



CELLULE DATA STEWARDSHIP RAPPORT ENTRETIENS CHERCHEUR·EUSES DOMAINE ÉCONOMIE ET SERVICES

Emilie Pasche, Data steward – Octobre 2024

TABLE DES MATIÈRES

I. INTRODUCTION	1
II. PRATIQUES EN MATIÈRE DE GESTION DES DONNÉES	2
III. BESOINS ET SOLUTIONS	10
IV. CONCLUSION ET RECOMMANDATIONS	12

I. INTRODUCTION

Afin de mieux comprendre les pratiques et les défis liés à l'Open research Data (ORD) du domaine Économie et Services (ES), des entretiens semi-dirigés ont été réalisés par Emilie Pasche en qualité de data steward auprès des chercheur·euses de différentes Hautes écoles du domaine ES. Ce rapport a pour objectif de résumer les principaux enjeux et pratiques en matière de gestion et d'ouverture des données de recherche, tels qu'ils ont été relevés lors des entretiens. Il met en avant les défis, les solutions mises en œuvre et les recommandations visant à améliorer la gestion des données au sein du domaine ES. Ce document vise à nourrir les réflexions internes en vue de l'élaboration de la stratégie et du plan d'action ORD de la HES-SO et peut également servir de support pour alimenter d'autres processus de planification stratégique au sein des Hautes écoles et/ou du domaine.

1. Profil des participant·es

En raison des difficultés à obtenir des réponses, certaines institutions sont sous-représentées dans ce rapport. En effet, sur 42 prises de contact, seul 12 entretiens ont pu être réalisés. Les entretiens se sont déroulés sur une période de deux mois, entre mars et mai 2024.

Répartition des entretiens par école :

- EHL : 3 entretiens
- HEG-ARC : 2 entretiens
- HEG-Fribourg : 1 entretien
- HEG-Genève : 3 entretiens
- HEG-Valais : 1 entretien
- HEIG-VD : 2 entretiens
-

Fonctions des personnes interrogées :

- Professeur·e ordinaire : 1





- Professeur·e associé·e : 2
- Professeur·e assistant·e : 4
- Adjoint·e scientifique : 1
- Collaborateur·trice scientifique : 1
- Assistant·e HES: 2
- Autre (Hors PER) : 1

Type de financement :

- FNS : 6
- Autres fonds externes : 2
- Fonds internes : 4
- Pas de fonds : 1

Disciplines concernées :

- Business intelligence : 1
- Économie : 4
- Informatique : 3
- Marketing : 2
- Sociologie : 2
- Système de santé : 1

II. PRATIQUES EN MATIÈRE DE GESTION DES DONNÉES

Cette section aborde les pratiques des participant·es en matière de gestion et d'ouverture des données et identifie les défis associés à ces pratiques.

2.1 Les types de production des données

Dans cette section seront présentés les différents types de production des données identifiés à partir des entretiens. Nous proposons une classification des projets en trois catégories : les données sont produites par les chercheur·euses de l'une des hautes écoles du domaine, les données sont produites conjointement par les chercheur·euses de l'une des hautes écoles du domaine et d'une autre institution, les données sont produites par des tiers. Pour chacune de ces catégories, des exemples sont proposés ci-dessous :

- Données produites par les chercheur·euses de l'une des écoles du domaine
 - Entretiens ou observations auprès d'employé·es et/ou entreprises en Suisse
 - Sondages auprès d'étudiant·es en Suisse et/ou en Europe
 - Développement de codes
- Données produites conjointement par les chercheur·euses de l'une des écoles du domaine et d'une autre institution
 - Sondages auprès d'étudiant·es de différentes institutions en Suisse ou en Europe
 - Sondages auprès d'employé·es et/ou entreprises à l'étranger (hors UE)
 - Entretiens auprès d'employé·es et/ou entreprises en Suisse
 - Analyses de données relatives à des entreprises à l'étranger (hors UE)
 - Génération de métadonnées associées à des contenus textuels
- Données produites par des tiers
 - Utilisation de jeux de données disponibles sous licences ouvertes





