



Résultats de l'enquête sur les usages et les besoins pour la gestion des données de la communauté des mathématiques

22 mars 2022

Table des matières

1	Introduction	2
2	Profils des répondants	3
3	Données de la recherche	5
3.1	Disciplines concernées par les données	5
3.2	Types de données	7
3.3	Volumes des données et solutions de stockage	10
3.4	Traitement des données	12
3.5	Données sensibles	13
4	Codes de recherche	14
4.1	Utilisation et développement de codes	14
4.2	Aspects techniques	17
4.3	Diffusion et partage	18
5	Diffusion des données de la recherche	21
5.1	Plans de Gestion des Données	21
5.2	Entrepôt de données	21
6	Reproductibilité et réutilisation	24
6.1	Reproductibilité des résultats de recherche	24
6.2	Réutilisation des données et des codes	24
7	Besoins exprimés	26
7.1	Besoins autour des données de recherche	26
7.2	Besoins autour des codes de recherche	26
8	Conclusions et perspectives	28
8.1	Synthèse	28
8.2	Quelques pistes pour aller plus loin	28
8.2.1	Au niveau local	28
8.2.2	Au niveau national	29
8.2.3	Au niveau international	29
A	Questions de l'enquête	30

1 Introduction

Les enjeux autour de la science ouverte mettent en exergue l'importance des productions scientifiques que sont les données de la recherche et les codes et logiciels de recherche.

Cette évolution impacte le travail au quotidien des communautés de recherche et implique que les organismes et établissements mettent en place les services et l'accompagnement nécessaires pour soutenir les laboratoires dans l'appropriation de ces transformations méthodologiques.

Ces pratiques sont très fortement disciplinaires. Les réseaux Mathrice (<https://mathrice.fr/>) et RNBM (<https://www.rnbn.org>) s'associent afin de proposer aux membres des laboratoires de mathématiques les réponses les plus pertinentes autour des données et des codes de la recherche.

Cette enquête a pour objectif de bien comprendre les usages et les besoins spécifiques de la communauté des mathématiques, et de proposer des recommandations afin de répondre aux problématiques exprimées.

Selon la définition proposée par l'ANDS (*Australian Research Data Commons*) et traduite par l'INIST-CNRS¹, « le terme de données de la recherche désigne les données sous forme de faits, d'observations, d'images, de résultats de programmes informatiques, d'enregistrements, de mesures ou d'expériences sur lesquelles un argument, une théorie, un test ou une hypothèse, ou un autre produit de la recherche est basé. Les données peuvent être numériques, descriptives, visuelles ou tactiles. Elles peuvent être brutes, nettoyées ou traitées, et peuvent être conservées dans tout format ou support. »

1. <https://www.inist.fr/>

2 Profils des répondants

Ce questionnaire a été largement relayé dans les laboratoires. Il a fait l'objet de 352 réponses.

Il est important de s'assurer d'une bonne représentation de la diversité des membres de la communauté mathématique, tant en terme de localisation, de statuts, de disciplines, et d'expérience professionnelle.

Les figures 1, 2, 3, 4 et 5 présentent différentes facettes des profils des répondants.

Ces éléments nous assurent une représentativité correcte du panel, et permettent de garantir que les résultats seront caractéristiques de la communauté mathématique française.

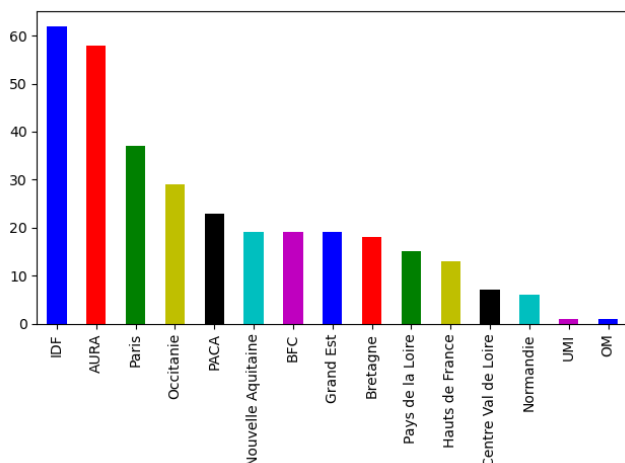


FIGURE 1 – Répartition régionale des répondants

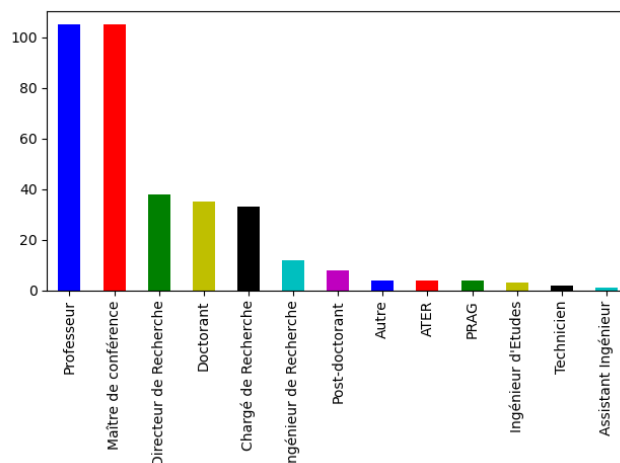


FIGURE 2 – Répartition des répondants selon leur statut

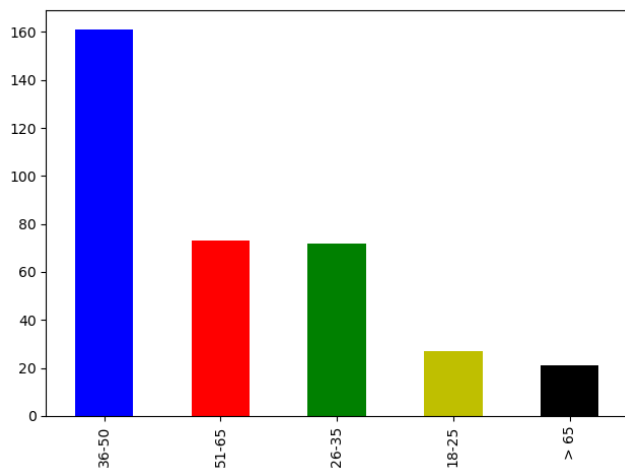


FIGURE 3 – Répartition des répondants selon leur tranche d'âge

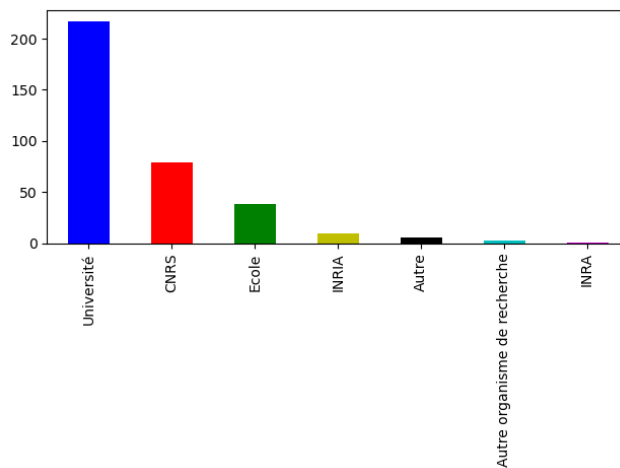


FIGURE 4 – Répartition des répondants selon leur employeur

La question de la thématique disciplinaire étant une question ouverte dans l'enquête, la répartition disciplinaire proposée en 5 se base sur différentes études (Rapport de perspectives CSI INSMI 2014², Etude de l'impact socio-économique des Mathématiques en France 2015³). Elle se divise en 16 catégories. La catégorie Probas-stats a été séparée en deux. Nous avons également ajouté une catégorie pour les personnes en soutien à la recherche.

2. <https://csi.math.cnrs.fr/rapports/>

3. <https://smf.emath.fr/actualites-smf/etude-de-limpact-socio-economique-des-mathematiques-en-france>

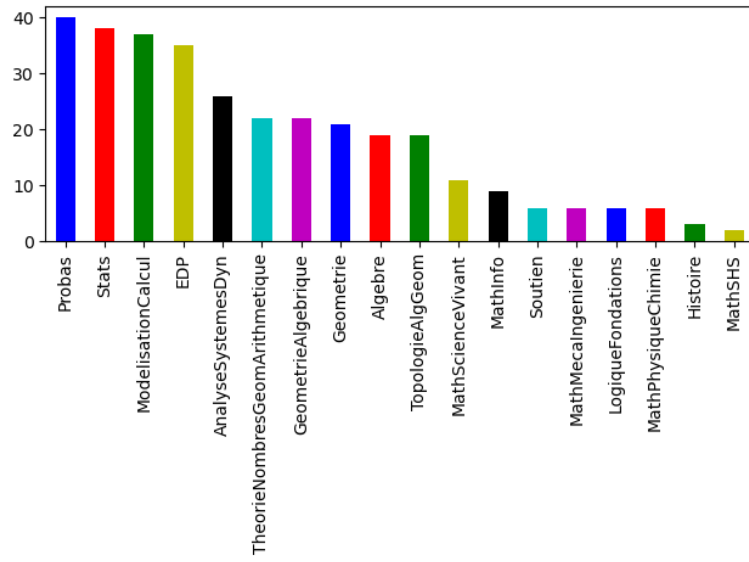


FIGURE 5 – Répartition des répondants selon leur thématique scientifique

3 Données de la recherche

L'objectif de cette partie est de comprendre quelles sont les données en jeu dans la communauté des mathématiques, quels sont les usages en terme d'infrastructures de stockage et de traitement et quelles sont les contraintes réglementaires si elles existent.

3.1 Disciplines concernées par les données

Le premier point est basique et vise à savoir qui manipule des données : près de 48% des répondants sont concernés par les données, contre 45 % pas du tout. 7% déclarent ne pas savoir, ce qui questionne sur la compréhension du terme « Données de la recherche ».

Si on croise ces éléments avec les différentes thématiques (figure 6, les réponses sont données en pourcentage du total de chaque thématique), les constats sont les suivants :

- les thématiques manipulant majoritairement des données sont essentiellement :
 - modélisation et calcul,
 - mathématique et informatique,
 - statistiques,
 - mathématiques et science du vivant,
 - mathématiques et mécanique, ingénierie,
 - mathématiques et SHS (Sciences Humaines et Sociales),
 - histoire,
 - mathématiques et physique, chimie,
- celles qui sont moins concernées :
 - géométrie algébrique,
 - théorie des nombres et géométrie arithmétique,
 - analyse et systèmes dynamiques,
 - géométrie,
 - EDP (Équations aux Dérivées Partielles),
 - algèbre,
 - topologie algébrique et géométrie,
 - logique et fondations.

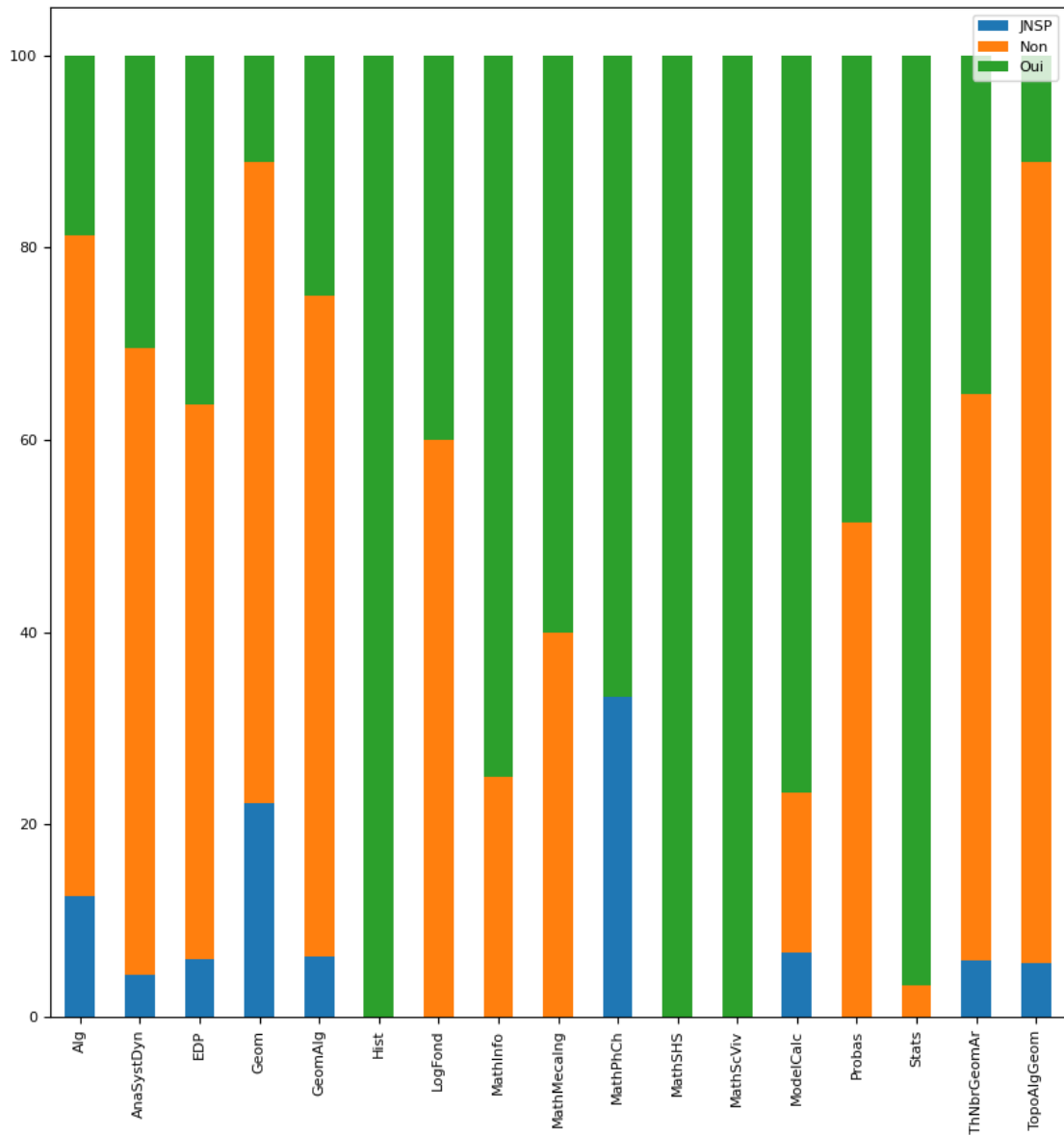


FIGURE 6 – Utilisation de données selon la thématique scientifique (en % des répondants par discipline)

