



# Synthèse de l'étude « Décliner la science ouverte »

Comité pour la science ouverte - Collège données de la recherche  
Groupe de travail « réussir l'appropriation de la science ouverte »

soutenu par



**MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR  
ET DE LA RECHERCHE**

*Liberté  
Égalité  
Fraternité*

## Table des matières

|  |    |
|--|----|
| 1   Contexte et objectifs de l'étude.....  | 3  |
| L'étude « Décliner la science ouverte ».....   | 3  |
| 2   Méthodologie.....  | 4  |
| 3   Résultats.....   | 6  |
| 3.1 Quels facteurs prendre en considération pour mieux appréhender la diversité de pratiques associées aux données en recherche ?.....   | 6  |
| Typologie de pratiques et personæ.....   | 6  |
| Naviguer finement dans les pratiques : démarche, outils et statut.....   | 9  |
| 3.2 Comment accompagner l'évolution des pratiques associées aux données en lien avec les incitations et les obligations portées par les politiques publiques de science ouverte ?..... | 12 |
| 4   Points d'attention.....  | 13 |
| 4.1 Comprendre en finesse les démarches de recherche.....  | 13 |
| Diversifier la terminologie employée autour des « données ».....   | 13 |
| Privilégier la notion de « qualité » en recherche plutôt que celle de « reproductibilité ».....  | 14 |
| Porter attention aux différentes formes de plus-values retirées du travail de recherche.....   | 14 |
| Démarches de recherche : spécificité clinique/ expérimentale/computationnelle.....   | 15 |
| 4.2 Appréhender différentes pratiques de mise à disposition des données.....   | 16 |
| Penser à la réutilisation de données et d'autres ressources et aux publics concernés.....  | 16 |
| Distinguer différentes limites à la mise à disposition et les leviers d'amélioration.....  | 17 |
| Mettre en avant les enjeux de conservation et de sécurité des données.....   | 17 |
| 4.3 Connaître les modalités d'apprentissage et les pratiques collaboratives.....   | 18 |
| Découverte et formation aux outils : un échange entre pairs.....   | 18 |
| Porter attention aux interfaces.....   | 18 |
| 4.4 Diversifier les types d'accompagnement.....  | 19 |
| Distinguer différents besoins d'accompagnement.....  | 19 |
| Élaborer un maillage des dispositifs d'accompagnement aux données au plus près des équipes.....  | 20 |
| Être vigilant aux enjeux de médiation au sein des équipes de recherche.....  | 20 |
| 4.5 Prendre en considération le statut et les enjeux de carrière.....  | 21 |
| 5   Limites.....   | 22 |
| 6   Conclusion.....  | 22 |
| 7   Annexes.....   | 23 |
| 7.1 Résultats principaux du questionnaire « données et science ouverte ».....  | 23 |
| 7.2 Accès aux ressources produites.....  | 27 |
| 7.3 Crédits.....   | 27 |

# 1 | Contexte et objectifs de l'étude

L'étude Décliner la science ouverte (Decliner-SO) a été réalisée dans le cadre du projet « **Réussir l'appropriation de la science ouverte** » mené au sein du Comité pour la science ouverte. Elle a été portée par un groupe de travail multi-disciplinaire et professionnel du collège Données de la recherche <sup>1</sup>. Ce projet s'est déroulé de mai 2020 à décembre 2021 et était composé de trois axes de travail :

- Le design et l'organisation d'Ateliers Juridiques Science Ouverte (AJSO)<sup>2</sup>
- La participation au groupe de travail cahiers de laboratoire électroniques du Comité pour la science ouverte<sup>3</sup>
- L'étude Décliner la science ouverte (Décliner-SO) faisant l'objet de la synthèse ci-dessous.<sup>4</sup>

## L'étude « Décliner la science ouverte »

L'étude « Décliner la science ouverte » constitue une enquête de terrain auprès de professionnel·le·s de la recherche dans diverses disciplines pour :

- mieux appréhender **les pratiques associées aux données et leurs évolutions avec la science ouverte** ;
- comprendre **les facteurs différenciant ces pratiques** (discipline, démarche de recherche, etc.)
- apporter **un accompagnement adapté** aux besoins de différentes communautés de recherche.

Deux problématiques ont été définies lors de l'étude :

1. Quels facteurs prendre en considération pour mieux appréhender la diversité de pratiques associées aux données en recherche ?
2. Comment accompagner l'évolution des pratiques associées aux données en lien avec les incitations / obligations portées par les politiques publiques de science ouverte ?

---

1 Le projet « Réussir l'appropriation de la Science Ouverte (SO) » a été initié à la suite des travaux d'un des groupes de travail de ce collège sur [l'usage et la gouvernance des données](#).

2 Les Ateliers Juridiques Science Ouverte (AJSO) *designés* et organisés en ligne pour faciliter le dialogue et les échanges entre juristes et professionnels de la recherche. L'objectif était collectivement de faire émerger des problématiques juridiques en lien à des thématiques associées aux données et à la science ouverte. Trois ateliers ont eu lieu de novembre 2020 à mai 2021 (1/images de la recherche, 2/cycle de vie de la donnée et 3/données personnelles) et ont regroupés plus de 150 personnes. Une [présentation de synthèse](#) et un [kit d'appropriation des ateliers](#) sont disponibles sur le site du Comité pour la science ouverte.

3 Voir : Gilles Mathieu, Dominique Pigeon, Tovo Rabemanantsoa, Christophe Chipeaux, Simon Duvillard, et al.. Rapport du groupe de travail sur les cahiers de laboratoires électroniques. [Rapport de recherche] Comité pour la science ouverte. 2021, 68 p. [DOI :10.52949/3](#)

4 Pour en savoir plus sur la mission « Réussir l'appropriation de la science ouverte » consultez la page d'ouvrirlascience et la liste des livrables produits : <https://www.ouvrirlascience.fr/reussir-l-appropriation-de-la-science-ouverte/>

## 2 | Méthodologie

L'étude réalisée s'est appuyée sur **des méthodes mixtes** avec **deux premières phases qualitatives** comprenant des **entretiens** (exploratoires, observation des pratiques), des **focus group**, et **une journée d'étude**. À la suite, une **phase quantitative** a été réalisée s'appuyant sur **la conception, diffusion et analyse d'un questionnaire** auprès des enseignant·e-chercheur·s, chercheur.e.s et personnels de la recherche en France (plus de 400 réponses). L'étude s'est finalisée par le **croisement des résultats** de ces phases<sup>5</sup> et une **démarche design** pour faciliter l'appropriation du contenu (cf. Figure 1).

Le travail de recherche s'est inscrit dans une **approche collaborative** entre les différentes membres du groupe de travail issues de disciplines (biologie, histoire de l'art, histoire, santé, *Science & Technologies Studies - STS*) et de domaines professionnels de recherche variés (archives, bibliothèques, recherche, direction et stratégie, etc.). **Une démarche de science ouverte** (cf. Figure 2.) a été également éprouvée afin de mettre à disposition les informations collectées (dans le respect du RGPD) pour faciliter le cheminement dans l'évolution de la recherche (partage des synthèses intermédiaires) et la reproductibilité des résultats quantitatifs (scripts, mises à disposition des données) (cf. 7.2 Accès aux ressources produites).

### Méthodologie de l'étude

**Groupe de travail (GT) :** multi-disciplinaire/professionnel  
**Méthodes mixtes** quali-quant

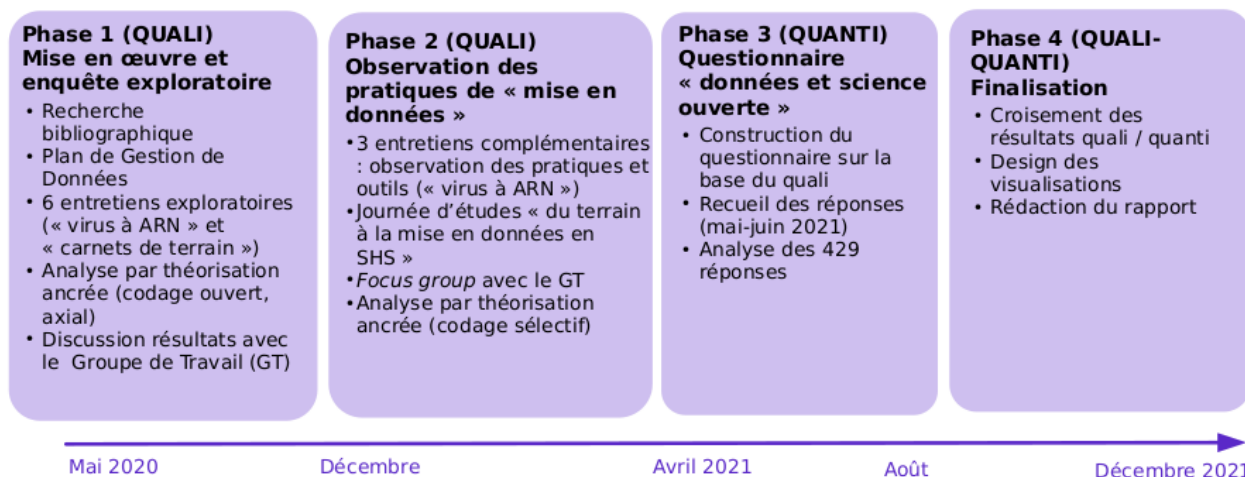


Figure 1: Résumé des différentes étapes de la méthodologie de l'étude Décliner-SO.