- Utilisation de jeux de données fournis par une autre institution
- Utilisation de données statistiques officielles
- Utilisation de données provenant des réseaux sociaux (X, Facebook)
- Utilisation de jeux de données commerciaux
- Utilisation de données récoltées auprès d'entreprises suisses
- Utilisation de données statistiques confidentielles (hors UE)

Ces différents types de production soulèvent plusieurs défis :

- a) Clarification des notions de propriété des données et propriété intellectuelle
On note une confusion entre les notions de propriété intellectuelle et de propriété des données. La propriété intellectuelle protège les créations originales, donnant à l'auteur des droits exclusifs d'exploitation. La question des droits de propriété intellectuelle est définie par la Convention intercantonale sur la Haute École Spécialisée de Suisse occidentale (HES-SO)¹. La propriété des données concerne la personne ou l'institution qui possède ou contrôle l'accès aux données. Elle dépend généralement des règlements internes des Hautes Écoles. Cependant, en Suisse, on considère que les données numériques ne peuvent pas être un objet de propriété. D'autres droits s'appliquent, comme la protection des données personnelles.
Certain·es chercheur·euses pensent, à tort, que si on possède les données, on en possède automatiquement les droits de propriété intellectuelle. Par exemple, un·e chercheur·euse peut avoir recueilli des données, mais il n'a pas forcément des droits exclusifs dessus.
- b) Propriété partagée des données
Dans les cas où la propriété des données est partagée entre plusieurs institutions académiques, il peut être complexe de définir clairement les droits et responsabilités de chaque partie. A l'exception des projets contenant des données sensibles, les entretiens montrent qu'il reste peu courant que les chercheur·euses établissent des contrats pour définir la propriété des données, ce qui pourrait conduire à des malentendus ou à des conflits potentiels sur l'utilisation et le partage ultérieur des données.
- c) Droits des participant·es
Lorsque des chercheur·euses collectent des données personnelles, les participant·es restent propriétaires de leurs données et conservent ainsi des droits selon la Loi sur la Protection des Données (LPD), notamment le droit d'accès, le droit de rectification et le droit d'effacement de leurs données. Ces droits peuvent poser des défis importants pour les chercheur·euses lors de l'ouverture des données : par exemple, comment garantir le droit à l'effacement lorsque le jeu de données a déjà été partagé ? Une autre problématique concerne l'utilisation de données provenant des réseaux sociaux. Bien que ces données soient publiquement accessibles, elles restent la propriété des individus. Il est donc essentiel d'obtenir le consentement des personnes concernées avant toute collecte ou partage de ces données.

¹ <https://www.lexfind.ch/fe/fr/tol/19537/fr>





d) Réutilisation des données appartenant à des tiers

L'utilisation de données provenant d'autres sources nécessite souvent de respecter des conditions spécifiques, telles que des licences d'utilisation ou des accords contractuels, qui peuvent restreindre la réutilisation des données. Le défi réside dans la gestion de ces conditions pour assurer une utilisation légale et éthique des données.

2.2 Types et formats des données collectées/produites

Les entretiens n'ont pas montré de défis particuliers concernant les types et formats de données. La plupart des formats utilisés sont ouverts, à l'exception de Microsoft Word et Microsoft Excel, dont l'ouverture reste discutable (*DOCX*, *XLSX*). A titre d'information, le tableau ci-dessous regroupe les formats mentionnés groupés par types de données. Les formats ouverts sont indiqués en gras.

