



HAL
open science

Décliner la science ouverte

Célyla Gruson-Daniel, Groupe Projet Réussir L'Appropriation De La Science Ouverte

► **To cite this version:**

Célyla Gruson-Daniel, Groupe Projet Réussir L'Appropriation De La Science Ouverte. Décliner la science ouverte. [Rapport de recherche] Comité pour la science ouverte. 2022, 149 p. hal-03798504

HAL Id: hal-03798504

<https://hal-lara.archives-ouvertes.fr/hal-03798504v1>

Submitted on 5 Oct 2022

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Décliner la science ouverte

Comité pour la science ouverte -
Collège des données de la recherche

Groupe de travail « Réussir
l'appropriation de la science ouverte »

Rapport final



soutenu par



Décliner la science ouverte

Réalisé par Inno³

Pour le Collège « Données de la recherche »
du Comité pour la Science Ouverte

Dans le cadre du groupe de travail « Réussir
l'appropriation de la science ouverte »

2022

Document sous licence CC BY 4.0

1. Préambule	3
• Contexte de la mission	4
• Problématiques et méthodologie	6
2. Appréhender la diversité des pratiques : facteurs différenciants	11
• Typologie de pratiques et personae	13
• Démarches, outils et statut	28
3. Accompagner l'évolution des pratiques : pistes d'orientation	35
• Comprendre en finesse les démarches de recherche	37
• Appréhender différentes pratiques de mise à disposition des données	47
• Connaître les modalités d'apprentissage et les pratiques collaboratives	54
• Diversifier les types d'accompagnement	63
• Prendre en considération les enjeux de carrière	72
4. Conclusion	75
5. Annexes	78
• comprenant notamment les trames d'entretiens, du questionnaire, le plan de gestion de données, etc.	78

Décliner la science ouverte

Rapport final



Préambule

1

PRÉAMBULE

Contexte de la mission



Mission « Réussir l'appropriation de la science ouverte »

- Réalisée dans le cadre du collège « données de la recherche » du Comité pour la science ouverte
- De Mai 2020 à Décembre 2021
- Menée par un groupe de travail multi-disciplinaire/professionnel
- Composée de 3 axes de travail :

1) Ateliers Juridiques Science Ouverte (AJSO) médiation

- 3 ateliers 2020/2021 regroupant plus de 150 personnes
- Une présentation de [synthèse](#)
- Un [kit d'appropriation des ateliers](#)

2) Groupe de travail cahiers de laboratoire électroniques expertise

- Participation au groupe de travail
- Rédaction et relecture du [rapport](#)

3) Décliner la science ouverte (Décliner-SO) recherche

Ce rapport concerne l'axe Décliner-SO

2

PRÉAMBULE

Problématiques et méthodologie

Problématique

#1

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

Par **pratiques associées aux données** nous entendons :

- L'ensemble **des étapes nécessaires** à la **production de données** et à **leur mise à disposition** (du partage restreint à l'*open data*)
- L'ensemble des **outils employés** dans un environnement de plus en plus numérique
- Les modalités d'apprentissages de ces pratiques et les besoins associés

Problématique

#2

Comment **accompagner l'évolution des pratiques associées aux données** en lien avec les incitations / obligations portées par les **politiques publiques « science ouverte »** ?

Concernant les **évolutions** de pratiques **en lien avec les incitations « science ouverte »**, 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**
- L'incitation à une plus grande **reproductibilité** des travaux de recherche
- La mise en place **d'accompagnements (formation, infrastructures)** pour faciliter ces évolutions

Deuxième axe

Structurer, partager et ouvrir les données de la recherche

12

Mesures

14

- 4 Mettre en œuvre l'obligation de diffusion des données de recherche financées sur fonds publics
- 5 Créer Recherche Data Gov, la plateforme nationale fédérée des données de la recherche
- 6 Promouvoir l'adoption d'une politique de données sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, interopérables et réutilisables (FAIR)

[Plan National pour la science ouverte 2 \(2021-2024\)](#)

Méthodologie de l'étude

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

Phase 1 (QUALI) Mise en œuvre et enquête exploratoire

- Recherche bibliographique
- Plan de Gestion de Données
- 6 entretiens exploratoires (« virus à ARN » et « carnets de terrain »)
- Analyse par théorisation ancrée (codage ouvert, axial)
- Discussion résultats avec le Groupe de Travail (GT)

Phase 2 (QUALI) Observation des pratiques de « mise en données »

- 3 entretiens complémentaires : observation des pratiques et outils (« virus à ARN »)
- Journée d'études « du terrain à la mise en données en SHS »
- *Focus group* avec le GT
- Analyse par théorisation ancrée (codage sélectif)

Phase 3 (QUANTI) Questionnaire « données et science ouverte »

- Construction du questionnaire sur la base du quali
- Recueil des réponses (mai-juin 2021)
- Analyse des 429 réponses

Phase 4 (QUALI-QUANTI) Finalisation

- Croisement des résultats quali / quanti
- Design des visualisations
- Rédaction du rapport

Mai 2020

Décembre

Avril 2021

Août

Décembre 2021

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

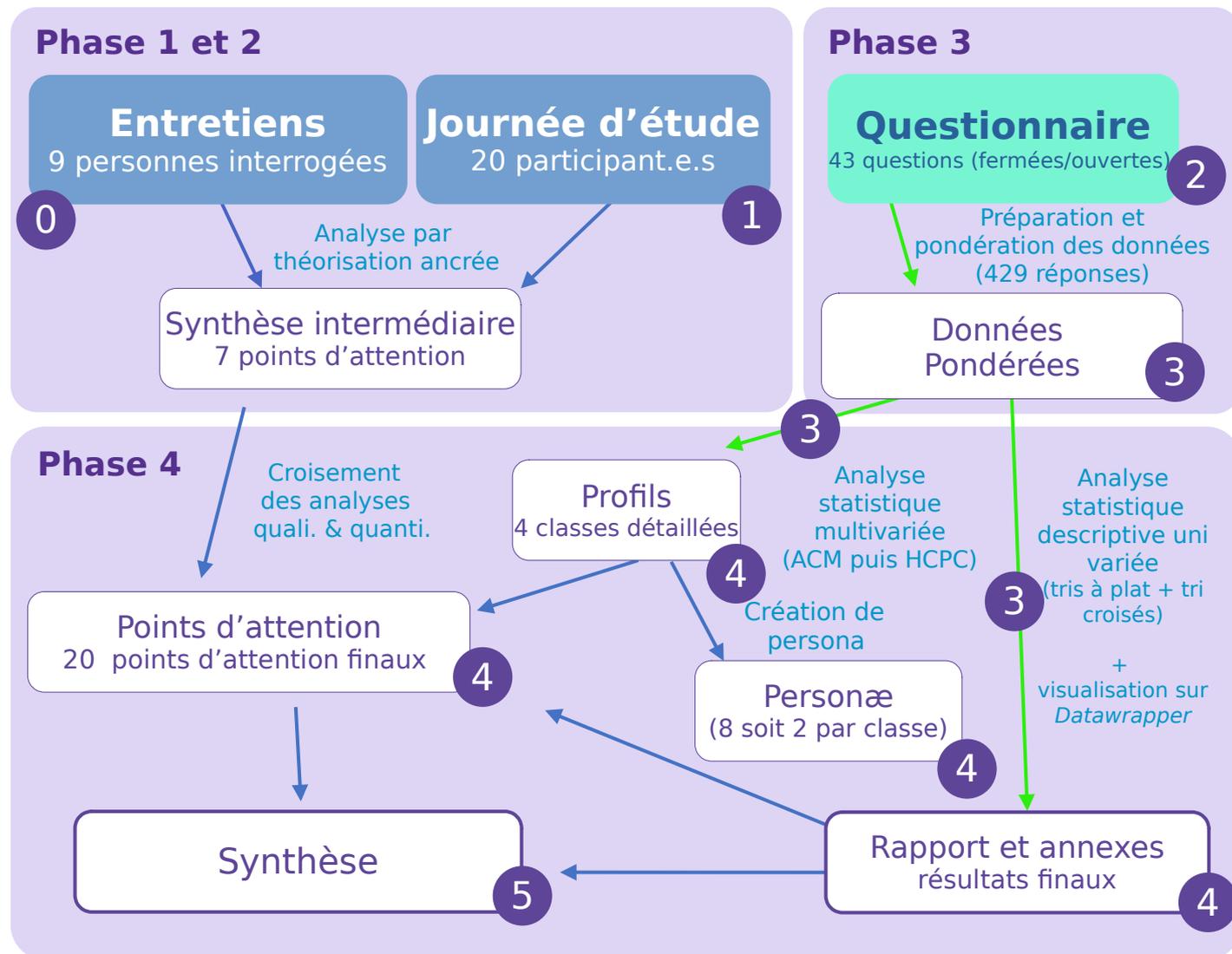
Livrables

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



Légende

-  Informations qualitatives
-  Données quantitatives
-  Méthodes
-  Résultats
-  Ressources



Décliner la science ouverte

Rapport final

2

Appréhender la
diversité des
pratiques :
facteurs
différenciants

Problématique #1

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

Par **pratiques associées aux données** nous entendons :

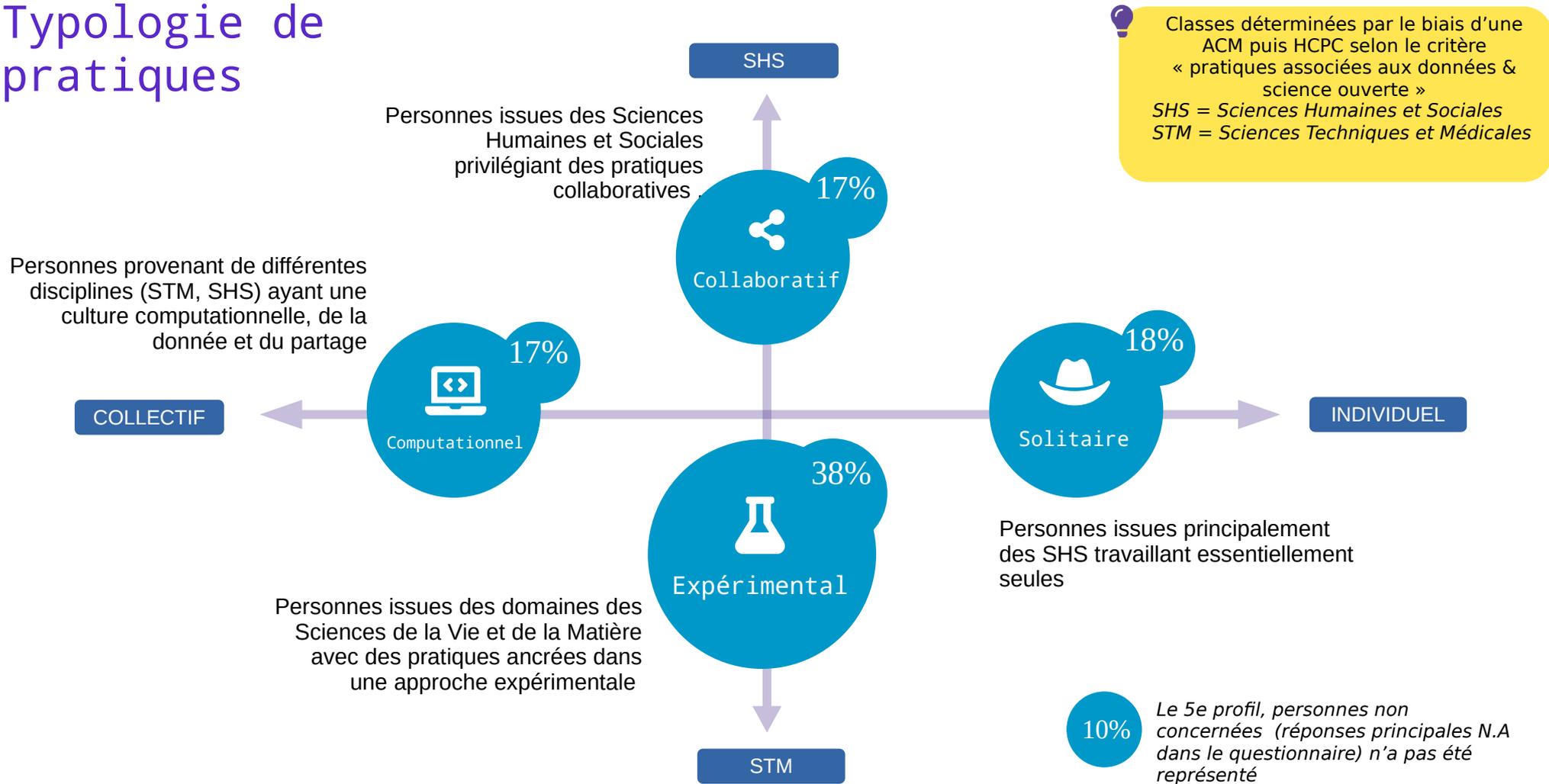
- L'ensemble **des étapes nécessaires** à la **production de données** et à **leur mise à disposition** (du partage restreint à *l'open data*)
- L'ensemble des **outils employés** dans un environnement de plus en plus numérique
- Les modalités d'apprentissages de ces pratiques et les besoins associés

1

FACTEURS DIFFÉRENCIANTS

Typologie de
pratiques et
personæ

Typologie de pratiques



Points de vigilance

Il est important de distinguer la méthodologie de construction des **profils** (typologie de pratiques) et des **personnæ** (personnages fictifs) :

4

Profils

Les profils se basent sur les résultats d'une ACM (Analyse des Correspondances Multiples) qui permettent de distinguer différentes classes.

Ces dernières indiquent des sur-représentations de modalités de variables du questionnaire pour une classe donnée.

Cela ne signifie pas que ces modalités sont majoritaires dans la classe.

8

Personnæ

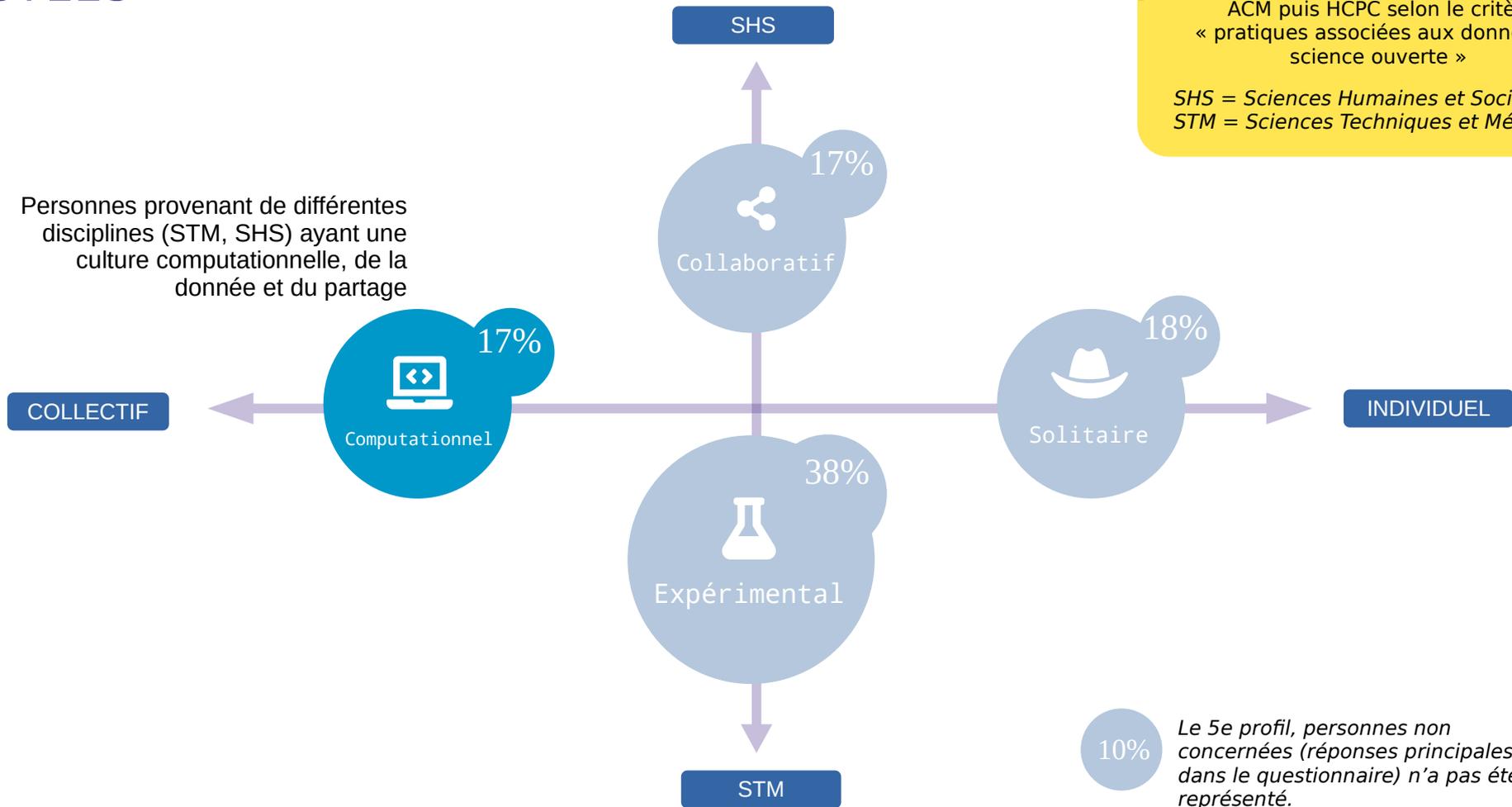
Les personæ sont des personnages fictifs, constitués à partir de l'exploration des réponses au questionnaire (les plus représentatives de chaque classe) en se nourrissant des résultats des phases d'analyse qualitative (journée d'étude, entretiens exploratoire et de pratiques) dans le but de mettre en contexte les profils et faire **ressortir plus spécifiquement des points saillants.**

Profils

Classes déterminées par le biais d'une ACM puis HCPC selon le critère « pratiques associées aux données & science ouverte »

SHS = Sciences Humaines et Sociales
STM = Sciences Techniques et Médicales

Personnes provenant de différentes disciplines (STM, SHS) ayant une culture computationnelle, de la donnée et du partage



10% Le 5e profil, personnes non concernées (réponses principales N.A dans le questionnaire) n'a pas été représenté.

Paula Letot

Domaine Linguistique informatique
Lieu Maîtresse de conférences à l'Université de Strasbourg
Expérience Depuis plus de 10 ans



Ordinateur analyse

Terminologie

Corpus de texte, ressources lexicales, données brutes et nettoyées, dictionnaires, jeux et bases de données

Explicabilité, recherche ouverte, reproductibilité

En bref

ID 1868530280

Familiarité avec la reproductibilité



Plus-value

Automatisation d'un *workflow*.
Constitution de données enrichies, Création d'une théorisation.

Story

Paula travaille en tant que maîtresse de conférences dans un laboratoire de sciences du langage où elle s'est spécialisée en traitement automatique des langues. Elle collabore régulièrement avec son équipe (5 personnes). Ses travaux consistent à constituer des corpus de textes augmentés d'informations syntaxiques et sémantiques pour aider à l'enrichissement d'analyses linguistiques. Les données collectées sont variées et proviennent notamment d'entretiens. Elle a l'habitude d'ouvrir ses données (avec l'attribution d'une licence). Les enjeux de reproductibilité sont connus au sein de son équipe. En plus des données, elle partage également le code source ayant servi aux analyses et prend soin de documenter les traitements effectués. Aujourd'hui, elle aurait besoin au sein de son équipe d'un référent « données » pour faciliter leur mise à disposition et traitement.

Mise à disposition des données

Conservation des données : sur le serveur de son laboratoire, son ordinateur professionnel et le service d'archivage de son institut.

Mise à disposition et réutilisation : partage souvent des données à l'aide de licences et parfois pour un public ciblé restreint. Partage sur un site web ou en complément des articles ainsi que sur des plateformes telles que Gitlab/Github. Réutilise fréquemment des données produites par d'autres personnes.

Autres ressources à partager : documentation des données et des codes sources, protocoles méthodologiques.

Limitations : des conditions éthiques qui limitent le partage, une mauvaise qualité des données, un manque d'habitude.

Obligations de partage : de la part d'un comité éditorial pour la publication d'un article et par les financeurs.

Pratiques, mise en données

Étapes de travail :

31-50 % analyser des données.

11-20 % nettoyer, modéliser, visualiser, décrire des données, former d'autres personnes.

Sources : bases de données partagées par des collègues ou en libre accès.

Ressources supplémentaires produites : documentation des données, normes, métadonnées, script et logiciels d'analyse, notes diverses, fichiers de données nécessitant un stockage central.

Environnement numérique

Outils utilisés : langage de programmation, tableurs génériques sous Windows.

Collaboration : travail collectif entre 2 et 5 personnes. Partage de documentation, d'outils de rédaction collectifs et de dossiers partagés.

Contexte d'initiation : par de la veille, des tutoriels, des temps informels et des colloques.

Accompagnement

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

Fréquence du soutien : ★★☆☆

Niveau de satisfaction : pas d'avis.



Besoins

Techniques et infrastructures numériques (aide au traitement des données).

Un référent au sein du laboratoire.

Alexandre Balankyen

Domaine Neurosciences cognitives
Lieu Post-doctorant (Centre de Recherche Clinique au GHU Paris)
Expérience Depuis moins de 5 ans



Cadre clinique
 Ordinateur analyse
 Ordinateur/papier réflexion

Terminologie

Sources, mesures comportementales, images, son, données expérimentales, pre-processing

Recherche ouverte, traçabilité

En bref

ID 2082305288

Familiarité avec la reproductibilité



Plus-value

Collecte de données nécessitant un temps important. Constitution d'une communauté de chercheurs et de contenu spécialisé et grand public. Supervision d'étudiant.e.s.

Story

Alexandre est post-doctorant dans un laboratoire de neurosciences cognitives au sein d'un centre hospitalier universitaire. Une grande partie de son travail consiste en l'acquisition de données expérimentales à la fois comportementales et de neuroimagerie (électrophysiologiques) pour des recherches sur le sommeil. Il travaille ensuite à traiter ces données (codage des expériences, traitement du signal, *pre-processing*) afin de les analyser statistiquement (tests bayésiens) et de les modéliser puis de réaliser les visualisations pour les figures d'articles scientifiques. Dans son équipe, Alexandre participe à la constitution d'un réseau de laboratoires interdisciplinaire afin de faciliter le recrutement de participant.e.s volontaires pour les expériences de recherche. Il s'intéresse aux enjeux des méta-analyses pour vérifier la qualité des travaux de recherche. Avec les étudiant.e.s de master, qu'il encadre, il s'applique à transmettre les bonnes pratiques de recherche et de la démarche scientifique. Les réunions de laboratoire sont un moment majeur pour le partage d'informations (méthodologie, outils, contenus, biblio, etc.).

Mise à disposition des données

Conservation des données : sur des supports externes, un ordinateur professionnel, un serveur institutionnel et interne du laboratoire et un service d'archivage.

Mise à disposition et réutilisation : partage toujours sans contrôle d'accès ou avec licences (open data). Met à disposition sur Open Science Framework, sur Google Drive, sur une plateforme partagée de codes (Gitlab/Github). Ne réutilise pas des données produites par d'autres mais pense que c'est possible.

Autres ressources à partager : logiciels et scripts d'analyse, métadonnées, protocoles méthodologiques.

Limitations : une volonté de valoriser (avantage compétitif), un temps nécessaire trop important, une mauvaise qualité des données.

Obligations de partage : de la part du financeur (ERC).

Pratiques, mise en données

Étapes de travail :

21-30 % analyser, modéliser, visualiser des données, former d'autres personnes.