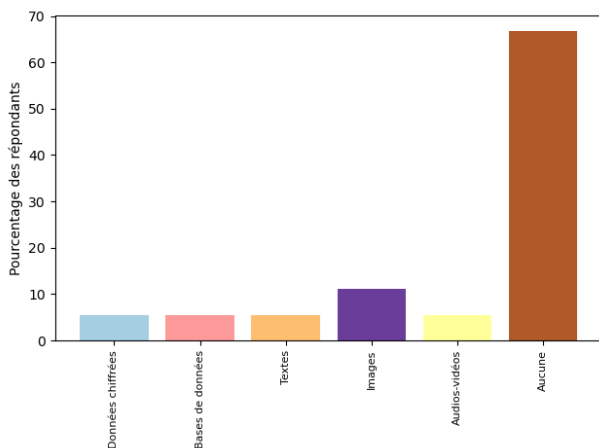
Rappel de la légende :

- Alg : algèbre, chimie.
- AnaSysDyn : analyse et systèmes dynamiques, — MathSHS : mathématiques et sciences humaines et sociales,
- EDP : équations aux dérivées partielles, — MathScViv : mathématiques et science du vivant,
- Geom : géométrie, — ModelCalc : modélisation et calcul,
- GeomAlg : géométrie algébrique, — Probas : probabilités,
- Hist : histoire, — Stats : statistiques,
- LogFond : logique et fondations, — ThNbrGeomAr : théorie des nombres et géométrie arithmétique,
- MathInfo : mathématique et informatique, — TopoAlgGeom : topologie algébrique et géométrie,
- MathMecaIng : mathématiques et mécanique, ingénierie,
- MathPhCh : mathématiques et physique,

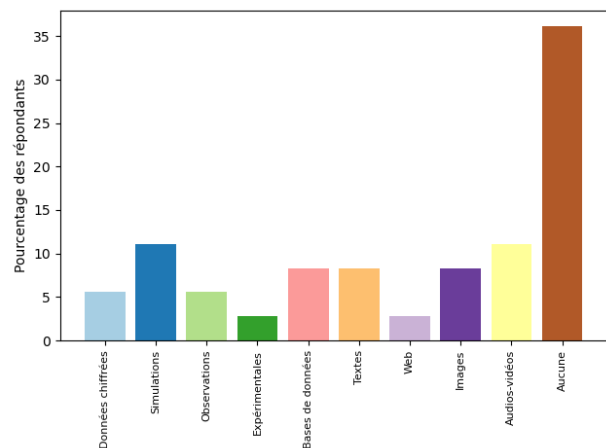
3.2 Types de données

Les figures ci-dessous illustrent les types de données utilisés dans les différentes disciplines des mathématiques. Certaines sont à prendre avec précaution compte tenu du faible nombre de réponses associées. Cependant, il est possible d'extraire quelques grandes tendances :

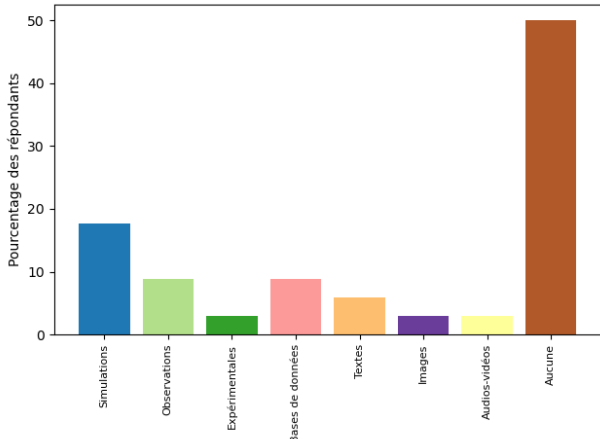
- On retrouve les données de simulations dans les disciplines en lien avec le calcul (« modélisation et calcul »), ainsi qu'à l'interface avec l'informatique (« mathématique et informatique »), l'ingénierie (« mathématiques et mécanique, ingénierie »), la physique (« mathématiques et physique, chimie ») et les sciences du vivant (« mathématiques et science du vivant »), dans une mesure un peu moindre en « probabilités ».
- Les données de type textes ont pu être comprises comme équivalentes aux publications, donc peut être non significatives dans certains cas.
- En « statistiques », la variété des données traitées est importante, de même qu'en « mathématiques et sciences du vivant ».
- En « histoire », on retrouve des données de type textes, images, audios et vidéos.
- L'usage de bases de données se retrouve en « mathématiques et informatique », « statistiques » et « probabilités », et dans une moindre mesure dans d'autres disciplines.



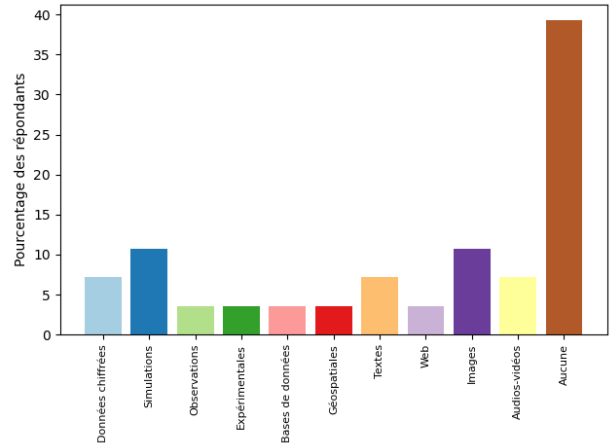
Type des données en algèbre



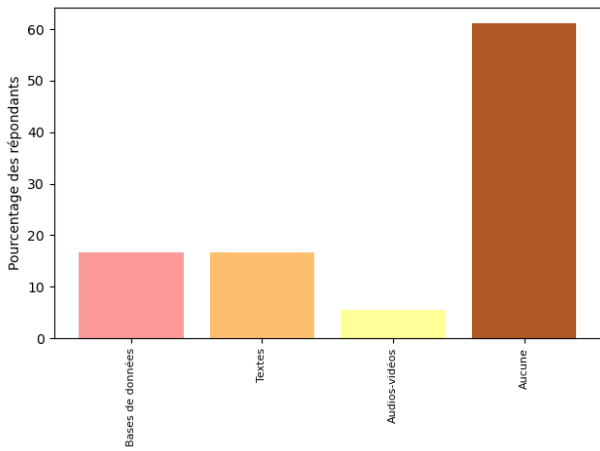
Type des données en analyse et systèmes dynamiques



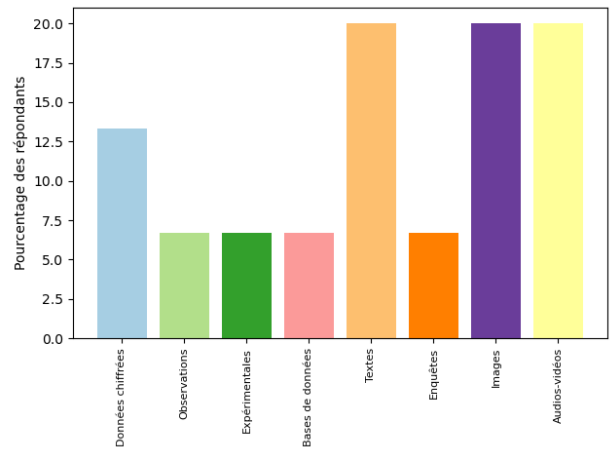
Type des données en EDP



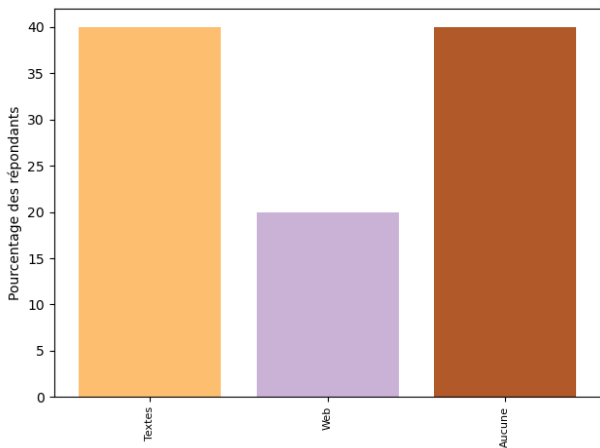
Type des données en géométrie



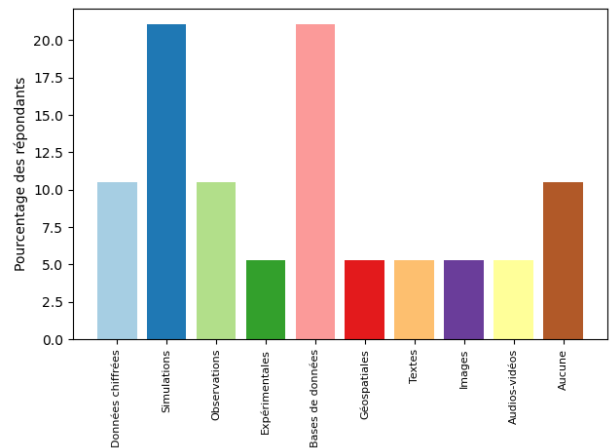
Type des données en géométrie algébrique



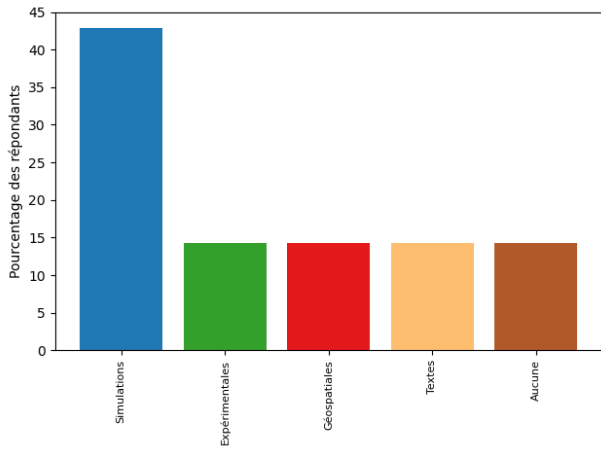
Type des données en histoire



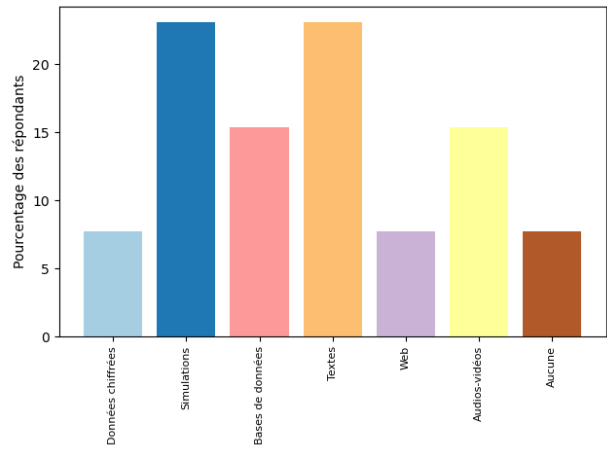
Type des données en logique et fondations



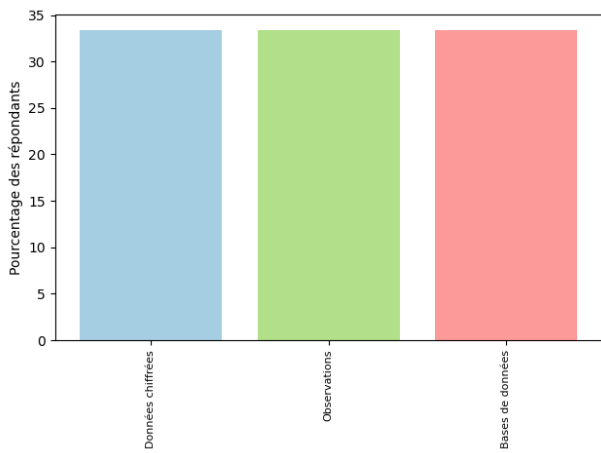
Type des données en mathématiques et informatique