Types de données	Formats
Documents textuels	TXT (texte brut) DOCX (Microsoft Word)
Formats (semi)-structurés	XML (eXtensible Markup Language) JSON (Javascript Object Notation) JATS (Journal Article Tag Suite) BioC CSV (Comma-Separated Values) XLSX (Microsoft Excel)
Enregistrements audio	WAV (Waveform Audio Format) MP3 (MPEG-1 Audio Layer 3) M4A (format natif iTunes)
Enregistrements vidéo	MP4 (MPEG-4 Part 14) AVI (Audio Video Interleave)
Données médicales	OMOP Common Data Model (Observational medical outcomes partnership – Common Data Model)
Codes informatiques	Python R
Questionnaires	Formats non spécifiés
Données de géolocalisation	Formats non spécifiés

2.3 Stockage des données

Concernant la pratique courante en termes de stockage des données, on note un manque d'information concernant les pratiques de stockages des données. La plupart des chercheur·euses n'ont pas reçu d'informations spécifiques concernant le stockage des données de leur projet de recherche. Cependant, la plupart affirment ne pas en avoir demandé. Dans la pratique, la majorité des chercheur·euses suivent les standards de l'école ou les pratiques de leurs collègues. Le stockage des données se fait généralement sur les ordinateurs de l'institution (5/12) ou les serveurs de l'institution (4/12). La distinction entre ces deux types de stockage est parfois floue : un·e chercheur·euse peut indiquer stocker les données sur son ordinateur de travail tout en mentionnant également que les données sont sauvegardées sur le serveur de l'école. Il est donc probable qu'il s'agisse en réalité d'un disque distant sur le serveur de l'école. Dans de rares cas, des chercheur·euses ont indiqué utiliser des disques durs externes. Pour l'un·e des chercheur·euses, cette pratique résulte de la spécificité de ses données : le·la chercheur·euse manipule des fichiers de grande taille, qui





nécessitent un accès local pour permettre leur édition. En effet, un accès distant à de tels fichiers rend toute manipulation trop fastidieuse en raison des temps de réponse élevés. Les disques durs externes sont cryptés et conservés au sein de l'institution. Une autre solution est actuellement envisagée par ce·tte chercheur·se. Un·e autre chercheur·euse utilisant des disques externes le fait à des fins de sauvegarde et travaille sur des données publiques et non sensibles.

On note également une méconnaissance des pratiques de sauvegardes. Peu de chercheur·euses sont au courant des pratiques de leur école à ce sujet. La plupart reconnaissent faire confiance au service IT et estiment que des sauvegardes sont effectivement faites.

Les défis liés au stockage des données sont multiples :

- a) Stockage de gros volumes de données ou de fichiers de grandes tailles
Certains projets manipulent des gros volumes de données ou des fichiers de grandes tailles. Ces données nécessitent des solutions de stockage particulière. Par exemple, dans le cas des fichiers de grandes tailles, les accès distants peuvent être problématiques pour des questions de performance lors de la manipulation des fichiers.
- b) Stockage partagé des données
Dans le cadre de projets de recherche impliquant plusieurs partenaires, les chercheur·euses ont besoin de solution de stockage partagé entre institutions. Trois chercheur·euses ont mentionné des solutions de stockage partagées au sein de l'une des institutions :
 - Un groupe stocke sur les serveurs du partenaire, qui ont mis en place des mesures sécurisées.
 - Un groupe a mis en place une solution de stockage sur un serveur du groupe, avec un accès au serveur via VPN.
 - Un groupe a indiqué que toutes les données collectées seront stockées au sein du serveur de la HES-SO.
- c) Stockage des données confidentielles
Dans le cadre de projets contenant des données confidentielles, les chercheur·euses ont besoin de solution de stockage sécurisé. Plusieurs chercheur·euses ont eu recours à des solutions personnalisées, généralement avec l'aide de leurs services informatiques respectifs :
 - Les fichiers sont stockés sur le serveur commun de l'école, mais avec accès limité aux parties prenantes du projet.
 - Le service IT a mis en place des mesures techniques pour permettre de sécuriser le partage des données. Les participant·es déposent les données sur un serveur sécurisé interne pour lequel il faut utiliser un VPN. Seuls deux chercheur·euses ont l'autorisation d'accéder aux données.
 - Dans un cas, il est prévu de stocker les données sur plusieurs SwitchDrive de manière à pouvoir contrôler les droits d'accès. Cette solution n'a pas été proposée par le service IT, mais résulte de la pratique courante au sein du groupe de recherche concerné.