5 Les résultats des phases qualitatives ont été obtenus par le biais d'une analyse par théorisation ancrée. En ce qui concerne les résultats quantitatifs, ils proviennent d'une analyse statistique univariée (tris à plat et tris croisés) et multivariée (ACM puis HCPC). Pour en savoir plus, consultez la note méthodologique (cf. 7.2 Accès aux ressources produites).

## Workflow de l'étude

### Liste des ressources

Documents de travail / sources

- 0 : Trames et verbatim des entretiens
- 1 : Synthèse journée d'études
- 2 : Trame questionnaire
- 3 : Données questionnaire et scripts

Livrables

- 4 : Rapport final et annexes (trames entretiens et questionnaire, Plan Gestion de Données)
- 5 : Synthèse

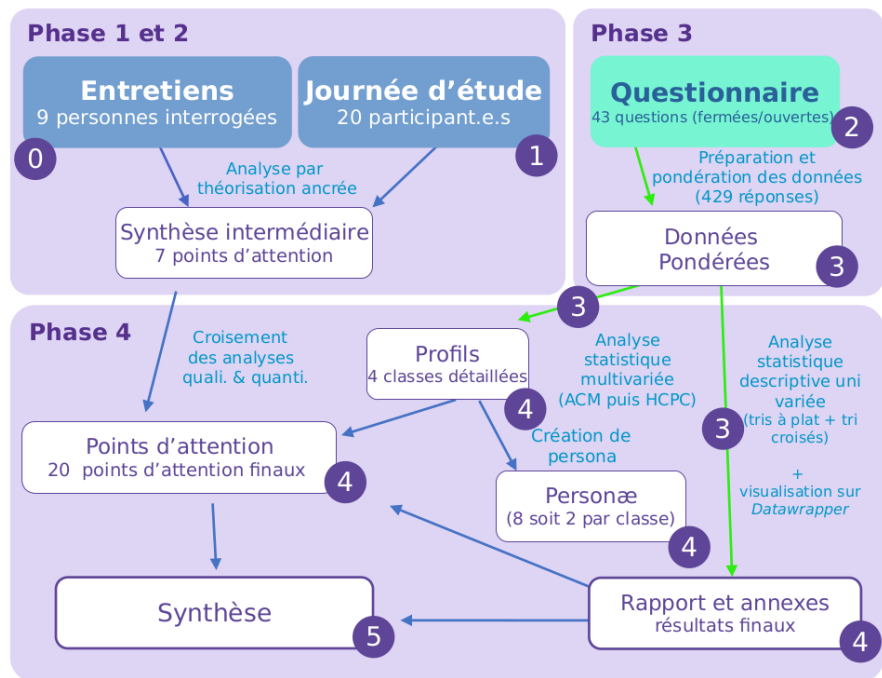
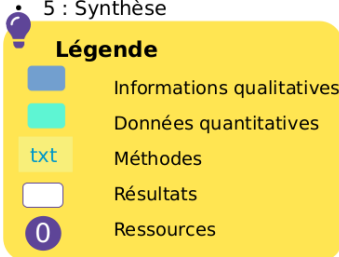


Figure 2: Suivi de la chaîne de traitement des informations recueillies (workflow) dans une démarche de science ouverte.

## 3 | Résultats

### 3.1 Quels facteurs prendre en considération pour mieux appréhender la diversité de pratiques associées aux données en recherche ?

#### Typologie de pratiques et personæ

Bien que les disciplines soient un facteur important pour différencier diverses pratiques associées aux données, l'étude montre qu'il est important d'**aller au-delà de la seule grille de lecture disciplinaire** et de distinguer d'autres facteurs différenciants.

Les premiers entretiens exploratoires nourris d'une revue de la littérature, ont amené à définir **les « pratiques associées aux données »**, comme l'ensemble des étapes nécessaires à la production de données (« mise en données »)<sup>6</sup> et à leur **mise à disposition** (allant du partage restreint à l'*open data*).

En plus des champs disciplinaires (sciences techniques et médecine – STM/ sciences humaines et sociales – SHS), un autre facteur pris en considération a été **la nature individuelle ou collective du travail de recherche**. Sur la base de ces facteurs, l'analyse statistique multivariée (Analyse des Correspondances Multiples - ACM) a permis de mettre en évidence **4 grands types de pratiques (expérimental, collaboratif, computationnel, solitaire)** (cf. Figure 3).

- Si l'axe « discipline » colore fortement ces typologies de pratiques – par exemple **le profil « expérimental »** est associé à des **personnes issues essentiellement des sciences de la vie et de la matière** – d'autres profils tel que **le « computationnel »** regroupe des individus provenant de différentes disciplines (de l'informatique à la linguistique) mais qui partagent **une même culture de la donnée** et souvent une **connaissance des logiciels libres et open source**.
- Par ailleurs, au sein des Sciences Humaines et Sociales, où la représentation d'un « chercheur solitaire » peut encore être dominante, **un profil « collaboratif » se distingue** avec des personnes mettant en œuvre des pratiques collectives à différentes étapes de leur recherche, ou tout au moins souhaitant se former à ces dernières.
- Enfin, **le profil « solitaire »** (non restreint aux SHS), regroupe des individus qui **mènent leur recherche seuls sans que cela soit forcément voulu** en fonction du statut ou des conditions de travail, par exemple dans le cas du doctorat.

---

<sup>6</sup> Le terme « mise en données » est issu des travaux en STS étudiant la production de connaissances et le travail des données, les données étant considérées comme un construit faisant l'objet de différentes étapes, échanges, utilisation d'outils, de processus, jusqu'à la production de ce que l'on appelle « données » dans le but notamment d'être partagé, échangé et avoir valeur de preuve. D'autres concepts au cœur de cette étude sont par exemple ceux de data journeys, de datafication ou encore de « public des données » (Jaton et Vinck 2016 ; Gruson-Daniel et De Quatrebarbes 2019 ; Gitelman 2013 ; Bowker et Star 2000 ; Heaton et Millerand 2013).

## Profils

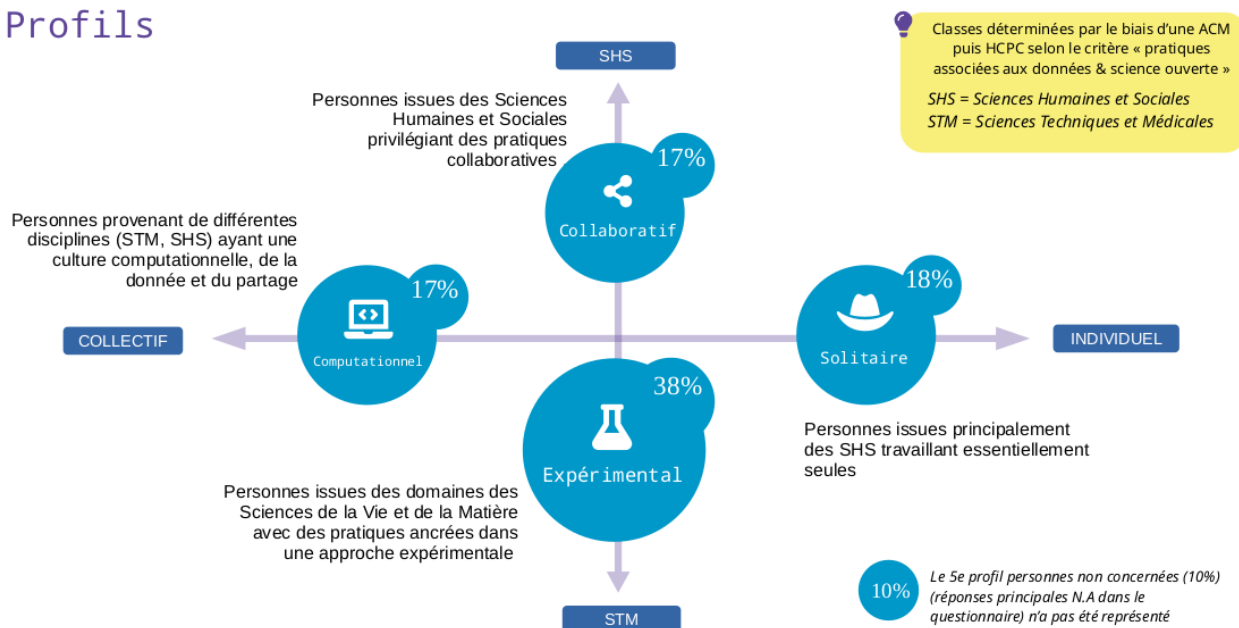


Figure 3: Présentation des 4 profils-types de pratiques associées aux données issus de l'analyse du questionnaire (ACM puis HCPC). Distinction des profils en fonction d'un axe discipline : Sciences, Techniques & Médecine (STM) / Sciences Humaines et Sociales (SHS) et nature du travail plus individuel ou collectif.