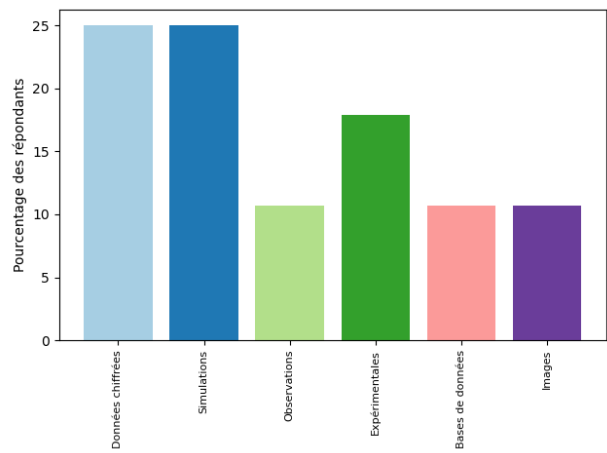
Type des données en mathématiques et mécanique, ingénierie



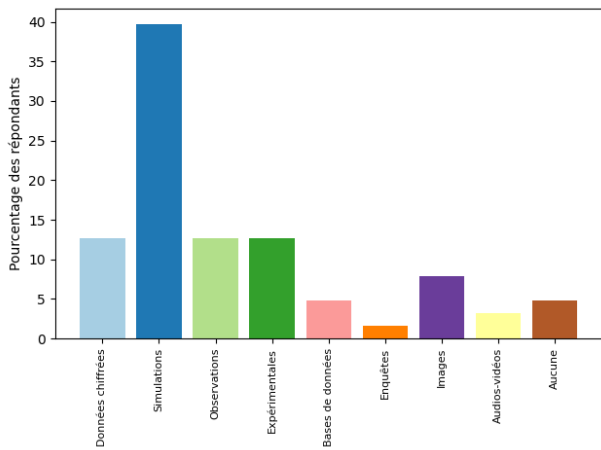
Type des données en mathématiques et physique, chimie



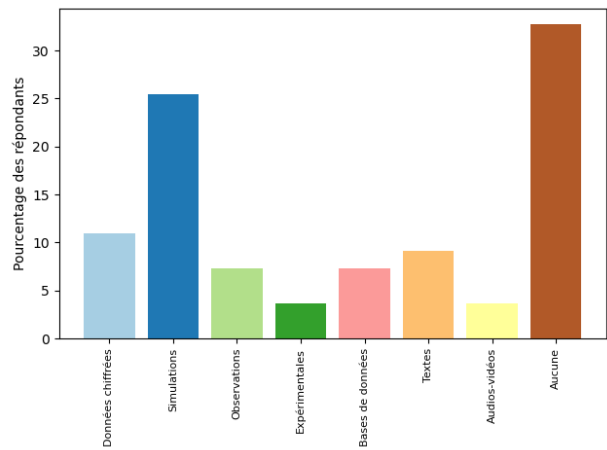
Type des données en mathématiques et SHS



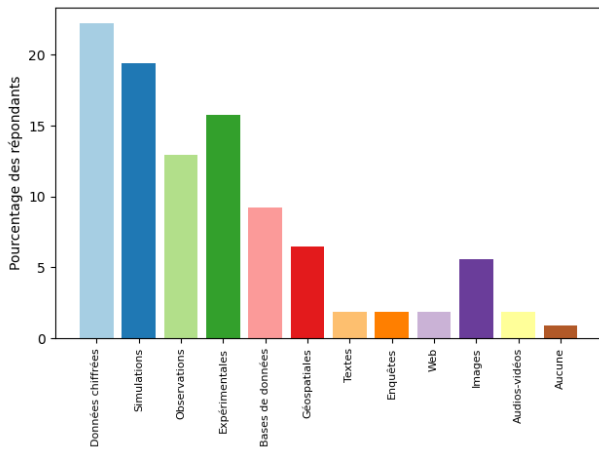
Type des données en mathématique et sciences du vivant



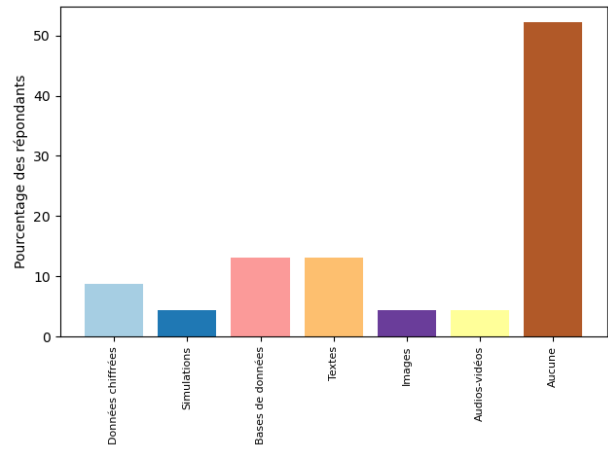
Type des données en modélisation et calcul



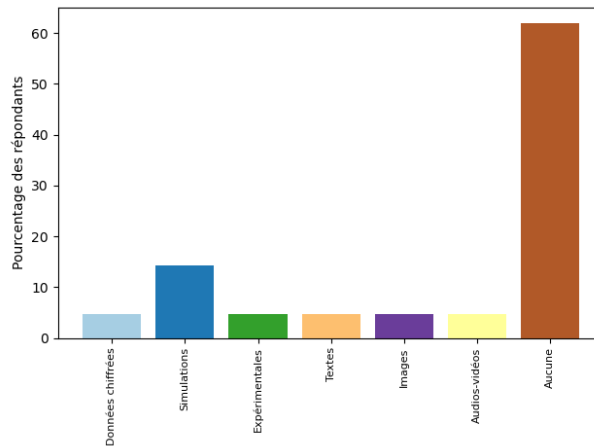
Type des données en probabilités



Type des données en statistiques



Type des données en théorie des nombres et géométrie arithmétique



Type des données en topologie algébrique et géométrie

Les réponses ouvertes associées à cette question sur les types de données donnent quelques compléments :

- Confirmation de l’ambiguïté sur la compréhension des données de type textes amalgamées avec les publications.
- Bases de données thématiques en ligne comme lmfdb.org, mais aussi données INSEE
-
- Certaines données sont fournies par les collaborateurs dans d’autres disciplines.

3.3 Volumes des données et solutions de stockage

La question sur les volumes des données manipulées a été laissée volontairement ouverte pour permettre à la fois des réponses précises mais aussi des éléments d’échelle moins techniques liés aux supports utilisés, de façon à faciliter l’appropriation de la question par tous. L’étude des réponses montre que très majoritairement, les volumes sont faibles voir très faibles :

- 20% pour des volumes inférieurs à quelques Go (Gigaoctets)
- 44% pour des volumes de plusieurs dizaines de Go
- 24% pour des volumes de 1 à 2 To (Téraoctets)
- 12 % pour des volumes supérieurs à 2 To.

Ces réponses influent assez directement les solutions de stockage utilisées et qui sont indiquées dans la figure 16 (exprimées en pourcentage) :

- Assez majoritairement l'ordinateur professionnel suffit pour stocker les données.
- Les serveurs de laboratoire sont également assez utilisés.

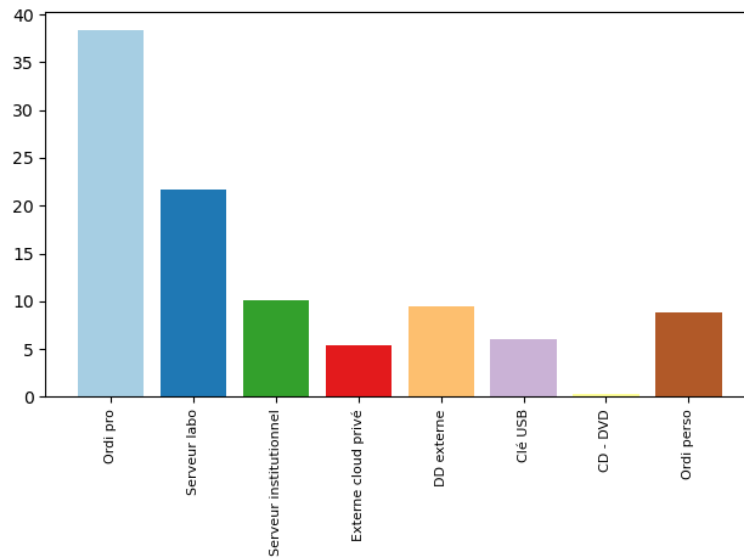


FIGURE 16 – Solutions de stockage des données

Les plateformes citées pour le stockage des données sont :

- les outils de la PLM⁴ : PLMBox, gitlab,
- les outils proposés par les établissements à base de nextcloud, ou de owncloud,
- les espaces de stockage des centres de calcul, mésocentres et centres nationaux,
- les outils commerciaux comme dropbox, AWS (Amazon), Google. Un des arguments pour cet usage concerne les collaborations avec des collègues à l'international.

Concernant la sauvegarde, c'est-à-dire la sécurisation des données sur un autre support que le support principal de stockage, les réponses montrent, dans la figure 17 (exprimées en pourcentage) que :

- l'ordinateur professionnel sert également pour la sauvegarde, de même que les serveurs de laboratoire.
- L'utilisation de disques durs externes est assez importante.

4. Plateforme en Ligne Mathrice

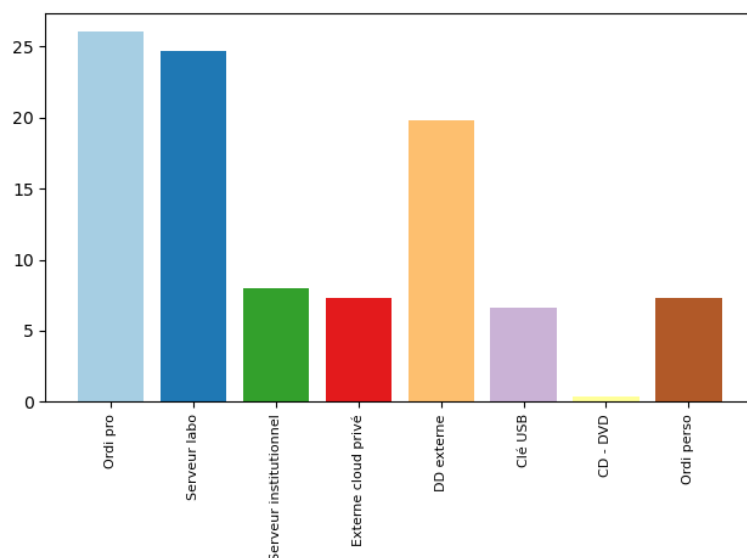


FIGURE 17 – Solutions de sauvegarde des données

Les plateformes indiquées pour la sauvegarde sont les outils institutionnels (dont ceux de la PLM) à base de technologie comme gitlab, nexcloud, owncloud, seafiler, mais également les espaces des centres de calcul et certains outils commerciaux comme dropbox et AWS. Afin de bien comprendre les usages et les besoins, des questions ouvertes ont porté sur les points forts et les points faibles de ces solutions de stockage et de sauvegarde. Les réponses laissent entrevoir un foisonnement de pratiques individuelles et de solutions techniques.

Les points faibles qui sont pointés sont majoritairement :

- le manque de simplicité (qui fait quelquefois se tourner vers des solutions privées), le manque d’automatisation (beaucoup de sauvegardes sont manuelles),
- les problématiques de sécurité physique des supports utilisés,
- les difficultés pour partager des données, en particulier avec des collègues étrangers,
- la question du respect du RGPD (Règlement Général sur la Protection des Données) pour certains types de données.

Globalement, les solutions proposées par les laboratoires ou les établissements, quand elles existent, ou les outils de la PLM semblent être satisfaisantes, tant en terme de facilité d’usage que sur la sécurisation proposée.

3.4 Traitement des données

L’enjeu des questions ouvertes sur les usages en matière de traitement de données est d’avoir une vision assez large des besoins en ressources de calcul. Les réponses sont très homogènes sur les traitements réalisés avec les données manipulées. Il s’agit principalement de :

- simulations,
- statistiques,
- représentation graphique,
- vérification de conjectures,
- post-traitement.

Pour exécuter ces différentes opérations, le besoin en terme de ressources est, dans une très grande majorité, très limité puisque la plupart des répondants utilisent leur ordinateur portable. Sont également cités les serveurs de calcul des laboratoires, et dans une moindre mesure les mésocentres. L’emploi de cartes accélératrices de type GPU (*Graphics Processing Unit*, carte graphique) est également cité par une très faible minorité de répondants.