11-20 % collecter, structurer et visualiser des données.

Sources : neuroimagerie (EEG, ECG), personne volontaire pour les expériences, articles scientifiques, Twitter.

Ressources supplémentaires produites : fichiers stockés sur ordinateur et centre de calcul, normes et métadonnées, scripts de programmation, logiciels d'analyse, notes diverses papier et numériques.

Environnement numérique

Outils utilisés : R, Python, Matlab, Excel, Inkscape.

Collaboration : travail collectif entre 2 et 5 personnes avec des enseignant.e.s-chercheur.e.s, des doctorant.e.s et étudiant.e.s en master. Utilise un cahier de laboratoire, des dossiers partagés, des plateformes de codes, des outils de gestion de projet. Échange lors des réunions de laboratoire.

Contexte d'initiation : par les autres membres de l'équipe, par des tutoriels (autodidacte).

Accompagnement

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

Fréquence du soutien : ★★☆☆

Niveau de satisfaction : ♥♥♥♥

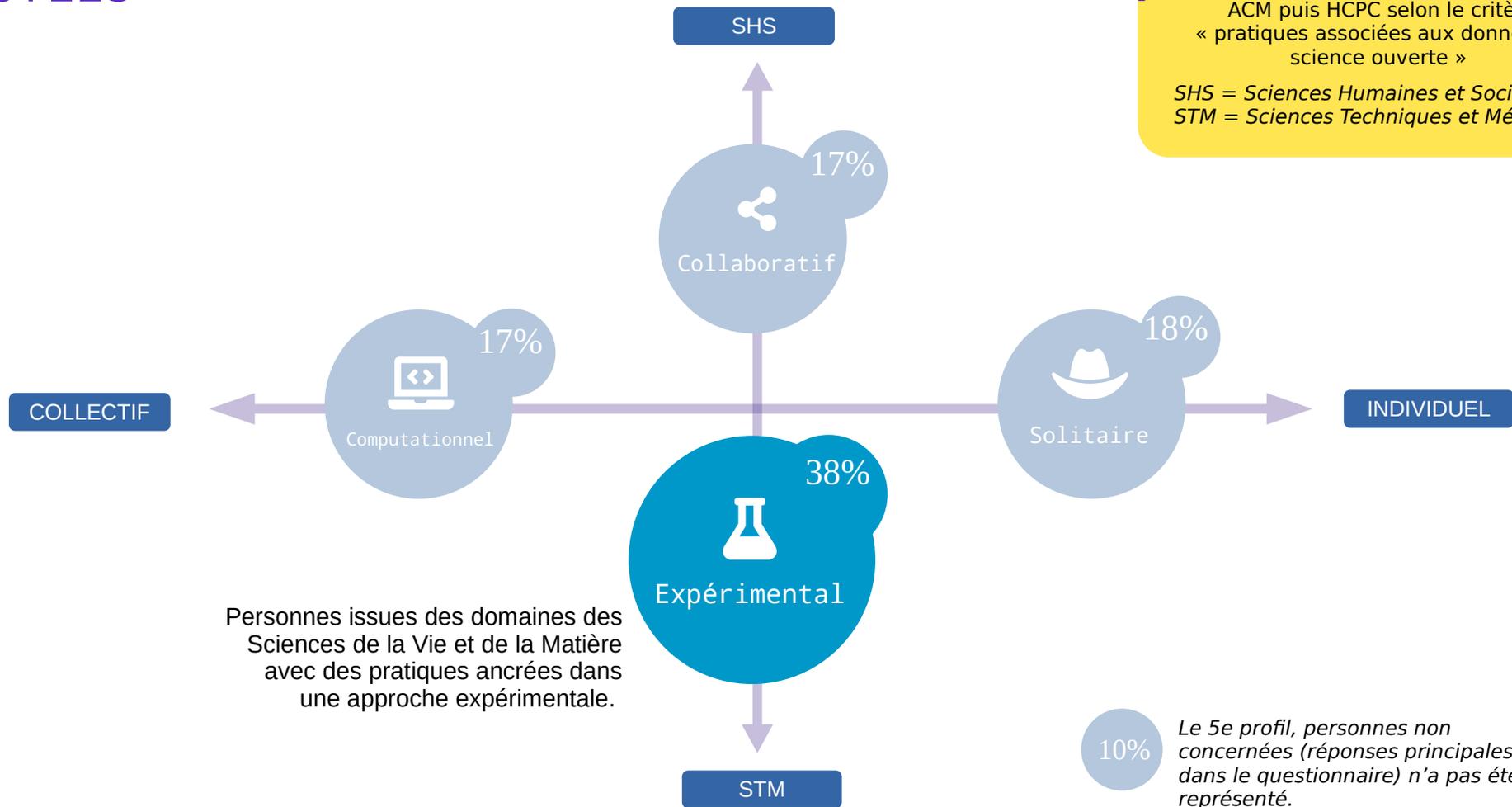


Besoins

Soutien financier et éthique (conscience politique).

Aide pour la valorisation des données et la publication à un large public.

Profils



Classes déterminées par le biais d'une ACM puis HCPC selon le critère « pratiques associées aux données & science ouverte »

SHS = Sciences Humaines et Sociales
STM = Sciences Techniques et Médicales

10% Le 5e profil, personnes non concernées (réponses principales N.A dans le questionnaire) n'a pas été représenté.



Experimental

Camille Bréchia

Domaine Pharmacologie

Lieu Ingénieure de recherche à l'INSERM

Expérience Depuis plus de 10 ans



Cadre clinique
Ordinateur analyse
Ordinateur/papier réflexion

Terminologie

Jeux de données, bases de données,
données brutes

Explicabilité, recherche ouverte,
reproductibilité, traçabilité

En bref

ID 774800833

Familiarité avec
la reproductibilité



Plus-value

Amélioration des techniques de collecte, de mesure et d'analyse. Standardisation de protocoles pour leur automatiser. Collecte de données rares, Création de modélisation et de théorisation.

Story

Camille est ingénieure de recherche dans une équipe de pharmacologie clinique au sein de l'INSERM. Elle est en charge à la fois du système d'information de l'équipe et de l'aide au suivi des projets de recherche clinique. Elle aide à vérifier la qualité des données produites et à améliorer les protocoles de collecte des données avec les chercheurs-cliniciens associés aux missions. Elle a l'habitude d'utiliser des langages de programmation et se forme depuis longtemps aux enjeux des données en autodidacte (veille à ce sujet, tutoriels). Le partage de données ne la concerne que peu de par la nature des données qu'elle traite (données de santé). Les données ne sont que partagées pour un public restreint et ciblé via un cloud institutionnel ou bien par e-mail pour des données non sensibles. Elle n'a jamais reçu d'aide concernant les pratiques associées aux données. Elle aimerait surtout recevoir un soutien financier et des ressources humaines supplémentaires afin d'être accompagnée sur l'amélioration des pratiques de collecte des données cliniques.

Mise à disposition des données

Conservation des données : sur son ordinateur professionnel, sur un serveur interne, sur un support externe.

Mise à disposition et réutilisation : partage restreint à un public ciblé par le biais d'un support amovible, de cloud institutionnel ou par mail. Réutilise (parfois) des données produites par d'autres.

Autres ressources à partager : protocoles méthodologiques, documentation des codes sources et données, logiciels et scripts d'analyse.

Limitations : une volonté de valoriser (avantage compétitif), barrières juridiques économiques, la peur d'un renforcement d'inégalités (production/analyse).

Obligations de partage : pas d'obligations de par la nature médicale des données.

Pratiques, mise en données

Étapes de travail :

31-50% collecte, description, contrôle qualité.

21-30 % mise à disposition, aide à la compréhension des données.

Sources : essais cliniques (données médicales).

Ressources supplémentaires produites : scripts et logiciels d'analyse, documentation du code source, des fichiers stockés sur son ordinateur.

Environnement numérique

Outils utilisés : R via R studio sous Linux.

Collaboration : travail collectif entre 2 et 5 personnes. Utilise des dossiers partagés.

Contexte d'initiation : se forme via des tutoriels et a découvert ses outils en faisant de la veille en ligne et sur les réseaux (autodidacte).

Accompagnement

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



Fréquence du soutien : non concernée.

Niveau de satisfaction : non concernée.

Besoins

Des spécialistes pour répondre à ses questions au sein de son établissement.

Des ressources humaines et une aide financière.

Une aide pour la collecte de données.

Experimental

Louis Tapures

Domaine Chimie

Lieu Directeur de recherche au CNRS

Expérience Depuis plus de 10 ans



Laboratoire
Ordinateur analyse
Ordinateur/papier réflexion

Terminologie

Mesures

Répliquabilité, 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 cible. 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 stockées 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 permanent.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 : ♥♥♥♥



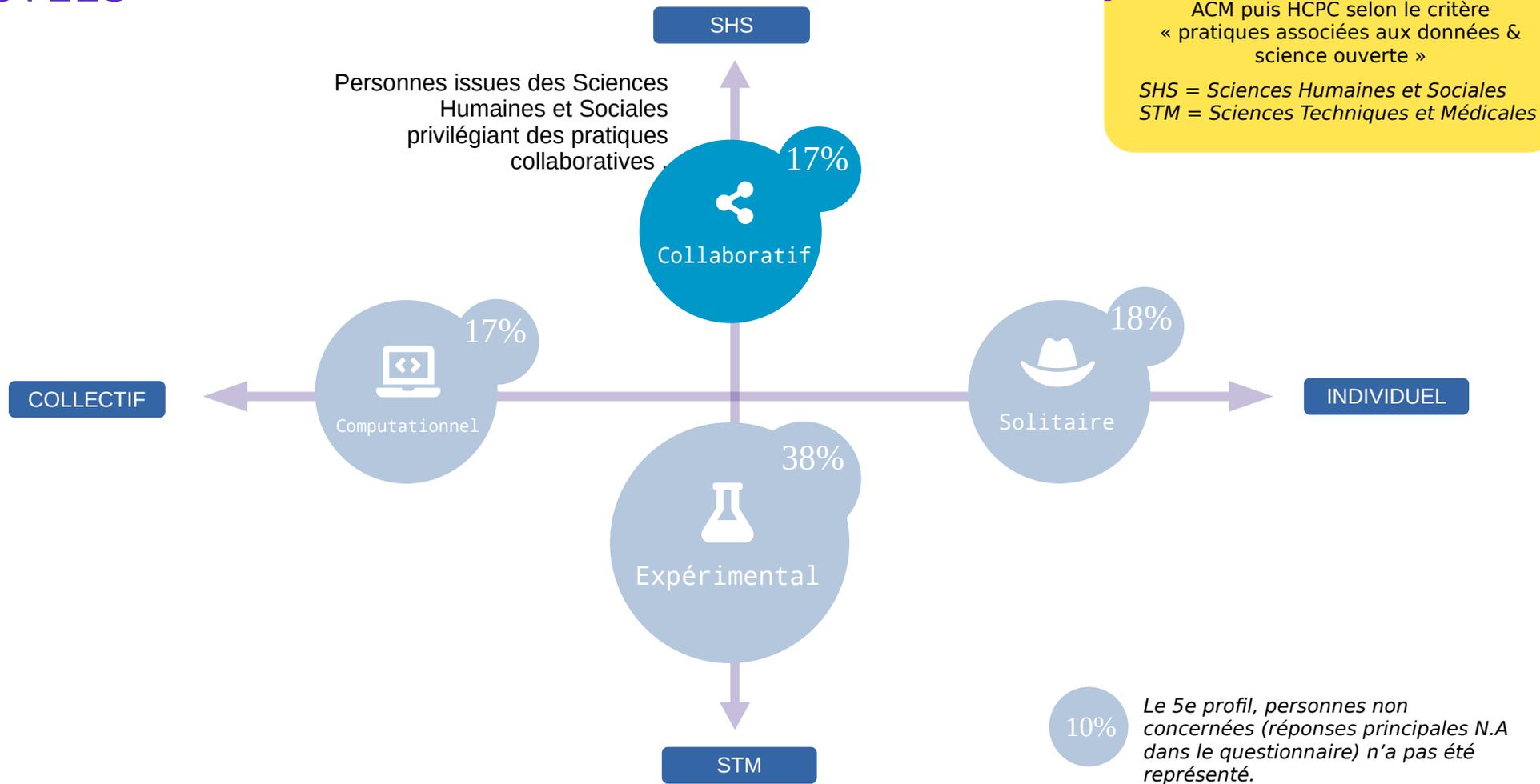
Besoins

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.

Profils



Collaboratif

Erwan GALINAND

Domaine Archéologie

Lieu Post-doctorant à l'Université de Rennes 1

Expérience Depuis moins de 5 ans



Bibliothèque ou archives
Ordinateur analyse

Terminologie

Corpus, sources, jeux et bases de données

Reproductibilité, traçabilité, transparence

En bref

ID 651511362

Familiarité avec la reproductibilité



Plus-value

Amélioration des techniques de collecte. Collecte de données rares et de données standards. Support à la recherche.

Story

Erwan effectue en ce moment un post-doctorat au sein de l'Université de Rennes 2. Il apprécie le travail dans une équipe interdisciplinaire où il apprend beaucoup de ses collègues historiens, géologues, écologues. Une partie conséquente de son travail implique de collecter des sources variées (issues d'archives ou d'autres terrains de collègues) pour constituer une base de données. Il est intéressé par les enjeux de partage des données et met déjà en ligne certaines informations de ses missions. Il reste néanmoins sceptique sur leur réutilisation possible car sa base est construite spécifiquement pour répondre à sa problématique de recherche. C'est plus par honnêteté intellectuelle qu'il souhaite les partager. Cela lui paraît aussi difficile d'obtenir les données des autres et de se les approprier (méthodologie spécifique). Pour lui, Twitter est une formidable façon de découvrir d'autres méthodologies issues d'une variété de disciplines. Pour l'instant, il n'a jamais reçu d'aide concernant ses pratiques associées aux données. Il est ouvert à une diversité de modalités d'accompagnement et souhaiterait notamment recevoir un soutien pour réaliser de courtes capsules vidéo pour expliquer son travail à un large public.

Mise à disposition des données

Conservation des données : sur un serveur externe, sur un ordinateur personnel et professionnel.

Mise à disposition et réutilisation : partage (souvent) en ligne sans contrôle d'accès. Partage (souvent) à un public cible restreint et connu. Met à disposition via un support amovible, par e-mail, via un site web et un entrepôt disciplinaire. Ne pense pas que les données soient réutilisables si partagées. Réutilise (rarement) des données déjà produites.

Mise à disposition possibles : notes de terrain, logiciels et scripts, métadonnées, protocoles, billets de blog, sites web grand public.

Limitations : une volonté de valoriser (avantage compétitif), la peur d'un renforcement d'inégalités (production/analyse), la peur de mésusages.

Obligations de partage : de la part du financeur.

Pratiques, mise en données

Étapes de travail :

>50 % collecter des données.

31-50 % constituer un corpus.

21-30 % analyser, visualiser des données.

Sources : bases de données en libre accès ou partagées par des collègues, articles scientifiques et fonds documentaires, terrains.

Ressources supplémentaires produites : des fichiers stockés sur son ordinateur, des notes diverses numériques et papier (notes de terrain).

Environnement numérique

Outils utilisés : logiciels de types tableurs et bases de données ainsi que visualisation et traitement d'images sous Windows.

Collaboration : taille de l'équipe entre 2 et 5 personnes. Travaille avec chercheur.e.s, doctorant.e.s. Utilise des dossiers partagés et des outils de rédaction partagée.

Contexte d'initiation : via réseaux sociaux et par des tutoriels (autodidacte) et par des formations, par des colloques et temps informels.

Accompagnement

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



Fréquence du soutien : non concerné.

Niveau de satisfaction : non concerné.

Besoins

Aide technique, juridique, infrastructures numériques.

Aide pour la mise à disposition la diffusion des données.

Par des spécialistes dans son université, un référent au sein de son équipe ou une plateforme dédiée.

Florence POIRAN

Domaine Histoire contemporaine

Lieu Maîtresse de conférences à l'Université de Clermont-Ferrand

Expérience Depuis 5 ans



Bibliothèque ou archives
Ordinateur analyse

Terminologie

Corpus, matériaux, bases de données, données brutes, données nettoyées, documents

Explicabilité, recherche ouverte, traçabilité, transparence

En bref

ID 1378924151

Familiarité avec la reproductibilité



Plus-value

Collecte de données rares et de données standards. Création d'une théorisation.

Story

Florence a été recrutée en tant que maîtresse de conférences à l'Université de Clermont-Ferrand en histoire contemporaine. Ses travaux de recherche l'amènent à consulter à la fois des bases de données en ligne (libre accès ou payant) ainsi que des fonds documentaires spécialisés. Aujourd'hui, elle collabore avec quelques chercheurs et partage sa base de données via e-mail ou sur Dropbox. L'habitude a été prise de rédiger collectivement des documents de synthèse via GoogleDocs. Les journées d'études sont des moments privilégiés pour découvrir d'autres projets et outils. Les temps informels de ces colloques lui permettent de recueillir certains conseils. Elle n'a jamais reçu d'aide sur la partie données et aurait besoin d'un soutien technique. Elle souhaite mieux connaître comment stocker et mettre à disposition ses données sur des infrastructures dédiées notamment pour son projet ANR. Pour l'instant, les données sont stockées sur un dossier partagé et sur son ordinateur personnel. Elle n'a en effet pas d'ordinateur professionnel mis à disposition. Même si elle souhaite se former, elle ne sait pas encore comment dans son emploi du temps chargé, elle pourrait consacrer plus de temps à la mise à disposition des données.

Mise à disposition des données

Conservation des données sur support externe (disque dur, clef USB) et sur son ordinateur personnel.

Mise à disposition et réutilisation : partage souvent à un public cible et connu. Partage parfois en ligne avec contrôle de l'accès ou avec l'attribution d'une licence. Utilise un site web, un cloud non institutionnel (GDrive). Ne sait pas si ses données sont réutilisables mais utilise souvent les données d'autres personnes.

Autres ressources à partager : la documentation des données, les protocoles méthodologiques, les logiciels d'analyse.

Limitations : un temps nécessaire trop important, un manque d'habitude, une volonté de valoriser (avantage compétitif).

Obligations de partage : non.

Pratiques, mise en données

Étapes de travail :

31-50 % collecter et analyser des données.

21-30 % nettoyer et structurer une base de données.

Sources : bases de données en libre accès et sous abonnement, données partagées par des collègues, archives et fonds documentaires.

Ressources supplémentaires produites : des vocabulaires contrôlés, des logiciels d'analyse, des notes diverses papier et numérique, des fichiers stockés sur son ordinateur.

Environnement numérique

Outils utilisés : tableurs génériques, logiciels de bases de données intégrées (FileMaker), logiciels d'analyse et de visualisation, langage de programmation.

Collaboration : taille de l'équipe entre 2 et 5 personnes. Travaille avec des chercheur.e.s et (post)-doctorant.e.s. Utilise des dossiers partagés, des outils de rédaction collaborative et un wiki.

Contexte d'initiation : via le réseau sociaux et lors de journées d'études et des temps informels d'échange. Se forme par des tutoriels et des ouvrages papier.

Accompagnement

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



Fréquence du soutien : non concerné.

Niveau de satisfaction : non concerné.

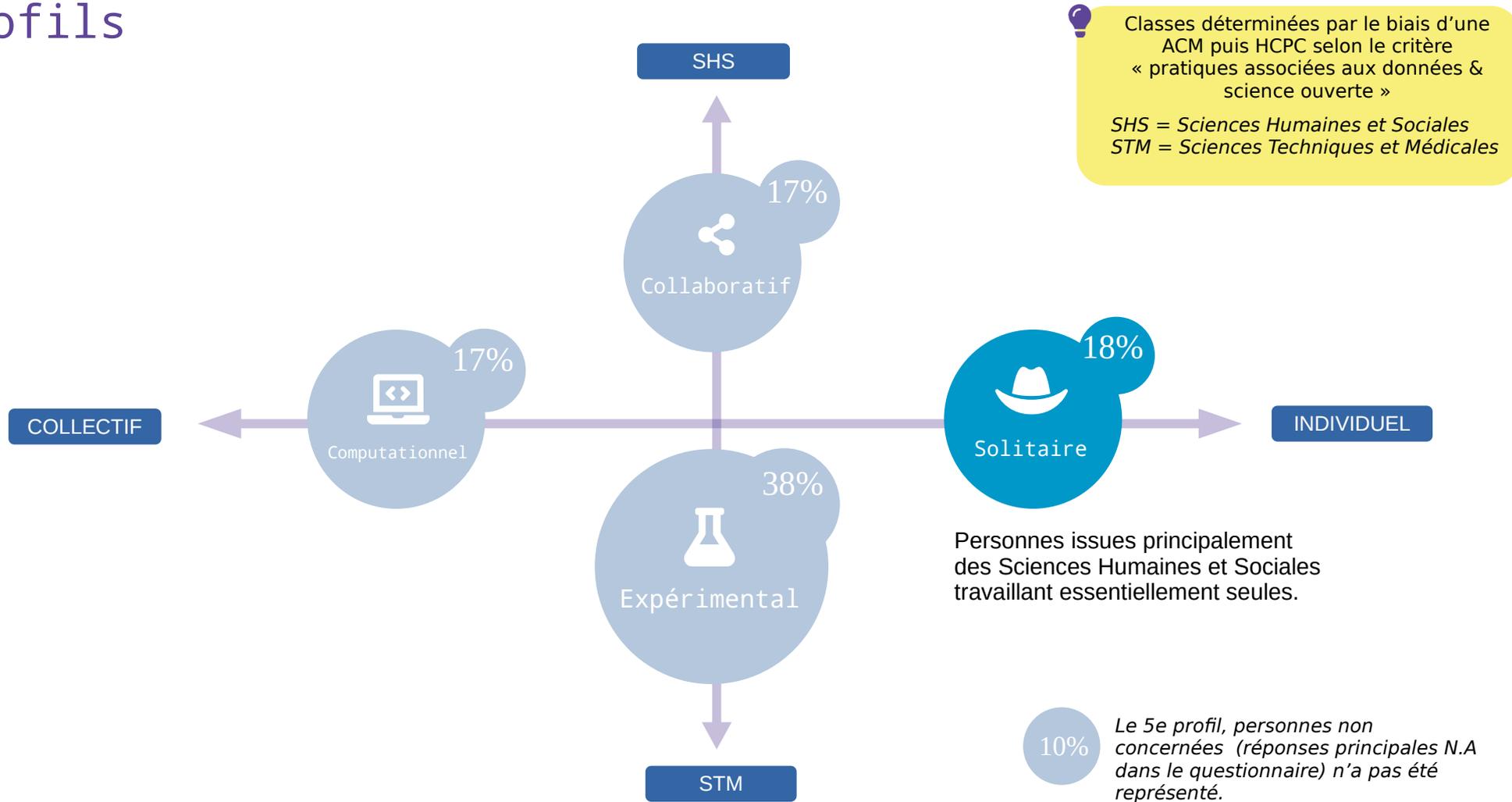
Besoins

Des ressources humaines, aide technique, des infrastructures numérique.

Aide pour le stockage et la mise à disposition ainsi que pour la publication à un large public.

Par des spécialistes disponibles dans son établissement, une personne soutien dans son équipe et des journées d'études.

Profils





Solitaire

Mehdi Ruschd

Domaine Sociologie

Lieu Doctorant à l'INSEE

Expérience Depuis 3 ans



Ordinateur analyse
Ordinateur/papier réflexion

Terminologie

Statistiques, jeux de données, bases de données

Explicabilité, transparence

En bref

ID 1772490178

Familiarité avec la reproductibilité



Plus-value

Création d'une théorisation, d'une interprétation d'un ensemble de faits observés.

