les exigences des financeurs de la recherche.

Les difficultés mentionnées pour la rédaction tendent aujourd'hui à s'estomper, en grande partie grâce à des outils comme DMP-Opidor et la possibilité d'accéder à une variété d'exemples de plus en plus large.

Il n'empêche que les besoins de formation et d'accompagnement sont importants afin d'aller vers une appropriation un peu générale de cet outil.

## 9 Personnes ressources pour l'aide à la gestion des données

L'objectif des questions associées aux personnes ressources pour l'aide à la gestion des données était de voir si les répondants savaient qui contacter sur cette problématique.

### 9.1 Analyse

Presque la moitié des répondants (45%) n'identifient pas de personne ressource pour les aider dans la gestion des données.

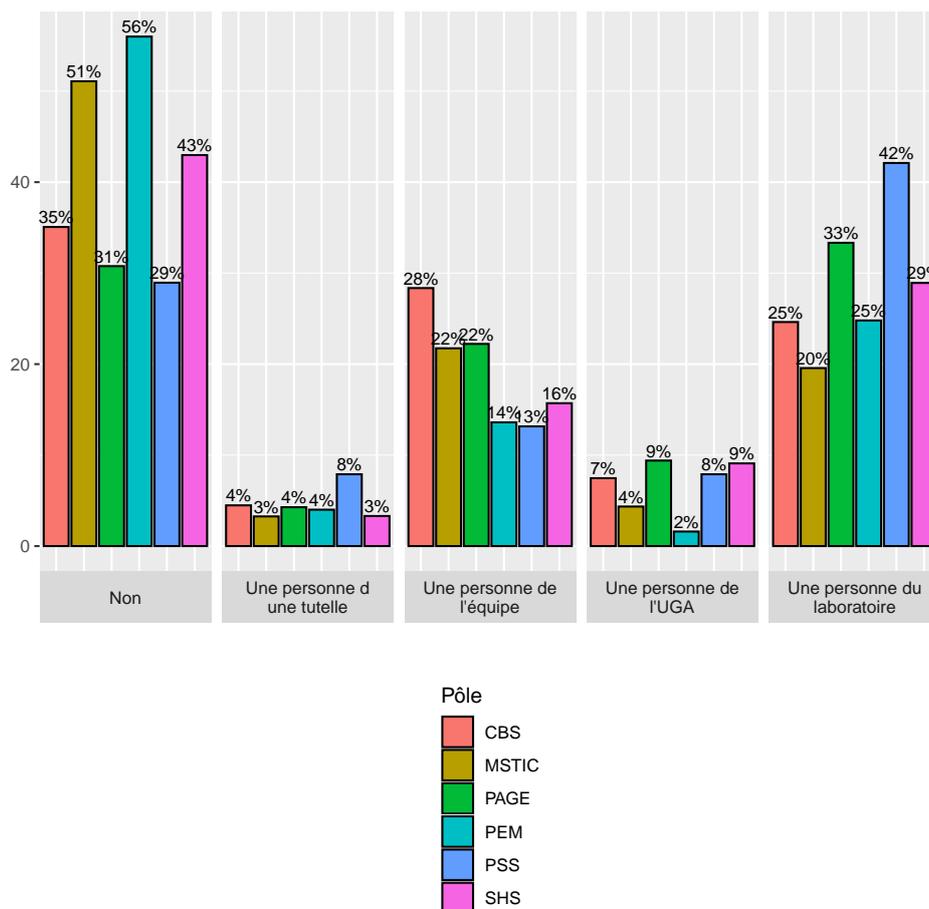


FIGURE 37 – Capacité à identifier une personne ressource par pôle de recherche

Selon la figure 37, il semble que les répondants issus des pôles PSS, PAGE et CBS soient plus nombreux à identifier une personne ressource.

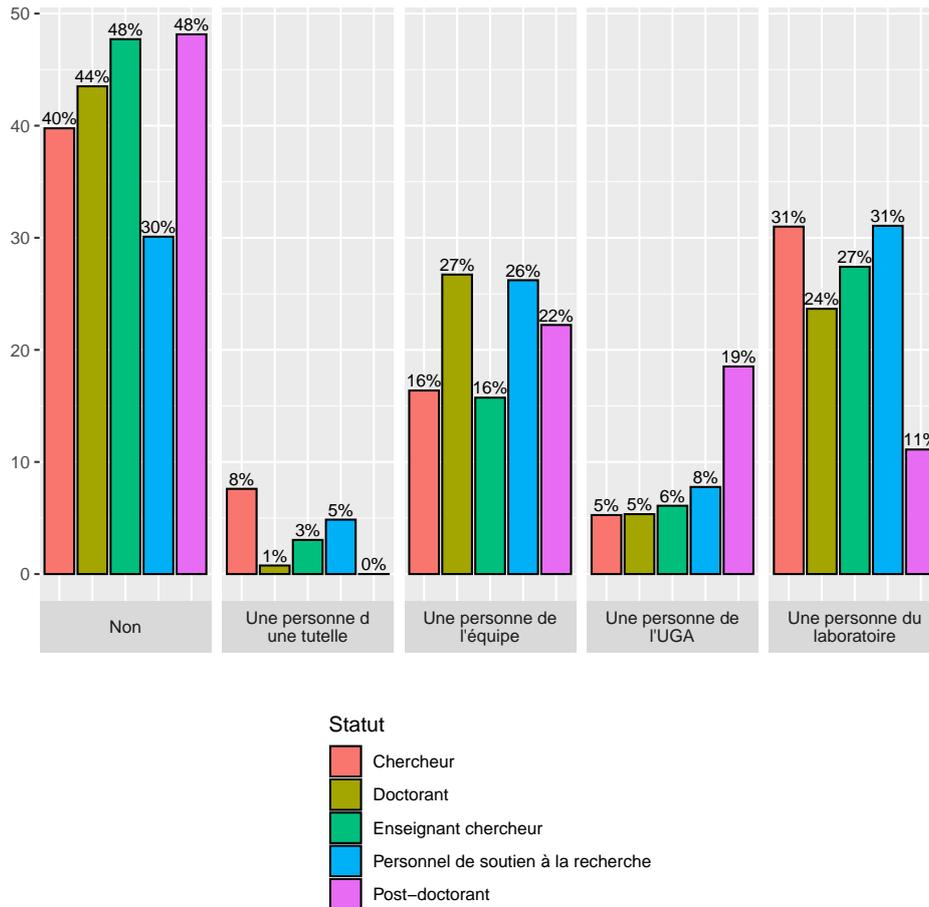


FIGURE 38 – Capacité à identifier une personne ressource par statut

La figure 38 montre que les doctorants ont recours plus largement à une personne de leur équipe.

Les répondants recourent principalement à une aide interne au laboratoire ou à l'équipe. La plupart du temps, et sans surprise, ce sont les services informatiques qui sont les plus sollicités avec des profils divers, datascientists, gestionnaires de BDD, responsable de sécurité informatique, etc.). Il peut s'agir également du directeur de laboratoire ou du responsable de l'équipe ou du programme de recherche, de post-doctorants.

Les doctorants citent évidemment leurs directeurs ou encadrants de thèse. Les documentalistes peuvent également être un recours. Ces pratiques semblent donc induire qu'il est important pour les répondants de disposer d'un recours de proximité.

On peut souligner également la place des réseaux, formels ou informels, au sein de l'UGA ou hors UGA : d'autres collègues chercheurs sont alors identifiés comme recours possibles. Il s'agit par exemple de personnes travaillant sur des données proches, des experts dans la discipline, d'autres doctorants. Cette forme d'entraide est intéressante à souligner. Parmi les répondants, quatre se déclarent personnes ressources pour leur laboratoire.

La cellule data ou la cellule HAL sont mentionnées par quelques personnes : si elles commencent à être connues, la proportion de répondants les citant reste faible.

## 9.2 Synthèse

Globalement, le constat est plutôt un manque important d'information sur les possibilités offertes d'accompagnement et d'aide sur la gestion des données.

Cet état des lieux est problématique et semble montrer la nécessité :

- de renforcer le réseau des référents données dans les laboratoires, et de faire en sorte que ceux-ci soient bien identifiés par leurs collègues,
- d'accroître la communication sur la Cellule Data Grenoble Alpes qui apporte un service d'aide à la gestion des données.

Plusieurs actions sont possibles pour améliorer le niveau d'information, en particulier les interventions dans les laboratoires (au plus près des communautés), l'organisation d'actions d'animation et de formation, et la diffusion d'informations via différents canaux : réseaux métiers (comme SARI, donc auprès des informaticiens de laboratoires qui peuvent relayer dans leurs unités), collège doctoral, directions de laboratoires ...