3.5 Données sensibles

Les données sensibles sont celles qui entrent dans le cadre du Règlement Général sur la Protection des Données (RGPD), et donc concernent les données à caractère personnel identifiantes. Les données de santé entrent dans ce cadre. Si une très grande majorité des répondants ne sont pas concernés par cette problématique, environ 20% (voir figure 18) disent manipuler ce type de données.

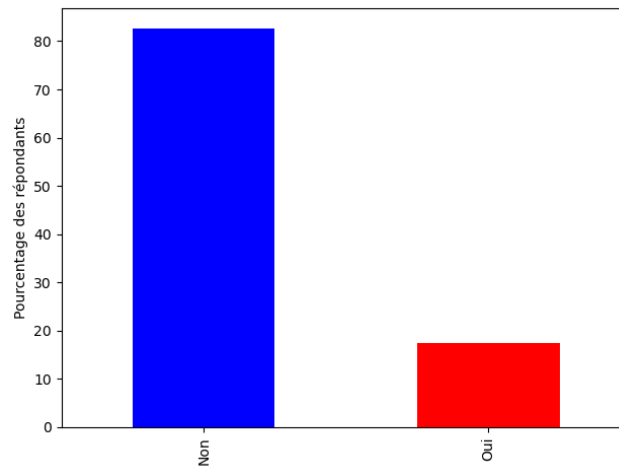


FIGURE 18 – Manipulation de données sensibles

Dans plusieurs cas, les données sont anonymisées donc sortent du cadre RGPD. Pour les autres, plusieurs difficultés sont relevées :

- le manque d’infrastructures adaptées pour stocker, sécuriser, partager, traiter correctement ce type de données,
- le manque d’informations et de formations sur la législation et les procédures à respecter,
- le manque de connaissances sur les processus d’anonymisation ou de pseudonymisation.

4 Codes de recherche

L'objectif de cette section est d'évaluer les pratiques autour de l'utilisation et du développement de logiciels pour la recherche, tant sur les aspects organisationnels, techniques que de diffusion.

4.1 Utilisation et développement de codes

Une très grande majorité des répondants utilisent des logiciels dans le cadre de leur recherche (voir figure 19).

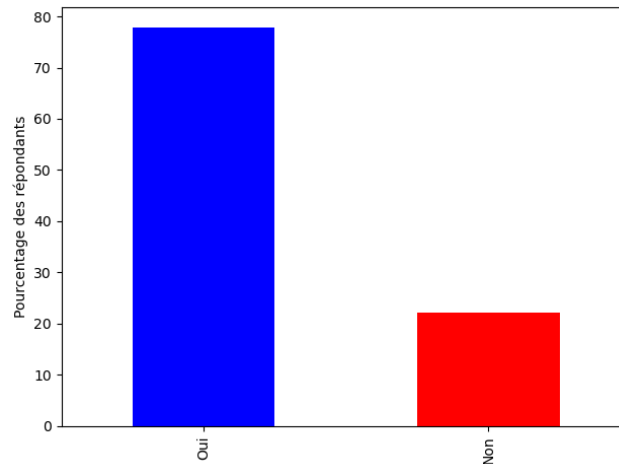


FIGURE 19 – Utilisation de codes

Le nuage de mot 20 illustre les logiciels les plus cités par les participants. En dehors des outils comme matlab et maple, il est intéressant de noter que l'immense majorité des logiciels utilisés sont des logiciels open source.

« Géométrie », « EDP », « Topologie algébrique et géométrie » et « Logique et fondations ».

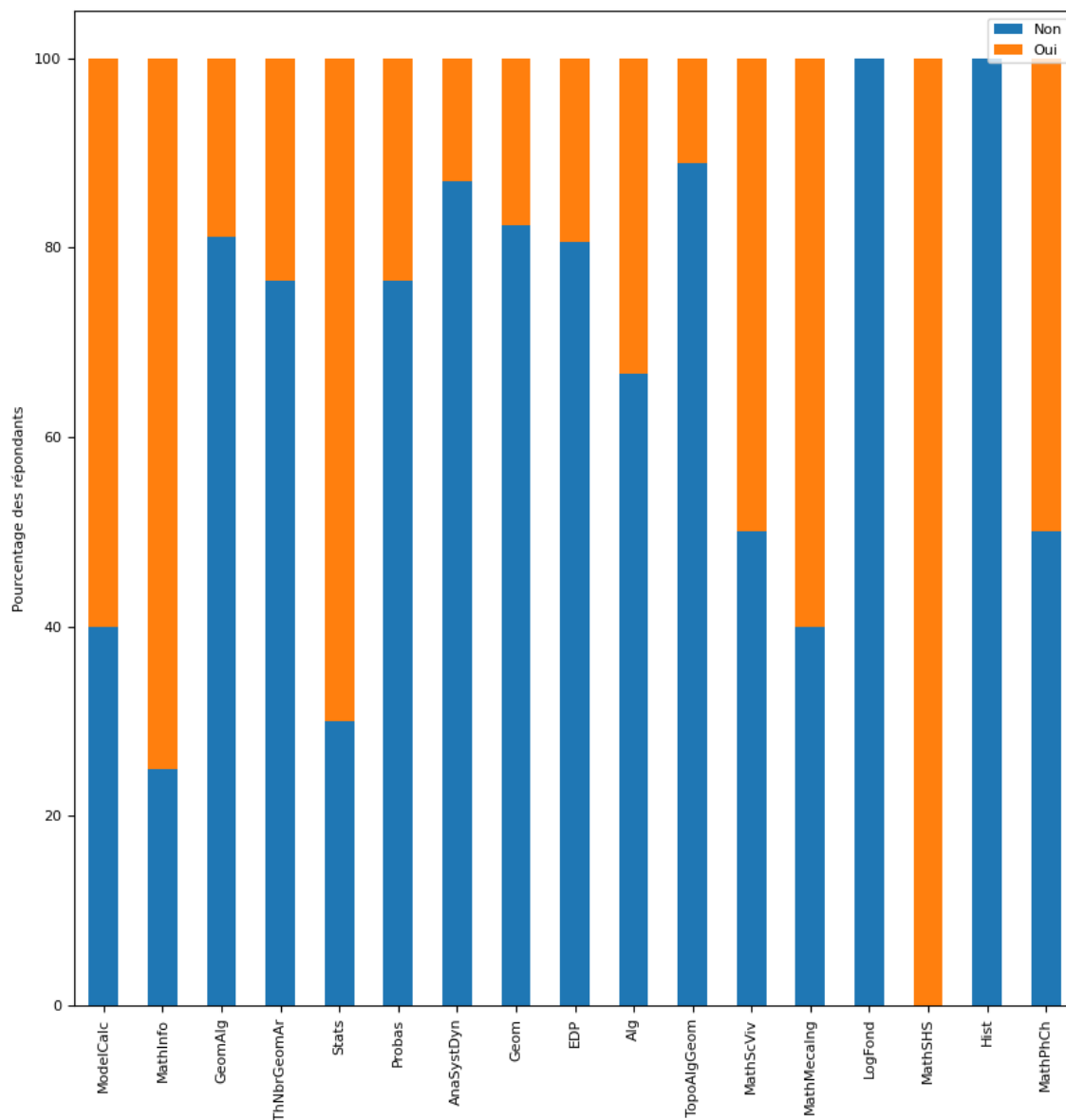


FIGURE 22 – Développement de codes parmi ceux qui utilisent du logiciel (en % des répondants par discipline). Voir la légende des disciplines section 3.1.

Dans une petite majorité des cas, le développement est plutôt réalisé de façon individuelle comme l'illustre la figure 23.

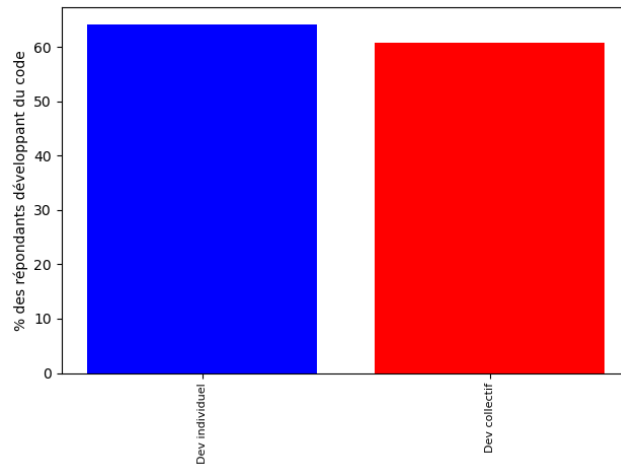


FIGURE 23 – Cadre des développements

4.2 Aspects techniques

Lorsque les participants ont une activité de développement, les langages de programmation les plus utilisés (voir figure 24) sont Python et C/C++ suivi par R.

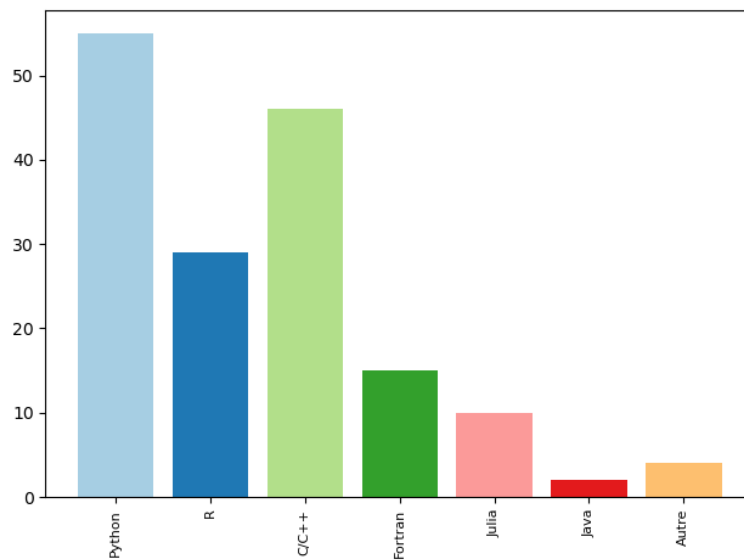


FIGURE 24 – Langages de programmation utilisés

Sur la question de l'utilisation d'un environnement de développement, beaucoup des répondants ont cité les notebooks (Jupyter), ou des IDE (*Integrated Development Environment*) comme VSCode et Rstudio. Assez peu d'entre eux mentionnent l'emploi de conteneur⁵.

Parmi les personnes qui développent des codes pour leur recherche, une majorité utilise une forge logicielle (voir figure 25), et parmi celles citées, elles sont quasiment toutes basées sur git et on retrouve souvent github ainsi que plmlab, et de façon plus ponctuelle gitlab INRIA.

5. voir Wikipedia

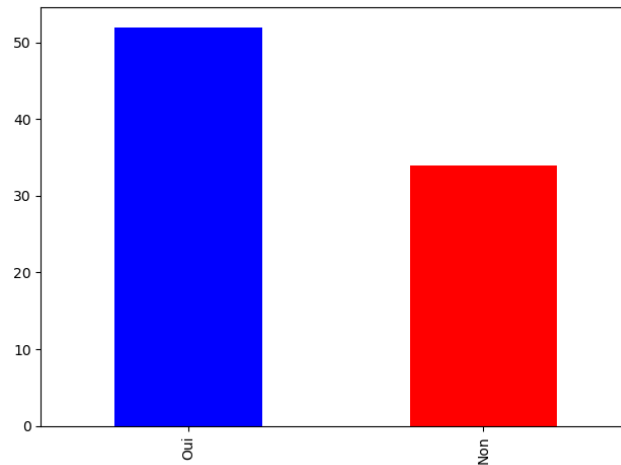


FIGURE 25 – Utilisation d’une forge

4.3 Diffusion et partage

La question du périmètre de l’utilisation des développements réalisés dans le cadre de la recherche est importante. Dans la figure 26, il est intéressant de noter que les codes développés dans les laboratoires de recherche sont majoritairement partagés, d’abord avec la communauté de recherche de la personne qui a développé le code et les partenaires de ses projets, puis avec son équipe. Dans les réponses ouvertes, on retrouve majoritairement le partage avec les doctorants et post-doctorants, mais aussi le cas ponctuel de logiciels très largement partagés au-delà des mathématiques.

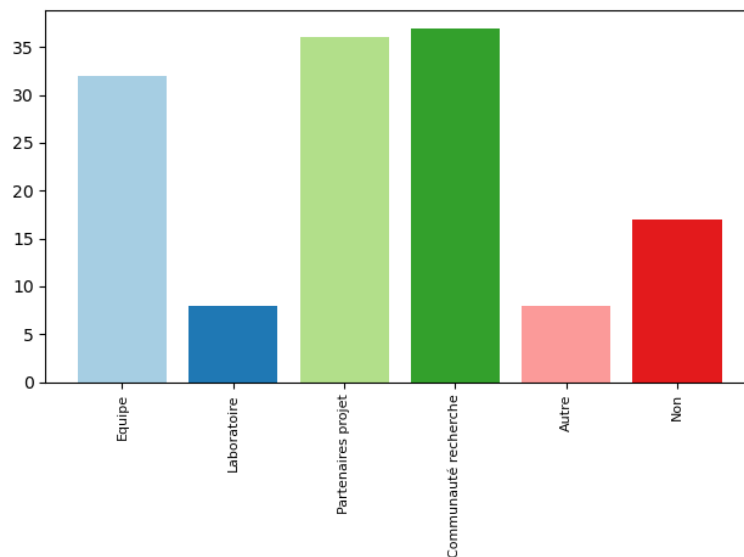


FIGURE 26 – Périmètre d’utilisation des développements

Si on croise ces réponses d’utilisation des développements par d’autres personnes que leur auteur (donc le cadre de codes partagés) avec le choix d’une licence ou la réalisation d’un dépôt à l’APP (Agence pour la Protection des Programmes), on obtient le graphe 27.