Story

Mehdi est en troisième année de doctorat à l'INED, il travaille d'arrache-pied pour finir ses dernières analyses avant de se consacrer à la rédaction de sa thèse. Il a travaillé pour sa thèse sur un ensemble d'enquêtes nationales de l'INSEE en libre accès. Il a été formé en master à Stata mais à découvert R par lui-même. Il a encore peu l'habitude de ces pratiques et aimerait bien être sensibilisé à tout cela mais ces problématiques ne lui paraissent peu prioritaires par rapport à son objectif premier de finir sa thèse. Il prévoit peut-être après sa thèse de trouver du temps pour remettre de l'ordre dans ses données, ses notes mais pour l'instant ces dernières lui paraissent impossibles à réutiliser. Le mieux pour lui serait d'avoir une personne au sein de l'équipe à qui il pourrait s'adresser pour toutes ces questions. Il a déjà perdu en début de thèse une partie de ses analyses et prend le soin maintenant d'avoir plusieurs copies (sur son ordinateur perso, son ordinateur pro mais aussi sur un serveur de l'INED).

Mise à disposition des données

Conservation des données : sur son ordinateur personnel et professionnel et sur un serveur institutionnel.

Mise à disposition et réutilisation : partage parfois des données à un public cible. Ne met pas à disposition ses données. Ne sait pas si ces données sont réutilisables. Réutilise déjà des données produites par d'autres.

Autres ressources à partager : documentation des données et des codes sources.

Limitations : un manque d'habitude, la peur de mésusages, un doute sur la capacité de réutilisation des données.

Obligations à partager : aucune.

Pratiques, mise en données

Étapes de travail :

> 50 % analyser des données.

11-20 % nettoyer, modéliser, visualiser des données.

Sources : bases de données disponibles en libre accès.

Ressources supplémentaires produites : des fichiers de données stockés sur son ordinateur ou sur un stockage central, des scripts de programmation, des notes papier et numériques.

Environnement numérique

Outils utilisés : logiciels de traitement statistiques (Stata) et langage de programmation (R) sous Windows.

Collaboration : travaille seul. N'utilise pas d'outils collaboratifs.

Contexte d'initiation : par des formations, par des tutoriels et des ouvrages papier.

Accompagnement

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



Fréquence du soutien : non concerné.

Niveau de satisfaction : non concerné.

Besoins

Infrastructures numériques, aide technique.

Aide pour la constitution, le stockage, la mise à disposition, la valorisation, la publication.

Par des référents au sein de l'équipe, des spécialistes dans l'établissement, des formations.

Périnne HOARAU

Domaine Anthropologie
Lieu Ingénieure d'étude à l'Université de La Réunion
Expérience Depuis plus de 10 ans



Travail de terrain
 Bibliothèque ou archives
 Ordinateur/papier réflexion

Terminologie

Corpus, sources, matériaux

Explicabilité, reproductibilité, traçabilité

En bref

ID 1099630968

Familiarité avec la reproductibilité



Plus-value

Amélioration des techniques de collecte. Collecte de données rares et de données standards. Support à la recherche.

Story

Périnne travaille en tant qu'ingénieure d'étude au sein d'une UMR d'anthropologie et de littérature à l'Université de La Réunion. Elle participe aux travaux de recherche aux côtés d'un chercheur permanent et d'un ingénieur de recherche. Elle aide essentiellement à la collecte et à la mise à disposition des données et publications (milieu académique et grand public) ainsi qu'à la sauvegarde et l'archivage des données. Avant d'obtenir ce poste, elle avait fait son master à l'Université de Montpellier et avait apprécié suivre des journées d'études. C'est là où elle a découvert plusieurs outils notamment Zotero pour la gestion bibliographique. Aujourd'hui, elle se sent peu soutenue dans ses tâches de collecte et de structuration des données. Peu d'offres de formations existent et elle se sent assez seule pour gérer cela au sein de son équipe.

Mise à disposition des données

Conservation des données : sur son ordinateur professionnel, sur un support externe, sur un serveur institutionnel, sur un service d'archivage.

Mise à disposition et réutilisation : partage souvent à un public cible et restreint. Partage parfois en ligne sans contrôle d'accès. Met à disposition via des dossiers partagés institutionnels. Réutilise parfois des données d'autres personnes.

Autres ressources à partager : les protocoles méthodologiques.

Limitations : un temps nécessaire trop important, des barrières juridiques et économiques, une mauvaise qualité des données.

Obligations à partager : aucune.

Pratiques, mise en données

Étapes de travail :

21-30 % collecter, mettre à disposition des données à d'autres professionnels.

11-20 % constituer un corpus, sauvegarder / archiver, mettre à disposition en accès libre, former et à aider à la compréhension des données.

Sources : issues de terrains d'enquête, des bases de données en libre accès ou sous abonnement, des articles scientifiques, des sites web.

Ressources supplémentaires produites : des fichiers stockés sur son ordinateur, des notes diverses papier et numériques.

Environnement numérique

Outils utilisés : tableurs génériques et logiciels bibliographiques sous Mac OS.

Collaboration : équipe entre 2 et 5 pers avec des ingénieur.e.s et des chercheur.e.s permanents. Travaille essentiellement seule. N'utilise pas d'outils collaboratifs.

Contexte d'initiation : par d'autres membre de l'équipe, lors de journée d'étude et des temps informels, par des tutoriels et par elle-même.

Accompagnement

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

Fréquence du soutien
 ★☆☆☆☆

Niveau de satisfaction
 ♥♥♥♥



Besoins

Formation.

Aide par des spécialistes dans l'établissement, une personne soutien au sein du labo, une plateforme en ligne, des journées d'études.

Aide pour la collecte et structuration des données.

2

FACTEURS DIFFÉRENCIANTS

Démarche, outils et
statut

Aller au-delà des disciplines

Les différences entre pratiques associées aux données ne dépendent pas seulement des disciplines



Cas d'étude « virus à ARN » : des disciplines communes mais des approches distinctes

Différents professionnels de la recherche ont été interrogés dans le cadre des phases qualitatives de l'enquête (entretiens exploratoires et d'observation de pratiques).

Ces personnes avaient en commun de travailler sur les « virus à ARN ». Même si ces personnes provenaient de disciplines communes, les approches étaient très différentes notamment entre recherche clinique vs. expérimentale / observationnelle (recherche fondamentale). Les distinctions portent sur

- Les démarches de recherche (épistémologie, méthodologie, scientificité, plus-value)
- Les environnements de travail (outils, communautés)
- Le cadre juridique, économique (cadre réglementaire recherche clinique, financement, etc.)

Une montée d'approche *data-driven* est présente de manière transversale.

Résultats
phase 1 et 2
qualitatives

Décliner la science ouverte en fonction de...

1

Démarches de recherche

2

Pratiques, outils et apprentissages

3

Statut et fonction en recherche

Décliner la science ouverte en fonction de...

1

Démarches de recherche

2

Pratiques, outils et apprentissages

3

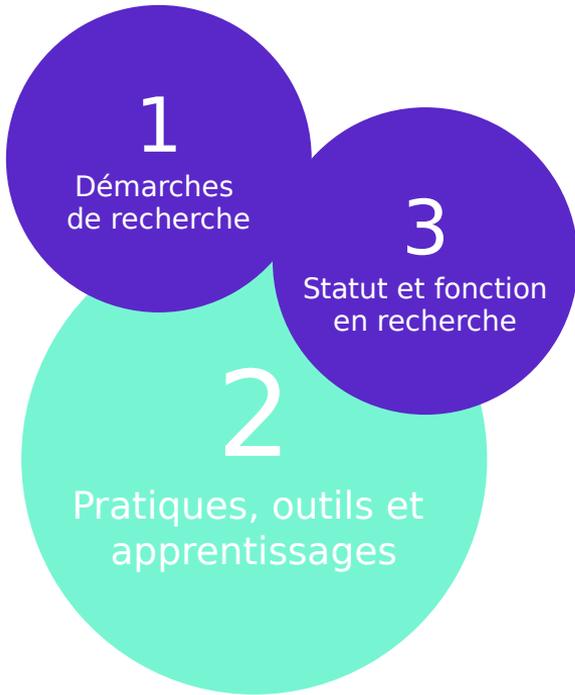
Statut et fonction en recherche

1. Démarches de recherche

- Environnement de recherche (laboratoire, clinique, terrains, etc.)
- Provenance (instruments de mesure, archives, etc.)
- Rapport aux données (qualificatifs associés)
- Critères associés à la qualité de la recherche (reproductibilité, transparence)
- Plus-value du travail de recherche
- Étapes de « mise en données »

⇒ Principes épistémologiques, méthodologiques, critères de scientificité (paradigmes de recherche, communautés épistémiques)

Décliner la science ouverte en fonction de...

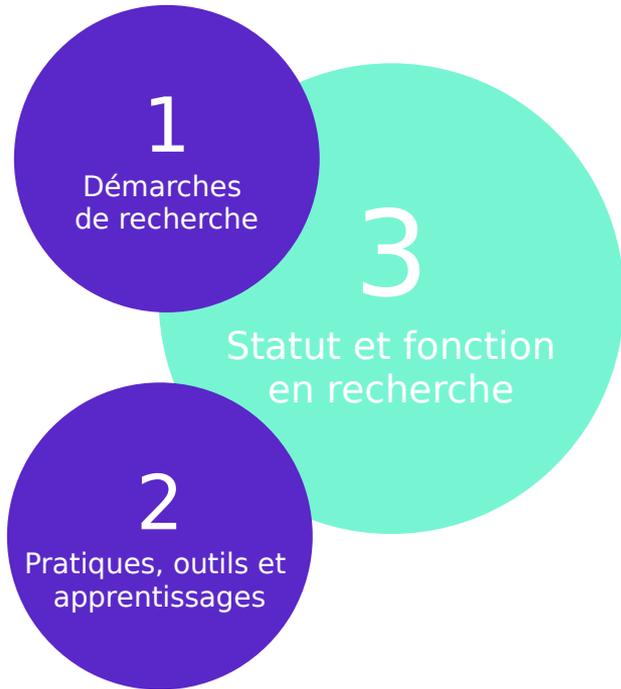


2. Pratiques, outils et apprentissages

- Environnement (numérique) de travail : outils employés en lien avec les données
- Pratiques collaboratives
- Modalités de découverte et d'apprentissage
- Types de besoins et d'accompagnement

⇒ Associées à la notion de communautés de pratiques et d'apprentissages

Décliner la science ouverte en fonction de...



3. Statut et fonction en recherche

- Catégorie professionnelle
- Statut, ancienneté

Décliner la science ouverte en fonction de...

1. Démarches de recherche

- Environnement de recherche (laboratoire, clinique, terrains, etc.)
- Provenance (instruments de mesure, archives, etc.)
- Rapport aux données (qualificatifs associés)
- Critères associés à la qualité de la recherche (reproductibilité, transparence)
- Plus-value du travail de recherche
- Étapes de « mise en données »

⇒ principes épistémologiques, méthodologiques, critères de scientificité (paradigmes de recherche, communautés épistémiques)

2. Pratiques, outils et apprentissages

- Environnement (numérique) de travail : outils employés en lien avec les données
- Pratiques collaboratives
- Modalités de découverte et d'apprentissage
- Types de besoins et d'accompagnement

⇒ Associées à la notion de communautés de pratiques et d'apprentissages

3. Statut et fonction en recherche

- Catégorie professionnelle
- Statut, ancienneté

Décliner la science ouverte

Rapport final

3

Accompagner
l'évolution des
pratiques :
pistes
d'orientation

Problématique

#2

Comment **accompagner l'évolution des pratiques associées aux données** en lien avec les incitations / obligations portées **par les politiques publiques « science ouverte » ?**

Pistes d'orientations :

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

1

PISTES D'ORIENTATION

Comprendre en finesse
les démarches de
recherche

Terminologie : rapport aux données

Outre le terme « données », d'autres sont employés pour décrire les éléments qui servent à l'obtention de résultats de recherche :

- Certains font consensus : bases de données, jeux de données
- D'autres sont plus spécifiques :
 - Corpus dans le cadre de recherche principalement en SHS sur des sources, des archives,
 - Matériaux en SHS
 - Mesures dans des approches expérimentales

« **Qu'appelle-t-on données à la fin ?** Je peux utiliser des éléments d'analyse repérés dans d'autres publications scientifiques, je peux utiliser des informations de sources différentes (site internet, réseaux sociaux, rapports, littérature scientifique, blog) que je transforme ensuite en données exploitables par le recoupement à d'autres informations. **Mais une donnée n'est pas une information en ce [que ce] n'est pas "digérée" par le chercheur dans son travail de recherche.** »

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.



Terminologie : rapport aux données

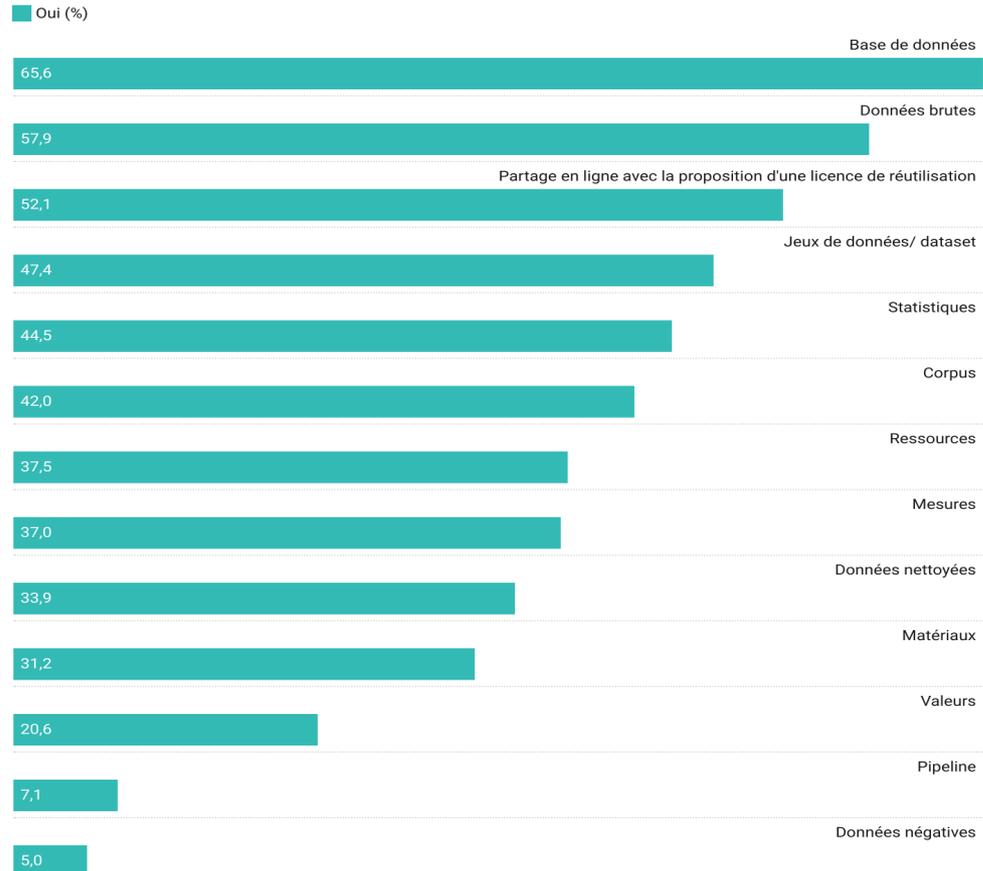
Le terme le plus couramment employé (65,6%) est « base de données »



Nous menons un travail de réflexion critique autour du terme de données, dans la mesure **où dans des approches très qualitatives les "traces" ne sont pas "données" mais font sens pour le chercheur** en fonction de son expérience et de son histoire, de l'expérience et de l'histoire du témoin, de la relation de recherche etc. »

Le terme de « données » est largement employé aujourd'hui. Y a-t-il d'autres termes (ou des qualificatifs supplémentaires) que vous utilisez pour décrire les éléments que vous manipulez dans votre travail et qui vous servent à l'obtention des résultats de

Réponse à choix multiples.



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

Se décentrer des enjeux de reproductibilité (1)

Le terme « reproductibilité » est plus favorablement employé dans le cadre de recherche impliquant des instruments de mesure. Il se rattache **à l'emploi de méthodes computationnelles** (vérification de calculs à partir de l'accès aux codes sources et aux « données brutes »).

D'autres termes sont plus inclusifs pour aborder la question de la qualité de la recherche

- La notion de **réplicabilité** est propre aux méthodes expérimentales et implique de considérer également l'accès aux protocoles méthodologiques (pas exclusivement aux données)
- La notion de **transparence** est plus corrélée à des cadres de recherche pluridisciplinaire et **l'explicitabilité** à des contextes SHS (travail sur des corpus et la construction de bases de données)

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



Se décentrer des enjeux de reproductibilité (2)

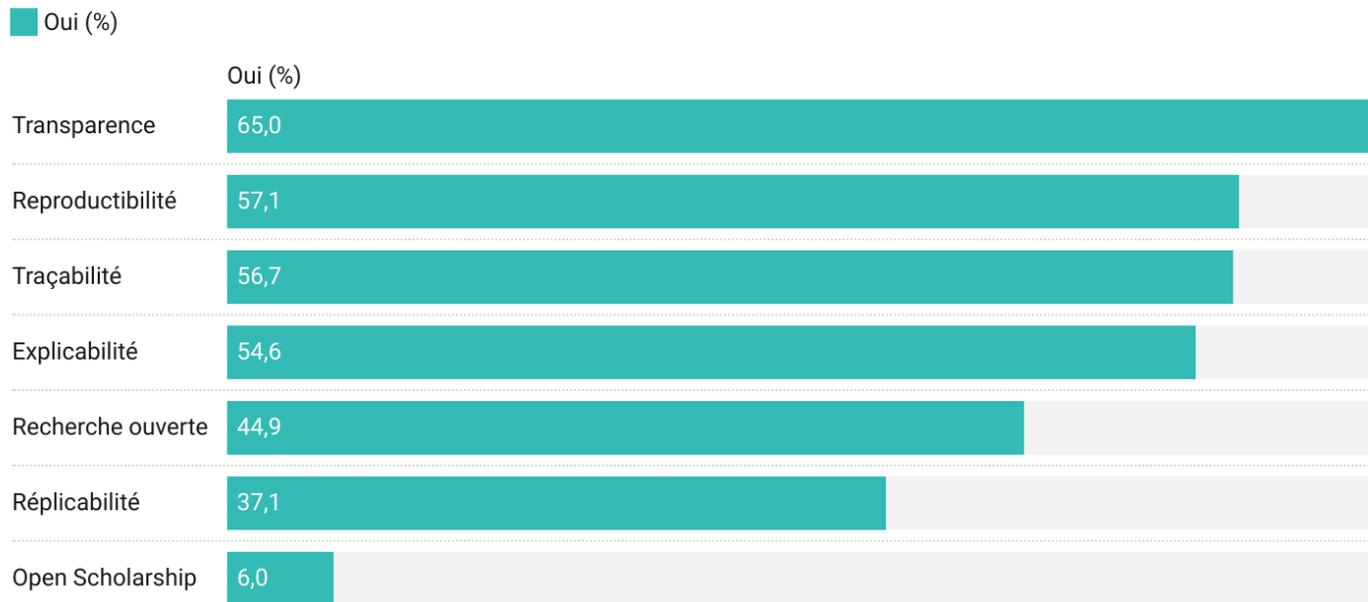
« Je ne l'utilise pas dans mon travail **et je ne cherche pas à ce qu'il soit reproductible (résultats) mais répliquable (méthode)** à un autre terrain d'enquête. »

« Je connais bien ce terme positiviste mais **je ne me reconnais pas là dedans.** »

« Dans ma discipline, toute recherche est personnelle, dans **laquelle le chercheur est nécessairement impliqué en tant que tel.** Un autre, avec le même sujet et les mêmes sources n'aboutira pas au même résultat. »

Quels mots représentent un ou des critères de qualité en recherche pour vous?

Réponse à choix multiples.



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

... pour aborder ceux de la qualité en recherche

Dans le questionnaire, les commentaires soulignent l'importance de considérer des valeurs éthiques (intégrité, honnêteté, etc.) pour la qualité en recherche



Nuage de mots clés réalisé sur la base des commentaires qualitatifs associés à la question portant sur la reproductibilité et la qualité de la recherche. Réalisé via voyant-tools

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.) .



Porter attention aux différentes formes de plus- values retirées du travail de recherche (1)

Différentes **plus-values** sont à distinguer en fonction des **disciplines** mais aussi et surtout des **démarches de recherche** (sources des données, environnement de travail, rapport aux données, etc.). Par exemple :

- La collecte de données rares ou nécessitant un temps de collecte important est corrélée à un travail de terrain ou sur des archives et fonds documentaires au sein des SHS
- La préparation d'échantillons et de 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 *workflow* et de la modélisation sur une grande quantité de données.

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.

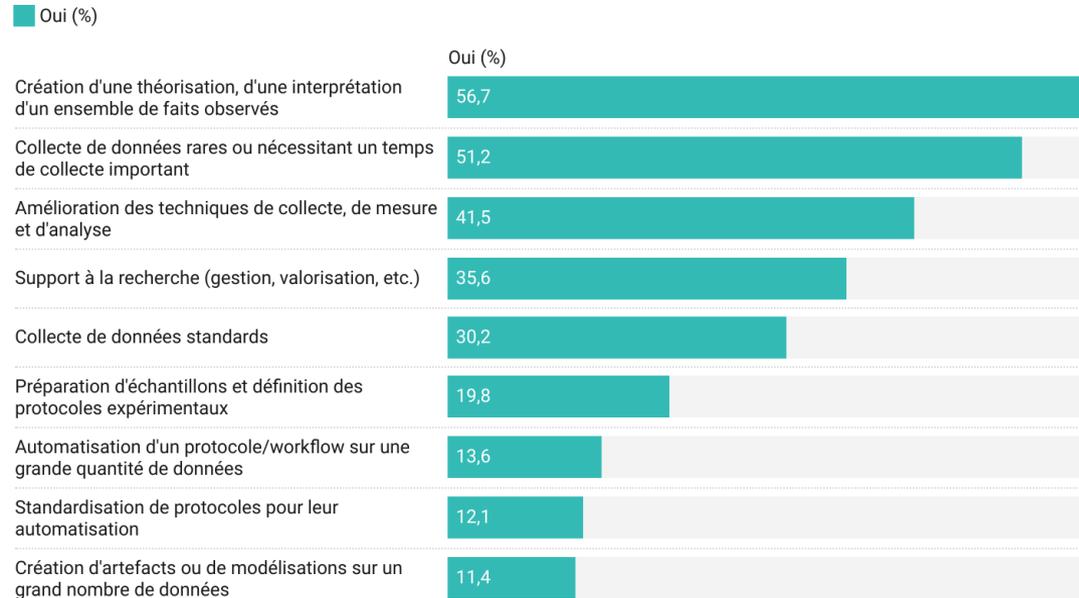


Porter attention aux différentes formes de plus-values retirées du travail de recherche (2)

⇒ Porter attention à ces différentes formes de plus-values permet de **repérer des blocages quant à la mise à disposition de données. Certaines formes peuvent l'inciter** (amélioration de technique, automatisation, modélisation) tandis que **d'autres la réfréner** (collecte de données rares, préparation d'échantillon chronophage).