## 10 Besoins exprimés

Les participants ont été invités à préciser leurs besoins, tant en terme d'accompagnement qu'en terme de services pour la gestion de leurs données, en priorisant des propositions qui leur ont été soumises.

### 10.1 Besoins sur l'accompagnement et la formation

Le besoin exprimé en priorité 1 (voir figure 39) est très majoritairement générique, tant sur la partie formation que sur l'accompagnement, avec également des demandes autour du stockage, de l'archivage et de l'aide sur les aspects juridiques.

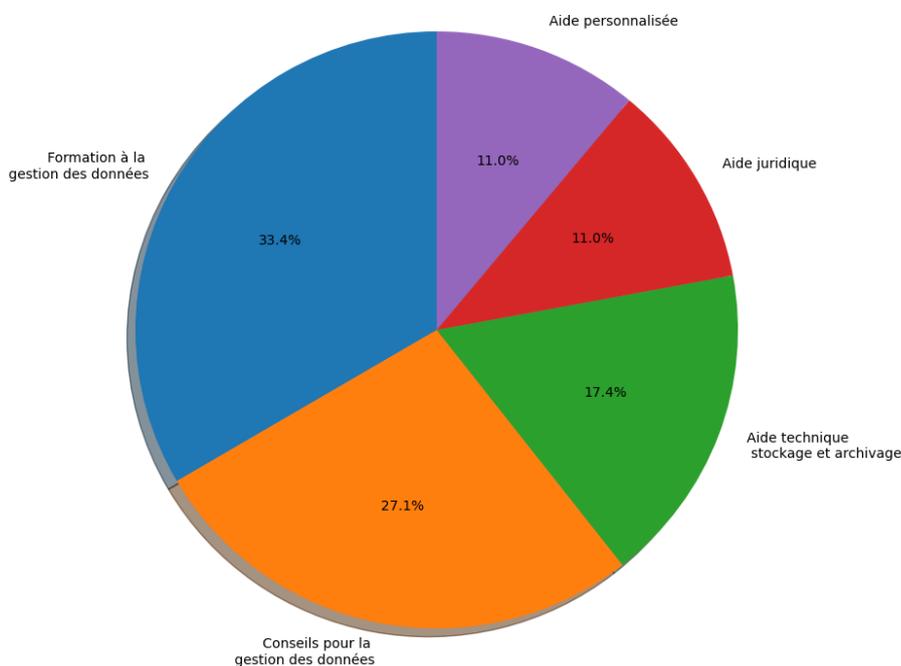


FIGURE 39 – Besoins en matière d'accompagnement et formations

Les réponses proposées en priorité suivante font apparaître en complément des besoins sur de l'accompagnement pour la publication et la citation de données, ainsi que pour le choix d'un entrepôt de données.

Enfin, de façon un peu moins prioritaire apparaissent les demandes sur de l'aide à la documentation et la description des données ainsi qu'à la rédaction de plans de gestion des données.

En complément des réponses qui étaient proposées, d'autres demandes ont été exprimées, concernant le besoin d'identifier les personnes ressources sur ces problématiques, et un point d'entrée unique pour l'accompagnement. Ces points recourent les constats évoqués dans la partie 9.

## 10.2 Besoin sur les services numériques

Concernant les besoins de priorité 1 en terme de services numériques (voir figure 40), les participants expriment la nécessité de disposer d'outils pour le stockage en général, avec différentes fonctionnalités, et le travail collaboratif.

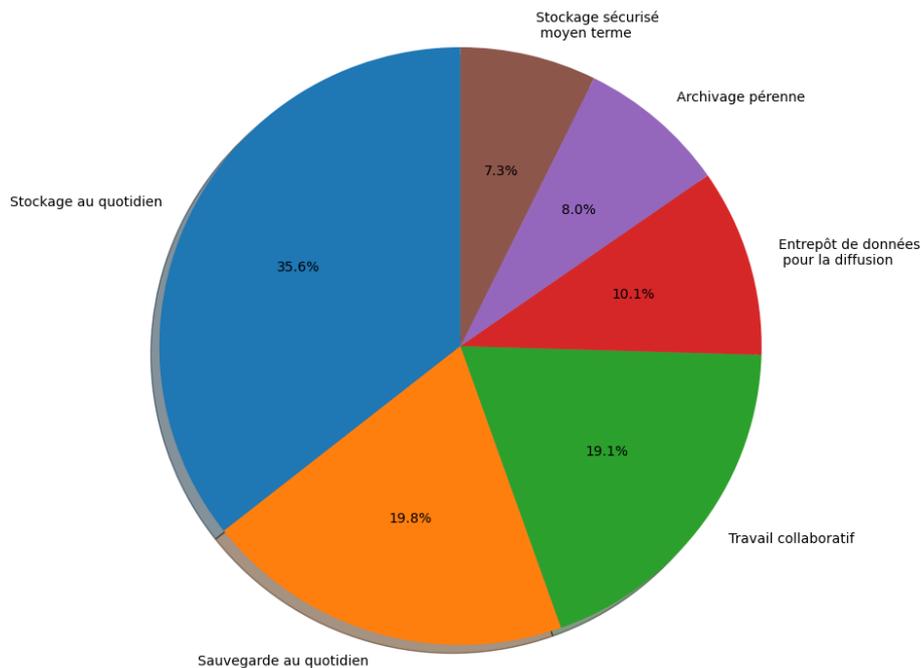


FIGURE 40 – Besoins en matière de services numériques

Parmi les réponses libres, plusieurs demandes sont exprimées, en particulier autour de plateformes d'expérimentation en ligne (enquêtes en ligne type sphinx ou limesurvey), de cahiers de laboratoire électroniques ou d'éditeur collaboratif (type overleaf).

Un besoin semble récurrent et concerne le partage de données, éventuellement de grande taille, avec des collaborateurs extérieurs.

## 10.3 Besoins sur les entrepôts de données

Les dernières questions concernaient les entrepôts de données, et leurs caractéristiques nécessaires pour une bonne adéquation avec les besoins.

La figure 41 montre les spécificités qui arrivent en priorité 1.

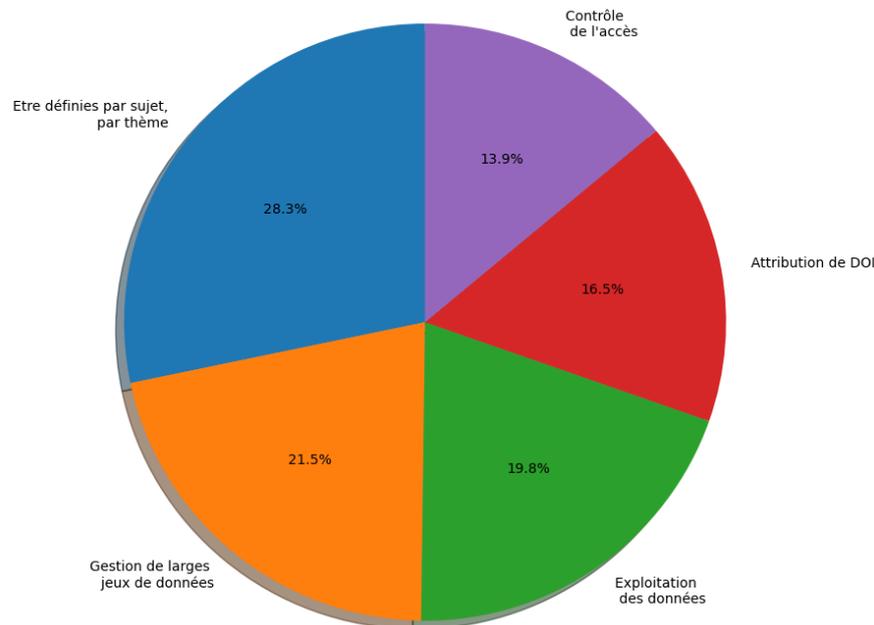


FIGURE 41 – Caractéristiques de l'entrepôt de données

Les besoins exprimés sont assez partagés et proches les uns des autres. En priorité 2 et suivantes, dans un ordre de grandeur similaire, sont indiqués les possibilités de versioning et le lien entre publications et jeux de données.

Les réponses libres apportent en complément des demandes sur le lien avec d'autres ressources comme les notebooks ou les cahiers de laboratoire électroniques.