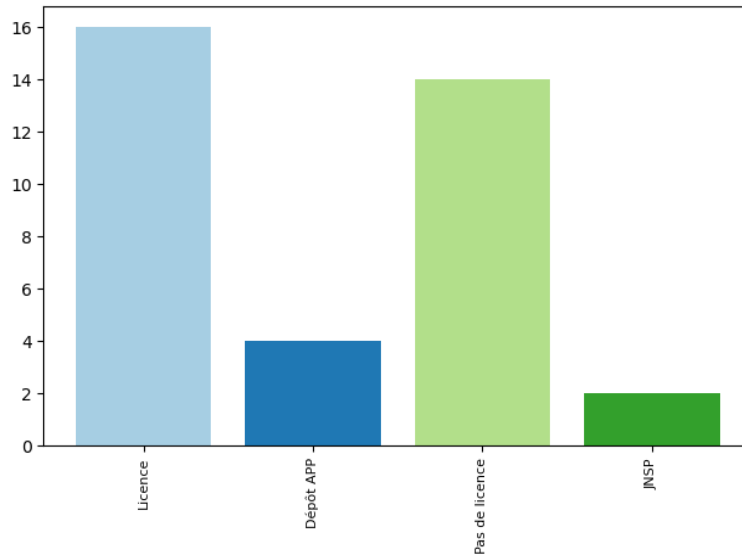


FIGURE 27 – Licences ou dépôt APP pour les développements partagés avec d’autres utilisateurs (JNSP = « Je ne sais pas »)

Si l’écrasante majorité des licences sont open source quand elles existent, il est aussi particulièrement problématique de constater que de nombreux codes sont partagés sans aucune licence. Le faible nombre de réponses concernant les dépôts APP questionne sur le manque d’information ou la difficulté de réaliser ces dépôts.

Les réponses à la question ouverte sur les freins rencontrés pour la déclaration d’une licence montrent :

- que le partage avec des collègues proches n’est pas considéré comme nécessitant la mise en place d’une licence,
- un manque de connaissance et de compréhension du sujet (par exemple, c’est public donc inutile de protéger ou un code pour un partenaire industriel n’a pas besoin de licence),
- une absence d’intérêt pour la problématique.

Le lien entre les publications et les logiciels a également été interrogé dans ce questionnaire, en particulier pour mesurer le niveau de citation des codes utilisés.

Comme l’illustre la figure 28, la situation est assez disparate, avec une majorité de répondants déclarant ne pas citer les logiciels qu’ils ont utilisés pour réaliser leur publication.

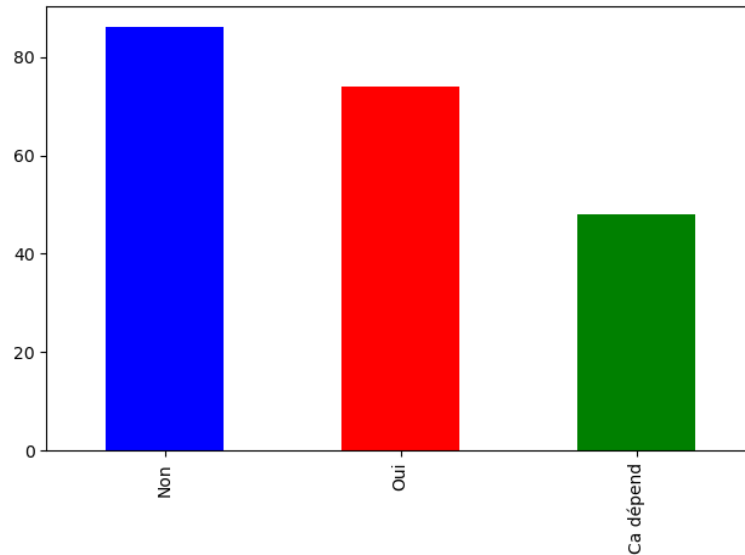


FIGURE 28 – Citation des logiciels dans les publications

Les commentaires accompagnant la question ont été essentiellement écrits par des personnes qui citent les codes employés :

- dépôt des codes,
- publication dans des journeaux spécialisés comme JOSS (Journal of Open Source Software),
- publication de notebooks pour pouvoir régénérer les résultats ...

Les réponses « Ça dépend » se réfèrent à des situations comme :

- importance du logiciel dans la publication,
- logiciel bien identifié vs outil générique,
- logiciel déjà diffusé ou pas, ...

5 Diffusion des données de la recherche

Cette section s'intéresse aux pratiques de diffusion des données de la recherche, en particulier autour des Plans de Gestion des Données (PGD) et des entrepôts de données.

5.1 Plans de Gestion des Données

Une très grande majorité des participants au questionnaire (voir figure 29) ne savent pas ce qu'est un Plan de Gestion des Données, et parmi ceux qui ont déjà été confrontés au sujet, seul un peu moins de 30% ont participé à leur rédaction (voir figure 30).

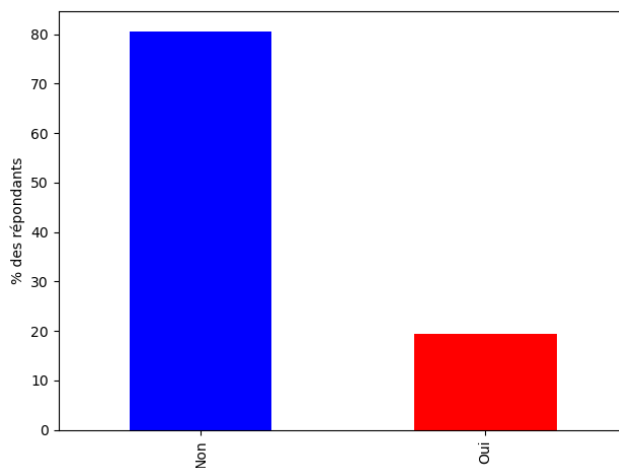


FIGURE 29 – Connaissance des PGD

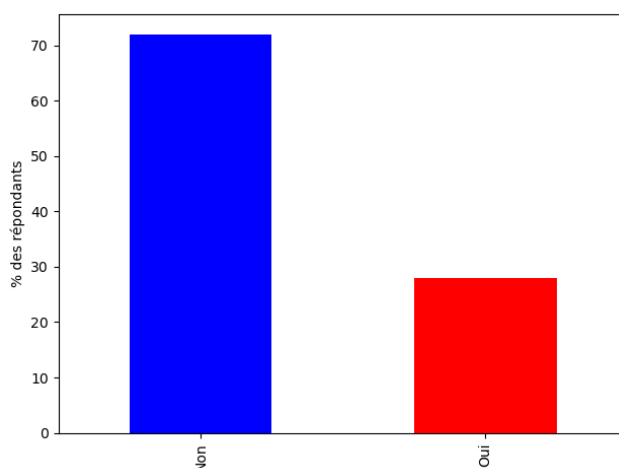


FIGURE 30 – Participation à la rédaction de PGD pour les répondants les connaissant déjà

Les principales difficultés relevées concernent essentiellement le fait que ce document n'est pas adapté à certaines thématiques des mathématiques, ainsi que le manque d'informations et de connaissances sur ce qui est attendu.

5.2 Entrepôt de données

Il s'agit ici d'évaluer le niveau de compréhension et les usages des entrepôts de données. Un entrepôt de données, dans la sémantique de la science ouverte, est une plateforme qui permet

le dépôt de jeux de données de recherche dans l'optique de leur diffusion, en général en open data.

Une grande majorité des répondants ne connaît pas d'entrepôt de données (voir figure 31).

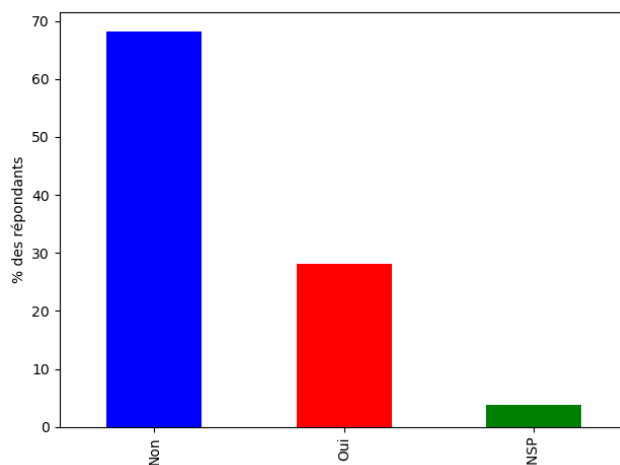


FIGURE 31 – Connaissance des entrepôts de données

Les commentaires associés à la question et qui permettaient de citer des entrepôts connus illustrent particulièrement bien le manque d'information sur le sujet puisqu'on y retrouve aussi bien des plateformes de type Gitlab, HAL, ArXiv, au côté de réponses plus attendues comme Zenodo⁶ et Dryad⁷.

Ce constat rend difficile l'analyse de la question sur le dépôt des données (voir 32) puisque ce concept est manifestement mal compris.

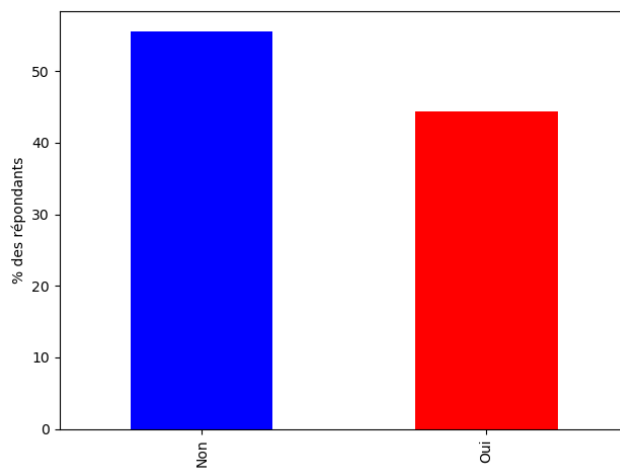


FIGURE 32 – Répondants ayant déjà déposé des données dans un entrepôt

Paradoxalement, la question ouverte sur l'intérêt de déposer des données apporte des réponses particulièrement intéressantes :

- sur le gain en visibilité,
- sur l'accès libre aux recherches pour tous,
- sur la possibilité de vérifier et reproduire les résultats de recherche.

Malgré ces bénéfices à la diffusion des données, un certain nombre de freins au dépôt sont identifiés par les participants au questionnaire :

6. <https://zenodo.org/>

7. <https://datadryad.org/stash>

- le manque d'informations sur les entrepôts disponibles et le processus de dépôt,
- les problématiques règlementaires et de propriété intellectuelle éventuellement associées,
- l'impression que les données ne sont pas pertinentes pour être diffusées,
- le manque de temps pour nettoyer les données et le désintérêt,
- le manque de valorisation de la mise à disposition des données,
- le désintérêt de la question du dépôt.

6 Reproductibilité et réutilisation

Cette section vise à évaluer les potentialités de reproductibilité et de réutilisation des productions scientifiques que sont les données et les codes.

6.1 Reproductibilité des résultats de recherche

La figure 33 illustre la perception des répondants concernant la possibilité de pouvoir reproduire leurs résultats de recherche. Il est particulièrement intéressant de constater que les participants estiment dans une très grande majorité que leurs résultats sont tout à fait reproductibles.

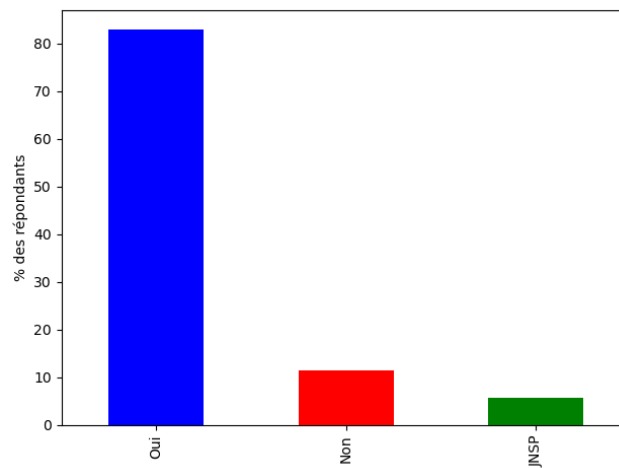


FIGURE 33 – Reproductibilité des résultats de leurs recherches d’après les répondants

Les commentaires associées à cette question sont un peu plus nuancés :

- beaucoup sont conscients de l’importance du sujet et font des efforts en ce sens sans pouvoir réellement garantir une bonne reproductibilité.
- le partage des données et des codes n’est bien souvent pas formalisé, mais ceux ci sont souvent disponibles sur demande.