Lesquelles de ces propositions décriraient le mieux la plus value de votre travail au sein de votre communauté de recherche

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 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
Laboratoire	Sciences, Techniques et Médecine	Reproductibilité, répliquabilité	↳ Mesures, valeurs, données brutes, ↳ Sources	Expérimentation	Instruments de mesure
Terrains, Archives,	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
Clinique	Sciences de la vie et de la santé	Résultats négatifs	↳ Pipelines, données nettoyées, données négatives ↳ corpus	Modélisation, automatisation de workflow	Instruments de mesure

Démarches de recherche : spécificités

Résultats phases 1 et 2 qualitatives



Recherche clinique

Une structuration des données dès le départ par le « *Case Report Form* », protocole déjà validé et à suivre (loi Jardé). ⇒ pratique hypothético-déductive.

Plus-value : amélioration de la précision des techniques de collecte, de mesure et d'analyse et leur automatisation (qualité des données, quantité des données et standardisation).



Recherche expérimentale / observationnelle

Structuration et choix des informations/sources recueillies de manière itérative. Importance d'un cadre malléable et souvent « visuel » pour l'articulation de différentes données obtenues ⇒ pratique inductive/abductive.

Plus-value : La collecte de données rares. Un « art de l'interprétation » aussi bien en biologie expérimentale qu'en SHS.



Recherche computationnelle (*data-driven*)

Travail de « nettoyage », analyse automatisée sur une large quantité de données déjà collectées nécessitant une standardisation ⇒ pratique *data driven/centric (computationnelle)*

Plus-value : automatisation des processus d'analyse (workflow/pipeline), quantité et qualité des données initiales.

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.



2

PISTES D'ORIENTATION

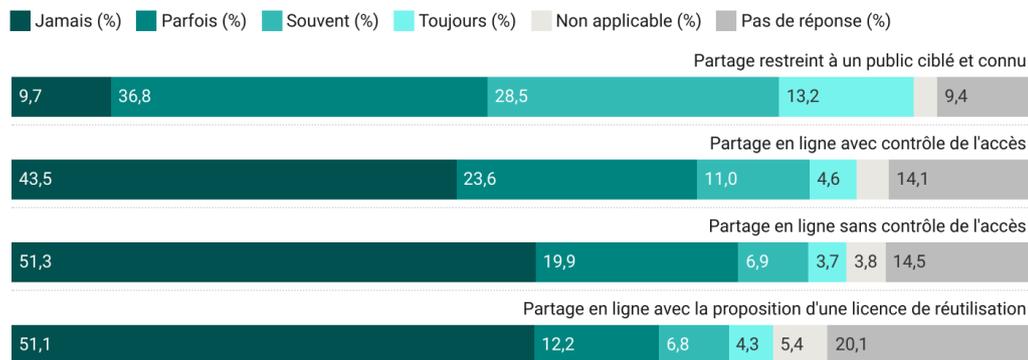
Appréhender différentes pratiques
de mise à disposition des données

Pratique majoritaire : partage restreint à un public ciblé

- Le **partage restreint pour un public ciblé et connu** reste la pratique majoritaire.
- Le **partage en ligne avec la proposition d'une licence ouverte (*open data*)** représente un peu plus de 20 % des pratiques.
- 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 (recherche biomédicale).

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

Le terme « mise à disposition » inclus :

- **L'ouverture des données au sens d'*open data*** ⇒ partage avec licences ouvertes
- **Le partage de données** avec différentes modalités (envoi à un public restreint, mise en ligne avec contrôle ou non de l'accès)

Inciter **les comités éditoriaux de revues** à s'appuyer sur les politiques nationales existantes concernant les données et les codes associés aux publications scientifiques

Réutilisation des données et partage d'autres ressources

- Près de 50 % des personnes interrogées **indiquent réutiliser parfois** des données déjà produites ou publiées
- Plus de 45 % de personnes interrogées considèrent que **leurs données seraient potentiellement réutilisables**
- **D'autres produits connexes à partager sont cités** (cahiers de laboratoire, protocoles méthodologiques, blogs, vidéos pour accompagner les résultats)
- **Les modalités de leur mise à disposition diffèrent en fonction du public** (des collègues au grand public) tout **autant que l'accompagnement nécessaire à leur partage** (aide à la structuration de données, diffusion de vidéos, etc.)

« [O]uvrons les codes sources, les méthodes, les protocoles, pour pouvoir les inspecter. Mais par contre, je n'ai pas mis les cahier[s] de laboratoire ou les notes de terrain, car **je ne pense pas qu'ils doivent être ouverts "à tout le monde"** et dans "n'importe quelles conditions. »

« Gestion des données : **conception des protocoles** dans la perspective d'un versement en Archive ouverte, finalisation des corpus pour archivage pérenne, gestion d'un site web pour un accès facile aux données. »

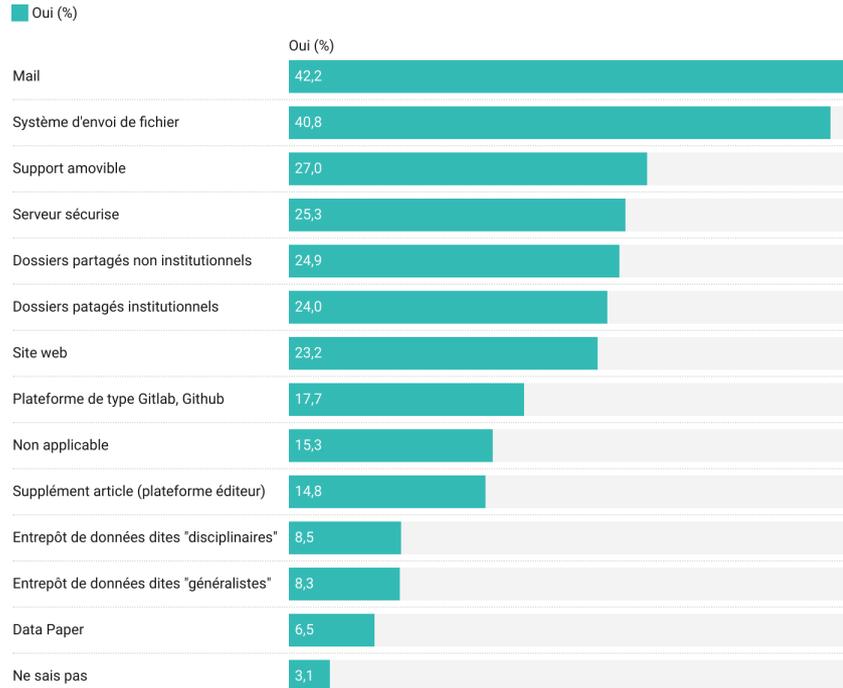
Mettre à disposition des données nécessite de **penser le « public des données »**, la **temporalité, les modalités de partage** (juridique, financier, technique, etc.) et l'accompagnement nécessaire.

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.

Modalités d'envoi

De quelles manières mettez-vous à disposition ces données?

Réponse à choix multiples.



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

- **L'envoi par mail ou par système d'envoi de fichier** sont les plus communément employés
- L'emploi **d'entrepôt de données généraliste ou disciplinaire** reste **minoritaire**
- L'utilisation **de dossiers partagés non institutionnels** (Gdrive, Dropbox) prévaut sur l'usage d'un système institutionnel



J'aimerais bien, **mais le Cloud du CNRS fonctionne une fois tous les jamais.**

Dans mon université, le cloud institutionnel est un drive Google. »



Focus sur les profils - spécificités

- La mise à disposition sur des plateformes de code, des entrepôts généralistes, le partage avec licences : sur-représentation dans le profil « computationnel ».
- Cloud non institutionnel : sur-représentation dans le profil SHS collaboratif.



Limites à la mise à disposition

- Les raisons principales limitant la mise à disposition des données sont essentiellement **le manque d'habitude de ces pratiques** (63%), un **temps trop important** nécessaire (49%) et une volonté de valoriser les données pour garder **un avantage compétitif** (48%)
- Les problématiques de bureaucratie, les enjeux de données personnelles et éthiques ont aussi été abordées.



D'autres critiques :

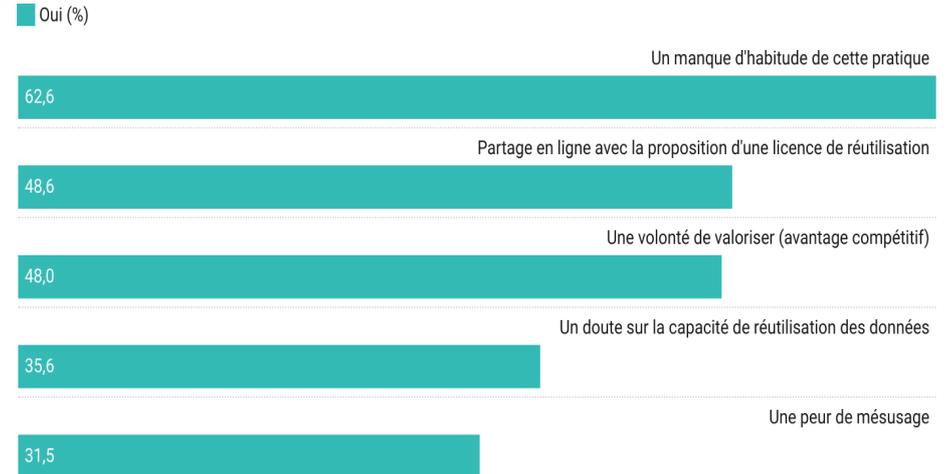
Les obligations naissantes en termes de partages de données de recherche semblent **difficilement conciliables avec les méthodes ethnographiques**. Elles constituent une contrainte qui n'est ni utile, ni réellement transparente, apportant pour l'essentiel dans **ma discipline un surcroît de bureaucratie.** »

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.



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

Réponse à choix multiples.



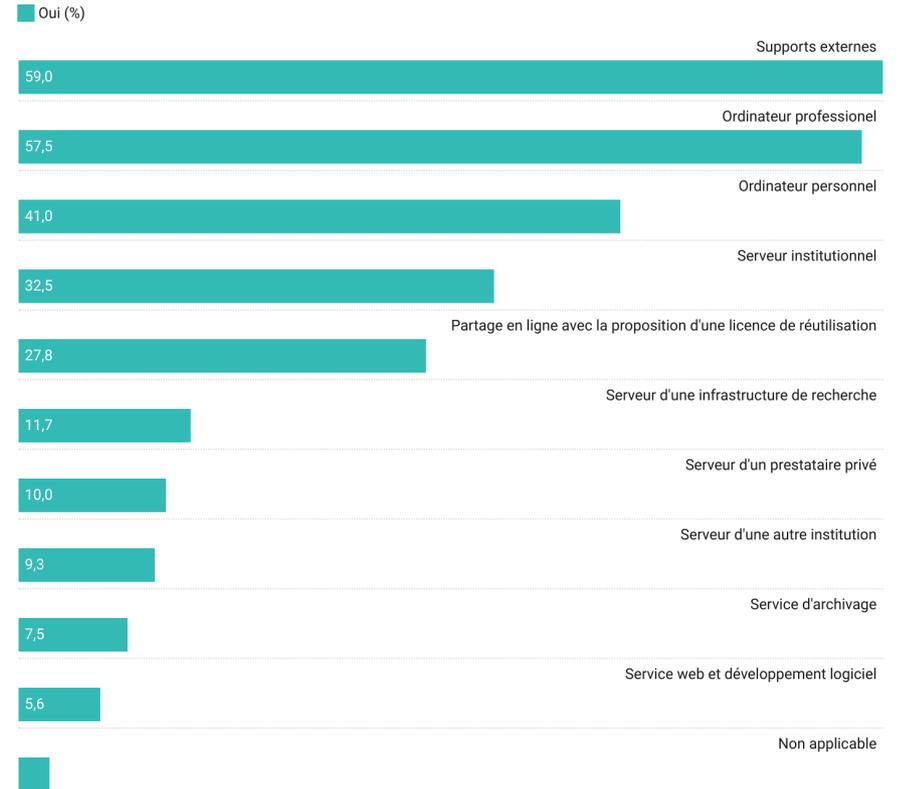
Graphique: Groupe de travail "Réussir l'appropriation de la science ouverte" • Source: Questionnaire "données et science ouverte" • Créé avec Datawrapper

Conservation et sécurité des données (1)

- Un stockage des données est majoritairement réalisé **sur des supports externes et des ordinateurs professionnels**
- Il y a peu d'appel à des services d'archives (7,5%)
- Les commentaires au questionnaire font ressortir **une attention et vigilance sur la sécurité des données** (données cryptées, risque de piratage, etc.) et un questionnement **sur la fiabilité des infrastructures institutionnelles**
- En SHS, il est fréquent de travailler sur son ordinateur personnel (pas d'ordinateur professionnel fourni) notamment pour les doctorant.e.s

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

Réponse à choix multiples.



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

Conservation et sécurité des données (2)

« Pas de personnel dans le laboratoire pour le support informatique. Cela a récemment engendré une faille de sécurité sur le lieu de travail, rendant inutilisable tout le parc informatique, avec les conséquences que cela implique pour les projets en cours. »

« Toujours. En copie de sécurité. Je n'ai qu'une confiance limitée dans la pérennité des serveurs institutionnels. »

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.

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.

3

PISTES D'ORIENTATION

Connaître les modalités
d'apprentissage et les
pratiques collaboratives

Découverte et formation aux outils : un échange entre pairs (1)

Outre **la découverte d'outils** par « soi-même », **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

« Je trouve que **Twitter est un formidable moyen de découvrir de nouvelles applications ou de nouvelles pratiques**, parce que ce réseau social permet de sortir facilement de ses domaines de recherche propres et donc de découvrir les solutions trouvées/mises au point par d'autres disciplines. Ce genre d'échanges méthodologiques transdisciplinaires est très difficile à avoir dans la vraie vie. »



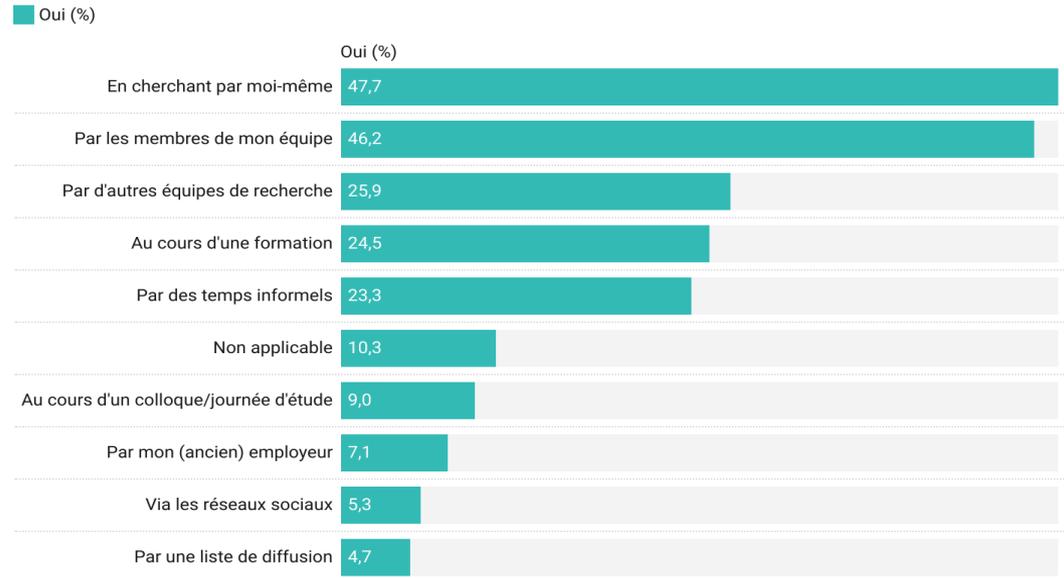
Focus sur les profils - spécificités

- Au sein des SHS, les temps informels lors de journées d'étude sont un élément clef.
- La prise en main d'outils s'effectue avec la « vie de laboratoire » (expérimental) dès les premiers stages avec une formation au sein des équipes (maître de stage).



Comment avez-vous découvert ces outils numériques de travail?

Réponse à choix multiples.



Découverte et formation aux outils : un échange entre pairs (2)

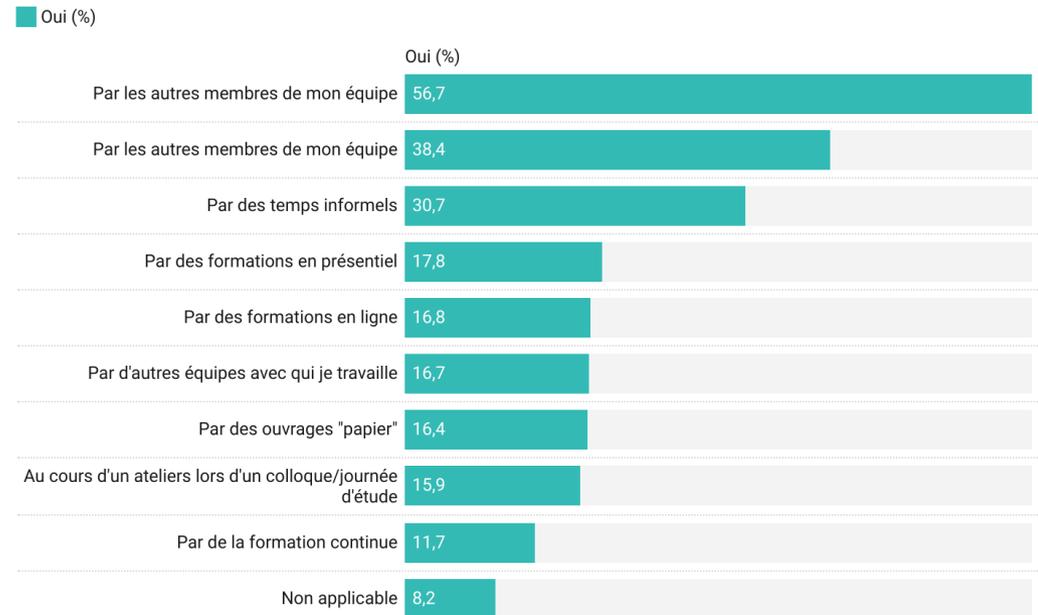
- De la même manière, **la formation à de nouveaux outils** s'opère majoritairement par des tutoriels et de la documentation en ligne **puis par le biais d'autres membres des équipes et de temps informels**. Les formations en présentiel et en ligne ne viennent qu'ensuite.
- **Les réseaux sociaux (Twitter)** sont cités comme espace d'échange et de découverte de pratiques



Pour la formation **c'est aussi pendant les études ou des écoles d'été.** »

Comment vous formez-vous habituellement à ces outils?

Réponse à choix multiples.



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

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

Résultats
phases 1 et 2
qualitatives

6a Journée d'étude : 20 personnes réunies, 12 présentations, 4 sessions

- 1 : Collecte et mise à disposition de données
- 2 : Travail collectif et pérennisation
- 3 : Bricolage et apprentissage
- 4 : Collecte et catégorisation

Trois thématiques clefs :

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

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.

**Synthèse des échanges disponibles
dans cette présentation**

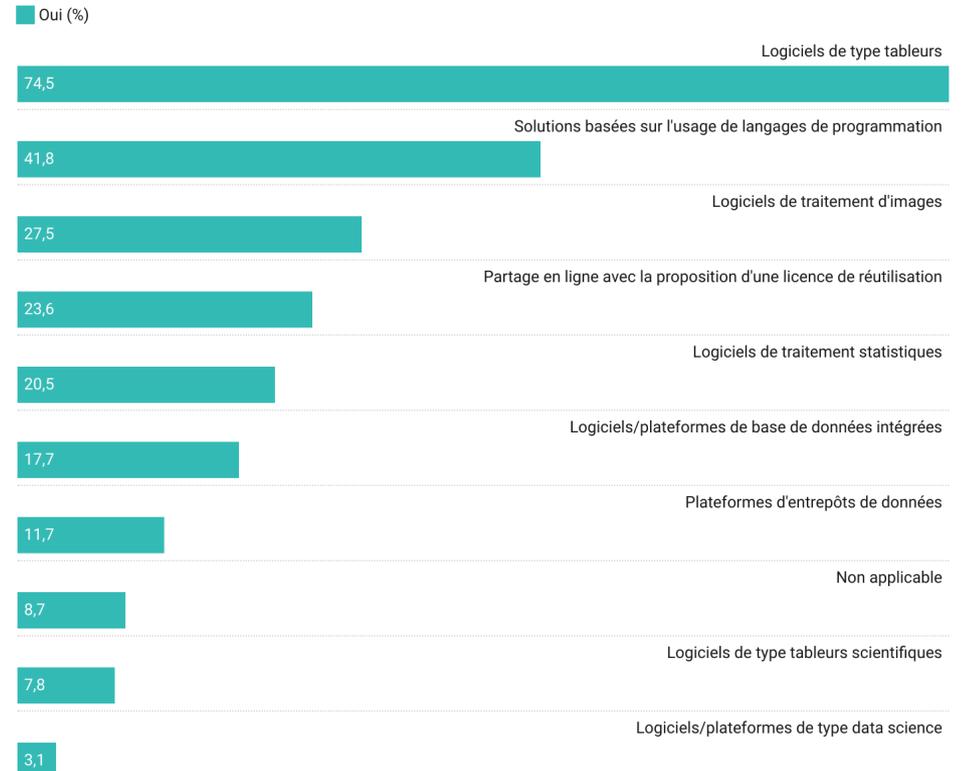
Pour faciliter l'appropriation de nouvelles pratiques, **prendre en considération les spécificités de rencontres et d'apprentissages des communautés** (vie de laboratoire, journées d'étude et de conférence, réseaux sociaux, etc.).

Outils employés avec les données (1)

- **Utilisation majoritaire de logiciels de type tableurs** (Excel, Calc) (75%)
- 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 en commentaire
- **Les logiciels/plateformes de base de données intégrées** (18%) cités sont par exemple FileMaker, PostgreSQL, MySQL
- **Les plateformes d'entrepôts de données** n'ont été utilisées que par 12 % des répondant.e.s

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.



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

Outils employés avec les données (2)

- **Le système d'exploitation** le plus largement employé est **Windows** (62%) contre 26 % pour MacOS et 12 % pour Linux et autres Unix

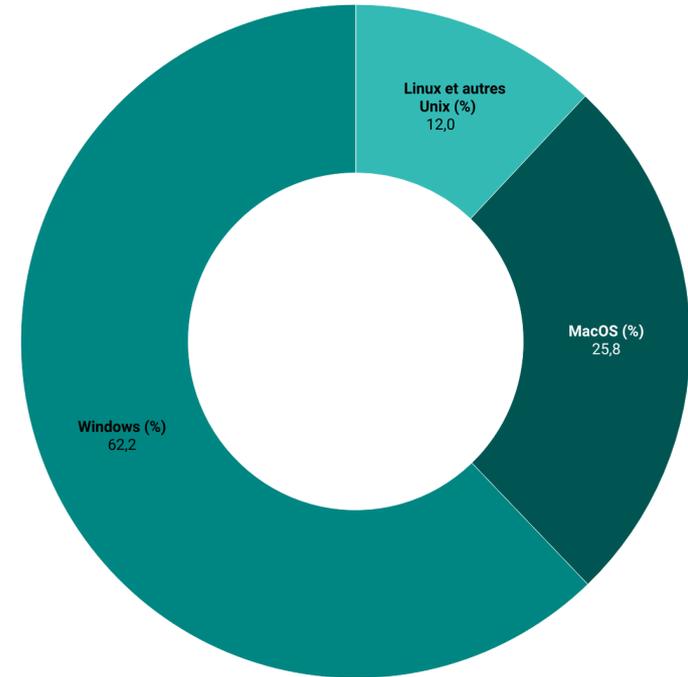
Autres « outils » cités :)

« Mon cerveau, des stylos et du papier. »

Quel(s) système(s) d'exploitation (OS) utilisez-vous principalement?

Réponse à choix multiples.