Pour mieux appréhender ces profils (typologie de pratiques) construits sur les résultats sur les résultats d'une ACM (Analyse des Correspondances Multiples), des personæ ont été proposés (cf. Figure 4 et 5). Les personæ sont des personnages fictifs, constitués à partir de l'exploration des réponses au questionnaire (les plus représentatives de chaque classe) en se nourrissant des résultats des phases d'analyse qualitative. Chaque persona est présenté sous forme de fiche descriptive et donne un aperçu de situations concrètes rencontrées par une diversité de professionnel·le·s de la recherche (maitre·sse de conférence, ingénieur·e de recherche ou d'étude, chercheur·e·s, etc.).





**Expérimental**

## Louis Tapures

Domaine Chimie  
Lieu Directeur de recherche au CNRS  
Expérience Depuis plus de 10 ans



Laboratoire  
Ordinateur analyse  
Ordinateur/papier réflexion

**Terminologie**

Mesures

Réplicabilité, reproductibilité, tracabilité

**En bref** ID 2113254999

Familiarité avec la reproductibilité ★★★★★

Plus-value Préparation d'échantillon et définition de protocoles expérimentaux.

**Story**

Louis travaille depuis maintenant 15 ans dans un laboratoire de chimie moléculaire où il dirige une équipe spécialisée dans la caractérisation de complexes organométalliques. Il est en contact au quotidien avec d'autres chercheurs.e.s, post-doctorant.e.s, doctorant.e.s et ingénieur.e.s qu'il conseille pour la définition des protocoles des expériences menées au sein du laboratoire. Il aide essentiellement à l'analyse des mesures produites. L'utilisation d'un tableur de type Excel lui est largement suffisant. Il prend soin également de former son équipe car c'est au sein du laboratoire que les bonnes habitudes se prennent. Il veille en plus à la sauvegarde des données du laboratoire (se faisant sur un serveur interne). Les enjeux de reproductibilité et de réplication lui sont familiers ; néanmoins, il préfère partager ses données de façon ciblée à ses collègues ou à d'autres équipes dans le cadre de coopération internationale. Aujourd'hui il cherche à être accompagné pour savoir comment mieux conserver les données et les mettre à disposition. Il espère avoir une aide financière pour recruter une personne à ce sujet.

**Mise à disposition des données**

**Conservation des données :** sur un serveur interne du laboratoire.

**Mise à disposition et réutilisation :** partage (parfois) ses données en accès restreint à un public ciblé. Réutilise (parfois) des données des autres. Ne sait pas si ses données pourraient être réutilisables. Partage les données sur des plateformes des éditeurs en complément d'un article.

**Autres ressources à partager :** protocoles méthodologiques.

**Limitations :** un temps nécessaire trop important, un manque d'habitude, des barrières juridiques et économiques.

**Obligations de partage :** de la part du financeur.

**Pratiques, mise en données**

**Étapes de travail :**

- 21-30 % analyser, sauvegarder/archiver, vérifier la qualité des données.
- 11-20 % structurer une base de données.

**Sources :** issues d'instruments de mesure, d'articles scientifiques et de bases de données partagées par ses collègues.

**Ressources supplémentaires produites :** notes diverses sous format papier et numérique notamment dans un cahier de laboratoire, logiciels d'analyse, fichiers de données stockés sur son ordinateur.

**Environnement numérique**

**Outils utilisés :** Logiciels de types tableurs sous Windows.

**Collaboration :** Taille de l'équipe de 6 à 10 personnes. Travail avec profils variés (IT, chercheur.e.s permanents.e.s, doctorant.e.s, etc.). Collabore à l'international essentiellement. Utilise des dossiers partagés.

**Contexte d'initiation :** par les autres membres ou dans le cadre de ses postes précédents (anciens employeurs).

**Accompagnement**

En contact avec des personnes accompagnant les pratiques liées aux données : ✕ ✓

Fréquence du soutien : ★★★★☆

Niveau de satisfaction : ★★★★☆

**Besoins**

Aide financière, technique et ressources humaines.

Des spécialistes pour répondre à ses questions dans son établissement, une plateforme dédiée en ligne.

Aide pour le stockage, la mise à disposition et la publication des données.

Figure 4: Exemple de deux fiches descriptives (personae) permettant d'illustrer sur la base de personnage fictif les profils-types ici « computationnel » et « expérimental ». Les photos sont sous licence Unsplash.





Figure 5: Exemple de deux fiches descriptives (personae) permettant d'illustrer sur la base de personnage fictif les profils-types ici « collaboratif » et « solitaire ». Les photos sont sous licence Unsplash.

## Naviguer finement dans les pratiques : démarche, outils et statut

Afin d'obtenir un niveau de granularité plus fin dans l'analyse de ces deux facteurs généraux (discipline et nature collaborative/solitaire), trois critères additionnels ont été explorés :

1. la démarche de recherche ;

2. les outils et modalités d'apprentissages ;
3. le statut et fonction en recherche.

**Pour l'axe « démarche de recherche »**, nous incluons différents éléments tels que **l'environnement de recherche** (laboratoire, clinique, terrains, etc.), la provenance des données (instruments de mesure, archives, etc.), le rapport aux données notamment dans **les qualificatifs employés**, les **éléments associés à une recherche de qualité**, la **plus-value** du travail de recherche ainsi que les **étapes de « mise en données »**<sup>7</sup>. Ces critères sur la base de l'analyse des réponses au questionnaire ont permis de caractériser différentes démarches de recherche influençant le rapport aux données (cf. Figure 6). Par exemple, dans le cadre de travaux en laboratoire, la plus-value des recherches est liée à l'expérimentation et l'on parle plus volontiers de mesures et valeurs, pour qualifier les données. Dans le cadre de travaux sur des archives ou des études de terrain, les termes de corpus et matériaux sont largement employés pour une plus-value principale associée à la collecte de données rares et d'un travail de théorisation. Pour des recherches cliniques entre autres, la publication de « résultats négatifs » avec le partage des « données négatives »<sup>8</sup> sont des pratiques encore peu fréquentes mais mises en avant dans le cadre de démarche de science ouverte.

## Les démarches de recherche

Comprendre en finesse les démarches de recherche pour s'adresser distinctement aux communautés de recherche

| Environnement de travail   | Discipline                             | Critères associés à la qualité de la recherche | Termes utilisés pour qualifier les données                 | Plus-value du travail de recherche                              | Provenance                   |
|----------------------------|--|--|--|---|------------------------------|
| <b>Laboratoire</b>         | Sciences, Techniques et Médecine       | Reproductibilité, réplicabilité                | ↗ Mesures, valeurs, données brutes, ↗ Sources              | Expérimentation   | Instruments de mesure        |
| <b>Terrains, Archives,</b> | Lettres, Sciences Humaines et Sociales | Explicabilité, transparence                    | ↗ Corpus, matériaux, sources ↗ valeurs et données brutes   | Collecte de données rares et de données standards, théorisation | Archives, site web, terrains |
| <b>Clinique</b>            | Sciences de la vie et de la santé      | Résultats négatifs                             | ↗ Pipelines, données nettoyées, données négatives ↗ corpus | Modélisation, automatisation de workflow                        | Instruments de mesure        |

Figure 6: Tableau récapitulatif de différentes démarches de recherche en fonction de l'environnement de travail (laboratoire, terrains/archives et clinique).

Dans la catégorie « **outils et modalités d'apprentissage** », le regard porte plus spécifiquement sur le **contexte matériel** (même si numérique) des pratiques par le biais

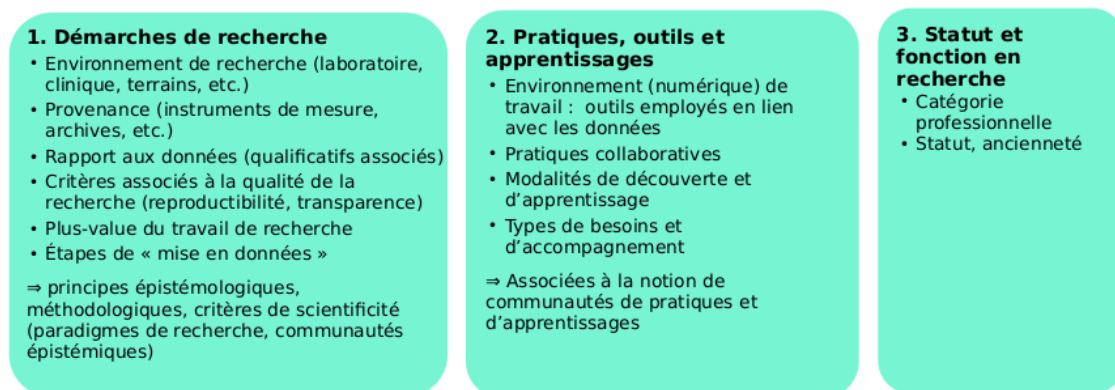
7 Cet axe se rattache ainsi à des principes épistémologiques, méthodologiques et de critères de scientificité associés à différents paradigmes de recherche et communautés épistémiques.