## 10.4 Synthèse

Les réponses des participants aux questionnaires illustrent très clairement un besoin important d'accompagnement et de formations sur la gestion des données en général dans toutes ses ramifications.

Le stockage, dans toute sa complexité (stockage, sauvegarde, archivage, entrepôt de données mais aussi partage de données) représente également un des points sur lesquels il est indispensable de proposer à la fois du conseil mais aussi des infrastructures adaptées.

Les efforts initiés dans les actions de sensibilisation et de formation sont évidemment à poursuivre et à amplifier dans les domaines où les besoins sont exprimés.

Un travail important reste à réaliser pour faire connaître les services numériques existants et faire en sorte qu'ils soient utilisables sans expertise technique trop forte.

Enfin, l'évolution des plateformes et la mise en place de nouveaux services doivent impérativement suivre les besoins exprimés par les communautés scientifiques. On peut par exemple citer le service de cahiers de laboratoire électronique qui a été déployé depuis la clôture de cette enquête, ou la participation active de l'UGA au projet Recherche Data Gouv qui permettra de

disposer d'une collection UGA sur cette plateforme nationale.

## 11 Focus doctorants

L'objectif de cette partie est de se concentrer sur les réponses issues des doctorants afin de bien comprendre leurs usages et leurs besoins spécifiques.

### 11.1 Profils

Les pôles CBS, MSTIC, PAGE et PEM (entre 13 et 19% des répondants doctorants) sont à peu près représentés de façon équivalentes dans les répondants doctorants. Le pôle SHS est plutôt sur-représenté (30%) alors que le pôle PSS l'est beaucoup moins (7%).

Cette répartition a un impact sur certaines des analyses proposées.

### 11.2 Type de données

Les types de données manipulées sont très variés.

Si on regarde les réponses par pôle de recherche (voir figures 42 et 43), on retrouve certaines tendances générales :

- les simulations et codes sources sont utilisées plutôt parmi les doctorants des pôles MSTIC, PAGE, PEM,
- les textes, les données web, les données de réseaux sociaux et les enregistrements numériques audio et vidéo sont utilisés plutôt parmi les doctorants du pôle SHS,
- les données qualitatives sont souvent mentionnées par les doctorants du pôle PSS,
- les données de mesures ou d'observations sont citées plus spécifiquement par les doctorants des pôles PAGE et PEM.

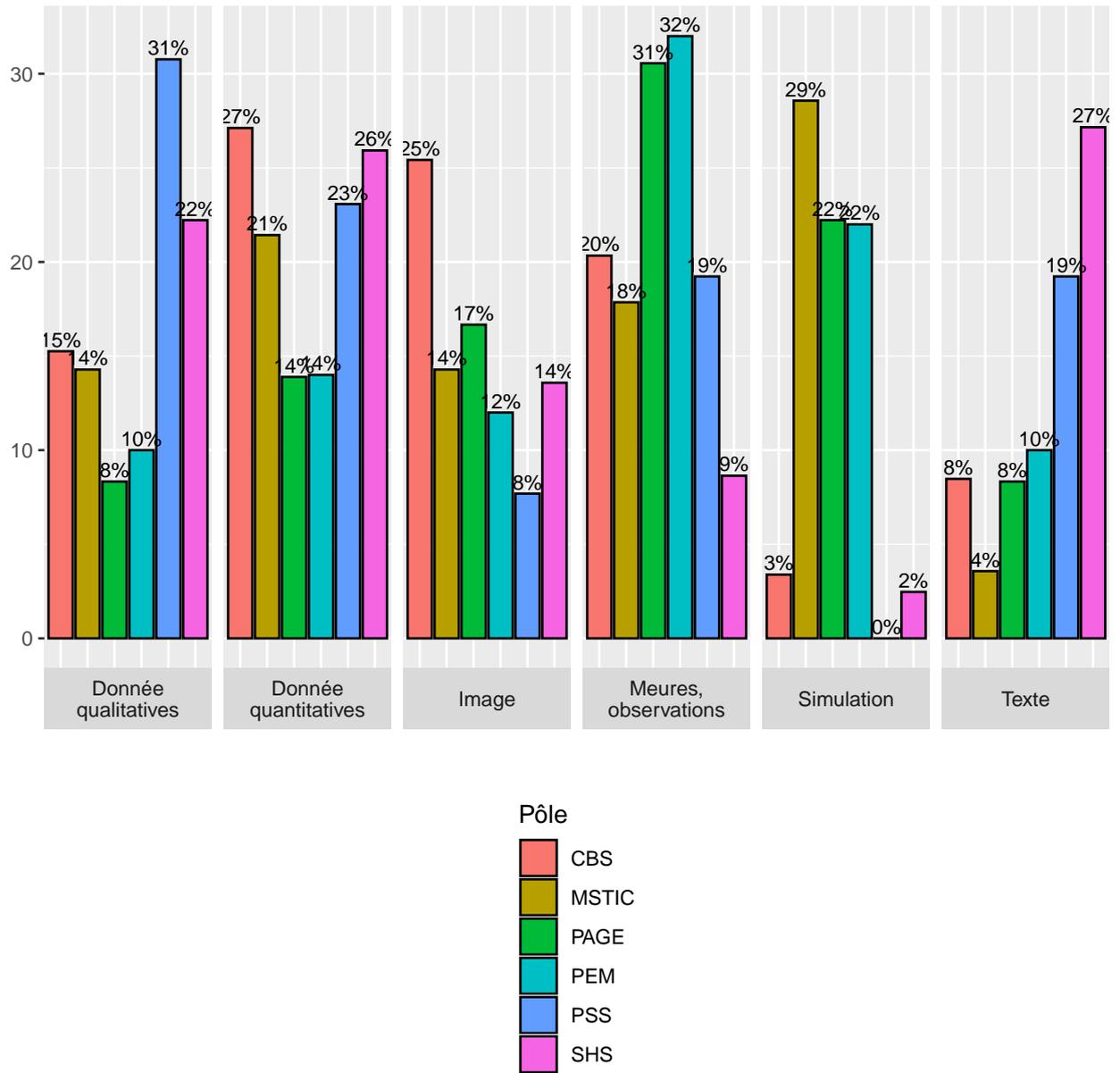


FIGURE 42 – Types de données par pôle de recherche pour les doctorants (partie 1/2)