■ Linux et autres Unix (%) ■ MacOS (%) ■ Windows (%)



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

Porter attention aux interfaces

Résultats phases 1 et 2 qualitatives

Phase 1 et 2 de l'étude : retour sur l'observation des pratiques (entretiens)

Différentes interfaces graphiques visuelles utilisées :

- Sciences de la vie/santé : GraphPad Prism, Excel, Tableau
- Sciences Humaines : Filemaker, SIG web, Excel

Pour quelles raisons ? **Un besoin de visualiser « rapidement »** pour analyser/interpréter de manière dynamique et collective :

- Aide au classement et à l'identification de catégories, aide à la prise de décision
- Aide au travail collectif : extraire des graphiques ou des éléments (de la base) des données, pour partager à l'équipe et analyser de manière collective

Un rôle majeur des interfaces pour à la fois :

- Filtrer des informations, visualiser de manière dynamique, manipuler aisément les données, faire des essais de configuration
- Créer une continuité entre différents acteurs avec des couches d'accès différenciés

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.).

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

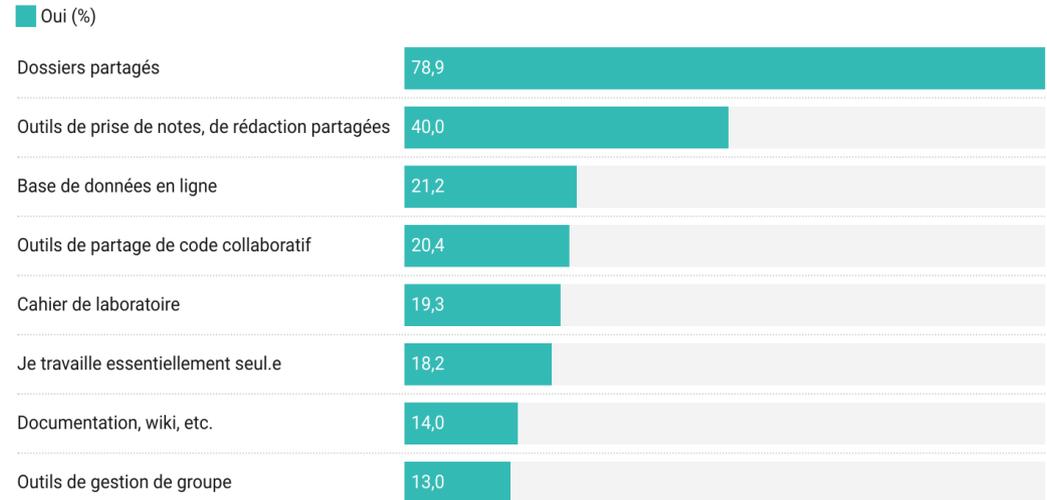
Former **aux pratiques computationnelles** sans toutefois vouloir faire de tous les acteurs de la recherche des *data scientists*. Donner un bagage suffisant pour être capable d'échanger et de comprendre les enjeux.

Place des pratiques et des outils collaboratifs (1)

- L'utilisation **de dossiers partagés** est à la base des pratiques collaboratives
- Des outils **de prises de notes partagées** sont employés par **40 % des personnes** ayant répondu au questionnaire
- Une utilisation d'outils non institutionnels fréquente (DropBox, GDrive, etc.)

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

Réponse à choix multiples.



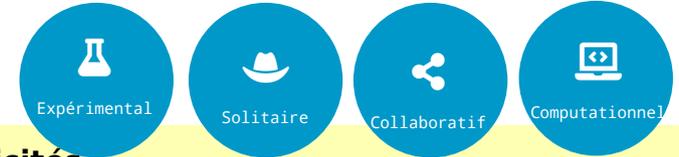
Graphique: Groupe de travail "Réussir l'appropriation de la science ouverte" • Source: Questionnaire "données et science ouverte" • Créé avec Datawrapper

Place des pratiques et des outils collaboratifs (2)

« [L'utilisation de bases de données en ligne] **permet de suivre le travail de chacun et de repérer les difficultés.** »

« Les outils collaboratifs en ligne (rédaction collective ...) **mis à disposition par mon université sont insuffisant[s]** (notamment peu ergonomique[s]) par **rapport à la qualité d'outils de plateformes commerciales** (ex: Google Doc). »

« Il me semble **important de revoir la culture du labo en terme de partage d'informations**, de communication et de transparence au sein d'une même équipe avant de penser au partage de ses données avec le reste de la communauté scientifique. »



Focus sur les profils - spécificités

Les profils sont associés à l'utilisation spécifique de certains outils (sur-représentation) :

- expérimental : cahier de laboratoire
- computationnel : outil de partage de code, documentation, wiki
- SHS collaboratif : bases de données en ligne
- SHS isolé : travail seul

4

PISTES D'ORIENTATION

Diversifier les types
d'accompagnement

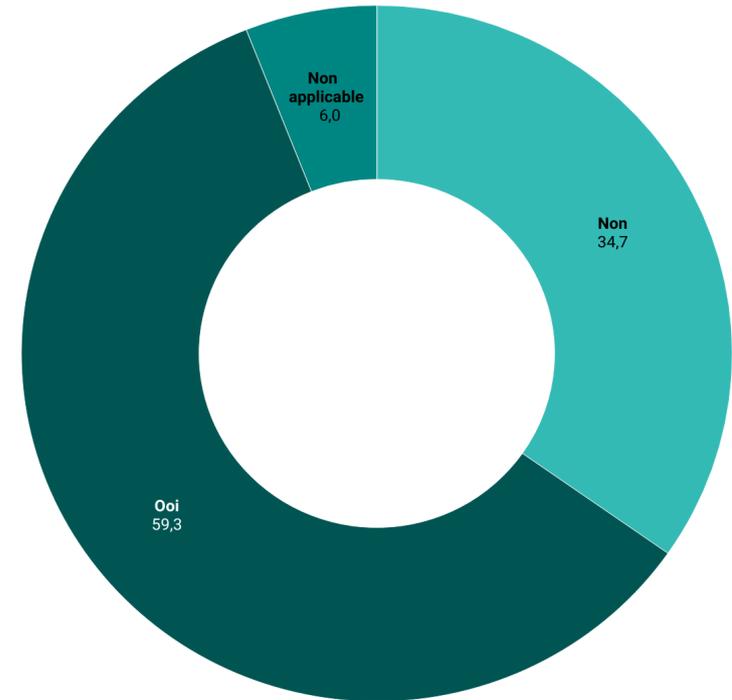
Connaissances des aides disponibles et satisfaction générale (1)

- Une majorité de personnes interrogées ont **connaissance d'aides disponibles** (plus de 55%)
- **Une satisfaction générale vis-à-vis des aides proposées** (plus de 60% pour les satisfaits très satisfaits) pour les personnes en ayant bénéficié
- Les **aides disponibles mentionnées en commentaires sont variées** et vont de l'aide des « collègues spécialisés », de « l'informaticien du labo », des services informatiques (avec une satisfaction parfois faible) jusqu'à l'existence de services « data » dédiés (à différentes échelles : GIR, etc.)
- Mention des DPO en commentaires pour la partie concernant les données personnelles

Connaissez-vous des personnes et services soutiens pour vous accompagner sur vos pratiques liées aux données au sein de votre structure

Réponse à choix multiples.

■ Non ■ Ooi ■ Non applicable



Connaissances des aides disponibles et satisfaction générale (2)



J'évolue dans un IFRE, où **un tel service ne peut être offert du fait de la petitesse de la structure**. En revanche, je peux m'appuyer sur des personnes et services proposés par le CNRS, ou m'enquérir auprès de membres des TGIR (comme Progédo). »



En général, et désolé de le dire comme ça, les services centraux de mon université me mettent plutôt des bâtons dans les roues qu'autre chose. »



Focus sur les profils - spécificités

Au sein des SHS, les aides apportées par les MSH, les TGIR (HumaNum), des réseaux (Mate-SHS, Visa), des projets tels que PROGEDO, DataBFC

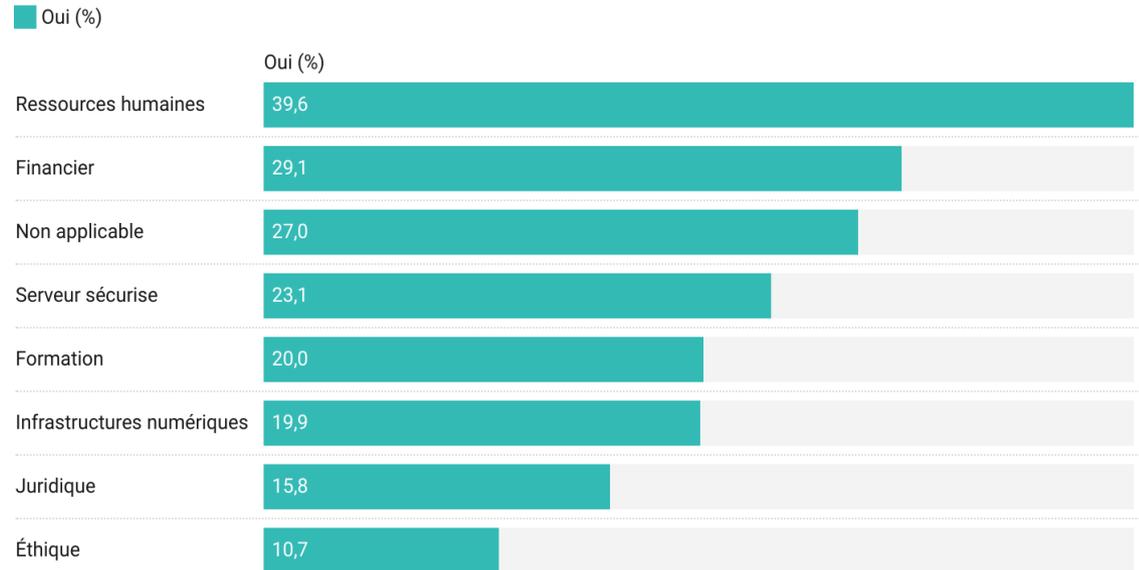


Besoins d'accompagnement (1)

- **Les besoins en ressources humaines et financières** sont prépondérants
- Les accompagnements nécessaires associés aux données concernent en premier lieu **les étapes de stockage et de conservation des données (48 %), et leur mise à disposition (40 %)**

De quels soutiens avez-vous le plus besoin?

Réponse à choix multiples.



Besoins d'accompagnement (2)



Les laboratoires de SHS **n'ont pas assez d'ingénieurs informatiques pérennes**, en particulier BDD; leur recrutement exclusif via des financements sur projet est un danger pour la pérennité des données. »



À la réflexion j'aurais aussi aimé que **l'aide juridico-éthique autour des données ait été plus facile** à trouver, mais ce point s'est beaucoup amélioré. »

- Les commentaires mentionnent le besoin de postes fixes ou leur renouvellement ainsi que des moyens pour accéder à des bases de données ou mener des projets de numérisation.
- D'autres abordent un nécessaire éclaircissement sur la propriété des données tout autant que les enjeux de données personnelles

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



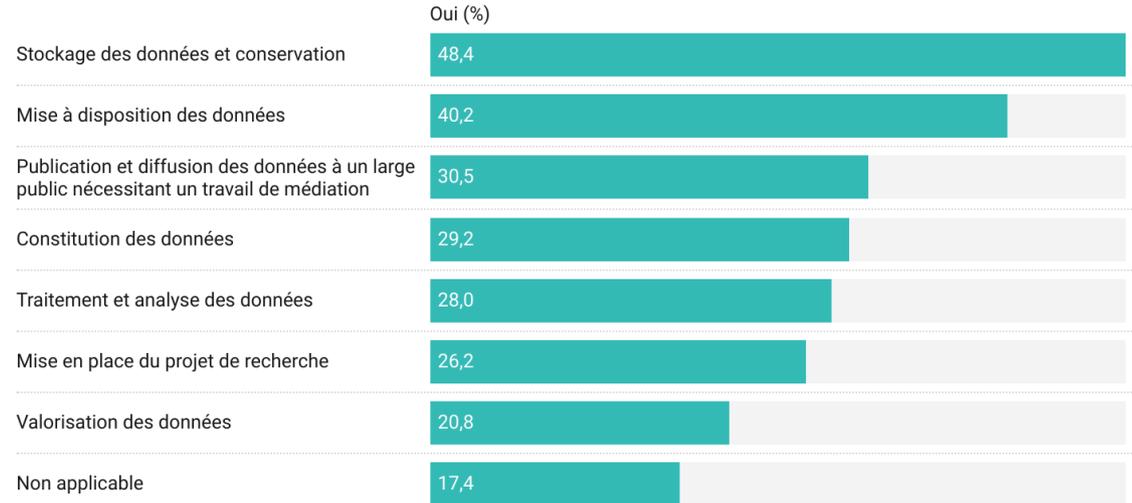
Besoins d'accompagnement (3)

- L'aide à la publication et à la diffusion des données à un large public est aussi une demande plébiscitée notamment au sein des SHS
- Les enjeux de valorisation ne représentent qu'une faible partie des besoins recensés (21 %)

À quelles étapes associées aux données souhaiteriez-vous recevoir de l'aide?

Réponse à choix multiples.

Oui (%)



Focus sur les profils - spécificités

- La question de l'aide à la diffusion à un grand public sous forme de vidéo, billets de blog constitue un enjeu important pour les SHS et qui n'est souvent pas pris en charge dans les budgets des projets de recherche.

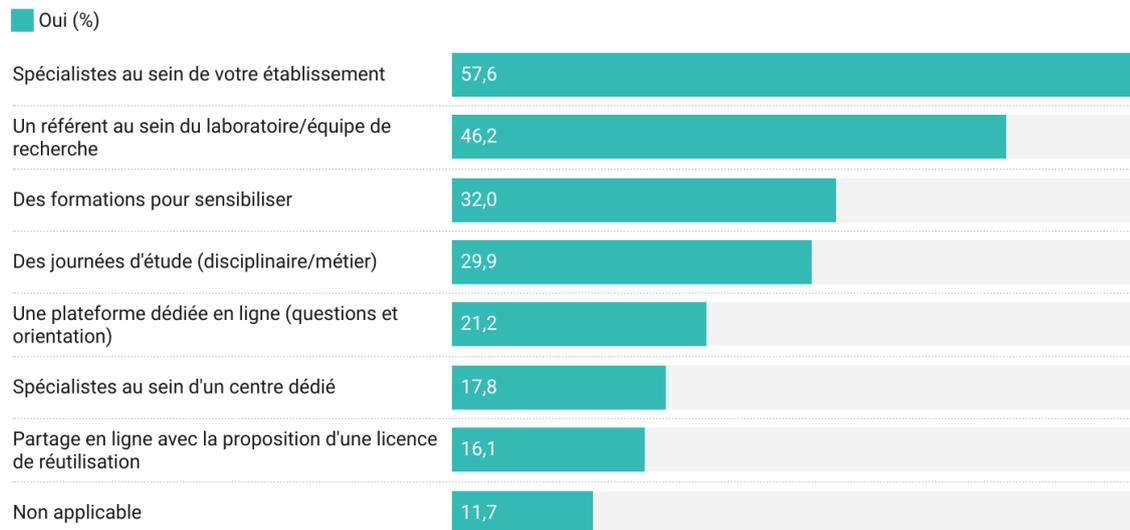


Un maillage des données au plus près des équipes (1)

- La **présence de spécialistes au sein d'établissement (58 %)** et de **référénts au sein du laboratoire/équipe de recherche (46 %)** sont les modalités d'accompagnement ayant les % les plus élevés
- La demande de plateforme dédiée en ligne ne représente que 21,2 % des réponses
- **Des relais stables et pérennes au sein des équipes sont plébiscités** (méfiance quant à l'ajout d'une étiquette « référent » dans les fonctions des personnes déjà en poste : IR/IE)

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

Réponse à choix multiples.



Un maillage des données au plus près des équipes (2)

« L'idée de référent est extrêmement mauvaise. Il y a déjà des référents pour à peu près toutes les tâches pour lesquelles les tutelles refusent de recruter du personnel. On ne peut pas charger la barque indéfiniment, il faut laisser du temps disponible pour faire les tâches pour lesquelles nous avons été recrutés au départ. »

Faciliter un « maillage des dispositifs d'accompagnement aux données » à différentes échelles **en diversifiant leurs modalités par** 1. le développement et maintien d'infrastructures, 2. l'acculturation au sein des équipes de recherche par le biais **de personnes soutien/relais dans le quotidien des équipes** (rôle de médiation, compréhension des besoins et de la culture du laboratoire/de l'équipe).

Faire attention aux **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 .



Résultats
phases 1 et 2
qualitatives

Considérer la culture des équipes de recherche pour faciliter un travail collectif sur les données (résultats des phases qualitatives)

La culture de l'équipe de recherche/du laboratoire :

- **Un apprentissage des pratiques et des outils au cœur des équipes de recherche** : connaissances tacites par la transmission des « savoir faire » (chercheurs - étudiants, réseau informel).
- **Des pratiques souvent ancrées et difficiles à changer** (reproduction des pratiques des « seniors », hiérarchie, etc.)
- **De nouvelles compétences *data science* néanmoins demandées et reconnues** (personne soutien en bio-info, compétences spécifiques pour les bases de données, formation nécessaires)

Des évolutions menant à des **tensions possibles** entre :

- Les ingénieur.e.s d'étude/ de recherche et les chercheur.e.s (aide à la standardisation, à la création de base de données)
- Les responsables de la recherche clinique et les autres chercheurs (sentiment de propriété sur les données, règles considérées comme restrictives, etc.)
- Les équipes de recherche et les DSI

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.



5

PISTES D'ORIENTATION

Prendre en considération le statut et les enjeux de carrière

Un travail solitaire : forcé ou souhaité ? (1)

Profil SHS solitaire : pour quelles raisons ?

- **Enseignant.e-chercheur.e / Chercheur.e. permanent.e** : figure de l'intellectuel indépendant (revendication de la liberté académique), critique de la bureaucratisation et de l'injonction à l'ouverture
- **Doctorant.e** : démarche solitaire, intéressé.e par le partage mais loin d'être la priorité
- **Post-doctorant.e** : recherche avant tout un poste avant de partager (critique de la précarité)
- **Ingénieur.e d'étude/de recherche** : isolement pas d'aide à disposition ou frictions au sein de collectifs de recherche reconfigurés (croisement des savoirs pratiques d'ingénierie et des connaissances du ou de la chercheuse, mise en application de politiques de recherche nécessitant des changements de pratiques, etc.)

Un travail solitaire : forcé ou souhaité ? (2)



Tout cela existe déjà, **il faudrait surtout que j'ai la perspective d'avoir un poste stable en sortant de doctorat** où j'aie la liberté de poursuivre mes différents axes de recherche et la production d'outils.

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.



Pour casser les résistances de certains collègues réticents au partage des données il faut :

1. **Que leur mise à disposition soit reconnue comme des publications qui puissent aider à la progression des carrières.** Que les auteurs de ces données ne puissent pas être pill[és].
2. Il faut des moyens humains, financiers et juridiques car cela représente un travail supplémentaire. »

Étude exploratoire sur
la « recherche sur la
recherche » :
acteurs et approches

Rapport final

4

Conclusion

Conclusion

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

Si les disciplines sont un facteur clef de différenciation de pratiques (notamment SHS/STM), l'enquête a permis de distinguer différents facteurs clefs :

- Les démarches de production de savoirs avec des plus-values distinctes (sources employées, critère de scientificité, méthodologie, rapport aux données, etc.)
- L'environnement (numérique) de travail et ses pratiques (outils, communautés d'apprentissage, etc.)
- Le statut socio-professionnel au sein du système de la recherche (statut, fonction, ancienneté, etc.)



Disciplines



Démarches de recherche
(épistémo., méthodo.)



Communautés de pratiques et
d'apprentissage



Statut et catégorie socio-
professionnelle

Conclusion

Comment **accompagner l'évolution des pratiques associées aux données** en lien avec les incitations / obligations portées **par les politiques publiques « science ouverte » ?**

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

Étude exploratoire sur
la « recherche sur la
recherche » :
acteurs et approches

Rapport final



Annexes

1 ANNEXES

Points d'attention

Points d'attention (1)

- 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.
- 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.).
- 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.

Au sein du rapport, les points d'attention issus des phases qualitatives sont encadrés en orange et ceux de la phase quantitative en bleu.

Points d'attention (2)

- 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.
- 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.
- 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.
- 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.

Au sein du rapport, les points d'attention issus des phases qualitatives sont encadrés en orange et ceux de la phase quantitative en bleu.

Points d'attention (3)

- 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.
- 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.).
- 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.

Au sein du rapport, les points d'attention issus des phases qualitatives sont encadrés en orange et ceux de la phase quantitative en bleu.

Points d'attention (4)

- 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 médiation des données à des publics variés** sous différents formats tout autant qu'aux enjeux juridiques et éthiques.
- Point d'attention 17: **Faciliter un « maillage des dispositifs d'accompagnement des 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 aux 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 .
- 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..
- 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.

Au sein du rapport, les points d'attention issus des phases qualitatives sont encadrés en orange et ceux de la phase quantitative en bleu.

Résumé des points d'attention

Lobbying	Enjeux de la mise a disposition des données dans les comités éditoriaux
Frein	Friction entre approche technico-industrielle vs. artisanale/creative
	Indépendance et liberté des chercheurs
	Ajout d'une fonction supplémentaire de « référent.e données »
Accélérateur	Utilisation de termes plus précis que celui de « données/data »
	Création de postes pérennes
	Sensibilisation à la transparence plutôt que la reproductibilité (qualité en recherche)
	Données partagées si formats interopérables et connus (outils, standardisation, protocoles, doc)
	Formation : aux pratiques computationnelles
	Sensibilisation : éthique *, SO**
	Médiation SO en prenant en compte les spécificités des « publics visés »
Définition de l'objet partagé	Mesures, corpus, matériaux
	Code source
	Cahiers de labo
	Protocoles

Mise en œuvre	Pour quel public ?
	Modalités de partage (juridique financier et technique)
	Accompagnement (poste perenne)
	Investissements à chaque étape de « mise en données »
	Plus-values créées
	Répercussions en termes d'évaluation
Traitement analyse	Vigilance sur les interfaces « boites noires » clefs en main
Maillage	Développement et maintien des infrastructures
	Sécurité et fiabilité des infrastructures
	Personnes relai/soutien dans les équipes et au sein des instituts (poste perenne)

* aux problématiques de valeurs scientifiques et éthiques (intégrité, honnêteté) et impact de la recherche (sociales, économiques, techniques)

** en prenant en considération les spécificités de rencontres et d'apprentissage des communautés (vie de labo, journée d'études, meetings internationaux)

Reformulation des points d'attention pour les ressources humaines

Formation	Méthode computationnelle
Sensibilisation (dans l'ESR)	Éthique
	SO (par discipline)
Méditation	à des publics variés (ESR non ESR)
Accompagnements	juridique
	financier
	technique
	numérique
Référents	SO labo
	SO central

2

ANNEXES

Précisions méthodologiques et bibliographiques

Précisions phases 1 et 2



Méthodologie axée science ouverte :

- Bibliographie sur le logiciel libre Zotero
- Analyse par théorisation ancrée (étapes de codage réalisées avec le logiciel de *mindmap* open source Freeplane)
- Journal de bord et mémo d'analyse sur Gitlab
- Documentation et compte-rendus sur pad collaboratif

Réalisation d'entretiens exploratoires

- Durée d'une heure environ par téléphone ou visio-conférence
- 4 entretiens avec différents profils en recherche : ingénieure de recherche, Professeur.e.s des universités (praticien.ne.s hospitalier), chargé de recherche clinique. Toutes et tous ayant pour objet d'étude les « virus à ARN »
- 2 entretiens avec un professeur des universités et un ingénieur de recherche en archéologie
- Transcription des entretiens sur la base des enregistrement

Réunions et *Focus group* avec le groupe de travail

- Échange, discussion et partages de pratiques tout au long de l'étude
- Sessions *focus group* de discussion et de retours d'expérience (1h30) sur la base de la présentation des résultats (intermédiaires et finaux)

Réalisation d'entretiens d'observation des pratiques

- Durée d'une heure environ par visio-conférence
- 3 entretiens avec un maître de conférences, une post-doctorante, un doctorant. Toute et tous travaillant sur les « virus à ARN »
- Envoi d'un article ou de données en amont servant de fil directeur pour la présentation de l'environnement de travail (partage d'écran)
- Transcription des entretiens sur la base des enregistrements

Journée d'études « du terrain à la mise en données en SHS »

- 20 participant.e.s, 12 présentations réparties sur 4 sessions
- Réalisation d'un compte-rendu sous forme de présentation relue et validée par les participant.e.s

Précisions phases 3 et 4



Méthodologie axée science ouverte :

- Utilisation du logiciel libre Limesurvey pour la réalisation du questionnaire
- Emploi de Libre Office Calc, de R et de datawrapper pour le traitement, l'analyse et la visualisation des données
- Documentation du contenu et des analyses sur pad collaboratif

Exportation des résultats (.csv) du questionnaire réalisé sur *limesurvey*.