6.2 Réutilisation des données et des codes

Comme préfiguré ci-dessus, le partage de données et de codes semble être une pratique courante d’après la figure 34.

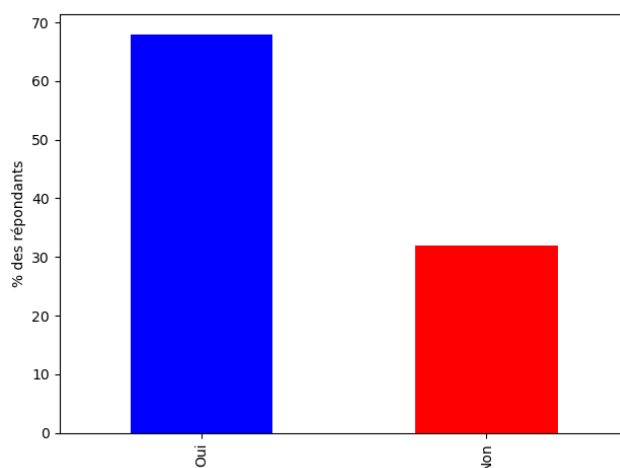


FIGURE 34 – Partage des données et des codes avec des collègues

Celui-ci se fait de différentes manières :

- partage par mail, par transfert de fichier, par clé USB ... sur demande,
- partage organisé via le dépôt des données et/ou des codes sur des plateformes adaptées (en particulier des cloud ou des forges logicielles).

De façon similaire, une majorité des répondants (voir figure 35) déclare avoir déjà réutilisé des données et/ou des codes, que ce soit leur propre production, ou celle de collègues.

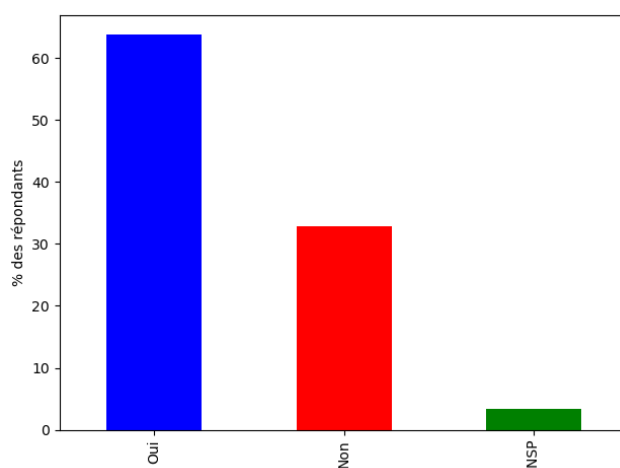


FIGURE 35 – Réutilisation de données ou de codes

Majoritairement (un peu plus de 60%), il n'y a pas eu de difficultés rencontrées dans cette réutilisation. Quand la réutilisation a été compliquée, les raisons invoquées sont :

- peu ou pas de documentation,
- des problèmes de compatibilités pour les codes, d'évolution des langages,
- l'utilisation de logiciels payants ou plus disponibles ou de dépendances obsolètes,
- des problèmes de formats,
- des problèmes de licence.

7 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.

7.1 Besoins autour des données de recherche

La figure 36 illustre une expression de besoins un peu globale sur la question de la gestion des données de la recherche, allant du Plan de Gestion des Données (*Data Management Plan, DMP*) aux aspect réglementaires et documentaires et aux problématiques techniques de stockage.

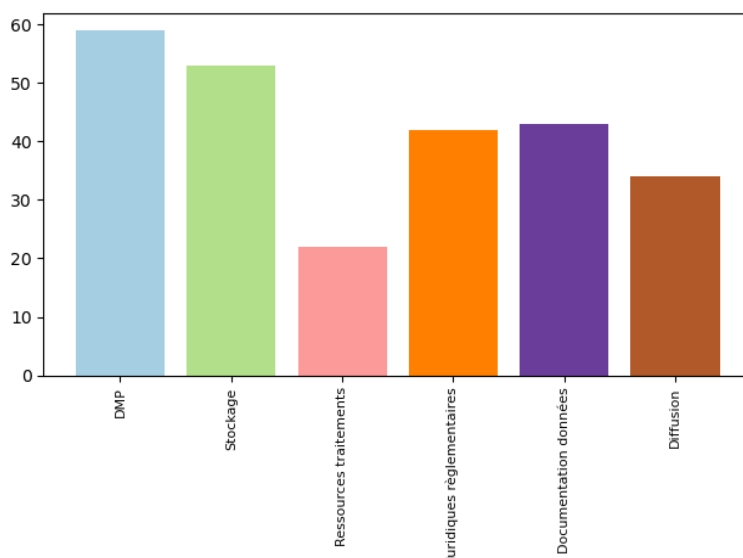


FIGURE 36 – Besoins en accompagnement et en formation sur les données

7.2 Besoins autour des codes de recherche

Comme pour les données ci-dessus, les besoins d'accompagnement et de formation autour des codes sont larges et couvrent à la fois les problématiques techniques et juridiques.

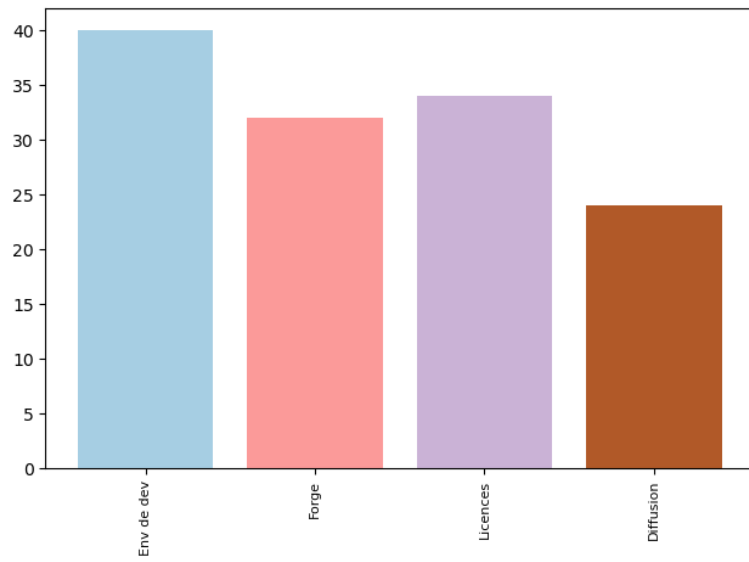


FIGURE 37 – Besoins en accompagnement et en formation sur les codes

8 Conclusions et perspectives

8.1 Synthèse

Les réponses au questionnaire permettent d'avoir un état des lieux assez représentatif de l'appropriation des problématiques sur les données de la recherche dans la communauté mathématique.

Un premier constat montre une compréhension très diverse de la notion de données de recherche, et une grande hétérogénéité entre les disciplines des mathématiques sur le fait ou non de manipuler des données. Certaines thématiques ne sont clairement pas concernées et les processus associés comme les demandes de Plans de Gestion de Données, ou l'ouverture des données ne sont absolument pas adaptés, et suscitent de l'incompréhension voire du rejet.

La problématique des codes de recherche est beaucoup plus partagée, même si elle est souvent associée aux questions des données.

Le sujet des ressources utilisées, en particulier pour le stockage des données, montre des usages extrêmement divers, avec beaucoup de pratiques individuelles et de système D. Le problème du partage des données avec des collaborateurs extérieurs (donc pas dans le référentiel des établissements) est régulièrement pointé comme délicat à réaliser. Les volumétries considérées sont généralement assez faibles, à de rares exceptions près. De même, les besoins de traitement sont assez limités en terme de ressources. Par ailleurs, peu de personnes sont concernées par des données sensibles (au sens du RGPD) mais dans ce cas, les difficultés rencontrées sont nombreuses, bien souvent par manque d'informations sur le sujet, et de connaissance des accompagnements existants.

L'activité autour des codes de recherche est importante, avec des pratiques certes à consolider mais globalement plutôt justes (en particulier sur l'utilisation des forges). Un point essentiel concerne l'ancrage dans le logiciel libre qui est extrêmement fort, tant sur l'utilisation des logiciels que sur la production de codes de recherche. Sur un plan similaire, le partage est très intégré dans les pratiques, partage vers sa communauté, mais également plus largement avec une conscience des enjeux d'ouverture bien présente. Si la notion de partage fait partie du processus de recherche, celui-ci n'est pas forcément formalisé et il concerne majoritairement les codes. La question des licences doit être adressée car il est nécessaire de faire évoluer certaines pratiques de partage sans licence. La citation des logiciels utilisés dans les publications, si elle est présente, est encore loin d'être majoritaire. Il existe également dans la communauté une sensibilité forte sur la reproductibilité des résultats de recherche.

Au sujet des données de recherche, le concept d'entrepôt de données est peu connu, de même que les Plans de Gestion de Données. Enfin, les besoins d'accompagnement et de formations sont très généraux, tant sur les données que sur les codes, et concernent l'ensemble des problématiques posées.

8.2 Quelques pistes pour aller plus loin

Des actions peuvent être initiées à plusieurs niveaux pour répondre aux différentes problématiques mises en exergue à travers cette enquête.

8.2.1 Au niveau local

Le soutien de proximité dans les laboratoires est le socle le plus important sur lequel doit s'appuyer tout ce qui pourra être mis en place. La présence de relais au plus près des chercheurs est une base indispensable, et les réseaux Mathrice et le RNBM pourront avoir un rôle de coordination, d'homogénéisation des pratiques et de mutualisation d'initiatives essentiel.

Dans un premier temps, des actions de sensibilisation adaptées aux besoins de la communauté pourront permettre de faire entrer le sujet dans les laboratoires, en particulier sur des points auxquels seront nécessairement confrontés les chercheurs :

- Les Plans de Gestion des Données, qui sont maintenant exigés par l'ensemble des financeurs. Ce thème a l'avantage de reposer l'ensemble de la sémantique (données, stockage, diffusion, archivage, ...).
- La problématique et l'importance des licences, avec une aide concrète à leur mise en place.
- La diffusion des données et des codes, et les obligations légales de la science ouverte.
- La question du stockage, de la sauvegarde et du partage de données, en particulier en insistant sur les services existants au niveau de la PLM.

8.2.2 Au niveau national

La dynamique autour de la science ouverte en lien avec les données et les codes est importante, tant au niveau national qu'international. Les mathématiques doivent trouver leur place dans ce mouvement et faire valoir leurs spécificités. En particulier, au niveau national, il est important de suivre les travaux des collègues du Comité pour la Science Ouverte (COSO), notamment le tout nouveau collègue Logiciels. Par ailleurs, l'intégration de la communauté mathématiques dans la dynamique de l'initiative Recherche Data Gouv⁸ pourrait être envisagée sous différents angles :

- via la structuration d'une collection spécifique sur l'entrepôt,
- via la réflexion autour d'un atelier de la donnée thématique.

8.2.3 Au niveau international

Des initiatives importantes se sont mises en place depuis plusieurs années au niveau européen et international, à travers l'European Open Science Cloud (EOSC)⁹, et la Research Data Alliance (RDA)¹⁰. Les mathématiciens sont peu présents dans ces grandes structurations, il semble important de fédérer la communauté internationale sur des projets communs, en particulier la question des standards de métadonnées pour les mathématiques.

8. <https://projet-recherchedatagv.ouvrirlascience.fr/>

9. <https://eosc.eu/>

10. <https://rd-alliance.org/>

A Questions de l'enquête

Les données de la recherche dans les laboratoires de mathématiques

Enquête proposée par les réseaux Mathrice et RNBM

Les réseaux Mathrice (<https://mathrice.fr/>) et RNBM (<https://www.rnbn.org>) s'associent afin de proposer aux membres des laboratoires de mathématiques les services et l'accompagnement les plus pertinents autour des données de la recherche.

Selon la définition proposée par l'ANDS (Australians Research Data Commons) et traduite par l'INIST-CNRS, "le terme de données de la recherche désigne les données sous forme de faits, d'observations, d'images, de résultats de programmes informatiques, d'enregistrements, de mesures ou d'expériences sur lesquelles un argument, une théorie, un test ou une hypothèse, ou un autre produit de la recherche est basé. Les données peuvent être numériques, descriptives, visuelles ou tactiles. Elles peuvent être brutes, nettoyées ou traitées, et peuvent être conservées dans tout format ou support."

Un groupe de travail commun a été constitué, et cette enquête a pour but de nous aider à bien identifier les problématiques auxquelles vous devez faire face.

Les résultats de cette enquête feront l'objet d'une analyse et de propositions afin de répondre aux questions et difficultés qui auront été exprimées.

Le temps nécessaire pour répondre est estimé autour de 15 mn.

Il y a 44 questions dans ce questionnaire.