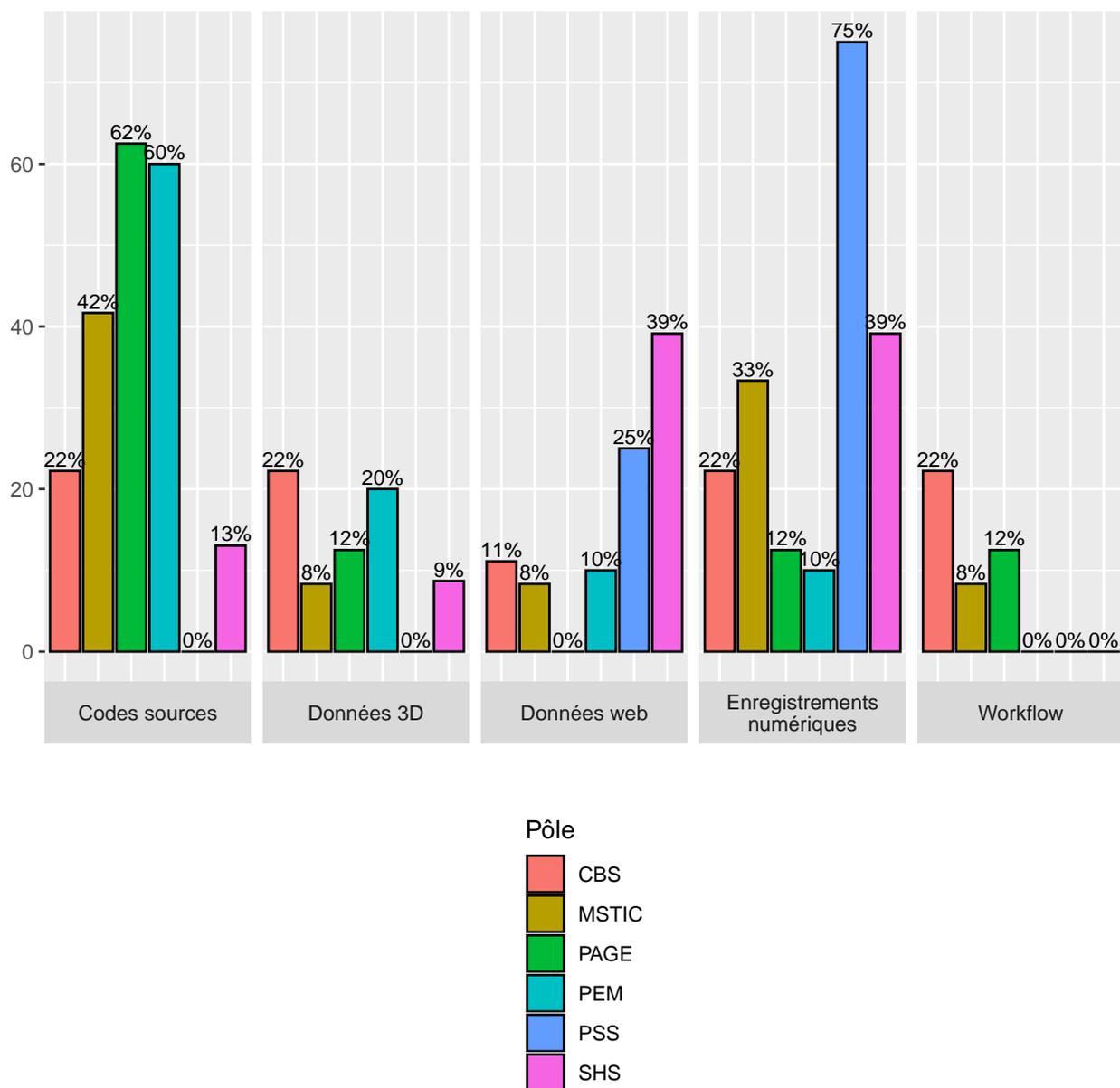


FIGURE 43 – Types de données par pôle de recherche pour les doctorants (partie 2/2)

### 11.3 Pratiques durant la recherche

Environ 65% des doctorants ont déclaré avoir déjà réutilisé des données publiées par d'autres chercheurs. Il y a une répartition entre tous les pôles avec une légère majorité dans les pôles PEM et SHS puis CBS, de l'ordre de 20%, les autres pôles étant à environ 10%.

Parmi les doctorants qui ont réutilisé des données, la très grande majorité (88%) a affirmé avoir eu des problèmes, en particulier :

- données inexistantes ou impossible à récupérer sur demande,
- problème d'accès aux données,
- problème pour retrouver des données,

et ce quel que soit leur pôle de rattachement.

## 11.4 Personnes ressources

Environ 44% des doctorants ont répondu ne pas connaître de personne ressource pour les aider. Les autres sollicitent quelqu'un de l'équipe ou du laboratoire (27% et 24% respectivement). Ces réponses sont assez similaires quelque soit le pôle d'appartenance.

Ce qui ressort des réponses ouvertes est que les doctorants demandent surtout de l'aide à leurs directeurs de thèse et aux collègues de leur propre équipe ou du laboratoire.

## 11.5 Stockage

Dans tous les pôles, l'utilisation d'ordinateurs personnels/professionnels et des disques durs externes personnels/professionnels sont les pratiques plus courantes. Si on croise le pôle d'appartenance avec les habitudes de stockage (voir figures 44 et 45) on retrouve :

- Les services de cloud privés (Dropbox, Google Drive ...) sont utilisés de façon non négligeable dans les pôles PSS et SHS.
- Les serveurs dédiés du laboratoire sont utilisés surtout dans les pôles PAGE et PEM.
- L'usage de matériels personnels est assez important, en particulier dans les pôles SHS, PSS et dans une moindre mesure PEM.

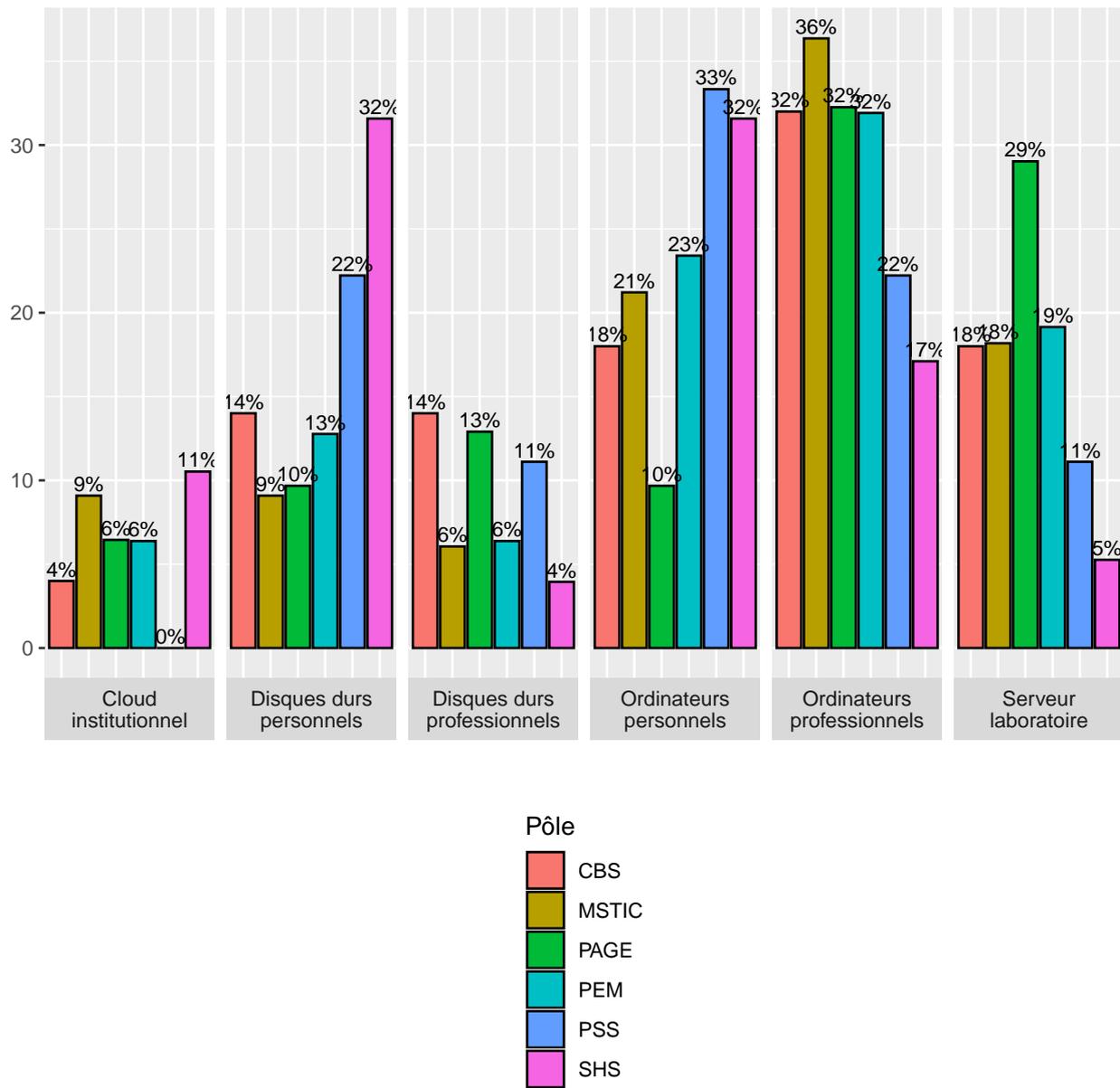


FIGURE 44 – Lieux de stockage par pôle de recherche pour les doctorants (partie 1/2)