- réponses complètes seulement (429 réponses)

Simplification/préparation des données

- sous *Libre Office Calc* avec détail en annexes
- pondération des données pour être représentatif de la répartition des disciplines en ESR

Analyse statistique descriptive univariée avec R

- tris à plat et tris croisés

Analyse statistique multivariée ACM avec R

- ACM puis HCPC selon deux critères (pratiques “données” et contexte de recherche)

Présentation des résultats (visualisation et personæ)

- diagramme, heatmap sur *Datawrapper*
- Méthodes des personæ

Concepts théoriques et ressources bibliographiques

- **Mise en (base de) données, travail des données, *data journey*, datafication** : des concepts issus des travaux en STS, SIC, anthropologie des connaissances
- **Les données : un construit.** Une « donnée » a une valeur, une finalité pour les personnes qui les manipulent. Une information peut-être considérée comme donnée (ou non) en fonction du projet de recherche (ex. images dans une publication en archéologie)
- **La mise en données : un ensemble d'étapes** qui permet à partir de sources/ressources de produire des jeux de données/bases de données sur lesquels s'appuie une recherche/étude (critère de scientificité) avec la production de visualisation et d'une « narration » sur les données
- Il existe différentes étapes dans la mise en donnée avec la production d'artefacts (version intermédiaire) qui **tendent souvent à être invisibilisés** (ainsi que les personnes y participant).

Ressources bibliographiques sur [Zotero](#)
concernant les concepts théoriques employés et les recherches bibliographiques menées

3

ANNEXES

Remerciements, groupe de
travail impliqué

Remerciements

Nous tenons à remercier :



- L'ensemble des personnes ayant consacré du temps et montré de l'intérêt pour cette étude que ce soit en participant à la journée d'étude « du terrain à la mise en données en SHS », aux entretiens qualitatifs ou en répondant au questionnaire « données et science ouverte »



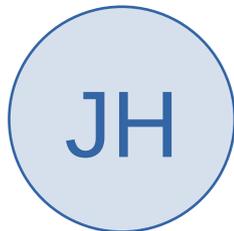
- L'Université Paris Cité : Sarah Pauloin, Magalie Moysan et Benjamin Rullier pour leur aide concernant la rédaction des consentements et du plan de gestion de données

- Le comité pour la science ouverte : les membres du SPSO et du collège « données de la recherche » notamment les coordinat.eur.rice.s Véronique Stoll et Pierre-Yves Arnould



- Inno³ : Maya Anderson-Gonzalez pour son aide lors de la journée d'étude, Ferial Saada pour les transcriptions, Benjamin Jean pour le suivi de la mission, Romain Rouyer pour le design des résultats et Tamara Glushetckaia pour l'éditorialisation des contenus.

Un groupe de travail multi-disciplinaire/ professionnel



Juliette Hueber

Ingénieure de recherche
InVisu (InHA/CNRS)
#Histoire de
l'art#Documentation#Édition



Anne VANET

Pilote du GT, Professeure des
Université, Vice-présidente
Science ouverte Université Paris
Cité
#Bio-informatique



Claire Lemerrier

Directrice de recherche
CNRS au CSO SciencePo
#Histoire



Marie HERBET

Conservatrice des
bibliothèques à l'Université
Lyon 1
#Documentation #Chimie



Célya Gruson- Daniel

Praticienne-chercheuse
(Inno³/UTC)
#STS



Hélène CHAMBEFORT

Responsable des Archives à
l'INSERM
#Archive #Santé

Mise à disposition des ressources et des livrables

- 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

Les contenus (texte et graphiques) sont mis à disposition sous licence Creative Commons CC BY 4.0.
Les photos des personae sont sous licence Unsplash.

4

ANNEXES

Trame d'entretiens phase 1 et 2

Phase 1 : entretiens exploratoires (1/4)

Déroulement possible (avec possibilité d'adaptation pour chaque participant.e)

Canevas général

Demandez à chacun.e de s'appuyer sur des exemples concrets et proposer de montrer son écran si la personne le souhaite

	Pratiques de recherche usuelles	Ouverture et partage de contenus	Évolutions récentes (politiques SO, covid19)
Production de savoirs (support, outils, process et leur représentation)	<ul style="list-style-type: none">- Quelles sont les différentes étapes (aller-retours) d'un projet de recherche (où se situe le démarrage la fin d'un projet de recherche) pour vous ?- Quels sont les différentes informations, objets, matériaux, supports employés manipulés produits au cours d'un projet de recherche ?- Que devient le projet lorsqu'il est fini (préservation, conservation, réutilisation) ?	<ul style="list-style-type: none">- Qu'est-ce que signifie « mettre à disposition » pour vous, partager ouvrir vos travaux ? À quelle étape cela peut-il arriver (en ne considérant pas seulement les publications) ?- Utilisez-vous des informations, données partagées par d'autres ? Aimerez-vous pouvoir le faire ?	<ul style="list-style-type: none">- Est-ce que les derniers événements (covid- 19) ou bien d'autres informations (science ouverte) ont fait évoluer vos pratiques ?- Avez-vous entendu parler de plateformes, de nouveaux outils/projets que vous souhaiteriez utiliser ?- Quels seraient vos intérêts à les utiliser, à vous former ?

Phase 1 : entretiens exploratoires (2/4)

	Pratiques de recherche usuelles	Ouverture et partage de contenus	Évolutions récentes (politiques SO, covid19)
Production de savoirs (support, outils, process et leur représentation)	<ul style="list-style-type: none">- Qu'est-ce que vous considérez comme des « données » ? En existe-t-il des différentes ? Pourquoi une information ou un matériel devient donnée ? Qu'est-ce qui statue cela ?- Quels sont les outils, logiciels informatiques, formats, standards employés communément (propriétaire/Ouvert) ?- Quels sont les passages de support papier à des supports numériques ? (circulation) ? Comment cela se passe-t-il (difficultés/avantages) ?- Que devient l'ensemble de ces informations/données, etc. à la fin d'un projet ?	<ul style="list-style-type: none">- Qu'est-ce qui est de l'ordre du partage entre équipe ? Qu'est-ce qui est mis à disposition à d'autres personnes (association, entreprises, écoles, etc.) ? Comment cela se concrétise-t-il ? Quels outils sont employés ? Quels processus avant de mettre à disposition, d'ouvrir partager (lissage, transformation, retravail des données) ?- Qu'est-ce qui est de l'ordre du partage entre équipe ? Qu'est-ce qui est mis à disposition à d'autres personnes (association, entreprises, écoles, etc.) ?- Comment cela se concrétise-t-il ? Quels outils sont employés ?- Quels processus avant de mettre à disposition, d'ouvrir partager (lissage, transformation, retravail des données) ?	

Phase 1 : entretiens exploratoires (3/4)

	Pratiques de recherche usuelles	Ouverture et partage de contenus	Évolutions récentes (politiques SO, covid19)
Social (collectif, interaction, communautés, organisation, gestion)	<ul style="list-style-type: none">- Comment s'organise le travail au sein du laboratoire sur les matériaux que vous collectez, les données (gestion particulière) ?- Est-ce un travail solitaire en équipe ? Y a-t-il des étapes plus solitaires que d'autres ?- Comment s'organise le travail à plusieurs au sein de l'équipe ? Quels outils sont employés pour échanger ? Y a-t-il des guides, des documents de références ?- Y a-t-il des lieux moments opportuns pour des échanges, du partage de bonnes pratiques ?- Avez-vous des recommandations, des exigences auxquelles répondre pour votre institut ? <p>Comment faites- vous à l'arrivée ou au départ de quelqu'un ?</p>	<ul style="list-style-type: none">- Pour qui partagez- vous, mettez à disposition vos travaux ? Aimerez- vous partager ces informations ?- Qu'est-ce qui est de l'ordre du partage entre équipe ? Ou bien à d'autres personnes ?- Est-ce que cela change la nature ou la manière de partager/ouvrir ?- Est-ce que cela nécessite des étapes de restructuration différentes supplémentaires- Qui s'en chargent ? Et comment (infrastructures, etc.) ?- Quels sont les bénéfices ? Quels risques/craintes à partager/mettre à disposition ?	<ul style="list-style-type: none">- Y a-t-il des formations, un partage d'info, des mesures prises au sein du labo, de l'institut ?- Avez-vous des retours d'expérience ? Des actions en cours ou futures ?- En discutez-vous au sein de l'équipe ? Y a-t-il des personnes qui s'y impliquent plus particulièrement ?

Phase 1 : entretiens exploratoires (4/4)

	Pratiques de recherche usuelles	Ouverture et partage de contenus	Évolutions récentes (politiques SO, covid19)
Régime des savoirs (rapport propriété, science société, conceptions sous-jacentes)	<ul style="list-style-type: none">- Y a-t-il des différences majeures dans la gestion des projets en fonction des types de recherche, du financement, de la valorisation de la recherche, du niveau de confidentialité ?- Comment disposez- vous des données avant, pendant, un projet et pendant ? Êtes-vous propriétaires des données ?- Quelles sont les obligations (stockage, archivage, conservation, partage) ?	<ul style="list-style-type: none">- Qui prend les décisions/statue sur ce qui peut être ouvert/mis à disposition ?- Quelles sont les contraintes (financières, techniques, juridiques, confidentialité, etc.) de cette mise à disposition, ouverture ?- À qui appartiennent les données lorsqu'elles sont partagées ?	<ul style="list-style-type: none">- Qu'est-ce que ces changements peuvent apporter pour votre recherche, votre équipe, d'autres sphères professionnelles, sociales ?- Quels risques, craintes pour le futur ?

Phase 2 : entretiens complémentaires d'observation des pratiques « virus à ARN » (1/3)

Déroulement possible (avec possibilité d'adaptation pour chaque participant·e)

1 Mise en données

1.1 Définition “données” et étapes de « mise en données »

Quelles sont les étapes successives avant d'arriver aux “résultats” présentés dans une publication ou bien à des données publiées/partagées ?

Qu'appelleriez-vous « données » dans votre pratique de recherche ?

Qu'est-ce qu'évoque la notion de données brutes, de données “propres” de nettoyage des données ?

1.2 Mise à disposition et publication

Avez-vous l'expérience de mise à disposition de corpus, de données ? Pour quels publics ?

Lesquelles de ces données devraient, selon vous, être rendues partageables (à supposer que les ressources pour cela existent) ? Quelles sont les motivations ou freins au partage ? •

Sous quels formats avez vous partager ces informations (site web, annexe de thèse, etc.) ?

1.3 Organisation du travail de « mise en données »

Comment s'organise cette “mise en données” ? Est-ce qu'il s'agit d'un projet collectif ou bien solitaire ?

Avez-vous des expériences de travail collectif pour la constitution de base de données, leurs mises à disposition à différents publics ? Quand s'insère le travail collectif ?

Quelles étapes sont faites de façon collective ? Quelles ont été les aides requises ?

Phase 2 : entretiens complémentaires d'observation des pratiques « virus à ARN » (2/3)

Déroulement possible (avec possibilité d'adaptation pour chaque participant.e)

2 Rapport aux outils, évolutions des pratiques et des usages

2.1 Outils : pratiques individuelles et collectives

- Quels sont les différents outils employés (outils papier, SIG, langages de programmation, Zotero, Nakala, tableur, FileMaker, etc.) au long de votre recherche pour une « mise en données » ?
- Quelles aides et quels outils utilisez-vous pour faire sens (intelligibilité) des informations pour vous et pour d'autres personnes ? Quels outils sont employés pour travailler collectivement ?

2.2 Pratiques et évolutions

- Quels sont les « petits trucs », les astuces découvertes au fil des recherches ? Quelles sont les difficultés rencontrées ?
- Vos pratiques avec ces outils ont-ils évolué au fur et à mesure d'une recherche ou de votre parcours personnel ? Avez-vous été confronté.e.s à des changements/découvertes de nouveaux outils ? Pour quelles raisons ?
- Y a-t-il un rapport spécifique aux personnes qui ont créé les outils ?

2.3 Traces de la recherche

- Que deviennent les fichiers intermédiaires de vos travaux ?
- Qu'est-ce que deviennent ces informations à la fin d'un projet ? Comment les conservez-vous ? Comment les retrouvez-vous si besoin ?

Phase 2 : entretiens complémentaires d'observation des pratiques « virus à ARN » (3/3)

Déroulement possible (avec possibilité d'adaptation pour chaque participant·e)

3 « Maillage des données » et science ouverte : acculturation et réseau socio-technique

3.1 Science ouverte, gestion des données : acculturation

- À l'heure de la science ouverte, comment vos pratiques et usages s'appliquent par exemple à la demande d'un plan de gestion de données ?
- Quels seraient vos besoins en termes de formations, d'informations ?
- Où trouvez-vous les informations nécessaires pour des problématiques méthodologiques, l'utilisation d'outils spécifiques ? (espace, personnes dédiées)
- Comment avez-vous entendu parler de ces ressources (infrastructures, outils, personnes soutien) ?

3.2 Normes et pratiques

- Quels sont les décalages entre normes de la science ouverte et demande de gestion des données et vos pratiques au quotidien ? Quelles sont les questions que cela vous pose (avantages/difficultés) ?

5

ANNEXES

Plan de gestion de données

Phase 1 : Informations sur le projet : contexte, acteurs, ressources

1 / Description du projet

Acronyme et nom complet du projet	Décliner la Science Ouverte (Decliner-SO)
Principaux bailleur-s / financeur-s	Commandité par l'Université Paris Cité. Dans le cadre des activités du Comité pour la science ouverte soutenues par le Ministère de l'enseignement supérieur et de la recherche.
Nom du programme de recherche	/
Identifiant de l'appel à projet	/
Référence de la convention de financement	/
Date et durée du projet	De mai 2020 à décembre 2021 (20 mois)

Objectifs du projet : L'objectif de cette étude par une enquête auprès de divers professionnel·le·s de la recherche consiste à mieux comprendre les pratiques de recherche au quotidien (observer la recherche « en train de se faire ») en étant au plus proche des personnes produisant les savoirs au sein des laboratoires, unités de recherche, universités. Il s'agit pour cela de suivre notamment la production, la (re)utilisation et la diffusion des données et autres artefacts (cahiers de laboratoire et carnets de terrain), d'en explorer toute la diversité et d'analyser les problématiques soulevées par des démarches de science ouverte. Cette étude vise également à alimenter les travaux et recommandations en cours du Comité pour la science ouverte quant à l'appropriation de la science ouverte au sein de la recherche.

Mots-clés : Science ouverte, données, recherche, open science, mise en données, méthodes mixtes, infrastructures, communautés, inclusivité

2/ Acteurs du projet

Institution coordinatrice du projet (nom, pays) :	Comité pour la science ouverte (Collège « données de la recherche » (France) Université Paris Cité (France)
Responsable scientifique du projet (Unité de rattachement) :	Équipe : Anne Vanet (vice-présidente numérique et science ouverte à l'Université de Paris) Hélène CHAMBEFORT (INSERM) Marie HERBET (Université Lyon 1) Juliette HUEBER (InVisu CNRS/INHA) Claire LEMERCIER (CNRS, CSO SciencePo) Célya GRUSON-DANIEL (Inno ³ /UTC)
Responsables de la gestion des données et de la mise à jour du DMP :	Célya Gruson-Daniel (Inno ³ /UTC)
Autres partenaires (nom, pays, rôle de chacun des partenaires en dehors de l'institution leader) :	Inno ³ (France)

3/ Ressources nécessaires

Besoins Matériels : Acquisition ou installation de matériel spécifiques ou de services (serveurs, machines virtuelles, infrastructures, archivage, etc.) ?	Les serveurs internes d'Inno ³ avec des instances installées de : NextCloud pour le stockage des données. BigBlueButton pour la réalisation des entretiens. Limesurvey pour collecter les données du questionnaire
Besoins humains (recrutements) et de formations	/
Estimation du surcoût attaché à la gestion de données, à leur diffusion et archivage	/
Couverture des besoins financiers	/

Phase 2 : Stockage, partage, protection et diffusion

4/ Informations générales sur les données (1/2)

Méthodes et processus de production des données utilisées au cours du projet	<p>Pour mener à bien cette étude, des méthodes mixtes quali/quantitative ont été employées.</p> <p>Pour la phase qualitative, les informations ont été recueillies par le biais d'entretiens et d'une journée d'étude. Les informations recueillies lors d'entretiens d'une heure environ ont fait l'objet d'un enregistrement audio et vidéo dont l'objectif est d'étudier les données et outils utilisés au cours des travaux de recherche (production, (ré)utilisation, partage, etc.) dans différentes disciplines. Les entretiens ont été retranscrits pour être analysés. La journée d'étude a fait l'objet d'une prise de note, d'enregistrements audio. Dans le cadre des entretiens et des journées d'observation participante, des captures d'écran des environnements numériques de travail ont été également réalisées.</p> <p>Pour la phase quantitative, un questionnaire (25 minutes pour y répondre) consistant en des questions à la fois ouvertes et fermées a été proposé. L'analyse a donné lieu à plusieurs jeux de données et scripts ainsi que la production de résultats sous forme de graphiques et de personæ.</p> <p>Pour plus de détails, voir ci-dessous un <i>workflow</i> de l'étude. (Cf. 1)</p>
Réutilisation de données existantes : source et origine	/
Propriété des données produites et collectées au cours du projet entre les partenaires	Comité pour la science ouverte (Collège « données de la recherche ») (France) Université Paris Cité (France)
Type et nature des données produites	<p>Pour les entretiens : vidéo enregistrement des entretiens, données textuelles des transcriptions d'entretiens, capture d'écran.</p> <p>Pour la journée d'étude : prise de note et synthèses des échanges.</p> <p>Pour le questionnaire : jeux de données issus de Limesurvey et traitement successifs, code source.</p>

4/ Informations générales sur les données (2/2)

Principaux formats de production de traitement ou de conversion des données, outils et logiciels de lectures associés	<p>L'ensemble des données sont numériques.</p> <p>Pour les entretiens et la journée d'étude :</p> <ul style="list-style-type: none">L'enregistrement audio et vidéo des entretiens existe au format.mp3.Les captures d'écran sont présentées au format.png.Les transcriptions sont au format.odt.L'analyse qualitative a été réalisée à l'aide de <i>mindmaps</i> en format. mm.Les prises de notes sont au format.md ou converties en.odt.Le compte-rendu de la journée d'étude est en format.md et pdf. <p>Pour le questionnaire :</p> <ul style="list-style-type: none">Les données sont brutes sous format.csv. Elles ont été traitées avec les langages de programmation Python et R avec la production de scripts et de Jupyter Notebook.Les prises de notes ont été réalisées sous format.md. Les graphiques sont en format.png.L'ensemble des présentations sont en format pdf.
Standards et /ou formats de métadonnées décrivant les données	<p>Nous utilisons des standards lors du dépôt sur hal-lara et recherche.data.gouv. Autres standards/formats employés en fonction des entrepôts : MARCXML, Dublin Core et DataCite Metadata Schema (selon les directives OpenAIRE).</p>

5/ Stockage et partage des données au cours du projet

Stockage : Supports utilisés pour les données (production ou échange)

L'Université Paris Cité a mandaté Inno³ en tant que prestataire externe pour la réalisation de la mission. Inno³ a proposé ses propres serveurs pour stocker et partager les données au cours du projet et ce jusqu'à un an après les dernières publications des résultats de recherche.

Lieux et types d'hébergement des données et modalités de stockage

Les données sont stockées sur un serveur virtuel dédié (Scaleway) qui héberge l'instance Cloudron avec différentes applications disponibles (Limesurvey, O'transcribe, Hedgedoc, Gitlab). L'instance Nextcloud d'Inno³ est hébergé chez Owncube dans un serveur français. Pour les enregistrements vidéo ils sont réalisés sur une instance BigBlueButton et hébergés sur un serveur OVH.

Stockage : types de flux empruntés par les données

Pour le questionnaire :

Les données sont issues de Limesurvey (Cloudron) et un export des données.csv a été fait puis stocké sur le Nextcloud.

Pour les entretiens et la journée d'étude :

Les enregistrements audio proviennent de l'instance BigBlueButton et ont été transféré sur le Nextcloud.

L'instance O'transcribe (Cloudron) a été employée pour faciliter les transcriptions.

L'ensemble des compte-rendus et prises de notes ont été réalisés sur une instance hedgedoc (Cloudron d'Inno³) ou bien sur le Gitlab d'Inno³ (carnet de bord)

Méthode d'accès et de partage des données entre les partenaires

Concernant Nextcloud, une fonction de partage interne est possible avec les membres d'Inno³ ainsi qu'un partage externe avec les membres de l'équipe projet.

Gestion de l'accès et du partage des données

Cf. Méthode d'accès et de partage des données entre les partenaires.

Volume prévisionnel de données produites

160 MB

Règles de nommage et de classement des informations et données produites

AAAA__Description-doc__Decliner-SO_version.

6/ Risques, sécurité et éthique des données

Risques et menaces qui pèsent sur les données	Les risques principaux seraient la perte de données présentes sur Nextcloud, Cloudron, ou BigBlueButton dépendant de serveur français (OVH, Scaleway) (risque d'incendie, d'inondations, etc.). Concernant les entretiens, le partage des entretiens avec une désanonymisation des données représente un risque.
Mesures et solutions mises en œuvre pour parer aux risques (politique de sécurité pour système d'information de l'hébergeur)	Pour ce qui est du Cloudron des <i>backups</i> automatiques sont réalisés. Pour Nextcloud, une historicisation du contenu est effectuée. Il existe un chiffrement de l'ensemble des communications avec les applications hébergées sur des serveurs distants (tout est en https). Les résultats de l'étude ainsi que les données ont fait aussi l'objet d'une sauvegarde par la pilote de la mission « réussir l'appropriation de la science ouverte » sur ses propres serveurs.
Précautions et moyens pour garantir la protection des données notamment en matière de confidentialité	Pour les entretiens, seul le responsable de projet Inno ³ détient la table de correspondance qui permet de faire le lien entre l'identité des personnes interrogées et le numéro aléatoire attribué dans les différents fichiers. Toutes les données présentées dans les scripts et les données du questionnaire ne possèdent aucun élément permettant l'identification des personnes (anonymisation). Les notes d'entretien, comptes-rendus d'entretien, notes d'observation, notes d'analyses et publications sont rendues confidentielles. Un consentement a été signé par les personnes ayant fait l'objet d'une interview ainsi que les personnes ayant répondu au questionnaire. Pour la journée d'étude, l'accord explicite de chaque participant-e a été demandé pour partager leur nom/prénom et un résumé des échanges.
Questions éthiques et dispositions prises pour y répondre	Cf. Précautions et moyens pour garantir la protection des données notamment en matière de confidentialité.
Confidentialité : objet d'échange ou de partage avec de tiers acteurs.	Aucun fichier ayant une identification des participants aux entretiens n'a été partagé à des tiers acteurs.
Sécurité – Confidentialité : détermination des droits d'accès aux données pendant les recherches et modalités d'accès	Pendant l'étude, l'équipe du projet a accès aux données via. NextCloud par des liens de partage envoyé à chaque membre de l'équipe. Un chiffrement de l'ensemble des communications est effectué avec les applications hébergées sur des serveurs distants (tout est en https).
Intégrité – Traçabilité : Mesures de protection mises en œuvre pour le suivi de la production et de l'analyse des données	Les données brutes partagées pour l'analyse avec les membres du projet étaient anonymisées (pas de collecte d'informations personnelles dans le cadre du questionnaire). Ces fichiers ont été partagés nominativement avec les personnes impliquées dans le projet (avec l'objet d'une clause de confidentialité).