Informations générales

- GREMAQ Groupe de recherche en économie mathématique et quantitative Toulouse
- ICJ Institut Camille Jordan Villeurbanne
- ICP Institut CNRS-PAULI Wien
- IDP Institut Denis Poisson Orléans
- IECL Institut Elie Cartan de Lorraine Vandoeuvre-lès-Nancy
- IF Institut Fourier Grenoble
- IFCAM Indo-French Center for Applied Mathematics Bangalore
- IHES Institut des hautes études scientifiques Bures-sur-Yvette
- IHP Institut Henri Poincaré Paris
- I2M Institut de Mathématiques de Marseille
- IMAG Institut Montpellierain Alexander Grothendieck Montpellier
- Imath Institut de Mathématiques de Toulon La Garde
- IMB Institut de mathématiques de Bordeaux Talence
- IMB Institut de Mathématiques de Bourgogne Dijon
- IMCCE Institut de mécanique céleste et de calcul des éphémérides Paris
- IMJ-PRG Institut de mathématiques de Jussieu - Paris Rive Gauche Paris
- IML Institut de Mathématiques de Luminy Marseille
- IMPA Institut National de Mathématiques Pures Rio de Janeiro
- IMT Institut de Mathématiques de Toulouse Toulouse
- INLN Institut Non Linéaire de Nice Sophia-Antipolis Valbonne
- IPRA Institut pluridisciplinaire de recherche appliquée dans le domaine du génie pétrolier Pau
- IRI Institut de recherche interdisciplinaire Lille
- IRIF Institut de Recherche en Informatique Fondamentale Paris
- IRMA Institut de recherche mathématique avancée Strasbourg Strasbourg
- IRMAR Institut de recherche mathématique de Rennes Rennes
- ISPED Institut de Santé Publique, d'Epidémiologie et de Développement Bordeaux
- IWD Institut Wolfgang Döblin VALBONNE
- Laboratoire Fibonacci Pisa
- Laboratoire Paul Painlevé Villeneuve d'Ascq
- LAG Laboratoire Alexander Grothendieck Bures-Sur-Yvette
- LAGA Laboratoire Analyse, Géométrie et Applications Villeneuve

Quel est votre laboratoire ? *

● Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- AGM Analyse, géométrie et modélisation Cergy-Pontoise
- Bibliothèque mathématique Jacques Hadamard Orsay
- BioSP Biostatistique et Processus Spatiaux Avignon
- CAMS Centre d'analyses et de mathématiques sociales Paris
- Cassiopée Laboratoire de Cosmologie, Astrophysique Stellaire et Solaire, de Planétologie et de mécanique des fluides
- Centre Borelli Gif sur Yvette
- CEREMADE Centre de recherches en mathématiques de la décision Paris
- Cermics Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique Marne la Vallée
- CERMSEM Centre de Recherche de Mathématiques, Statistique et Économie Mathématique Paris
- CES Centre d'économie de la Sorbonne Paris
- CIRM Centre International de Rencontres Mathématiques Marseille
- CMAP Centre de mathématiques appliquées Palaiseau
- CMLS Centre de mathématiques Laurent Schwartz Palaiseau
- CMM Centre de Modélisation Mathématique / Centro de Modelamiento Matemático Santiago de Chile
- CPT Centre de Physique Théorique Marseille
- CRAN Centre de Recherche en Automatique de Nancy Vandoeuvre Cedex
- CREST Centre de Recherche en Economie et en Statistique Palaiseau
- CRM Centre de recherches mathématiques Montréal
- DMA Département de mathématiques et applications Paris
- EM2C Laboratoire Énergétique Moléculaire et Macroscopique, Combustion Châtenay Malabry
- ESILV Ecole Supérieure d'Ingénieurs Léonard de Vinci Paris La Défense
- EURANDOM European Institute for Statistics, Probability, Stochastic Operations Research and their Applications Eindhoven
- GAATI Géométrie algébrique et applications à la théorie de l'information Tahiti, Polynésie Française

- LAMA Laboratoire d'analyse et de mathématiques appliquées de Marne-la-Vallée Marne-la-Vallée
- LAMA Laboratoire de Mathématiques de l'USMB Le Bourget du Lac
- LAMAV Laboratoire de Mathématiques et leurs Applications de Valenciennes Valenciennes
- LAMFA Laboratoire amiénois de mathématique fondamentale et appliquée Amiens
- LAMIA Laboratoire de Mathématiques Informatique et Applications Pointe à Pitre
- LaMME Laboratoire de Mathématiques et Modélisation d'Évry Évry
- LAP Laboratoire Analyse et Probabilités Évry
- LAREMA Laboratoire angevin de recherche mathématiques Angers
- LATP Laboratoire d'Analyse, Topologie, Probabilités Marseille
- LATP Laboratoire d'analyse, topologie, probabilités Marseille
- LBBE Laboratoire Biométrie et Biologie Evolutive Villeurbanne
- LDAR Laboratoire de Didactique André Revuz Paris
- LFC Laboratoire de Thermodynamique et Énergétique des Fluides Complexes Pau
- LIAFA Laboratoire d'Informatique Algorithmique: Fondements et Applications Paris
- LIFR-MI2P Laboratoire J.-V. Poncelet Moscow
- LIGM Laboratoire d'Informatique Gaspard-Monge Marne la Vallée
- LIM Laboratoire d'Informatique et de Mathématiques Sainte Clotilde
- LIP Laboratoire d'informatique du parallélisme Lyon
- LIX Laboratoire d'Informatique de l'Ecole Polytechnique Palaiseau
- LJAD Laboratoire Jean-Alexandre Dieudonné Nice
- LJK Laboratoire Jean Kuntzmann Grenoble
- LJLL Laboratoire Jacques-Louis Lions Paris
- LMA Laboratoire de Mathématiques d'Avignon Avignon
- LMA Laboratoire de mathématiques et applications de Poitiers Futuroscope
- LMAC Laboratoire de Mathématiques Appliquées de Compiègne Compiègne
- LMAH Laboratoire de Mathématiques Appliquées du Havre Le Havre
- LMAM Laboratoire de mathématiques et applications de Metz Metz
- LMAM Mathématiques et Applications des Mathématiques Vannes
- LMAP Laboratoire de mathématiques et de leurs applications de Pau Pau
- LMB Laboratoire de mathématiques de Besançon Besançon
- LMBA Laboratoire de mathématiques de Bretagne Atlantique Brest

- LMBP Laboratoire de Mathématiques Blaise Pascal Aubière
- LMI Laboratoire de Mathématiques pour l'Ingénieur Valenciennes
- LMI Laboratoire Mathématique de l'INSA Saint-Etienne-du-Rouvray
- LMIA Département de Mathématiques de l'institut IRIMAS Mulhouse
- LMJL Laboratoire de Mathématiques Jean Leray Nantes
- LML Laboratoire de Mathématiques de Lens Lens
- LMM Laboratoire Manceau de Mathématiques Le Mans
- LMNO Laboratoire de mathématiques Nicolas Oresme Caen
- LMO Laboratoire de mathématiques d'Orsay Orsay
- LMPA Laboratoire de Mathématiques Pures et Appliquées Joseph Liouville Calais
- LMR Laboratoire des Mathématiques de Reims Reims
- LMRS Laboratoire de mathématiques Raphaël Salem Saint-Étienne-du-Rouvray
- LMV Laboratoire de mathématiques de Versailles Versailles
- LORIA Laboratoire Lorrain de Recherche en Informatique et ses Applications Vandoeuvre-lès-Nancy
- LPMA Laboratoire de probabilités et modèles aléatoires Paris
- LPSC Laboratoire de physique subatomique et de cosmologie Grenoble
- LPSM Laboratoire de Probabilités, Statistique et Modélisation Paris
- LSTA Laboratoire de Statistique Théorique et Appliquée Paris
- LTCI Laboratoire Traitement et Communication de l'Information Paris
- M2N Laboratoire Modélisation mathématique et numérique Paris
- M2N-LaSIE Méthodes mathématiques et numériques pour les phénomènes de transfert La Rochelle
- MAATICAH Mathématiques et Algorithmique Appliquées aux Technologies de l'Information et de la Communication, Approche Historique Saint-Denis
- MADS Mathématiques de la Décision et Statistique Toulouse
- MaIAGE Mathématique et Informatique Appliquées du Génome à l'Environnement Jouy-en-josas
- MAP5 Mathématiques appliquées Paris 5 Paris
- MAS Mathématiques Appliquées aux Systèmes Châtenay Malabry
- Mathdoc Cellule de Coordination Documentaire Nationale pour les Mathématiques Grenoble
- MEPS Laboratoire de Mathématiques Et Physique pour les Systèmes Perpignan
- MIA Mathématiques et Informatique Appliquées Paris

- MIA Mathématiques, Image et Applications La Rochelle
- MIAT Mathématiques et Informatique Appliquées de Toulouse CASTANET TOLOSAN
- MIGP Laboratoire de Modélisation et Imagerie en Géosciences de Pau PAU
- MIPA Laboratoire Mathématiques, Informatique, Physique et Applications Nîmes
- MISTEA Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie Montpellier
- MODALX Laboratoire de Modélisation Aléatoire Nanterre
- MSSMat Mécanique des Sols, Structures et Matériaux Châtenay-Malabry
- PIMS - Europe The Pacific Institute for the Mathematical Sciences Vancouver
- POEMS Propagation des ondes : étude mathématique et simulation Palaiseau
- PPS Preuves, Programmes et Systèmes PARIS
- REHSEIS Recherches Epistémologiques et Historiques sur les Sciences Exactes et les Institutions Scientifiques Paris
- SMT Sciences de la Matière et Technologies Dijon
- STIX Sciences et Technologies de l'Information et de la Communication à Polytechnique Palaiseau
- UMPA Unité de mathématiques pures et appliquées Lyon
- UTT-Maths Université de Technologie de Troyes Troye
- XLIM Département Mathématiques et Sécurité de l'information du laboratoire XLIM Limoges
- Autre

Quel est votre employeur ? Merci de préciser si besoin dans le champ commentaire (l'université, l'école, l'organisme ...). *

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- CNRS
- Université
- Ecole
- INRIA
- INRA
- Autre organisme de recherche
- Autre

Faites le commentaire de votre choix ici :

Quel est ou quels sont vos domaines de recherche ? Merci d'indiquer quelques mots clés. *

Veuillez écrire votre réponse ici :

Quel est votre statut ? *

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Professeur
- Maître de conférence
- Directeur de Recherche
- Chargé de Recherche
- Ingénieur de Recherche
- Ingénieur d'Etudes
- Doctorant
- Post-doctorant

Autre

Quelle est votre tranche d'âge ? *

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- 18-25
- 26-35
- 36-50
- 51-65
- > 65

Informations sur les données que vous manipulez

Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement.

A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme Imfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.

*

➊ Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas / je ne comprends pas la question

Faites le commentaire de votre choix ici :

Quelles sont les données de la recherche que vous manipulez ? N'hésitez pas à compléter si la liste proposée ne vous semble pas correspondre à votre cas ou à préciser. Plusieurs choix sont possibles. *

➋ Ajoutez un commentaire seulement si vous sélectionnez la réponse.

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

- Des données chiffrées (tableaux ...)
-
- Des données de simulations
-
- Des données d'observations (données de capteurs, remontées de terrain ...)
-
- Des données expérimentales (en médecine, psychologie...)
-
- Des bases de données
-
- Des données géospatiales
-
- Des données textuelles (issues de corpus de textes, d'archives...)
-
- Des données d'enquêtes, de sondages (entretiens)
-
- Des données web (données d'usages, cartographies de sites, données de réseaux sociaux...)
-
- Des images ou des photos
-
- Des enregistrements audios ou vidéos
-

Aucune

Autre :

A combien estimez vous les volumes de données que vous manipulez ? Si vous ne savez pas évaluer la taille de vos données, vous pouvez par exemple dire si elles tiennent sur une clé USB, sur votre disque dur d'ordinateur ou si vous avez des difficultés à ouvrir les fichiers sur votre ordinateur ...

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

La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme Imfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

Veillez écrire votre réponse ici :

Où stockez-vous vos données ? Merci de préciser la plateforme utilisée si besoin. Plusieurs réponses sont possibles.

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

La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme Imfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

➋ Ajoutez un commentaire seulement si vous sélectionnez la réponse.

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

- en local, sur mon ordinateur professionnel
-
- en réseau, sur le serveur du laboratoire / de mon employeur
-
- sur un serveur externe hébergé par une autre institution
-
- sur un serveur externe hébergé par un prestataire privé (cloud)
-
- sur un disque dur externe
-
- sur une clé USB
-
- sur un CD ou un DVD
-
- en local, sur mon ordinateur privé
-

Autre :

Où sauvegardez-vous vos données (pour les sécuriser en cas de problèmes sur le stockage principal ou d'erreur de manipulation par exemple) ? Plusieurs réponses sont possibles. Merci de préciser la plateforme utilisée si besoin.