8 Pour en savoir plus, consultez l'article « Les données négatives : la partie immergée de l'iceberg des publications scientifiques » : <https://www.dataacc.org/bonnes-pratiques/diffuser-des-resultats-negatifs/les-donnees-negatives-la-partie-immergee-de-liceberg-des-publications-scientifiques/>

de l'usage d'un ensemble d'outils. Il s'agit de considérer **les modalités de découverte et d'apprentissage d'outils, l'appropriation d'environnements numériques de travail** et de s'intéresser **aux habitudes plus ou moins collaboratives** de travail et aux **besoins recensés en accompagnement**. Cet axe offre une distinction de différentes communautés de pratiques et d'apprentissage. En dernier lieu, une dernière catégorie est celle du **statut et des fonctions allouées** aux individus en recherche par exemple **la catégorie professionnelle** ou encore le **statut et l'ancienneté** au sein de la recherche (doctorant, fonctionnaire, etc.).

Pour chaque axe, l'ensemble des critères détaillés (cf. Figure 7) a pour but d'apporter une compréhension fine des divers types de rapport aux données et de représentation de celles-ci que peuvent avoir les professionnel·le·s de la recherche. Ces critères peuvent également influencer sur leurs appréhensions et/ou motivations quant à l'ouverture ou au partage des données dans une démarche de science ouverte et, en conséquence, sont à prendre en compte dans les aides et accompagnements à privilégier.

### Décliner la science ouverte en fonction de ...



*Figure 7: Résumé des facteurs différenciant les pratiques associées aux données (démarches de recherches, pratiques/outils/apprentissages, statut et fonction en recherche) et leurs caractéristiques.*

## 3.2 Comment accompagner l'évolution des pratiques associées aux données en lien avec les incitations et les obligations portées par les politiques publiques de science ouverte ?

Pour répondre à cette problématique, cinq pistes d'orientation pour des mesures d'accompagnement ont été déterminées sur la base des enseignements des phases qualitatives (entretiens, observations des usages, journée d'étude) et des résultats aux questionnaires « données et science ouverte » (cf. annexe).

- Piste d'orientation 1 : Comprendre en finesse les démarches de recherche ;
- Piste d'orientation 2 : Appréhender différentes pratiques de mise à disposition des données ;
- Piste d'orientation 3 : Connaître les modalités d'apprentissages et les pratiques collaboratives ;
- Piste d'orientation 4 : Diversifier les types d'accompagnement ;
- Piste d'orientation 5 : Prendre en considération le statut et les enjeux de carrière.

À propos **des mesures incitatives à la science ouverte dans le domaine des données de recherche**, nous incluons par exemple l'**application des principes FAIR** pour les données (*Findable, Accessible, Interoperable, Reusable*), la mise en œuvre de **plans de gestion de données (PGD)**, l'incitation à **une plus grande reproductibilité** des travaux de recherche, la mise en place d'accompagnements et le déploiement d'infrastructures pour la mise à disposition de données.

Pour chaque piste, différentes thématiques ont été distinguées chacune associée à des points d'attention. Les 20 points d'attention visent à faciliter l'évolution des pratiques associées aux données et aux incitations de politiques publiques de science ouverte tout en s'adaptant aux contextes variés de la recherche académique.

## 4 | Points d'attention

Les points d'attention sont présentés en fonction de 5 pistes d'orientations définies en amont et de sous-thématiques associées. Ils ont été formulés, de manière inductive à partir des résultats de l'étude<sup>9</sup>, et reflètent les cadres théoriques engagés et les terminologies des personnes enquêtées.

Un éclairage sur chaque point d'attention est proposé par la suite. L'accès à l'ensemble des résultats de l'étude et des explications est disponible dans la présentation générale (cf.7.2 Accès aux ressources produites).

### 4.1 Comprendre en finesse les démarches de recherche

#### Diversifier la terminologie employée autour des « données »

**Point d'attention 1 :** Ne pas en rester à la notion de « données/data ». Utiliser un ensemble de termes plus précis et spécifiques pour désigner les différents objets manipulés et produits au cours de la recherche.

**Explication :** Le terme « données » est employé majoritairement dans les recommandations/incitations de politiques publiques de science ouverte afin de désigner toutes informations menant à la production de résultats scientifiques. Les données de la recherche ont été définies par l'OCDE : comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. En dehors de cette définition, d'autres sont employés plus communément au sein des communautés de recherche pour décrire les éléments qui servent à l'obtention de résultats de recherche. Ainsi, il est recommandé de ne pas en rester à la notion de « données/data », mais d'utiliser des termes plus précis et spécifiques pour désigner les différents objets manipulés et produits au cours de la recherche. En fonction des communautés de recherche auxquelles on s'adresse, plusieurs expressions peuvent être utilisées : « bases de données », « jeux de données », « corpus », « archives », « sources », « matériaux », « mesures ».

---

9 Les points d'attention sont issus des résultats qualitatifs et quantitatifs de l'étude, c'est-à-dire les entretiens en sciences de la vie et en archéologie, la journée d'étude avec des professionnels de la recherche en sciences humaines et sociales ainsi que le questionnaire « données et science ouverte ». Bien qu'une majorité de répondant.e.s étaient en sciences humaines et sociales, les résultats et l'analyse qui en a été faite se basent sur une pondération des disciplines pour avoir un échantillon représentatif de la répartition des disciplines de l'ESR (cf. note méthodologique sur Gitlab : <https://code.inno3.eu/ouvert/decliner-so>).

## Privilégier la notion de « qualité » en recherche plutôt que celle de « reproductibilité »

**Point d'attention 2 :** Élargir les enjeux de reproductibilité à ceux de qualité en recherche. Employer d'autres termes tels que ceux de transparence, de traçabilité, d'explicabilité notamment avec des communautés SHS.

**Point d'attention 3 :** Intégrer dans les réflexions sur la qualité en recherche et la mise à disposition des données de recherche, les problématiques de valeurs scientifiques et éthiques (intégrité, honnêteté) et d'impact de la recherche dans ses différentes dimensions (sociales, économiques, techniques, etc.).

**Explication :** Aujourd'hui, les enjeux de reproductibilité font partie intégrante des discours et incitations de politiques publiques de science ouverte<sup>10</sup>. Or il s'agit de se détacher du terme reproductibilité pour aborder plus largement la question de la « qualité » en recherche. En effet, **la notion de « reproductibilité » s'applique plus spécifiquement dans le cadre de recherche impliquant des instruments de mesure et l'emploi de méthodes computationnelles** (vérification de calculs à partir de l'accès aux codes sources et aux données de départ). D'autres termes sont plus inclusifs pour aborder plus globalement la question de la qualité de la recherche au sein de milieux de recherche divers. Par exemple, dans les résultats du questionnaire, **le principe de « transparence »** était privilégié **dans des contextes de recherche en droit et économie**<sup>11</sup>. Le concept **« d'explicabilité »** s'utilise dans le cadre de travaux en SHS qui nécessitent la constitution de corpus, ou la construction de bases de données. Par ailleurs, la notion de **« répliquabilité »** peut être utilisée préférentiellement dans le cadre de recherches expérimentales lorsqu'il s'agit de reproduire une expérience. Ceci implique de considérer l'accès aux protocoles méthodologiques (pas exclusivement aux données et aux codes sources). Plusieurs commentaires ont fait également ressortir l'importance d'associer aux réflexions sur la question de la qualité en recherche, **les principes et valeurs éthiques (intégrité, honnêteté, etc.) de la recherche et ses impacts**(sociaux, économiques, techniques, etc.).

## Porter attention aux différentes formes de plus-values retirées du travail de recherche

**Point d'attention 4 :** Faciliter la mise à disposition des données implique de prendre en compte, de manière différenciée, l'investissement de travail nécessaire à différentes étapes de la recherche, la plus-value créée en fonction des démarches de recherche et les répercussions en termes d'évaluation et de carrière.

10 Le deuxième Plan national de la science ouverte met ainsi en avant l'importance d'une science reproductible, transparente, plus efficace et cumulative. Les enjeux de reproductibilité des résultats scientifiques sont abordés notamment en lien avec l'ouverture des codes sources (troisième axe). L'accès aux codes sources et aux données étant des éléments essentiels à des démarches reproductibles.