7/ Dissémination/PARTAGE et archivage des données (1/3)

Potentiel de réutilisation des données en dehors du projet et publics cibles	<p>Dans le cadre d'une démarche de science ouverte et, nous facilitons la réutilisation des données de l'enquête dans le respect des conditions expliquées ci-dessous (confidentialité, données personnelles, licences, etc.)</p> <p>Pour une reproductibilité des résultats quantitatifs, un partage des données brutes du questionnaire et des scripts utilisées lors de l'analyse est effectué. Pour les informations qualitatives, notre approche vise à documenter notre démarche pour assurer une traçabilité du raisonnement.</p> <p>L'ensemble des données et informations de l'étude peuvent intéresser des chercheurs travaillant sur les terrains de la science ouverte mais aussi les chargé-e-s « science ouverte » ou « gestion des données » au sein des établissements. Ce partage est aussi à destination des personnes ayant participé à l'étude pour qu'il puisse avoir un retour sur les résultats obtenus.</p>
Données vocation à être publiées. Principes de diffusion édictés par le financeur ou autre engagement	<p>Les résultats de cette recherche seront diffusés de façon anonyme dans des colloques professionnels et scientifiques, dans des rapports destinés aux autorités, dans des revues professionnelles et académiques et dans des médias destinés au grand public.</p>
Modalités de partage des données (support)	<p>Les données seront présentées sur plusieurs plateformes telles que HAL, PubPub, GitLab. La plateforme dépendra du format des données.</p> <p>Sur HAL est placé une synthèse en français et en anglais (<i>executive summary</i>) et un rapport final en pdf accompagné d'annexes (trames d'entretiens, trame du questionnaire, plan de gestion des données.</p> <p>Sur Gitlab, seront partagés les scripts ayant servi à la reproductibilité de l'analyse du questionnaire ainsi que la documentation pour présenter l'étude et la méthodologie.</p> <p>La page sur la plateforme PubPub contient le contenu de la synthèse et un référencement de tous les liens vers l'ensemble des ressources : la synthèse, le rapport final et les annexes, les résultats de la phase qualitative sous forme de quelques verbatims des entretiens et une <i>mindmap</i>, la synthèse de la journée d'étude et de la phase quantitative avec les données de questionnaire, la méthodologie de l'étude.</p> <p>Concernant les retranscriptions des entretiens anonymisées, celles-ci ne seront pas diffusées en accès libre, mais elles pourront faire l'objet sur demande d'une communication auprès d'autres équipes de recherche publique et membres du Comité pour la science ouverte.</p> <p>Cf. Fig.2</p>

7/ Dissémination/PARTAGE et archivage des données (2/3)

Détail du recours à un logiciel ou outil spécifique pour accéder aux données (INRA)	Cf. modalités et partage des données
Licences employées pour le partage	<p>Les résultats (rapport final, synthèse, graphiques) ainsi que les transcriptions et notes seront disponibles sous une licence Creative Commons – CC-BY 4.0. Seules les photos employées pour les personae sont sous licence Unsplash.</p> <p>Les données du questionnaire et les scripts associés à leur analyse seront partagées sous licence ouverte 2.0 (Etalab) https://www.etalab.gouv.fr/wp-content/uploads/2017/04/ETALAB-Licence-Ouverte-v2.0.pdf</p>
Temporalité du partage et conservation (à partir de quand, et combien de temps)	Les données ne sont pas conservées en base courante au-delà d'une durée de deux ans à compter de la dernière publication des résultats de la recherche ou, en cas d'absence de publication, jusqu'à la signature du rapport final de la recherche.
Identifiants pérennes pour les données et organismes en charge de la demande	<p>Pour HAL, les DOI seront fournis par CrossRef.</p> <p>Pour Zenodo, des DOI sont associés pour chaque fichier partagé.</p>
Modalités et garanties d'archivage à l'issue du projet (plateformes partenaires, durée de conservation, etc.)	<p>Les services d'Université Paris Cité sont en charge de garantir les modalités d'archivage du projet..</p> <p>Les données font ensuite l'objet d'un archivage au sein d'Université Paris Cité pendant une durée de cinq ans.</p>

7/ Dissémination/PARTAGE et archivage des données (3/3)

HAL

- résumé directif FR et EN
- rapport synthétique sous forme de présentation

Annexes

- trame questionnaire
- trame entretiens
- DMP

Recherche Data Gouv

- données questionnaire

GitLab

- scripts (R/Python)
- documentation dont note méthodologique
- graphiques

PubPub

- Éditorialisation du contenu
- à propos
 - recommandations
 - synthèse
 - méthodo de l'étude

Schéma de stockage des données.

Phase 3 : archivage des données

8 / Tri, archivage et conservation des données (1/2)

Données conservées sur le moyen ou long terme (plus de 10 ans) critères de sélection de réutilisation scientifique, valeur de preuve, Données à détruire	<p>Les données à archiver privilégiées sur une durée de 5 ans sont :</p> <ul style="list-style-type: none">La synthèse en anglais et françaisle rapport sous forme de présentationles transcripts des entretiens,les comptes-rendus de la journée d'étudeles données du questionnaire et les scripts associéesla note méthodologiqueles trames d'entretiens et de questionnaire,le DMP/PGD <p>Les données à détruire sont :</p> <ul style="list-style-type: none">les enregistrements audio et vidéo des entretiens et de la journée d'étudeles consentementsla présentation intermédiaire des résultatsla <i>mindmap</i> présentant quelques résultats de la phase qualitativeles graphiques réalisés.
Durée de conservation de chaque donnée	<p>Les documents seront conservés pendant la durée réglementaire (dépendantes du types de données cfttps://doranum.fr/stockage-archivage/referentiel-de-gestion-des-archives-de-la-recherche_10_13143_pcqd-hy47/documents-a-detruire/).</p> <p>La destruction fera l'objet d'un bordereau d'élimination.</p>
Volume de données à archiver	<p>En totalité cela représente environ 16,5 MB dont voici le détail :</p> <ul style="list-style-type: none">le résumé directif en anglais et français (3MB)le rapport sous forme de présentation (5,2MB)les transcripts des entretiens (450 kB),la synthèse de la journée d'étude (5,5MB),les données du questionnaire (1,1MB)les scripts associés (800kB)la note méthodologique (520kB)les trames d'entretiens (80kB) et de questionnaire (300kB)le DMP (200kB)

8 / Tri, archivage et conservation des données (2/2)

Plateforme d'archivage pérenne choisie ou autres procédures	À ce stade, l'Université Paris Cité ne dispose pas d'une solution d'archivage pérenne des données. Un stockage des données sera effectué sur le serveur dédié à l'archivage géré par le département avec répllication sur un disque dur externe, tout en assurant un plan de préservation « manuel » (migration de supports et conversions de formats réalisés ponctuellement, par exemple tous les 5 ans).
Responsable de la conservation à long terme	Le département des archives d'Université Paris Cité assurera la conservation à long terme des données ayant vocation à être archivées dans le temps long. En parallèle, l'entrepôt qui hébergera les données partagées assurera la conservation pérenne de ces données.
Garantie de financement de la conservation à long terme	L'université de Paris Cité ne garantit pas de financement de la conservation à long terme (par exemple CINES). Les données et documents publiés sur des entrepôts de données pourront bénéficier d'un archivage pérenne (en fonction de la politique des entrepôts).

6

ANNEXES

Trame du questionnaire

Données et science ouverte

Ce questionnaire s'intègre dans l'étude du Comité pour la Science Ouverte (CoSO) « réussir l'appropriation de la science ouverte (<https://www.ouvrirlascience.fr/reussir-lappropriation-de-la-science-ouverte/>) » Son objectif est de mieux comprendre.

1. Les **étapes associées à la production et diffusion des données** (collecte, analyse, mise à disposition, etc.) dans un projet de recherche.
2. Les **différentes personnes** qui y participent.
3. **L'accompagnement nécessaire** à ces pratiques

Cette étude est menée par une équipe de recherche pilotée par l'Université de Paris (Mme Anne Vanet). **Les résultats de cette étude viendront alimenter les travaux ainsi que les recommandations du CoSO et seront publiés sur le site ouvrir la science (<https://www.ouvrirlascience.fr/reussir-lappropriation-de-la-science-ouverte/>).**

Le questionnaire représente la 3ème et dernière phase de l'étude et se nourrit de résultats préliminaires issus d'entretiens dans le domaine particulier de la biologie/santé (recherche clinique/fondamentale traitant des « virus à ARN ») et d'une journée d'étude regroupant différentes disciplines de sciences humaines et sociales (archéologie, histoire, sociologie, etc.). Nous souhaitons continuer avec ce questionnaire **à mieux connaître les pratiques de ces communautés tout autant que de recueillir les retours d'une plus grande diversité de disciplines.**

Ce questionnaire **s'adresse à toute personne au sein de l'ESR** (chercheuses et chercheurs, ingénieur.e.s, technicien.ne.s, documentalistes, post-doctorant.e.s, doctorant.e.s, support et soutien à la recherche, etc.) **amenée à effectuer des activités impliquant des données** : collecte, traitement, gestion, structuration, nettoyage, mise en forme, visualisation, transformation, analyse, modélisation, sauvegarde, archivage, etc.

Répondre au questionnaire vous prendra environ 25 minutes. **La date limite de réponse est fixée au 15 juin 2021.**

Merci pour votre participation

L'équipe de recherche (<https://www.ouvrirlascience.fr/reussir-lappropriation-de-la-science-ouverte/?menu=2>)

Il y a 43 questions dans ce questionnaire.

Partie 1 : Cadre et démarche de recherche

1. Environnement de travail ?

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

👉 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

Laboratoire nécessitant l'emploi de différents instruments de mesures (par exemple travail à « la paillasse », expérimental, etc.)

Cadre clinique/thérapeutique (centres hospitaliers, psychologiques, etc...)

Travail de « terrain » et d'enquête (par exemple observation sur le terrain, réalisation d'entretiens)

Bibliothèque ou archives (numériques également)

Sur un ordinateur pour des travaux d'analyse de données de modélisation (par exemple, en bio-informatique, en épidémiologie, etc...)

Sur un ordinateur ou sur papier pour effectuer différentes tâches de réflexions théoriques, de rédaction

Autre :

2. Le terme de « données » est largement employé aujourd'hui. Y a-t-il d'autres termes (ou des qualificatifs supplémentaires) que vous utilisez pour décrire les éléments que vous manipulez dans votre travail et qui vous servent à l'obtention des résultats de recherche ?

🗳️ Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

🗨️ Ajoutez un commentaire seulement si vous sélectionnez la réponse.

<input type="checkbox"/> Corpus	<input type="text"/>
<input type="checkbox"/> Ressources	<input type="text"/>
<input type="checkbox"/> Valeurs	<input type="text"/>
<input type="checkbox"/> Sources	<input type="text"/>
<input type="checkbox"/> Matériaux	<input type="text"/>
<input type="checkbox"/> Mesures	<input type="text"/>
<input type="checkbox"/> Statistiques	<input type="text"/>
<input type="checkbox"/> Jeux de données/dataset	<input type="text"/>
<input type="checkbox"/> Base de données	<input type="text"/>
<input type="checkbox"/> Pipeline	<input type="text"/>
<input type="checkbox"/> Données brutes	<input type="text"/>
<input type="checkbox"/> Données nettoyées	<input type="text"/>
<input type="checkbox"/> Données négatives	<input type="text"/>
Autre : <input type="text"/>	<input type="text"/>

3. Lesquelles de ces propositions décriraient le mieux la plus-value de votre travail au sein de votre communauté de recherche ?

📌 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

🗨 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

<input type="checkbox"/> Amélioration des techniques de collecte, de mesure et d'analyse	<input type="text"/>
<input type="checkbox"/> Préparation d'échantillons (lames biologiques, semi-conducteurs, produits de synthèse, etc.) et définition des protocoles expérimentaux	<input type="text"/>
<input type="checkbox"/> Standardisation de protocoles pour leur automatisation	<input type="text"/>
<input type="checkbox"/> Collecte de données rares ou nécessitant un temps de collecte important, étude de cas d'étude spécifique	<input type="text"/>
<input type="checkbox"/> Collecte de données standards (sans amélioration de protocoles ou de méthodologies)	<input type="text"/>
<input type="checkbox"/> Création d'une théorisation, d'une interprétation d'un ensemble de faits observés	<input type="text"/>
<input type="checkbox"/> Automatisation d'un protocole/workflow sur une grande quantité de données déjà collectées	<input type="text"/>
<input type="checkbox"/> Création d'artefacts ou de modélisations sur un grand nombre de données	<input type="text"/>
<input type="checkbox"/> Support à la recherche (gestion, valorisation, etc...)	<input type="text"/>
Autre : <input type="text"/>	<input type="text"/>

4. Familiarité avec la notion de reproductibilité en recherche

🔗 Précisez en commentaires si vous le souhaitez vos pratiques/connaissances/opinions à ce sujet.

📌 Veuillez sélectionner une réponse ci-dessous

- Tout à fait
- Assez
- Peu
- Pas du tout

Veuillez saisir votre commentaire ici:

5. Quels mots représentent un ou des critères de qualité en recherche pour vous ?

🔗 Cochez toutes les cases qui s'appliquent.

📌 Cochez la ou les réponses

- Traçabilité
- Réplicabilité
- Reproductibilité
- Explicabilité
- Recherche ouverte
- Open scholarship
- Transparence
- Autre :

Partie 2.1 Pratiques de « mises en données »

Dans l'étude «réussir l'appropriation de la science ouverte», nous souhaitons comprendre en détail l'ensemble des étapes nécessaires à la production de données qui soutiennent les résultats scientifiques d'une recherche. Le terme « mise en données » est ici employé pour décrire les étapes en amont de la mise à disposition des données : des étapes qui sont souvent invisibilisées et/ou le temps qui y est consacré n'est pas forcément comptabilisé. La suite des questions vous permettent de décrire ces actions (vocabulaire employé, temps consacré, etc.)

6. Étapes associées au travail avec des données

6.1 : Quelle part de votre temps de recherche (sur l'année) consacrez-vous aux étapes détaillées ci-dessous?

6.2 : Réalisez-vous ces étapes seule ou à plusieurs ?

Indiquez « sans réponse » aux étapes qui ne correspondent pas à vos pratiques.

	Moins de 10% du temps	11-20%	21-30%	31-50%	Plus de 50%	Seul.e	À plusieurs	Sans réponse
Collecter des données	<input type="radio"/>	<input checked="" type="radio"/>						
Numériser des sources papier	<input type="radio"/>	<input checked="" type="radio"/>						
Constituer un corpus de sources documentaires	<input type="radio"/>	<input checked="" type="radio"/>						
Nettoyer des données	<input type="radio"/>	<input checked="" type="radio"/>						
Structurer une base de données	<input type="radio"/>	<input checked="" type="radio"/>						
Analyser des données	<input type="radio"/>	<input checked="" type="radio"/>						
Modéliser des données	<input type="radio"/>	<input checked="" type="radio"/>						
Visualiser des données	<input type="radio"/>	<input checked="" type="radio"/>						
Qualifier/ décrire vos données (ontologies, métadonnées)	<input type="radio"/>	<input checked="" type="radio"/>						
Développer des interfaces de base de données	<input type="radio"/>	<input checked="" type="radio"/>						
Sauvegarder et/ou archiver vos données	<input type="radio"/>	<input checked="" type="radio"/>						
Construire un plan de gestion de données	<input type="radio"/>	<input checked="" type="radio"/>						
Vérifier les données, leur qualité	<input type="radio"/>	<input checked="" type="radio"/>						
Mette en conformité avec les mesures juridiques et éthiques (RGPD, DPI, etc...)	<input type="radio"/>	<input checked="" type="radio"/>						
Mettre à disposition des données à d'autres professionnels de la recherche	<input type="radio"/>	<input checked="" type="radio"/>						
Mettre à disposition en ligne en accès libre (grand public) des données	<input type="radio"/>	<input checked="" type="radio"/>						
Valoriser les données d'un point de vue économique	<input type="radio"/>	<input checked="" type="radio"/>						
Aider à la compréhension des données (communication, médiation, etc...)	<input type="radio"/>	<input checked="" type="radio"/>						
Former d'autres personnes sur ces différentes étapes (associées aux données)	<input type="radio"/>	<input checked="" type="radio"/>						

7. Sources de vos données

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

📌 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

- Terrain (fouilles archéologiques, entretiens, etc...)
- Instruments de mesure (ex. IRM, PCR, etc...)
- Bases de données disponibles en libre accès
- Bases de données disponibles via un abonnement
- Données/bases de données partagées par des collègues
- Articles scientifiques
- Sites web et réseaux sociaux
- Archives ou fonds documentaires
- Autre :

8. Données et ressources connexes produites dans le cadre de la recherche

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

📌 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

- Des fichiers de données que je peux stocker et analyser sur mon ordinateur
- Des fichiers de données qui nécessitent généralement un stockage central et une installation de calcul à haute performance pour l'analyse
- Des normes pour les données et les métadonnées telles que des terminologies (vocabulaires contrôlés, ontologies, etc.), modèles, schémas et formats d'échange.
- Des scripts de programmation, des workflows
- Des logiciels d'analyse
- Des notes diverses (textes, graphiques, etc.) sous format numérique (cahier de laboratoire, fichiers, wiki)
- Des notes diverses (textes, graphiques, etc.) sous format papier (cahier de labo, carnet de terrain, etc...).
- Autre :

9. Souhaitez-vous ajouter un commentaire général concernant vos pratiques de « mise en données » ?

Partie 2.2 : Mise à disposition de données et réutilisation.

Par la suite, nous employons le terme de « mise à disposition » de données pour inclure à la fois le partage de données à un groupe restreint de personnes jusqu'à l'ouverture des données (open data).

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

📌 Par mise à disposition, nous entendons des pratiques allant du partage de données à un public cible jusqu'à l'ouverture des données en ligne impliquant le choix d'une licence de réutilisation (*open data*).

	Jamais	Parfois	Souvent	Toujours	Non applicable	Sans réponse
Partage restreint à un public ciblé et connu (envoi spécifique, partage off-line, etc.)	<input type="radio"/>	<input checked="" type="radio"/>				
Partage en ligne avec contrôle de l'accès (demande d'accès nécessaire, etc.)	<input type="radio"/>	<input checked="" type="radio"/>				
Partage en ligne sans contrôle de l'accès	<input type="radio"/>	<input checked="" type="radio"/>				
Partage en ligne avec la proposition d'une licence de réutilisation (<i>open data</i>)	<input type="radio"/>	<input checked="" type="radio"/>				

11. Avez-vous déjà eu des obligations en termes de mise à disposition de données de la part :

📌 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

📌 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

D'un comité éditorial pour la publication d'un article dans une revue

D'un comité éthique

D'un financeur de votre projet (UE, ANR, etc.)

D'un partenaire privé ou associatif

Non

Non applicable

12. De quelles manières mettez-vous à disposition ces données ?

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

👉 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

- Par le biais d'un support amovible (disque dur, externe clef usb, etc...)
- Par un système d'envoi de fichier (FileSender, etc...)
- Par mail
- Par le biais d'un serveur sécurisé
- Par le biais de data paper
- Par le biais des plateformes éditeurs en complément d'un article
- Sur un site web
- Par le biais de dossiers partagés cloud non institutionnels (Gdrive, Dropbox, Owncube, etc...)
- Par le biais de dossiers partagés cloud institutionnels (ShareDoc, etc...)
- Par le biais d'entrepôt de données dites « généralistes » (Zenodo, FigShare, etc...)
- Par le biais d'entrepôt de données dites « disciplinaires » (NCBI, Nakala, etc...)
- Par des plateformes de type Gitlab, Github
- Je ne sais pas
- Non applicable

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

👉 Classez jusqu'à 4 éléments du plus important (haut de la liste) au moins important (bas de la liste).

Vous pouvez également indiquer « non applicable ».

Effectuez un double-clic ou glissez/déposez le/les élément(s) de la liste de gauche vers la liste de droite.

👉 Veuillez sélectionner de 1 à 4 réponses.

Vos choix

Un manque d'habitude de cette pratique dans les communautés de recherche (savoir faire, etc.)

Un doute sur la capacité de réutilisation des données par d'autres (contextualisation et documentation nécessaire).

Une volonté de valoriser le travail effectué avant un partage pour garder un avantage compétitif.

La peur d'un renforcement d'inégalités entre des personnes produisant des données d'une part et celles les analysant et en tirant des publications d'autre part.

Une perspective de mise à disposition qui vient limiter, freiner le travail interprétatif possible.

Une peur de mésusage des données par les personnes les réutilisant.

Une mauvaise qualité des données en l'état.

Un temps trop important nécessaire à leur mise à disposition.

Des barrières juridiques et économiques contraignant leurs partage (contrat de consortium, valorisation, exclusivité, droit d'auteur).

Des conditions éthiques limitant leur partage.

Non applicable.

Votre classement

14. Selon vous, si vous n'aviez pas de restrictions à la mise à disposition de données, seraient-elles en l'état potentiellement réutilisables par d'autres ?

📌 Cochez la case avec commentaire libre.

📌 Veuillez sélectionner une réponse ci-dessous

- Oui
- Non
- Je ne sais pas
- Non applicable

Veillez saisir votre commentaire ici:

16. Dans quelle mesure réutilisez-vous des données déjà produites ou publiées ?

📌 Choix unique. Vous pouvez donner un exemple en commentaire.

📌 Veuillez sélectionner une réponse ci-dessous

- Jamais
- Parfois
- Souvent
- Toujours
- Je ne sais pas
- Non applicable

17. Outre les données, lesquels de ces éléments devraient, selon vous, être mis à disposition (à supposer que les ressources pour cela existent) ?

📌 📌 Cochez toutes les cases qui s'appliquent avec un commentaire/exemple possible pour les cases cochées si vous le souhaitez.

📌 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

- Cahier de laboratoires ou notes de terrain
- Logiciels et scripts d'analyse
- Documentation des données, des codes sources
- Metadonnées, thesaurus, vocabulaire contrôlé.
- Protocole méthodologique
- Billets de blog, vidéo, site web, etc. nécessaires à la compréhension des travaux de recherche.

Autre :

17. Outre les données, lesquels de ces éléments devraient, selon vous, être mis à disposition (à supposer que les ressources pour cela existent) ?

👉 Cochez toutes les cases qui s'appliquent avec un commentaire/exemple possible pour les cases cochées si vous le souhaitez.