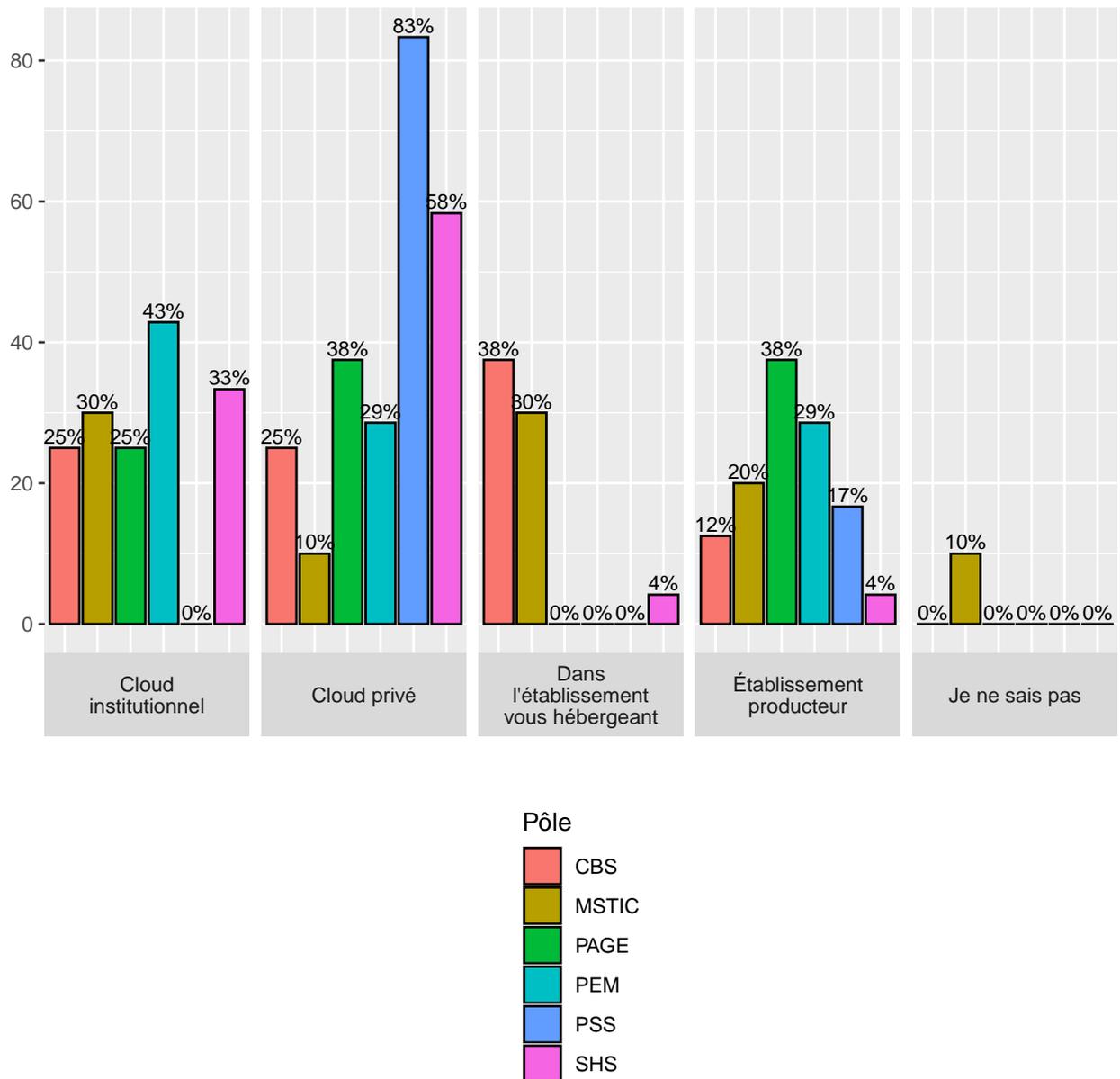


FIGURE 45 – Lieux de stockage par pôle de recherche pour les doctorants (partie 2/2)

## 11.6 Pratiques collaboratives

Environ 70% des répondants doctorants travaillent dans le cadre de projets collaboratifs, 37% sur des collaborations ponctuelles et 33% sur des collaborations régulières.

Le pôle SHS est celui où les répondants ont déclarés le plus de travaux individuels (54%).

## 11.7 Besoins exprimés

Les besoins exprimés par les doctorants, tant en termes d'accompagnement qu'en terme de services pour la gestion de leurs données, sont très variés.

Dans l'analyse des réponses, on trouve que la priorité est la formation en termes de gestion, de publication et de citation de données. Arrivent ensuite l'aspect stockage et archivage de données et les problématiques juridiques. Il n'y a pas réellement de différences entre les disciplines.

Les solutions numériques dont les doctorants ont besoin pour la gestion de données en priorité sont celles associées au stockage, à la sauvegarde et à l'archivage pérenne.

Concernant les caractéristiques pour un entrepôt, les priorités sont surtout :

- la définition par sujet et par thème,
- la certification,
- l'attribution de DOI,
- l'exploitation des données.

Si on croise le pôle d'appartenance avec les besoins concernant l'entrepôt on trouve :

- la définition par sujet et par thème, ainsi que la gestion de larges jeux de données (> 1 To) est prioritaire pour les doctorants de CBS et SHS,
- la gestion de larges jeux de données (> 1 To) et la possibilité de versioning est prioritaire pour les doctorants de MSTIC,
- la certification est importante pour les doctorants de PAGE.

## 11.8 Diffusion des données

La pratique de diffuser ses propres jeux de données n'est pas très courante (14%). Pour ceux qui ont déjà diffusé des jeux de données, les lieux de diffusion cités sont plutôt des articles, ou des plateformes comme gitlab.

## 11.9 Archivage

Les doctorants déclarent avoir besoin d'aide pour archiver les données de façon pérenne. Pour la pratique de l'archivage pérenne des données, les réponses libres sont assez disparates, ce qui montre un besoin de clarification des notions.

## 11.10 Synthèse

L'analyse des réponses des doctorants permet de mieux spécifier leurs usages et leurs besoins.

Même si les grandes tendances sont celles de l'ensemble des réponses au questionnaire, il ressort des spécificités liées :

- au manque d'information sur les différentes ressources auxquels les doctorants peuvent avoir accès, tant sur les possibilités d'accompagnement que sur les plateformes et services disponibles,
- au fort besoin de formation sur toutes les questions liées aux données.

Il est important de noter que les services existants sur le site sont d'abord à destination des membres permanents des laboratoires, et difficilement utilisables par des doctorants.

La mise en place d'un accompagnement spécifique et adapté à cette population est donc crucial.

## 12 Conclusions, perspectives et recommandations

### 12.1 État des lieux

La population des participants à l'enquête couvre bien l'ensemble des disciplines présentes sur le site de l'UGA, ainsi qu'une diversité suffisante de profils différents en terme de fonction et d'expérience professionnelle.

Les pratiques de stockage sont assez fortement disciplinaires. Les grosses volumétries poussent certaines communautés à se tourner, pour certaines depuis longtemps, vers des plateformes institutionnelles ou dédiées.

La problématique des données sensibles concernent directement plusieurs pôles de recherche. Les contraintes techniques et juridiques associées sont importantes. L'aide juridique fait d'ailleurs partie des besoins exprimés les plus importants.

La majorité de la recherche se fait dans un cadre collaboratif avec quelques disparités liés à la discipline, comme pour le pôle SHS, et à la fonction, comme pour les doctorants, exemples pour lesquels on note une tendance vers le travail individuel.

Les méthodes de travail s'accompagnent de l'usage d'outils collaboratifs. L'information la plus importante est sans aucun doute l'utilisation plus ou moins importante d'outils commerciaux.

La réutilisation de données est une pratique assez courante, en particulier dans certaines disciplines mais elle s'accompagne de difficultés liées à différents facteurs : des données difficiles à trouver, non accessibles, peu interopérables et donc en définitive malaisément ré-utilisables. On retrouve ainsi l'importance des principes FAIR.

La diffusion des données n'est pas encore une activité intégrée dans le processus de recherche, en particulier pour certains pôles de recherche, mais également chez les jeunes chercheurs comme les doctorants.

Les réponses aux questions concernant l'archivage, et en particulier les réponses libres, nous permettent de mesurer un manque d'information très majoritairement partagé sur ces concepts.

Les plans de gestion de données sont encore largement méconnus de la majorité des répondants, et très peu ont été réellement confrontés à leur rédaction.

### 12.2 Sensibilisations et information

Un travail important reste à réaliser pour faire connaître les services numériques existants et faire en sorte qu'ils soient utilisables sans expertise technique trop forte.

Les outils institutionnels peuvent être quelque fois difficiles d'accès et pas assez bien connus. Cela pose en particulier des questions autour des enjeux de souveraineté et de protection des données et du patrimoine scientifique lors de l'utilisation de services privés.

Il est donc nécessaire, d'une part de faire connaître les outils existants, d'aider à leur usage, mais aussi d'identifier les freins à leur utilisation. Cela ne pourra se faire que par le biais d'un accompagnement de proximité et de la mise en place d'échanges simples et faciles avec les utilisateurs.

### 12.3 Formations et accompagnements à prévoir

Il est important de s'assurer de la mise en place de formations adaptées aux différentes communautés, afin en particulier de les aider à surmonter la problématique technique associée à l'utilisation des divers outils et plateformes à leur disposition.

La mise en place d'un accompagnement spécifique aux doctorants est cruciale.