11 Résultats de l'analyse statistique multivariée du questionnaire, corrélation entre le terme « transparence » et les autres variables du questionnaire cf. le script des résultats : <https://code.inno3.eu/ouvert/decliner-so/>

**Explications :** Lors de travaux de recherche, différentes étapes sont nécessaires à l'obtention de résultats partageables à la communauté de pairs. Ces étapes génèrent **une plus-value plus ou moins importante en fonction du temps consacré à leur réalisation ou encore au degré de reconnaissance attribué à ce travail** par la communauté. Différents types de plus-values ont été distingués puis corrélés à des critères en lien avec la démarche de recherche. Par exemple :

- **la collecte de données rares ou nécessitant un temps de collecte important** est principalement associée à **un travail de terrain ou sur des archives** et fonds documentaires en sciences humaines et sociales ;
- **la préparation d'échantillons** et la définition de protocoles expérimentaux sont des activités associées à des **travaux de recherche en laboratoire** ;
- un **cadre de recherche clinique** est plus fortement corrélé avec **une plus-value tirée de l'automatisation de processus de travail en chaîne (workflow)** et de la modélisation sur une grande quantité de données.

Porter attention à ces différentes démarches de recherche ainsi qu'aux formes de plus-values générées en fonction des contextes est important pour repérer des blocages quant à la mise à disposition de données. Certaines démarches de recherche (amélioration de technique, automatisation, modélisation) peuvent inciter à la mise à disposition tandis que d'autres démarches peuvent la réfréner (collecte de données rares, préparation d'échantillon chronophage).

## **Démarches de recherche : spécificité clinique/expérimentale/computationnelle**

**Point d'attention 5 :** Prendre en compte des éléments de friction propres au milieu de la recherche. Ce milieu est constitué de cultures qui cohabitent tout en étant parfois en opposition. D'une part, une culture technico-industrielle vise à l'amélioration des processus, de la gestion du risque, et de l'efficacité. D'autre part, il existe une culture d'indépendance et de liberté des chercheurs associée à la revendication d'une posture d'« artisanat » et de créativité dans la démarche scientifique.

**Explications :** Même lorsque des travaux de recherche s'intéressent aux mêmes objets d'étude, cela ne signifie pas que les démarches de recherche soient identiques. Par exemple, des études portant sur les virus à ARN (VIH, SARS, etc.) peuvent faire l'objet de recherches dans le cadre d'études cliniques d'une part et dans un cadre de recherche fondamentale à la paillasse ou computationnelle d'autre part. Dans le cas de la recherche clinique par exemple, une structuration des données (avec une réglementation stricte sur leurs usages) est prévue dès le départ (*Case Report Form*) et la plus-value du travail provient d'une automatisation et d'une amélioration des protocoles. Alors que dans le cadre de travaux en biologie fondamentale, la collecte de données rares obtenues à la suite d'expériences chronophages tout autant que leur

interprétation sont au cœur de la démarche scientifique avec la défense d'une posture d'artisanat et de création. Ces différentes démarches et plus-values se rencontrent parfois au sein de mêmes projets de recherche avec la critique d'une part d'une « ingénierisation » de la recherche et d'autre part d'un manque de contrôle de la qualité de la recherche d'autre part.

## 4.2 Appréhender différentes pratiques de mise à disposition des données

### Penser à la réutilisation de données et d'autres ressources et aux publics concernés

**Point d'attention 6 :** Mettre à disposition des données nécessite de penser le « public des données », la temporalité, les modalités de partage (juridique, financier, technique, etc.) et l'accompagnement nécessaire.

**Point d'attention 7 :** Outre les données, d'autres objets (protocoles, codes source, etc.) peuvent faire aussi l'objet d'une mise à disposition dans des conditions spécifiques à définir.

**Explications :** Concernant la réutilisation des données, l'objectif affirmé des politiques de science ouverte, il s'agit de prendre en considération les « publics des données » qui influencent la manière dont les données sont mises à disposition, mais également les modalités de partage et de contextualisation de ces informations. Un partage de données entre collègues (communauté de pairs) pour garantir une reproductibilité n'impliquera pas le même travail en matière d'explicitation et de contextualisation des données qu'une mise à disposition pour un public plus vaste dans l'objectif d'une diffusion plus large des résultats de recherche. Cela implique de réfléchir également à l'accompagnement nécessaire à leur partage (aide à la structuration de données, diffusion de vidéos de vulgarisation des travaux de recherche) tout comme les problématiques éthiques et juridiques associées. De plus, le partage des données n'est pas le seul élément à inclure dans une démarche de science ouverte. Ainsi pour le cas de la répliquabilité d'expérience, le partage du protocole est essentiel tout comme la mise à disposition des codes sources pour reproduire l'analyse des données employées.

Le terme de « mise à disposition » est employé dans l'étude pour distinguer différentes pratiques comprenant :

- le partage restreint à un public ciblé et connu (via mail par exemple) ;
- la mise en ligne des données sur un site/entrepôt avec un contrôle ou non de l'accès ;
- l'ouverture des données sur un site/entrepôt avec l'apposition d'une licence ouverte (*open data*).



## Distinguer différentes limites à la mise à disposition et les leviers d'amélioration

**Point d'attention 8 :** Différencier les raisons limitant la mise à disposition des données (temps trop important nécessaire, manque d'habitude, avantage compétitif à ne pas partager) pour y apporter des réponses adaptées.

**Point d'attention 9 :** Inciter les comités éditoriaux de revues à s'appuyer sur les politiques nationales existantes concernant les données et les codes associés aux publications scientifiques.

**Explications :** Les raisons principales limitant la mise à disposition des données sont essentiellement le manque d'habitude de ces pratiques, un temps trop important nécessaire à leur mise à disposition et une rétention des données pour garder un avantage compétitif. Parmi les raisons secondaires figurent des questionnements concernant les risques de bureaucratie supplémentaire engendrée par la mise à disposition ainsi que des enjeux juridiques et éthiques autour de l'accès aux données personnelles. Les obligations à la mise à disposition sont peu connues et proviennent surtout des comités éditoriaux de revues ou des comités éthiques. Sensibiliser les comités à mentionner les enjeux de mise à disposition est un élément clef pour la prise en compte de ces pratiques dans l'évaluation et la reconnaissance des travaux de recherche, car leur rôle dans ce processus est important<sup>12</sup>.

## Mettre en avant les enjeux de conservation et de sécurité des données

**Point d'attention 10 :** Sensibiliser à la distinction entre stockage et archivage des données qui impliquent des services et des infrastructures différentes ainsi qu'une sélection possible de données à opérer, notamment pour distinguer les données à conserver et à détruire.

**Point d'attention 11 :** Privilégier et/ou mettre en avant les fonctionnalités de sécurité et les éléments de fiabilité proposés par les infrastructures de recherche mises à disposition pour le stockage des données.

**Explications :** Concernant le stockage des données<sup>13</sup>, il est majoritairement réalisé sur des supports externes et des ordinateurs professionnels. Néanmoins, en sciences humaines et sociales, l'usage d'ordinateur personnel est fréquent notamment pour les

12 Par ailleurs, la loi pour une République numérique (2016) peut être rappelée au sein de ces comités avec l'article 30 qui empêche les éditeurs de limiter la diffusion et la réutilisation des données associées aux publications dès lors qu'elles sont issues d'une recherche financée pour moitié au moins sur fonds publics.

13 Par stockage des données, nous incluons les données manipulées par les chercheurs à l'étape de la collecte et de l'analyse (données en cours de traitement) ainsi que les données produites pour une mise à disposition (partage/ouverture). La notion de « cycle de vie des données » n'a pas été abordée par les répondant·e·s et les processus d'archivage des données (choix de données à détruire ou à archiver) n'étaient que peu ou pas connus.

doctorant·e·s, ce qui ne facilite pas le suivi des données, leur sécurité ou encore leur réutilisation à la fin d'un projet. Une attention et vigilance toute particulière est portée par les communautés sur la sécurité des données (données cryptées, risque de piratage, etc.) avec un questionnement sur la fiabilité des infrastructures institutionnelles. Les solutions de *cloud* de type Google Drive ou Dropbox sont majoritairement employées pour le partage de fichiers. Par ailleurs, à l'heure actuelle, la différence entre stockage et archivage reste floue pour les communautés de chercheur·e·s. L'appel aux services d'archivage restent rares, car le stockage des données peut sembler pour les professionnel·le·s de la recherche une action suffisante pour conserver les données.

## 4.3 Connaître les modalités d'apprentissage et les pratiques collaboratives

### Découverte et formation aux outils : un échange entre pairs

**Point d'attention 12 :** Pour faciliter l'appropriation de nouvelles pratiques, prendre en considération les spécificités de rencontres et d'apprentissages des communautés (vie de laboratoire, journées d'étude et de conférences, réseaux sociaux, etc.).