📝 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

Cahier de laboratoires ou notes de terrain

Logiciels et scripts d'analyse

Documentation des données, des codes sources

Metadonnées, thesaurus, vocabulaire contrôlé.

Protocole méthodologique

Billets de blog, vidéo, site web, etc. nécessaires à la compréhension des travaux de recherche.

Autre :

19. Que deviennent les fichiers intermédiaires de vos travaux (versions intermédiaires d'articles, données non prises en compte dans les résultats de recherche, figures non publiées dans un article, etc.) Comment les retrouvez-vous si besoin ?

👉 Commentaires libres

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

👉 Cochez toutes les cases qui s'appliquent avec un commentaire/exemple possible pour les cases cochées si vous le souhaitez.

📝 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

Sur un ordinateur personnel

Sur un ordinateur professionnel

Sur des supports externes (disques durs externes, clés USB, etc.)

Sur un serveur interne de votre service, laboratoire

Sur un serveur institutionnel de votre structure

Sur un serveur hébergé par une autre structure/institution

Sur un serveur intégré à une infrastructure de recherche

Sur un serveur hébergé par un prestataire privé

Sur un service web d'hébergement et de gestion de développement de logiciels ou une forge

Sur un service d'archivage par le biais de votre structure

Non applicable

Partie 3 : Collectifs de recherche et environnement numérique de travail

20. Habituellement, pour un projet de recherche dans lequel vous êtes impliqué.e, avec combien de personnes travaillez-vous ?

🔒 Choix unique.

📌 Veuillez sélectionner une réponse ci-dessous

- Je travaille seul.e
- Nous sommes entre 2 et 5 personnes
- Nous sommes entre 6 et 10 personnes
- Nous sommes entre 11 et 20 personnes
- Nous sommes entre 21 et 50 personnes
- Nous sommes plus de 50 personnes

21. De qui s'agit-il essentiellement ?

📌 Cochez toutes les cases applicables.

📌 Cochez la ou les réponses

- Permanent de l'enseignement supérieur et la recherche (enseignant.e- chercheur.se, chercheur.se, etc. responsable de laboratoire)
- Non-permanent de l'enseignement supérieur et la recherche (ATER, doctorant.e, post-doctorant.e, etc.)
- Ingénieur.e, technicien.ne., assistant.e de recherche et de formation
- Personnel support des services systèmes d'information (SI)
- Personnel soutien des services de bibliothèques, documentation, archives et musées
- Personnel soutien en affaires juridiques
- Personnel de support des services de valorisation de la recherche
- Personnel de support des services de transformation numérique et science ouverte
- Personnel de support en management, pilotage, contrôle, aide au montage de projets
- Personnel de support en communication ou édition
- Collaboration internationale
- Autre :

22. Quel(s) système(s) d'exploitation informatique (OS) utilisez-vous principalement ?

📌 Veuillez sélectionner une réponse ci-dessous

- Windows
- MacOS
- Linux
- Autres Unix
- autres OS (Android)
- Je ne sais pas

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

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

🗨️ Ajoutez un commentaire seulement si vous sélectionnez la réponse.

Logiciels de type tableurs (Excel, Calc, etc.)

Logiciels de type tableurs scientifiques (Origin, Igor, QtiPlot,...)

Logiciels/plateformes de base de données intégrées (Access, Libre office Base, FileMaker, etc.)

Logiciels d'analyse et de visualisation de données (GraphPadPrism, Gephi, Power BI, QGIS, etc.)

Logiciels de traitement statistiques (Stata, SAS, SPSS, etc.)

Logiciels de traitement d'images (Photoshop, Gimp, etc.)

Logiciels/plateformes de type data science (Weka, workbenchdata, Tableau, etc.)

Solutions basées sur l'usage de langages de programmation (R, python, SQL, Matlab, etc.)

Plateformes d'entrepôts de données

Non applicable

Autre :

23. Comment avez-vous découvert ces outils numériques de travail (questions précédentes) ?

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

🗨️ Ajoutez un commentaire seulement si vous sélectionnez la réponse.

En cherchant par moi-même

Par les autres membres de mon équipe ou d'équipes précédentes

Par d'autres équipes de recherche avec qui je travaille ou ai travaillé

Par mon (ancien) employeur

Via les réseaux sociaux numériques (Twitter...)

Par une liste de diffusion

Au cours d'une formation

Au cours d'une présentation lors d'un colloque, une journée d'étude

Par des temps informels d'échanges avec des collègues (laboratoire, journée d'études, etc...)

Par une veille sur ces questions

Par démarchage commercial

Non applicable

Autre :

24. Comment vous formez-vous habituellement à ces outils ?

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

📌 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

- Par les autres membres de mon équipe.
- Par d'autres équipes avec qui je travaille.
- Au cours d'un atelier lors d'un colloque, une journée d'étude.
- Par des temps informels d'échanges avec des collègues (laboratoire, journée d'études, etc...).
- Par des formations en ligne (MOOC).
- Par des formations en présentiel.
- Par de la formation continue
- Par des tutoriels, documentations diverses en ligne.
- Par des ouvrages « papier »
- Non applicable

Autre :

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

👉 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

📌 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

- Cahier de laboratoire
- Base de données en ligne
- Dossiers partagés
- Outils de prise de notes, de rédaction partagées
- Outils de partage de code collaboratif (Github, Gitlab)
- Documentation, wiki, etc.
- Outils de gestion de projets
- Je travaille essentiellement seule.

Autre :

26. Les évènements associés à la covid-19 ont-ils fait évoluer vos pratiques ? Si oui, comment?

📍 Commentaire libre.

27. Souhaitez-vous ajouter un commentaire général concernant les questions de cette partie « Collectifs de recherche et environnement numérique de travail » ?

📍 Commentaire libre.

Partie 4 : Accompagnement et besoins

28. Connaissez-vous des personnes et services soutiens pour vous accompagner sur vos pratiques liées aux données au sein de votre structure ?

📌 Si oui, indiquez en commentaire les personnes et services concernés.

- Oui
- Non
- Non applicable

Veillez saisir votre commentaire ici:

29. Dans quelle mesure avez-vous bénéficié de ces aides ?

📌 Cochez une seule case ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple.

📌 Veuillez sélectionner une réponse ci-dessous

- Jamais
- Parfois
- Souvent
- Toujours
- Je ne sais pas
- Sans réponse

Veillez saisir votre commentaire ici:

30. Dans quelle mesure ces aides vous conviennent-elles aujourd'hui ?

📌 Cochez une seule case ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple.

- Très satisfait
- Plutôt satisfait
- Ni satisfait, ni insatisfait
- Plutôt insatisfait
- Très insatisfait
- Sans réponse

Veillez saisir votre commentaire ici:

31. Quels soutiens n'arrivez-vous pas à trouver pour le moment ?

📌 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

- Financier
- Ressources humaines
- Infrastructures numériques (entrepôts, puissance de calcul, outils, etc...).
- Technique
- Formation
- Juridique
- Ethique
- Non applicable

32. À quelles étapes associées aux données souhaiteriez-vous recevoir de l'aide ?

📌 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

- Mise en place amont du projet de recherche (stratégie d'analyse, méthodologie de collecte, etc.)
- Constitution des données (collecte de données, recueil d'informations, structuration, etc.)
- Traitement et analyse des données (calculs, traitements statistiques, interprétation)
- Stockage des données et conservation
- Mise à disposition des données
- Valorisation des données (socio-économiques)
- Publication et diffusion des données à un large public nécessitant un travail de médiation (site web, vidéos, podcast, etc.)
- Non applicable

33. De quels soutiens avez-vous le plus besoin ?

🕒 Classez jusqu'à 4 éléments du plus important (haut de la liste) au moins important (bas de la liste). Vous pouvez également indiquer « non applicable ».

Effectuez un double-clic ou glissez/déposez le/les élément(s) de la liste de gauche vers la liste de droite.

🕒 Veuillez sélectionner de 1 à 4 réponses.

Vos choix

Financier
Ressources humaines
Infrastructures numériques (entrepôts, puissance de calculs, outils, etc...).
Technique
Formation
Juridique
Éthique
Non applicable

Votre classement

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

🕒 Cochez toutes les cases qui s'appliquent ; si vous le souhaitez, vous pouvez ajouter un commentaire/exemple pour les cases cochées.

🕒 Ajoutez un commentaire seulement si vous sélectionnez la réponse.

Des spécialistes disponibles pour répondre à vos questions et aider à la formalisation des pratiques au sein de votre établissement

Des spécialistes situés dans des centres dédiés

Un référent (personne/soutien) au sein de mon laboratoire /équipe de recherche

Une communauté d'aide via une liste de diffusion

Une plateforme dédiée en ligne pour répondre à mes questions et m'orienter (guichet unique de la donnée)

Des formations pour être sensibilisé.e.s, à la gestion, diffusion

Des journées d'étude avec d'autres membres de votre discipline/métier

Non applicable

Autre :

35. Souhaitez-vous ajouter un commentaire général concernant les questions de cette partie « Accompagnement et besoins » ?

📌 Commentaire libre.

Partie 5 : Affiliations, fonctions et champs disciplinaires.

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

📌 Choix unique.

📌 Veuillez sélectionner une réponse ci-dessous

- Moins de cinq ans
- Cinq à dix ans
- Plus de dix ans

37. Quel est le nom de votre principale structure de rattachement ?

38. Quelle est votre fonction principale au sein de cette structure ?

 Cochez toutes les cases applicables.

 Cochez la ou les réponses

Enseignement-supérieur et de la recherche permanent (enseignant.e- chercheur.se, chercheur.se, etc. responsable de laboratoire)

Enseignement supérieur et de recherche non permanent (ATER, doctorant.e, post-doctorant.e, etc.)

Ingénieur.e, technicien.ne., assistant.e de recherche et de formation

Systèmes d'information (SI) : support à l'enseignement supérieur et à la recherche

Bibliothèques, documentation, archives et musées : support à l'enseignement supérieur et à la recherche

Affaires juridiques : soutien et support à l'enseignement supérieur et à la recherche

Valorisation de la recherche : soutien et support à l'enseignement supérieur et à la recherche

Transformation numérique et science ouverte : soutien et support à l'enseignement supérieur et à la recherche

Management, pilotage, contrôle, aide au montage de projets : soutien et support à l'enseignement supérieur et à la recherche

Communication - édition : soutien et support à l'enseignement supérieur et à la recherche

Autre :

39. Quel est votre statut professionnel ?

📌 Choix unique.

📌 Veuillez sélectionner une réponse ci-dessous

- Fonctionnaire ou assimilé
- En CDD
- En CDI
- Contrat doctoral
- Travailleur/travailleuse indépendante
- Autre (précisez en commentaire)

Veuillez saisir votre commentaire ici:

40. Lesquelles des catégories suivantes décrivent le mieux votre domaine de recherche ?

📌 Choisissez toutes les cases qui s'appliquent.

📌 Cochez la ou les réponses

- Droit, économie et gestion
- Informatique et science des données
- Lettres, sciences humaines et sociales
- Mathématiques
- Pluridisciplinaires
- Sciences de la vie et de la santé
- Sciences de l'ingénieur
- Sciences de la Terre et de l'univers
- Sciences de la matière (physique-chimie)
- Autre :

41. Quelle est votre discipline de spécialité (par exemple biologie structurale, épidémiologie, physique fondamentale, socio-anthropologie, etc...) ?

42. Vous êtes :

i Veuillez sélectionner une réponse ci-dessous

Une femme

Un homme

Autre :

Sans réponse

43. Souhaitez-vous ajouter un commentaire général concernant le questionnaire ?

7

ANNEXES

Résultats du questionnaire

Population



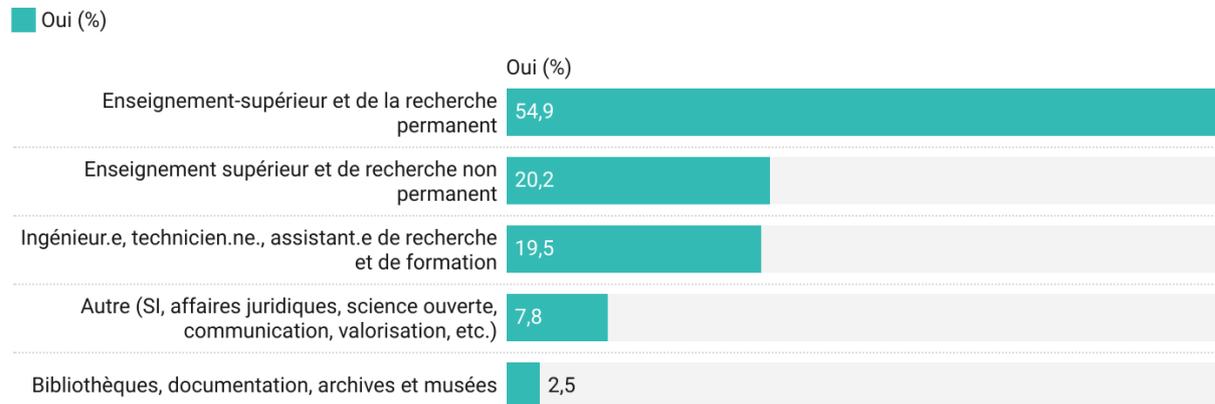
Genre : 👤 47,8 % et 👤 44,8 % (autre : 7,4 %)

Fonction principale : essentiellement des enseignant.e.s-chercheur.se.s « permanent » (55 %) et « non permanent » (20%)

Ancienneté : Majoritairement plus 10 ans (65,5%) avec un statut de fonctionnaire (68,7%)

Quelle est votre fonction principale au sein de cette structure?

Le total des réponses est supérieur à 100% car les réponses pouvaient être multiples (combinaison de plusieurs fonctions)



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

Catégories disciplinaires

- Sur-représentation de la catégorie « Lettres et Sciences Humaines et Sociale » (61,8% avec principalement de la sociologie et de l'histoire)
- Sous-représentation des autres disciplines notamment le domaine des sciences de la vie et de la santé (11%)
- Pondération des résultats effectuée en fonction des catégories disciplinaires

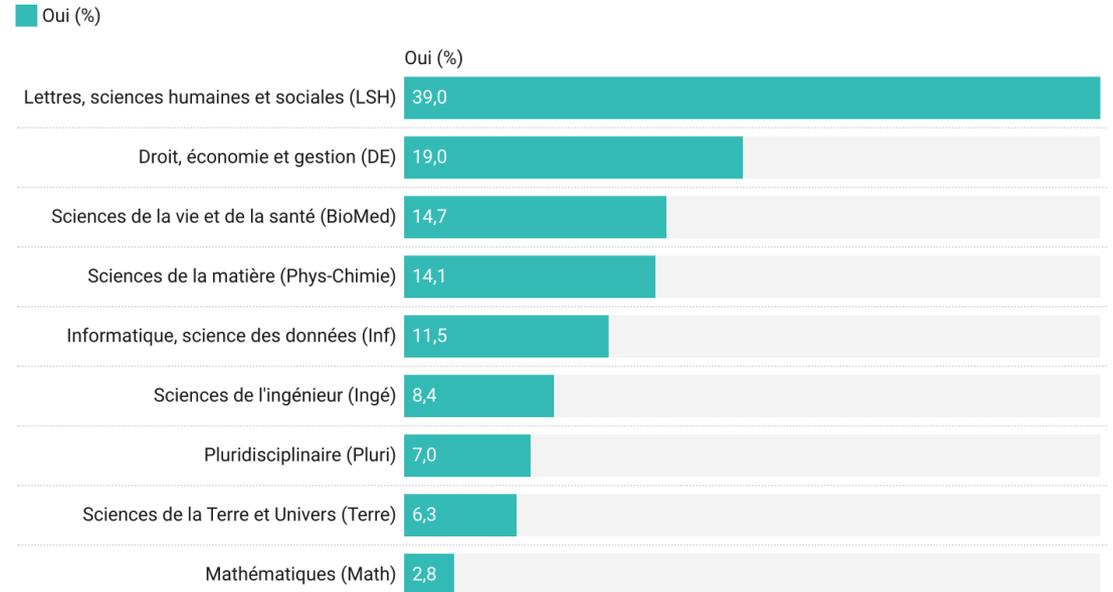


Pondération des résultats effectuée sur la base des Données 2019-2020 ESR *L'état de l'Enseignement supérieur, de la Recherche et de l'Innovation en France* (n°14 - Avril 2021)

Catégories disciplinaires représentées après pondération des résultats

Lesquelles des catégories suivantes décrivent le mieux votre domaine de recherche?

Les réponses pouvaient être multiples, d'où un total supérieur à 100%. Ce résultat a été obtenu après pondération des résultats pour être au plus proche représentatif de la répartition des discipline au sein de l'ESR.



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

Heatmap

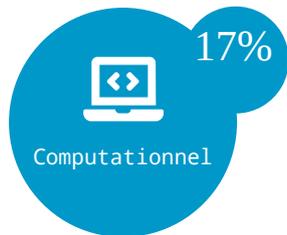
Visualisation (heatmap) des modalités des variables expliquant les distinctions principales entre les profils

Les coefficients de corrélation (positifs ou négatifs) indiquent une sur-représentation d'une modalité par rapport aux autres profils.



Modalités des variables	Interdisciplinaire : computationnel	STM : expérimental	SHS : collaboratif	SHS : solitaire
Emploi d'outils collaboratifs : partage de code	11		-5	-6
Modalités de mise à disposition des données : plateformes de type Gitlab, Github	12		-5	-6
Modalités de mise à disposition des données : entrepôts généralistes	8		-2	-4
Autres éléments à mettre à disposition : documentation (données, codes sources)	8			-6
Obligation pour la mise à disposition de données : par des financeurs	8			-4
Emploi d'outils collaboratifs : documentation, wiki, etc.	8	-4		-3
Qualificatif pour une recherche de qualité : recherche ouverte	7			-4
Données et ressources connexes produites : normes pour données et métadonnées	7	-3		-4
Autres éléments à mettre à disposition : métadonnées, thesaurus, vocabulaire contrôlé	7	-4		-3
Modalités de conservation des données : serveur intégré à une infrastructure de recherche	7	-3		-2
Habitue de partager en ligne avec licences : souvent	7			-4
Sources des données : base de données partagées par des collègues	7			-3
Plus-value du travail : automatisation d'un protocole/workflow	6		-3	-5
Collaboration avec : services système d'information (SI)	6			-3

Personnes provenant de différentes disciplines (STM, SHS) ayant une culture computationnelle, de la donnée et du partage



Classes déterminées par le biais d'une ACM puis HCPC selon le critère « pratiques associées aux données & science ouverte ».

Heatmap

Personnes issues des domaines des Sciences de la Vie et de la Matière avec des pratiques ancrées dans une approche expérimentale



Classes déterminées par le biais d'une ACM puis HCPC selon le critère « pratiques associées aux données & science ouverte ».

Visualisation (heatmap) des modalités des variables expliquant les distinctions principales entre les profils

Les coefficients de corrélation (positifs ou négatifs) indiquent une sur-représentation d'une modalité par rapport aux autres profils.



Modalités des variables	Interdisciplinaire : computationnel	STM : expérimental	SHS : collaboratif	SHS : solitaire
Sources des données : instrument de mesure		11	-5	-6
Environnement de recherche : laboratoire		10	-5	-6
Plus-value du travail : préparation échantillon et protocoles expérimentaux	-2	9	-3	-5
Emploi du terme mesures	2	9	-4	-5
Emploi du terme valeurs		8	-3	-4
Qualificatif pour une recherche de qualité : reproductibilité	6	8	-5	-7
Emploi d'outils collaboratifs : cahier de laboratoire		8	-3	-4
Logiciels utilisés pour les données : tableurs scientifiques	-3	7	-3	-2
Domaine de recherche : sciences de la matière		7	-4	-4
Domaine de recherche : sciences de la vie et de la santé		6	-3	-5
Modalités de mise à disposition des données : système d'envoi de fichier		6		-5
Logiciels utilisés pour les données : langage de programmation	7	5	-3	-5
Emploi du terme données brutes	5	5	-3	-5
Emploi du terme jeux de données	7	5	-3	-6
Collaboration avec : ingénieur.e technicien.ne	6	5	-4	-5
Collaboration avec : permanent de l'ESR	4	5		-4
Environnement de recherche : ordinateur pour analyse de données, de modélisation	4	5		-6

Heatmap

Personnes issues des Sciences Humaines et Sociales privilégiant des pratiques collaboratives .



Classes déterminées par le biais d'une ACM puis HCPC selon le critère « pratiques associées aux données & science ouverte »

Visualisation (heatmap) des modalités des variables expliquant les distinctions principales entre les profils

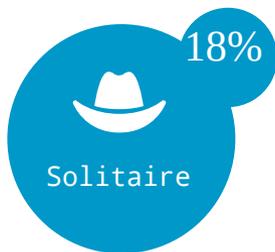
Les coefficients de corrélation (positifs ou négatifs) indiquent une sur-représentation d'une modalité par rapport aux autres profils.



Modalités des variables	Interdisciplinaire : computationnel	STM : expérimental	SHS : collaboratif	SHS : solitaire
Soutien souhaité : technique	-2	-2	7	
Type d'aide souhaitée : publication et diffusion à un large public		-3	6	-2
Type d'aide souhaitée : valorisation des données		-2	5	-3
Type d'aide souhaitée : mise en place du projet de recherche			5	
Emploi du terme corpus		-12	6	4
Modalités de mise à disposition des données : cloud non institutionnel	3	-2	5	-3
Modalités de conservation des données : support externe	-3		5	
Modalités de conservation des données : ordinateur personnel	-4	-6	5	4
Domaine de recherche : lettres et sciences humaines		-11	6	6
Environnement de recherche : bibliothèques ou archives		-10	5	4
Sources des données : archives		-10	5	2
Type d'aide souhaitée : mise à disposition des données			4	-3
Environnement de recherche : travail de terrain, d'enquête		-6	4	5
Source des données : terrain		-6	4	4
Formation aux outils : échanges informels			4	
Découverte des outils : par moi-même	-2		3	-2

Heatmap

Personnes issues principalement des SHS travaillant essentiellement seules



Classes déterminées par le biais d'une ACM puis HCPC selon le critère « pratiques associées aux données & science ouverte »

Visualisation (heatmap) des modalités des variables expliquant les distinctions principales entre les profils

Les coefficients de corrélation (positifs ou négatifs) indiquent une sur-représentation d'une modalité par rapport aux autres profils.



Modalités des variables	Interdisciplinaire : computationnel	STM : expérimental	SHS : collaboratif	SHS : solitaire
Habitude de partager en ligne avec contrôle d'accès : jamais	-5			10
Habitude de partager en ligne avec une licence : jamais	-7	3		9
Habitude de partager en ligne sans contrôle d'accès : jamais	-4	2		8
Habitude de partager restreint : jamais		-3		4
Obligation pour le mise à disposition de données : aucune	-5			7
Spécialité disciplinaire : sociologie	-3	-4		6
Emploi du terme matériaux		-5		5
Fréquence de réutilisation de données produites ou publiées : jamais			-3	4
Taille de l'équipe : travail solitaire	-4	-6		4
Limites à la mise à disposition : un doute sur la capacité de réutilisation	-4			4
Découverte des outils : non-applicable	-3	-3		3

Présentation (c) Inno³ 2022 diffusée sous licence CC-BY-4.0, réalisée par Célya Gruson-Daniel (Inno³) avec la contribution de Claire Lemercier (analyse quanti, conseils, relecture), d'Hélène Chambefort, de Juliette Hueber, d'Anne Vanet (conseils, relecture), de Romain Rouyer et de Tamara Glushetckaia (design et mise en forme). Photos sous licence Unsplash.

Contact : cgruson-daniel@inno3.fr