Il y a besoin de sensibilisation aux différentes notions liées à la gestion des données, en particulier autour des questions d'archivage.

Les réponses des participants aux questionnaires illustrent très clairement un besoin important d'accompagnement et de formations sur la gestion et le stockage des données durant tout leur cycle de vie.

## 12.4 Nouveaux outils ou services à prévoir

Il est indispensable de pouvoir identifier sur le site les expertises permettant de répondre aux questionnements sur les plans juridique et d'archivage.

L'évolution des plateformes et la mise en place de nouveaux services doivent impérativement suivre les besoins exprimés par les communautés scientifiques. On peut par exemple citer le service de cahiers de laboratoire électronique qui a été déployé depuis la clôture de cette enquête, ou la participation active de l'UGA au projet Recherche Data Gouv qui permettra de disposer d'une collection UGA sur cette plateforme nationale.

Il est fondamental de poursuivre l'amélioration des services académiques afin de détourner les scientifiques de l'usage de clouds privés qui peuvent se révéler très problématique pour la souveraineté des données en particulier dans le cas des données sous RGPD.

## 12.5 Mise en Perspective

Pour rappel, l'enquête avait pour objectif d'identifier les pratiques et les besoins de la communauté scientifique UGA pour pouvoir au mieux adapter les services d'appui à la gestion des données que propose la cellule data Grenoble Alpes.

Depuis la passation de l'enquête, la cellule a pu mettre en place quelques outils pouvant répondre aux besoins exprimés dans l'enquête comme :

- un Site web Science Ouverte UGA qui relaie informations et bonnes pratiques (<https://scienceouverte.grenoble-alpes.fr>)
- un guichet unique, lié à un système de ticket permettant le suivi des demandes, est désormais disponible pour toute question sur les données ([sos-data@univ-grenoble-alpes.fr](mailto:sos-data@univ-grenoble-alpes.fr)),
- un séminaire de formation à l'adresse des doctorants complété par des modules spécifiques concernant le RGPD et la diffusion des données et plus largement, des séminaires organisés régulièrement pour toute la communautés
- un accompagnement à la rédaction de plans de gestion des données pour les porteurs de projet
- etc.

Pour renforcer le développement de la Science ouverte en son sein, l'établissement a par ailleurs constitué une commission Science ouverte et rédige une charte Science ouverte.

Les actions de la cellule data et de l'UGA se construisent enfin en cohérence avec le Plan National Science Ouverte 2021-2024 qui fait de la gestion des données et de la prise en compte des codes et logiciels des axes prioritaires national. Dans ce cadre, l'UGA s'implique notamment dans le projet de plateforme nationale fédérée des données de la recherche [Recherche.data.gouv](https://recherche.data.gouv.fr) et s'inscrit dans le dispositif d'accompagnement de proximité des ateliers de la donnée.

## A Questions de l'enquête

# Enquête sur les usages et les besoins pour la gestion des données de la recherche

Ce questionnaire est proposé par la Cellule Data Stewardship de l'UGA afin de connaître les pratiques et besoins des chercheurs et ingénieurs de l'UGA en matière de gestion des données de la recherche. La cellule a pour but de proposer un accompagnement, des solutions logicielles, ainsi que des formations qui répondent aux besoins immédiats de la communauté scientifique grenobloise.

Le temps estimé pour répondre est de dix minutes.

Dans cette enquête, les termes "données de la recherche" font référence à l'ensemble des informations collectées, produites et utilisées lors de l'activité de recherche. Cela comprend à la fois les données analysées lors de la recherche (mesures, simulations, textes, images, sons, échantillons, etc.), mais aussi le code informatique élaboré pour réaliser cette analyse.

Il y a 56 questions dans ce questionnaire.

## Informations personnelles

[Q1] Quel est votre pôle de rattachement ? \*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Chimie biologie santé (CBS)
- Mathématiques, sciences et technologies de l'information et de la communication (MSTIC)
- Physique des particules, astrophysique, géosciences, environnement et écologie (PAGE)
- Physique, ingénierie, matériaux (PEM)
- Sciences sociales (PSS)
- Sciences humaines et sociales (SHS)
- Autre

[Q2] Quelle(s) est/sont votre/vos discipline(s) de recherche ? \*

Veuillez écrire votre réponse ici :

[Q3] Quel est votre statut ? \*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Doctorant
- Post-doctorant
- Chercheur
- Enseignant-Chercheur
- Personnel de soutien à la recherche
- Autre

[Q4] Depuis combien de temps travaillez-vous dans la recherche ? (Depuis l'inscription en thèse ou équivalent) \*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- 0-5 ans
- 5-10 ans
- Plus de 10 ans

## Les données de la recherche

[DR1] Avez-vous déjà réutilisé des données publiées par d'autres chercheurs ? \*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Non concerné.e

**Données de la recherche** : ensemble des informations collectées, produites et utilisées par la recherche scientifique.

[DR2] Avez-vous déjà rencontré des problèmes liés à l'utilisation des données de la recherche ? \*

● Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- Données inexistantes ou impossibles à récupérer sur demande
- Données ne correspondent pas à ce qui a été publié
- Pas de métadonnées/annotations utilisables
- Problème d'accès aux données
- Problème pour retrouver des données
- Problème de format des données
- Pas accès à un logiciel permettant de lire les données
- Manque d'éléments descriptifs permettant de réutiliser les données
- Perte de données suite à une suppression des données
- Perte de données suite à une destruction matérielle
- Non
- Non concerné.e
- Autre:

[DR3]

Connaissez-vous une personne que vous pouvez contacter afin d'obtenir de l'aide pour la gestion des données ? Il peut s'agir d'une demande informelle ou non.

Veuillez indiquer sa fonction dans le champ commentaire associé à votre ou vos réponse(s).

\*

● Ajoutez un commentaire seulement si vous sélectionnez la réponse.

Veuillez choisir toutes les réponses qui conviennent et laissez un commentaire :

Une personne de l'équipe

Une personne du laboratoire

Une personne de l'UGA

Une personne d'une tutelle

Non

Autre :

**La gestion de données de recherche** se réfère à l'ensemble des opérations de collecte, documentation, analyse, stockage et mise à disposition de l'ensemble des informations produites au cours d'un projet de **recherche** (*La gestion des données de la recherche : enjeux et contraintes*, A. Rivet, JRES, Nantes, 2017, p3).

## Pendant la recherche

[PDT1] Quel(s) type(s) de données vous manipulez ? \*

☛ Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

- Mesures ou observations (par exemple données de capteurs)
- Simulations
- Données quantitatives
- Données qualitatives
- Images
- Textes
- Codes sources
- Enregistrements numériques (audio, vidéo, etc)
- Données web (données d'usage, cartographie de sites, données de réseaux sociaux, etc.)
- Données 3D
- Workflow
- Aucune donnée
- Non concerné.e
- Question non comprise
- Autre:

**Données de la recherche** : Ensemble des informations collectées, produites et utilisées par la recherche scientifique.

[PDT2] Y a-t-il des données sensibles au sens juridique du terme ? \*

☛ Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

- Santé
- Personnel
- Non
- Autre:

**Données sensibles** : "Le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits." (RGPD, Article 9)

[PDT3] Quelle est la volumétrie approximative de l'ensemble de vos données de recherche ? \*

☛ Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

- < 1 Go
- Entre 1 et 20 Go
- Entre 20 et 100 Go
- Entre 100 Go et 1 To
- > 1 To
- Je ne sais pas
- Non concerné.e

**Volumétrie** : ici, mesure de la quantité de données ou de traitement des données.

[PDT4] Quel(s) est/sont le(s) lieu(x) de stockage de vos données de recherche ? \*

\*

☛ Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

- Ordinateurs professionnels
- Ordinateurs personnels
- Disques durs externes professionnels
- Disques durs externes personnels
- Serveur dédié du laboratoire
- Dans l'établissement vous hébergeant
- Au sein du laboratoire ou de l'établissement ayant produit les données (CNES, CERN, Météo-France, etc.)
- Service de cloud institutionnel (UGA, tutelles, etc.)
- Service de cloud privé (Dropbox, Google Drive, etc.)
- Je ne sais pas
- Autre:

**Stockage** : rassemblement de données déposées afin de retrouver ces données et de les partager.

[PDT5] Quelle(s) pratique(s) du travail collaboratif autour des données avez-vous ? \*

☛ Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

- Collaborations ponctuelles (par exemple pour un projet multipartenarial impliquant plusieurs équipes)
- Collaborations régulières (projets récurrents, annuels, etc.)
- Aucun travail collaboratif
- Non concerné.e
- Question non comprise
- Autre:

**Projet multipartenarial** : projet associant plusieurs partenaires, pouvant être internes ou externes à l'unité.

**Collaboration externe à l'unité** : collaboration avec un partenaire n'appartenant pas à votre composante

[PDT6] Quel(s) type(s) de logiciel(s) ou service(s) utilisez-vous dans le cadre de vos travaux collaboratifs ? \*

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

La réponse était à la question '13 [PDT5]' (Quelle(s) pratique(s) du travail collaboratif autour des données avez-vous ?) et La réponse était à la question '13 [PDT5]' (Quelle(s) pratique(s) du travail collaboratif autour des données avez-vous ?)