**Explications :** Outre la découverte d'outils en autodidacte, le rôle d'autres personnes au sein des équipes de recherche (membres de l'équipe ou autres équipes) est primordial pour constituer son environnement numérique de travail. Les habitudes se prennent souvent dès les premiers stages de recherche en master avec une formation au sein des équipes (maitre de stage, vie de « laboratoire » pour les travaux à « la paillasse », etc.). Au sein des sciences humaines et sociales, les journées d'étude et les temps informels jouent un rôle important pour découvrir de nouveaux outils, partager ses pratiques. Les réseaux sociaux représentent aussi des espaces d'échange et de découverte de pratiques, jugés utiles notamment lorsque se croisent différentes communautés.

#### **Journée d'étude : du terrain à la mise en données en SHS**

Dans le cadre de l'enquête (phase 2), une journée d'étude a été dédiée à l'étude des pratiques de « mise en données » en SHS et a permis de faire ressortir trois problématiques clefs :

- Problématiques communes aux pratiques de la mise en données ;
- Reconfiguration des collectifs de recherche ;
- Environnement et valorisation du travail de mise en données.

### Porter attention aux interfaces

**Point d'attention 13 :** Porter une attention toute particulière aux interfaces de traitement et d'analyse des données pour qu'elles ne deviennent pas des « boîtes noires » et des « culs-de-sac » (manque d'interopérabilité, formats propriétaires, etc.).

**Point d'attention 14 :** Être vigilant quant aux nouvelles solutions clefs en main qui se développent pour l'analyse et la manipulation de données.

**Point d'attention 15 :** Former aux pratiques computationnelles sans toutefois vouloir faire de tous les acteurs de la recherche des data scientists. Donner un bagage suffisant pour être capable d'échanger et de comprendre les enjeux.

**Explications :** Les interfaces graphiques sont essentielles dans les étapes de traitement et d'analyse de données. Elles aident à manipuler aisément les données, filtrer des informations, les classer, les visualiser rapidement de manière dynamique et faciliter une continuité entre différents acteurs avec des couches d'accès différenciés. Pour le travail d'équipe, des extractions de graphiques ou d'éléments d'une base de données facilitent les échanges et l'analyse collective. Or des outils clefs en main, aussi appelés « *click and play* », viennent invisibiliser l'ensemble des traitements algorithmiques sous-jacents. À la montée de l'offre de plateformes *data science* (outils statistiques de sciences des données orientées IA), une littératie aux principes computationnels est nécessaire dès les premiers cycles de l'enseignement supérieur pour cultiver un regard critique concernant ces plateformes clefs-en-main.

## 4.4 Diversifier les types d'accompagnement

### Distinguer différents besoins d'accompagnement

**Point d'attention 16 :** Outre les aides au stockage, à l'archivage et à la mise à disposition des données, proposer des aides quant à la description et à la médiation des données auprès de publics variés sous différents formats tout autant qu'aux enjeux juridiques et éthiques.

**Explications :** Les demandes d'aides formulées par les communautés de recherche consistent en premier lieu en des demandes de ressources humaines et financières : obtention ou renouvellement de postes fixes, aide financière pour l'accès à des bases de données, ou la numérisation. Concernant les données, les aides demandées portent sur le stockage en cours de traitement des données, la mise à disposition de données visant à être ouvertes ou partagées et l'archivage. Pour les sciences humaines et sociales, le soutien à la diffusion pour le grand public sous forme de vidéo ou billets de blog (médiation scientifique) constitue un enjeu important et qui n'est souvent pas pris en charge dans les budgets des projets de recherche.

## Élaborer un maillage des dispositifs d'accompagnement aux données au plus près des équipes

**Point d'attention 17 :** Faciliter un « maillage des dispositifs d'accompagnement aux données » à différentes échelles en diversifiant leurs modalités par 1. le développement et le maintien d'infrastructures, 2. l'acculturation au sein des équipes de recherche par le biais de personnes soutien déjà présentes dans le quotidien des équipes pour jouer un rôle de médiation, comprendre les besoins et la culture du laboratoire ou de l'équipe.

**Point d'attention 18 :** Faire attention à l'ajout de fonctions supplémentaires « référent données » s'ajoutant aux charges de travail de personnes déjà en poste au détriment de la création de postes stables et pérennes dédiés aux missions de mise à disposition des données.

**Explications :** En plus de la mise en œuvre de guichets uniques et d'infrastructures nationales pour accompagner les pratiques associées aux données, un maillage au plus près des équipes est plébiscité par les personnes interrogées dans le cadre de l'étude. Des relais stables et pérennes au sein des équipes sont demandés avec une méfiance néanmoins quant à la surcharge de travail qu'engendrerait l'ajout d'une nouvelle fonction « référent données » aux personnes déjà en poste, notamment aux ingénieurs de recherche ou d'étude (IR/IE).

### Être vigilant aux enjeux de médiation au sein des équipes de recherche

**Point d'attention 19 :** Porter attention aux enjeux de traduction et de médiation nécessaires à la gestion et à la mise à disposition des données par le biais de pratiques collectives au sein des équipes de recherche. Ceci implique de trouver des « dénominateurs communs » parmi les outils et la documentation employés, les protocoles et les processus de standardisation des données.

**Explications :** Pour beaucoup, l'adaptation à de nouvelles pratiques de traitement, d'analyse, et de partage des données s'accompagne de processus et d'environnements de travail nouveaux et/ou complémentaires à s'approprier. Cela vient également reconfigurer les modalités de travail entre différents membres d'équipe (services informatiques, ingénieurs, chercheurs, etc.) avec un ensemble de frictions possibles. La constitution de bases de données entre différents profils disciplinaires ou métiers tout comme leur mise à disposition dans des entrepôts de données (partage ou ouverture) cristallisent des tensions (constitution de vocabulaires, réduction de la complexité d'une étude, reconnaissance des personnes ayant participé à la création de la base, etc.). Néanmoins, ce sont également un moyen de construire de nouvelles pratiques adaptées

aux compétences de chacun·e. Le dialogue et la compréhension à construire entre différents profils (traduction de vocabulaire spécifique, facilitation des échanges par des processus de médiation, etc.) nécessitent du temps et parfois des aides financières, matérielles, ou encore organisationnelles pour faciliter leur développement.

## 4.5 Prendre en considération le statut et les enjeux de carrière

**Point d'attention 20 :** Mieux considérer dans l'évolution des carrières et l'évaluation des professionnels de la recherche le travail de « mise en données » et de mise à disposition des données.

**Explications :** Les travaux de « mise en données » et de mise à disposition des données nécessitent souvent du temps, par exemple la collecte de données rares, la mise en forme de données, l'ajout de documentation, l'ajout de métadonnées, la mise en ligne sur des entrepôts. Il est important de reconnaître le temps passé à ces activités dans l'évolution des carrières, notamment dans le cas de personnes aux statut et fonction pouvant mener au travail solitaire, contexte dans lequel ces tâches sont encore plus invisibilisées. En effet, si des chercheur·e·s préfèrent travailler seul·e·s et ne pas changer leurs pratiques par choix ou par positionnement politique, d'autres ont une démarche solitaire et de « non-partage » imposée. C'est le cas par exemple des doctorant·e·s intéressé·e·s par les sujets de science ouverte, mais pour qui les activités de mise en données ne sont pas une priorité, ni pour leur·s encadrant·e·s. Pour les post-doctorant·e·s, de la même manière, la recherche d'un poste apparaît souvent prioritaire par rapport au développement de ces pratiques, même si certain·e·s peuvent en tirer parti pour élaborer une stratégie de visibilité et de réseautage.

## 5 | Limites

Une première analyse des résultats du questionnaire a montré une surreprésentation des communautés de recherche en Sciences Humaines et Sociales. En réponse à cela, une pondération des résultats en fonction de la répartition actuelle des chercheur·e·s dans différentes catégories disciplinaires a été effectuée<sup>14</sup>. Les modalités de diffusion du questionnaire expliquent certainement cette surreprésentation. Le questionnaire a été partagé sur des listes de discussion et des réseaux sociaux suivis par les membres du groupe de travail Décliner-SO. Plusieurs listes étaient associées aux sciences humaines et sociales (histoire, sociologie, économie, etc.) et l'annonce a plus largement circulé dans ces communautés. Au vu des résultats, le questionnaire gagnerait à être partagé plus largement au sein d'établissements pour affiner les résultats concernant les disciplines actuellement sous-représentées et confirmer la pertinence des facteurs différenciant les pratiques mises en évidence.

## 6 | Conclusion

Cette étude visait à étudier les pratiques actuelles associées aux données dans diverses communautés de recherche et accompagner au mieux leurs évolutions dans un contexte numérique et de politique publique favorable à la science ouverte. L'objectif a été de présenter avec une granularité fine des éléments expliquant la diversité de pratiques de recherche au sein de ce que l'on dénomme la « Science » afin de mieux décliner et adapter les mesures de science ouverte en fonction de communautés épistémiques ou de pratiques. Plus qu'un simple regard disciplinaire, la typologie de pratiques mise en évidence – et leur illustration par des personæ (profil type)– montre l'importance de considérer la nature solitaire ou collaborative du travail qui s'inscrit dans des tissus sociaux, méthodologiques, et techniques divers.