Répondre à cette question seulement si les conditions suivantes sont réunies : La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme lmfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

➊ Ajoutez un commentaire seulement si vous sélectionnez la réponse. Veuillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- en local, sur mon ordinateur professionnel
 en réseau, sur le serveur du laboratoire / de mon employeur
 sur un serveur externe hébergé par une autre institution
 sur un serveur externe hébergé par un prestataire privé (cloud)
 sur un disque dur externe
 sur une clé USB
 sur un CD ou un DVD
 en local, sur mon ordinateur privé
Autre :

Quels sont les points forts des solutions de stockage que vous utilisez ?

Répondre à cette question seulement si les conditions suivantes sont réunies : La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme lmfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

Veuillez écrire votre réponse ici :

Quels sont les points faibles des solutions de stockage que vous utilisez ? Quels sont les besoins qui ne sont pas couverts ?

Répondre à cette question seulement si les conditions suivantes sont réunies : La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme lmfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

Veuillez écrire votre réponse ici :

Quel type de traitements faites-vous sur ces données ? Par exemple, traitement statistique, simulation ...

Répondre à cette question seulement si les conditions suivantes sont réunies : La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme lmfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

Veuillez écrire votre réponse ici :

Quelles sont les ressources nécessaires pour ces traitements ? Par exemple, votre portable, un cluster de calcul, un serveur avec un accélérateur ...

Répondre à cette question seulement si les conditions suivantes sont réunies : La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme lmfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

Veuillez écrire votre réponse ici :

Manipulez-vous des données sensibles (c'est-à-dire soumises à des réglementations comme le RGPD, ou des données de santé ...) ?

Répondre à cette question seulement si les conditions suivantes sont réunies : La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme lmfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

Veuillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Si vous manipulez des données sensibles, quelles sont les difficultés que vous avez pu rencontrer et l'organisation que vous avez du mettre en place ?

Répondre à cette question seulement si les conditions suivantes sont réunies : La réponse était 'Oui' à la question '15 [sensible]' (Manipulez-vous des données sensibles (c'est-à-dire soumises à des réglementations comme le RGPD, ou des données de santé ...) ?)

Veuillez écrire votre réponse ici :

Informations sur les logiciels nécessaires à votre recherche

Utilisez-vous des logiciels existants pour votre recherche ?
Si vous ne pensez pas être concerné, n'hésitez pas à considérer des logiciels comme les tableurs, les logiciels de visualisation ou de conception de schéma, courbes ... ainsi que les langages et outils comme R, python, matlab ... *

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Si oui, quels sont ces logiciels ? Merci aussi de préciser le type de licence.

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '17 [utilisationlogiciel]' (Utilisez-vous des logiciels existants pour votre recherche ? Si vous ne pensez pas être concerné, n'hésitez pas à considérer des logiciels comme les tableurs, les logiciels de visualisation ou de conception de schéma, courbes ... ainsi que les langages et outils comme R, python, matlab ...)

Veillez écrire votre réponse ici :

Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ? *

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Ces développements sont-ils individuels (uniquement vous) ou collectifs (équipe, communauté plus importante ...) ? Si besoin préciser en commentaire.

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

➊ Ajoutez un commentaire seulement si vous sélectionnez la réponse.

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

développement individuel

développement collectif

Utilisez-vous un environnement de développement particulier (par exemple, des notebooks, des containers, des machines de calcul, un cloud ...) ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

Veillez écrire votre réponse ici :

Quels langages de programmation utilisez-vous pour vos développements ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

➊ Cochez la ou les réponses

Veillez choisir toutes les réponses qui conviennent :

- python
 R
 C / C++
 Fortran
 Julia
 Java
 Autre:

Utilisez-vous une forge ou un gestionnaire de version pour vos développements ? Merci de préciser quel outil dans le commentaire.

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

➊ Veuillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Faites le commentaire de votre choix ici :

Vos développements sont-ils utilisés par d'autres que vous ? Préciser.

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

➊ Ajoutez un commentaire seulement si vous sélectionnez la réponse.
Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

Oui, dans mon équipe
 Oui, dans mon laboratoire
 Oui, parmi les partenaires du projet
 Oui, dans ma communauté de recherche
 Oui, autres (à préciser)
 Non

Sous quelle licence vos développements sont-ils diffusés ? Sont-ils protégés par un dépôt APP (Agence pour la Protection des Programmes) ? N'hésitez pas à préciser les types de licence utilisés.

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

➊ Ajoutez un commentaire seulement si vous sélectionnez la réponse.
Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

Licence
 dépôt APP
 Pas de licence
 Je ne comprends pas la question

Si vos développements ne sont pas diffusés sous licence, pouvez-vous préciser pourquoi ? Quels sont les freins rencontrés ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
----- Scenario 1 -----
La réponse était à la question '25 [quellicence]' (Sous quelle licence vos développements sont-ils diffusés ? Sont-ils protégés par un dépôt APP (Agence pour la Protection des Programmes) ? N'hésitez pas à préciser les types de licence utilisés.)
----- ou Scenario 2 -----

La réponse était à la question '25 [quellicence]' (Sous quelle licence vos développements sont-ils diffusés ? Sont-ils protégés par un dépôt APP (Agence pour la Protection des Programmes) ? N'hésitez pas à préciser les types de licence utilisés.)

Veillez écrire votre réponse ici :

Les logiciels que vous utilisez dans vos recherches sont-ils cités dans vos publications ? Précisez si besoin. Par exemple, n'hésitez pas à signaler si vous déposez des programmes avec les articles.

Répondre à cette question seulement si les conditions suivantes sont réunies :
----- Scenario 1 -----

La réponse était 'Oui' à la question '17 [utilisationlogiciel]' (Utilisez-vous des logiciels existants pour votre recherche ? Si vous ne pensez pas être concerné, n'hésitez pas à considérer des logiciels comme les tableurs, les logiciels de visualisation ou de conception de schéma, courbes ... ainsi que les langages et outils comme R, python, matlab ...)

----- ou Scenario 2 -----
La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

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

- Oui
 Non
 Ca dépend

Faites le commentaire de votre choix ici :

Diffusion et partage

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

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Avez-vous déjà rédigé ou participé à la rédaction d'un Plan de Gestion de Données ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '28 [dmp]' (Savez-vous ce qu'est un Plan de Gestion de Données (PGD) ou Data Management Plan (DMP) ?)

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Si oui, avez-vous éprouvé des difficultés ? Lesquelles ? De quel type d'aide auriez-vous eu besoin ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '29 [redacDMP]' (Avez-vous déjà rédigé ou participé à la rédaction d'un Plan de Gestion de Données ?)

Veillez écrire votre réponse ici :

Connaissez-vous des entrepôts pour la diffusion des données de recherche ? Si oui, pouvez-vous les citer ?

Un entrepôt de données est une plateforme permettant le dépôt de données à des fins de diffusion comme HAL l'est pour les publications.

● Veillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non
 Je ne sais pas / je ne comprends pas la question

Faites le commentaire de votre choix ici :

Avez-vous déjà déposé des données dans un entrepôt ? Par exemple, les données associées à une publication. Si oui, le ou lesquels ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '31 [entrepots]' (Connaissez-vous des entrepôts pour la diffusion des données de recherche ? Si oui, pouvez-vous les citer ? Un entrepôt de données est une plateforme permettant le dépôt de données à des fins de diffusion comme HAL l'est pour les publications.)

● Veillez sélectionner une réponse ci-dessous

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Faites le commentaire de votre choix ici :

Pourquoi avez-vous déposé des données dans un entrepôt ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '32 [depot]' (Avez-vous déjà déposé des données dans un entrepôt ? Par exemple, les données associées à une publication. Si oui, le ou lesquels ?)

Veillez écrire votre réponse ici :

Le ou les entrepôts utilisés fournissent-ils des identifiants uniques (DOI) pour les jeux de données ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '32 [depot]' (Avez-vous déjà déposé des données dans un entrepôt ? Par exemple, les données associées à une publication. Si oui, le ou lesquels ?)

Veillez sélectionner une seule des propositions suivantes :

- Oui
 Non

Quels sont à votre avis les freins principaux à la diffusion de vos jeux de données, de manière restreinte ou en open data ?

Veuillez écrire votre réponse ici :

Avez-vous cité les jeux de données diffusés dans vos publications ?

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

La réponse était 'Oui' à la question '32 [depot]' (Avez-vous déjà déposé des données dans un entrepôt ? Par exemple, les données associées à une publication. Si oui, le ou lesquels ?)

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui. Préciser comment cela a été fait (DOI par exemple, autre ?)
- Non

Faites le commentaire de votre choix ici :

Réutilisation

Les résultats de vos recherches sont-ils selon vous reproductibles ? Par exemple, les codes et/ou les données sont facilement accessibles, les calculs sont bien documentés pour être reproduits ... N'hésitez pas à apporter des précisions.

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

----- Scenario 1 -----

La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme Imfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

----- ou Scenario 3 -----

La réponse était 'Oui' à la question '17 [utilisationlogiciel]' (Utilisez-vous des logiciels existants pour votre recherche ? Si vous ne pensez pas être concerné, n'hésitez pas à considérer des logiciels comme les tableurs, les logiciels de visualisation ou de conception de schéma, courbes ... ainsi que les langages et outils comme R, python, matlab ...)

----- ou Scenario 4 -----

La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas / je ne comprends pas la question

Faites le commentaire de votre choix ici :

Avez-vous déjà partagé des données ou des logiciels avec des collègues qui vous l'auraient demandé ? N'hésitez pas à préciser de quelle façon ce partage s'est réalisé. *

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

----- Scenario 1 -----

La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme Imfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

----- ou Scenario 3 -----

La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui
- Non

Faites le commentaire de votre choix ici :

Avez-vous déjà réutilisé des données et/ou des codes ? (vos propres données / codes ou ceux de quelqu'un d'autre ?) N'hésitez pas à préciser.

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

----- Scenario 1 -----

La réponse était 'Oui' à la question '6 [besoindata]' (Manipulez-vous des données pour votre recherche (utilisation de données existantes ou création de données) ? Si oui, merci de préciser succinctement. A noter qu'il faut prendre le terme données dans un sens assez large : par exemple, on peut inclure les bases de données disponibles en ligne (comme Imfdb.org ...). Même si vous avez un doute, n'hésitez pas à répondre.)

----- ou Scenario 2 -----

La réponse était 'Oui' à la question '17 [utilisationlogiciel]' (Utilisez-vous des logiciels existants pour votre recherche ? Si vous ne pensez pas être concerné, n'hésitez pas à considérer des logiciels comme les tableurs, les logiciels de visualisation ou de conception de schéma, courbes ... ainsi que les langages et outils comme R, python, matlab ...)

----- ou Scenario 3 -----

La réponse était 'Oui' à la question '19 [devlogiciel]' (Développez-vous des logiciels ou des bibliothèques / modules / packages dans le cadre de votre recherche ?)

📌 Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- Oui
- Non
- Je ne sais pas / je ne comprends pas la question

Faites le commentaire de votre choix ici :

Avez-vous rencontré des difficultés ? Lesquelles ?

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Oui' à la question '39 [réutilisation]' (Avez-vous déjà réutilisé des données et/ou des codes ? (vos propres données / codes ou ceux de quelqu'un d'autre ?) N'hésitez pas à préciser.)

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

- Oui
 Non

Faites le commentaire de votre choix ici :

Si non, voyez-vous des freins à la réutilisation de données ou de logiciels ? Merci de préciser.

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'Non' ou 'Je ne sais pas / je ne comprends pas la question' à la question '39 [réutilisation]' (Avez-vous déjà réutilisé des données et/ou des codes ? (vos propres données / codes ou ceux de quelqu'un d'autre ?) N'hésitez pas à préciser.)

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

- Oui
 Non

Faites le commentaire de votre choix ici :

Vos besoins en accompagnement et formation

Quels sont vos besoins en accompagnement et en formation sur les données ? N'hésitez pas à compléter les propositions.

➌ Ajoutez un commentaire seulement si vous sélectionnez la réponse.
Veuillez choisir toutes les réponses qui conviennent et laissez un commentaire :

Plan de Gestion de Données
 Stockage des données
 Ressources pour le traitement des données
 Aspects juridiques et réglementaires
 Description et documentation des données, métadonnées
 Diffusion des données

Autre :

Quels sont vos besoins en accompagnement et en formation sur les logiciels et codes ? N'hésitez pas à compléter les propositions.

➍ Ajoutez un commentaire seulement si vous sélectionnez la réponse.
Veuillez choisir toutes les réponses qui conviennent et laissez un commentaire :

Environnement de développement (notebook, container ...)
 Forge logiciels
 Licences et dépôt APP
 Diffusion des logiciels

Fin du questionnaire

Merci de votre participation à cette enquête. Vos réponses nous seront précieuses pour nous permettre de déployer les services et l'accompagnement les plus pertinents.

Si vous souhaitez nous contacter : gt_donnees@listes.rnbnm.org

Si vous le souhaitez, vous pouvez nous laisser votre mail afin de nous permettre de vous contacter ultérieurement pour que nous approfondissions vos réponses

Veuillez écrire votre réponse ici :

Merci pour le temps consacré à cette enquête. Vos réponses seront précieuses pour nous permettre d'avancer et pour vous proposer les services et l'accompagnement les plus pertinents.

Envoyer votre questionnaire.