☛ Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

- Services institutionnels (Gitlab UGA, Cloud UGA, Renavisio, Evento, etc.)
- Services privés (Dropbox, Google Drive, Doodle, Skype, etc.)
- Aucun
- Non concerné.e
- Question non comprise
- Autre:

[PDT8] Veuillez préciser le(s) lieu(x) de stockage de vos données dans votre établissement d'accueil. \*

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

La réponse était à la question '11 [PDT4]' (Quel(s) est/sont le(s) lieu(x) de stockage de vos données de recherche ?)

Veillez écrire votre réponse ici :

**[PDT7]Quelles sont les conditions d'accès aux données lors du travail collaboratif ? \***

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '13 [PDT5]' (Quelle(s) pratique(s) du travail collaboratif autour des données avez-vous ?) et La réponse était à la question '13 [PDT5]' (Quelle(s) pratique(s) du travail collaboratif autour des données avez-vous ?)

Veillez choisir toutes les réponses qui conviennent :

- Identification par un service d'authentification institutionnel (ex. eduGAIN)
- Identification par un service privé
- Aucune
- Question non comprise
- Autre:

**Diffusion et valorisation**

**[DV1]Avez-vous déjà diffusé vos données ? \***

● Veuillez sélectionner une réponse ci-dessous  
Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Non concerné.e

**[DV1A]Lors de la diffusion de vos données, avez-vous constitué un jeu de données spécifique, différent des données stockées lors de la recherche (par exemple à cause de données sensibles ou confidentielles) ?**

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse n'était PAS 'Non' à la question '16 [DV1]' (Avez-vous déjà diffusé vos données ?)

● Veuillez sélectionner une réponse ci-dessous  
Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas
- Question non comprise

**Jeu de données** : ensemble cohérent de données produites dans le cadre d'un même projet, sur un même objet d'étude et/ou recueillies sur un même lieu.

**[DV1B]Où avez-vous diffusé vos données ?**

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse n'était PAS 'Non' à la question '16 [DV1]' (Avez-vous déjà diffusé vos données ?)

● Cochez la ou les réponses  
Veillez choisir toutes les réponses qui conviennent :

- Entrepôt thématique (Pangaea, Seano, etc...)
- Entrepôt généraliste (Harvard Dataverse, Zenodo, etc...)
- Datapapers
- Je ne sais pas
- Autre:

**[DV1B1]Veillez préciser le nom de l'entrepôt thématique dans lequel vous avez déposé vos données.**

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '18 [DV1B]' (Où avez-vous diffusé vos données ?)

Veillez écrire votre réponse ici :

**[DV1B2]Veillez préciser le nom de l'entrepôt généraliste dans lequel vous avez déposé vos données.**

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '18 [DV1B]' (Où avez-vous diffusé vos données ?)

Veillez écrire votre réponse ici :

**[DV1B3]Veillez préciser le nom du Datapapers dans lequel vous avez déposé vos données.**

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '18 [DV1B]' (Où avez-vous diffusé vos données ?)

Veillez écrire votre réponse ici :

**[DV2]Êtes-vous satisfait de la solution retenue pour la diffusion de vos données ? Pourquoi (efficacité, pertinence, simplicité, etc.) ? \***

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse n'était PAS 'Non' à la question '16 [DV1]' (Avez-vous déjà diffusé vos données ?)

Veillez écrire votre réponse ici :

**[DV3]Quelle personne a réalisé le dépôt des données ? Veuillez préciser sa fonction ou son rôle. \***

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse n'était PAS 'Non' à la question '16 [DV1]' (Avez-vous déjà diffusé vos données ?)

Veillez écrire votre réponse ici :

**[DV4]À l'avenir, aurez-vous besoin d'aide pour diffuser vos données ? \***

● Veuillez sélectionner une réponse ci-dessous  
Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas

[DV4A] Veuillez préciser le type d'aide que vous attendez pour la diffusion de vos données.

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse n'était PAS 'Non' à la question '24 [DV4]' (À l'avenir, aurez-vous besoin d'aide pour diffuser vos données ?)

● Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

Aide au choix d'un entrepôt de données

Aide au dépôt de vos données

Autre:

[DV5] Quelle(s) est/sont votre/vos pratique(s) du stockage sécurisé des données ? \*

● Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

Dans votre établissement

Dans une plateforme nationale

Aucune

Question non comprise

Autre:

**Stockage** : rassemblement de données déposées afin de retrouver ces données et de les partager.

[DV5A] Veuillez préciser le nom du service de stockage sécurisé que vous utilisez au sein de votre établissement.

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '26 [DV5]' (Quelle(s) est/sont votre/vos pratique(s) du stockage sécurisé des données ?)

Veuillez écrire votre réponse ici :

[DV5B] Veuillez préciser le nom de la plateforme nationale de stockage sécurisé que vous utilisez.

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '26 [DV5]' (Quelle(s) est/sont votre/vos pratique(s) du stockage sécurisé des données ?)

Veuillez écrire votre réponse ici :

[DV6] À l'avenir, aurez-vous besoin d'aide pour stocker les données de façon sécurisée ? \*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

Oui

Non

Je ne sais pas

[DV6A] Veuillez préciser le type d'aide que vous attendez pour le stockage sécurisé de vos données.

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse n'était PAS 'Non' à la question '29 [DV6]' (À l'avenir, aurez-vous besoin d'aide pour stocker les données de façon sécurisée ?)

● Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

Aide au choix d'une plateforme de stockage sécurisé

Aide au dépôt de vos données

Autre:

[DV7] Quelle(s) est/sont votre/vos pratique(s) de l'archivage pérenne des données ? \*

● Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

Dans votre établissement

Dans une plateforme nationale

Aucune

Question non comprise

Autre:

**Archivage pérenne** : rassemblement de données afin de les conserver et de garantir leur accès et leur utilisation sur le long terme. Cela implique, dans la majorité des cas, la conservation de données supplémentaires pour garantir la réutilisation de l'archive, tels que des logiciels.

[DV7A] Veuillez préciser le nom du service d'archivage pérenne que vous utilisez au sein de votre établissement.

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '31 [DV7]' (Quelle(s) est/sont votre/vos pratique(s) de l'archivage pérenne des données ?)

Veuillez écrire votre réponse ici :

[DV7B] Veuillez préciser le nom de la plateforme nationale d'archivage pérenne que vous utilisez.

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse était à la question '31 [DV7]' (Quelle(s) est/sont votre/vos pratique(s) de l'archivage pérenne des données ?)

Veuillez écrire votre réponse ici :

[DV8] À l'avenir, aurez-vous besoin d'aide pour archiver les données de façon pérenne ? \*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

Oui

Non

Je ne sais pas

[DV8A] Veuillez préciser le type d'aide que vous attendez pour l'archivage pérenne de vos données.

Répondre à cette question seulement si les conditions suivantes sont réunies :  
La réponse n'était PAS 'Non' à la question '34 [DV8]' (À l'avenir, aurez-vous besoin d'aide pour archiver les données de façon pérenne ?)

● Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

Aide au choix d'une plateforme d'archivage pérenne

Aide au dépôt de vos données

Autre:

[DMP2]

Avez-vous déjà participé à un projet de recherche ayant nécessité la rédaction d'un DMP ?

\*

● Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

Oui

Non

Je ne sais pas

## Plan de gestion des données (PGD) - Data Management Plan (DMP)

[DMP1]

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

\*

● Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

Oui, je sais déjà ce que c'est (bonnes ou très bonnes notions)

Oui, j'en ai déjà entendu parlé (vagues notions)

Non

Non concerné.e

**Data management plan (DMP) ou Plan de gestion des données (PGD)** : document évolutif qui aide le(s) chercheur(s) ou les chercheuse(s) ou le ou la chargé(e) de projet de la recherche à définir un plan pour gérer les données utilisées et générées dans le cadre de son activité ou de son projet de recherche.