Une meilleure appropriation de nouvelles pratiques associées aux données par les communautés passe par une appréhension en finesse de différentes démarches de recherche, tout comme un regard sur les outils et dispositifs employés et leurs modalités d'apprentissage et de découverte. Par les facteurs différenciants définis, les pistes d'orientations et les Point d'attentions proposées, cette étude souhaite aider les personnes impliquées dans des politiques et projets de science ouverte à mieux dialoguer avec les professionnel·le·s de recherche qu'elles sont amenées à accompagner tout autant qu'à diversifier les types d'aides proposées.

Pour les personnes concernées par ses pratiques et soumises à leurs évolutions, l'étude souhaite participer à une prise de recul et de réflexivité. Il s'agit de mieux comprendre

---

14 Nous avons pris comme référentiel les données 2019-2020 de l'état de l'Enseignement supérieur, de la Recherche et de l'Innovation en France – n°14 – Avril 2021 [https://publication.enseignementsup-recherche.gouv.fr/eesr/FR/T579/les\\_personnels\\_enseignants\\_de\\_l\\_enseignement\\_superieur\\_public\\_sous\\_tutelle\\_du\\_mesri/](https://publication.enseignementsup-recherche.gouv.fr/eesr/FR/T579/les_personnels_enseignants_de_l_enseignement_superieur_public_sous_tutelle_du_mesri/)

« nos pratiques » et/ou avoir un cadre d'explication sur les pratiques d'autres collègues. Loin de vouloir décider ou de juger la qualité des normes à appliquer au sein d'équipes ou de collectifs de recherche, il est plutôt question dans cette étude de donner des pistes pour adapter les modalités d'interaction entre professionnel·le·s de recherche, comprendre les raisons de frictions ou de blocages aux mesures de science ouverte et leurs incitations, ainsi que de mettre à disposition des éléments d'argumentation et de débat pour que ces changements de pratiques soient un acte éclairé et souhaité.

## 7 | Annexes

### 7.1 Résultats principaux du questionnaire « données et science ouverte »

429 réponses ont été obtenues au questionnaire « données et science ouverte » donnant un aperçu des pratiques actuelles des professionnel·le·s de la recherche<sup>15</sup>.

#### Population

- Genre : 47,8 % et 44,8 % (autre : 7,4 %).
- Fonction principale : essentiellement des enseignant·e·s-chercheur·e·s « permanent » (55 %) et « non permanent » (20 %).
- Ancienneté : Majoritairement plus de 10 ans (65,5 %) (voir graphique ci-dessous) avec un statut de fonctionnaire (68,7 %).
- Pondération des résultats effectuée en fonction des catégories disciplinaires pour être représentatif des données 2019-2020 (état de l'enseignement supérieur, de la Recherche et de l'Innovation en France n°14)<sup>16</sup>.

#### Depuis combien de temps travaillez-vous dans la recherche ou l'enseignement supérieur?

Réponse à choix multiples.

■ Cinq à six ans (%) ■ Moins de cinq ans (%) ■ Plus de dix ans(%)



Graphique: Groupe de travail "Réussir l'appropriation de la science ouverte" • Source: Questionnaire "données et science ouverte" • Créé avec Datawrapper

#### Pratiques de « mise à disposition » de données et obligations

- Pour les obligations de mise à disposition, celles-ci sont peu connues et concernent surtout des obligations de la part d'un comité éditorial ou éthique (par exemple, en recherche biomédicale).

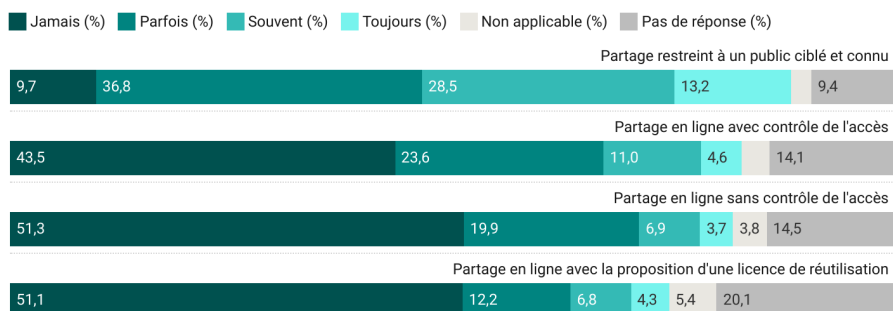
15 Les résultats présentés sont ceux après pondération des disciplines afin d'arriver à un échantillon représentatif de la répartition disciplinaire dans l'ESR sur la base des Données 2019-2020 ESR – L'état de l'Enseignement supérieur, de la Recherche et de l'Innovation en France (n°14 – Avril 2021).

16 SIES, Sous-direction des systèmes d'information et des études statistiques, eds. 08. L'État de l'Enseignement supérieur, de la Recherche et de l'Innovation en France 2021. Paris : SIES.

- Le partage restreint pour un public ciblé et connu reste la pratique majoritaire (78,5 %) (voir graphique ci-dessous).
- Le partage en ligne avec la proposition d'une licence ouverte (open data) représente un peu plus de 20 % des pratiques (voir graphique ci-dessous).

### Dans quelle mesure mettez-vous à disposition actuellement des données de recherche ?

Réponse à choix multiples.



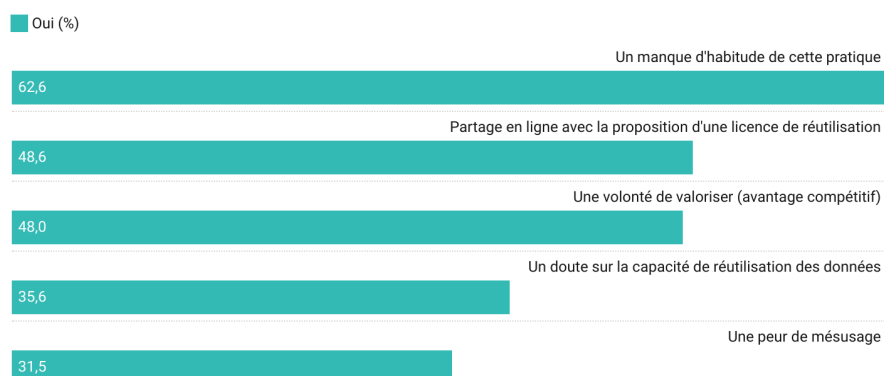
Graphique: Mission "réussir l'appropriation de la science ouverte" • Source: Questionnaire "données et science ouverte" • Créé avec Datawrapper

### Réutilisation des données et limites pour la mise à disposition

- Près de 50 % des personnes interrogées indiquent réutiliser souvent et/ou parfois des données déjà produites ou publiées.
- Plus de 45 % de personnes interrogées considèrent que leurs données seraient potentiellement réutilisables.
- Les raisons principales limitant la mise à disposition des données sont essentiellement (voir graphique ci-dessous) :
  - le manque d'habitude de ces pratiques (63 %) ;
  - un temps trop important nécessaire (49 %) ;
  - une volonté de valoriser les données pour garder un avantage compétitif (48 %).

### Selon vous, quelles sont les raisons principales qui limitent la mise à disposition des données?

Réponse à choix multiples.



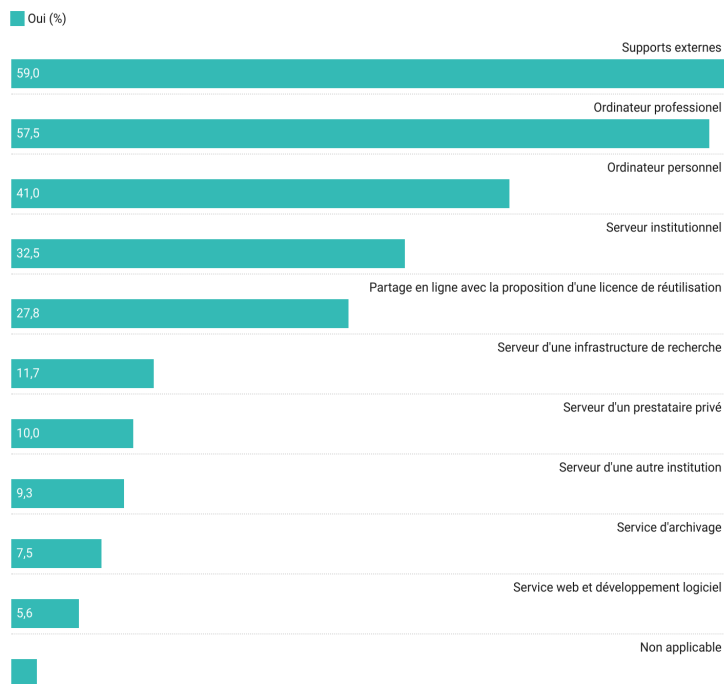


## Stockage des données

- Le stockage des données est majoritairement réalisé sur des supports externes (59 %) et des ordinateurs professionnels (57,5 %).
- Il y a peu d'appels à des services d'archives (7,5 %) (voir graphique ci-dessous).

### Actuellement, comment conservez-vous vos données à la fin d'un projet?

Réponse à choix multiples.

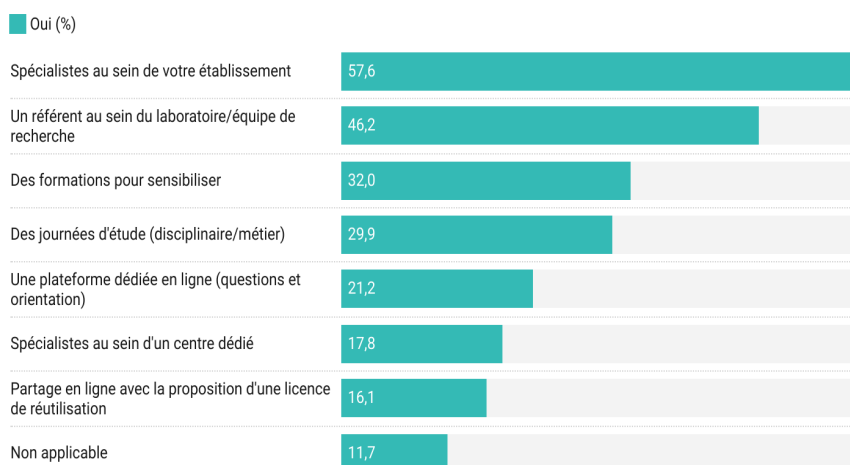


## Besoins et accompagnement

- Les accompagnements nécessaires associés aux données concernent en premier lieu les étapes de stockage et de conservation des données (48 %), puis leur mise à disposition (40 %). La majorité des personnes interrogées a connaissance d'aides disponibles (plus de 55 %).

### Par qui souhaiteriez-vous obtenir de l'aide sur ces différentes étapes?

Réponse à choix multiples.



Graphique: Groupe de travail "Réussir l'appropriation de la science ouverte" • Source: Questionnaire "données et science ouverte" • Créé avec Datawrapper

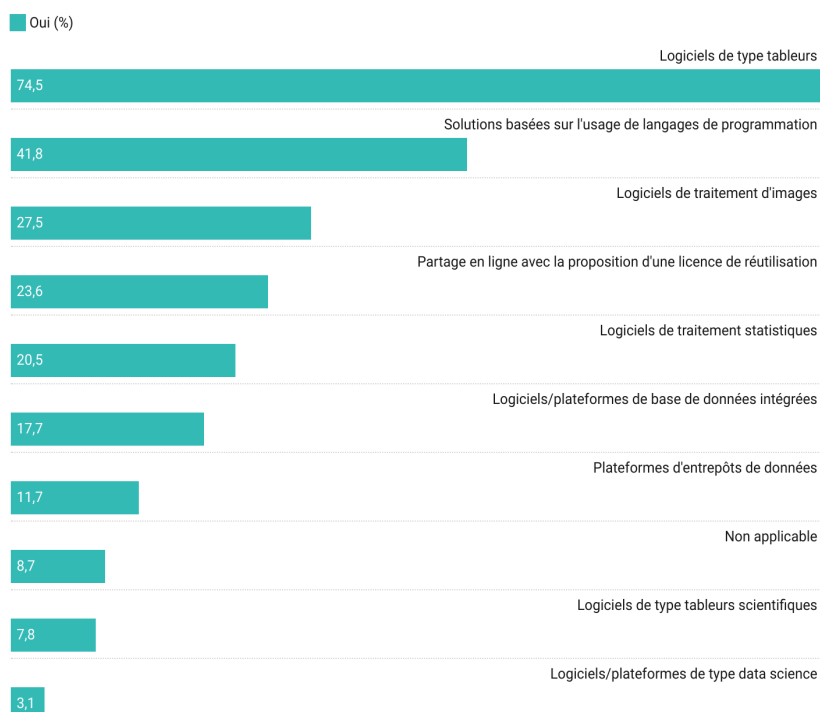
- Les personnes ayant bénéficié d'aides sont généralement satisfaits (plus de 60 %).
- La présence de spécialistes au sein d'établissement (58 %) et de référents au sein de laboratoires/équipes de recherche (46 %) sont les modalités d'accompagnement les plus plébiscitées (voir graphique ci-dessous).

## Outils employés associés aux données

- Utilisation majoritaire de logiciels de type tableurs (Excel, Calc) (74,5 %).
- Plus de 40 % utilisent des solutions basées sur l'usage de langage de programmation (R, Python).
- QGIS est un des logiciels d'analyse et de visualisation de données (24 %) fréquemment cité.
- Les logiciels/plateformes de base de données intégrées (18 %) souvent cités sont FileMaker, PostgreSQL, MySQL.
- Les plateformes d'entrepôts de données n'ont été utilisées que par 12 % des répondant·e·s.
- Le système d'exploitation le plus largement employé est Windows (62 %) contre 26 % pour MacOS et 12 % pour Linux et autres Unix.

### Dans les 12 derniers mois, quels sont les outils que vous utilisez régulièrement pour traiter, structurer, analyser, visualiser, partager, etc. des données?

Réponse à choix multiples.



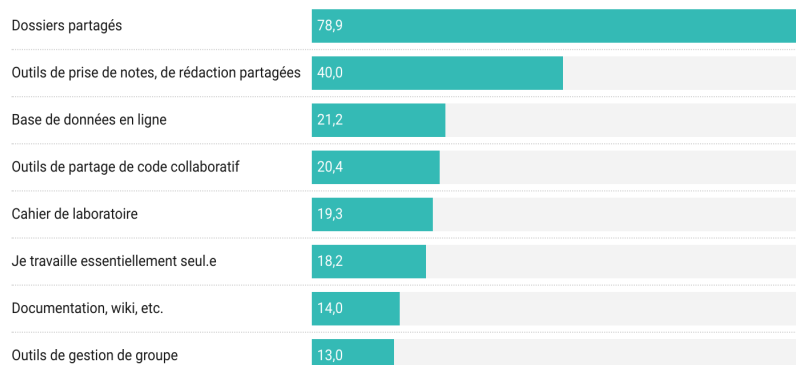
## Pratiques collaboratives

- Des outils de prises de notes partagées sont employés par 40 % des personnes ayant répondu au questionnaire.
- L'utilisation d'outils non institutionnels est fréquente (DropBox, GDrive, etc.).

### Quels types d'outils collaboratifs utilisez-vous lorsque vous travaillez avec d'autres personnes?

Réponse à choix multiples.

■ Oui (%)



## 7.2 Accès aux ressources produites

- Rapport final et Annexes (trames d'entretiens, trame du questionnaire, plan de gestion de données - PGD) : sur [HAL](#) (DOI : 10.52949/27)
- Synthèse en français : sur [HAL](#) (DOI : 10.52949/28)
- Synthèse en anglais : sur [HAL](#) (DOI : 10.52949/29)
- Données « brutes » du questionnaire sur [Recherche.data.gouv](#) (DOI 10.57745/V64RYT)
- Script de reproductibilité de l'analyse du questionnaire et note méthodologique : sur [Gitlab](#)
- Éditorialisation de l'ensemble du contenu : sur [PubPub](#)

## 7.3 Crédits

L'étude a été menée par un groupe de travail interdisciplinaire dans lequel étaient représentées différentes professions de l'enseignement supérieur et de la recherche, ce qui a permis de réunir des expertises variées et des compétences diverses (analyse quantitative, qualitative, retours d'expérience variés, etc.) Ont pris part au projet :

- Anne Vanet : Vice-Présidente Numérique et science ouverte (Université de Paris), Directrice du pôle de génoinformatique Institut Jacques Monod
- Hélène Chambefort : Responsable des Archives (INSERM)
- Marie Herbert : Chargée de mission Collex/Persée (Université Lyon 1)

- Juliette Hueber : Responsable éditoriale et de l'ingénierie documentaire (InVisu CNRS/INHA)
- Claire Lemerrier : Directrice de recherche CNRS au CSO SciencePo

L'équipe a été accompagnée par le cabinet de conseil inno<sup>3</sup>.

- Célya Gruson-Daniel : Chercheure associée au COSTECH (UTC), consultante a pris en charge la gestion de projets, le design de l'étude, la collecte et l'analyse des informations qualitatives ainsi que la rédaction des contenus (présentation, résumé).
- Benjamin Jean : Juriste, président Inno<sup>3</sup>.
- Romain Rouyer a apporté ses compétences de design pour les visualisations et graphiques.
- La mise en page et la réalisation des graphiques ont été effectuées par Tamara Glushetckaia.
- Le travail de reproductibilité de l'analyse du questionnaire a été réalisé par la suite par Émilien Schultz (Medialab SciencePo).

La synthèse (texte et graphiques) est publiée sous [licence Creative Commons CC-BY 4.0](#).  
Les photos des personae sont sous licence [Unsplash](#).