[DMP2A] Veuillez préciser le nom du ou des projets auxquels vous avez participé et qui ont rédigé un DMP.

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

La réponse était 'Oui' à la question '37 [DMP2]' (Avez-vous déjà participé à un projet de recherche ayant nécessité la rédaction d'un DMP ? )

Veillez écrire votre réponse ici :

[DMP2B] Est-ce que la rédaction du DMP a posé des problèmes ?

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

La réponse était 'Oui' à la question '37 [DMP2]' (Avez-vous déjà participé à un projet de recherche ayant nécessité la rédaction d'un DMP ? )

● Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

Oui

Non

Je ne sais pas

[DMP2B1] Quels problèmes ont été rencontrés ?

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

La réponse était 'Oui' à la question '39 [DMP2B]' (Est-ce que la rédaction du DMP a posé des problèmes ?)

Veillez écrire votre réponse ici :

[DMP3B] Pensez-vous que la rédaction du DMP pourra poser des problèmes ?

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

La réponse était 'Oui' à la question '41 [DMP3]' (Allez-vous participer à un projet de recherche qui nécessitera la rédaction d'un DMP ? )

● Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

Oui

Non

Je ne sais pas

[DMP3]

Allez-vous participer à un projet de recherche qui nécessitera la rédaction d'un DMP ?

\*

● Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

Oui

Non

Je ne sais pas

[DMP3C] Quels problèmes vont être rencontrés à votre avis ?

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

La réponse était 'Oui' à la question '43 [DMP3B]' (Pensez-vous que la rédaction du DMP pourra poser des problèmes ?)

Veillez écrire votre réponse ici :

[DMP3A] Veuillez préciser le nom du ou des projets auxquels vous allez participer et qui vont rédiger un DMP.

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

La réponse était 'Oui' à la question '41 [DMP3]' (Allez-vous participer à un projet de recherche qui nécessitera la rédaction d'un DMP ? )

Veillez écrire votre réponse ici :

[DMP4] Avez-vous déjà rédigé un DMP ? \*

● Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

Oui

Non

Je ne sais pas

[DMP5] Allez-vous rédiger un DMP ? \*

➊ Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui  
 Non  
 Je ne sais pas

[DMP6]

À combien de formations ou de sensibilisations à la gestion de données avez-vous participé ?

\*

➋ Seuls des nombres peuvent être entrés dans ce champ.

Veuillez écrire votre réponse ici :

La **gestion de données de recherche** se réfère à l'ensemble des opérations de collecte, documentation, analyse, stockage et mise à disponibilité de l'ensemble des informations produites au cours d'un projet de recherche (*La gestion des données de la recherche : enjeux et contraintes*, A. Rivet, JRES, Nantes, 2017, p3).

## Besoins concernant la gestion

[BGD1] De quel(s) accompagnement(s) ou formation(s) avez-vous besoin dans le cadre de votre gestion de données ?

➌ Vos réponses doivent être différentes, et vous devez les classer dans l'ordre.

➍ Veuillez sélectionner 11 réponses maximum

Numérotez chaque case dans l'ordre de vos préférences de 1 à 11

- Formation / Sensibilisation à la gestion des données
- Conseils d'ordre général pour la gestion des données
- Conseils sur la publication et la citation de données
- Aide juridique (restrictions d'accès, traitement des données sensibles, utilisation de licences, etc.)
- Aide technique sur le stockage et l'archivage des données (métadonnées, normes, formats, etc.)
- Aide à la documentation et la description des données
- Aide au choix d'un entrepôt de données
- Aide au dépôt de données dans un entrepôt
- Aide à la rédaction d'un plan de gestion des données
- Participer à / Solliciter un réseau de référents
- Bénéficier d'une aide personnalisée

La **gestion de données de recherche** se réfère à l'ensemble des opérations de collecte, documentation, analyse, stockage et mise à disponibilité de l'ensemble des informations produites au cours d'un projet de **recherche** (*La gestion des données de la recherche : enjeux et contraintes*, A. Rivet, JRES, Nantes, 2017, p3).

[BGD2] Avez-vous besoin d'accompagnement(s) ou de formation(s) non listé(s) ? \*

➎ Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui  
 Non  
 Je ne sais pas

[BGD2A] Veuillez préciser le type d'accompagnement ou de formation que vous souhaiteriez avoir.

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

La réponse n'était PAS 'Non' à la question '49 [BGD2]' (Avez-vous besoin d'accompagnement(s) ou de formation(s) non listé(s) ?)

Veuillez écrire votre réponse ici :

[BGD3] De quelle(s) solution(s) numérique(s) avez-vous besoin dans le cadre de votre gestion de données ?

➏ Vos réponses doivent être différentes, et vous devez les classer dans l'ordre.

➐ Veuillez sélectionner 6 réponses maximum

Numérotez chaque case dans l'ordre de vos préférences de 1 à 6

- Stockage au quotidien
- Sauvegarde au quotidien
- Travail collaboratif (ex. cloud)
- Entrepôt de données pour la diffusion
- Stockage sécurisé moyen terme
- Archivage pérenne

**Entrepôt** : archive numérique qui collecte et diffuse des jeux de données et leurs métadonnées.

**Stockage** : rassemblement de données déposées afin de retrouver ces données et de les partager.

**Archivage pérenne** : rassemblement de données afin de les conserver et de garantir leur accès et leur utilisation sur le long terme. Cela implique, dans la majorité des cas, la conservation de données supplémentaires pour garantir la réutilisation de l'archive, tels que des logiciels.

[BGD4] Avez-vous besoin de solution(s) numérique(s) non listée(s) ?

\*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui  
 Non  
 Je ne sais pas

[BGD4A] Veuillez préciser le type de solution(s) numérique(s) que vous souhaiteriez avoir.

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

La réponse n'était PAS 'Non' à la question '52 [BGD4]' (Avez-vous besoin de solution(s) numérique(s) non listée(s) ? )

Veuillez écrire votre réponse ici :

[BGD5] Quelle(s) sont les caractéristique(s) de l'entrepôt de données dont vous avez besoin ?

● Vos réponses doivent être différentes, et vous devez les classer dans l'ordre.

● Veuillez sélectionner 14 réponses maximum

Numérotez chaque case dans l'ordre de vos préférences de 1 à 14

Certification

Être définies par sujet, par thème

Être orienté métier

Utilisation d'un vocabulaire contrôlé

Citations des jeux de données dans le style des références

bibliographiques

Attribution de DOI (Digital Object Identifier) à chaque jeu de données

Possibilité de versioning

Gestion de larges jeux de données (> 1 To)

Exploitation des données (outils d'exportation, de visualisation, etc.)

Possibilité de moissonner les métadonnées

Lien entre une ou plusieurs publications et le jeu de données

Lien entre jeux de données

Choix des licences

Contrôle de l'accès aux données (embargo, accès restreint, etc...)

**Entrepôt** : archive numérique qui collecte et diffuse des jeux de données et leurs métadonnées.

**DOI** : *Digital Object Identifier* (littéralement « identifiant d'objet numérique »). Numéro international identifiant et localisant un objet documentaire publié sur Internet.

**Versioning** : fonctionnalité permettant d'accéder à l'ensemble des différentes versions d'un document lors de ses modifications.

**Jeu de données** : ensemble cohérent de données produites dans le cadre d'un même projet, sur un même objet d'étude et/ou recueillies sur un même lieu.

[BGD6] Avez-vous besoin de caractéristiques pour un entrepôt de données non listés ? \*

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui  
 Non  
 Je ne sais pas

[BGD6A] Veuillez préciser le type de caractéristiques pour un entrepôt de données que vous souhaiteriez avoir.

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

La réponse n'était PAS 'Non' à la question '55 [BGD6]' (Avez-vous besoin de caractéristiques pour un entrepôt de données non listés ?)

Veuillez écrire votre réponse ici :

**Merci à tous pour votre participation à cette enquête !**

Pour en savoir plus, le site web de la Cellule Data Stewardship est disponible ici (<https://gricad.gricad-pages.univ-grenoble-alpes.fr/cellule-data-stewardship/web/>).

Pour toute question, demande d'information : un seul point d'entrée (mailto:uga-cellule-data@univ-grenoble-alpes.fr).

Il est également possible de s'abonner à la liste de diffusion de la cellule data en cliquant ici (<https://listes.univ-grenoble-alpes.fr/sympa/info/uga-cellule-data>).

13.12.2021 – 08:20

Envoyer votre questionnaire.

Merci d'avoir complété ce questionnaire.