2.4 Aspects juridiques

L'ouverture des données n'est généralement pas abordée avec les juristes. Certains chercheur·euses ont eu des contacts avec les juristes de leur école, mais plutôt pour des aspects tels que les formulaires de consentements ou de copyrights. Aucun·e chercheur·euse n'a mentionné une aide relative à l'ouverture des données par exemple. Les chercheur·euses font généralement signer des formulaires de consentements à leurs participant·es. Cependant, l'aspect du partage des données n'est généralement pas abordé dans ces formulaires de consentement. Ces aspects juridiques nécessitent dans certains cas une expertise externe. En effet, un cas a montré que la spécificité des besoins était telle que cela sortait du domaine de compétence des juristes de l'école et a nécessité des aides externes.

2.4.1 Données personnelles/sensibles

Environ la moitié des chercheur·euses interrogé·es manipulent des données personnelles. La présence de données sensibles est moins fréquente. Dans ces cas, les chercheur·euses sont plutôt conscient·es de l'importance de bien gérer de telles données. Toutefois, la gestion des données personnelles et sensibles soulève deux défis majeurs :

a) Anonymisation

L'anonymisation des données personnelles est un processus complexe, nécessitant des ressources ou compétences que les chercheur·euses n'ont parfois pas. Dans le cas des chercheur·euses interrogé·es, l'anonymisation a été effectuée par les chercheur·euses eux-mêmes. Dans tous les cas, cette anonymisation se base sur les pratiques courantes du domaine et généralement de manière manuelle. Des contrôles qualité sont mis en place, tels que la validation par un·e deuxième chercheur·euse. Dans quelques cas, les chercheur·euses ont reçu des recommandations du rectorat. Certains chercheur·euses souhaiteraient de l'aide pour l'anonymisation.

b) Dépersonnalisation

Dans certains cas, il n'est pas possible d'anonymiser les données comme les vidéos ou des photographies. Un·e chercheur·euse a eu recours à la dépersonnalisation : les visages sont floutés ou pixellisés et les noms supprimés. Cependant la personne reste reconnaissable dans certains contextes (par exemple par un·e collègue ou un·e proche). L'ouverture de ces données est donc problématique.

2.5 Data Management Plan (DMP)

La plupart des chercheur·euses connaissent le DMP (10/12), soit parce qu'ils ont dû en rédiger un dans le cadre d'une demande de fonds (principalement FNS), soit parce que les spécificités de leur projet les ont amenés à devoir décrire comment ils allaient gérer leurs données. C'est généralement lorsqu'ils doivent écrire leur premier DMP que les chercheur·euses ont réfléchi à des stratégies tels que le stockage sécurisé. Dans deux cas, les chercheur·euses mentionnent avoir bénéficié de l'expérience de collègues en ayant déjà écrit. Certains chercheur·euses ont lu des guidelines, ont suivi des webinaires ou ont demandé de l'aide (par exemple au rectorat, au FNS ou à DLCM).





2.6 Codes et logiciels

Parmi les chercheur·euses interrogé·es, un peu moins de la moitié (5/12) crée des codes ou logiciels. Le partage des codes est une pratique qui reste plutôt rare (1/5). Cela peut s'expliquer par un sentiment que les codes ne sont pas suffisamment matures ou documentés pour être partagés, ou ne sont pas perçus comme utiles à la communauté. Cependant la plupart des chercheur·euses seraient prêts à partager leurs codes. Dans trois cas, les codes sont déjà déposés dans GitHub mais en accès privé. Concernant le cas où le code est partagé, le code est réécrit et amélioré par un autre groupe avant d'être mis à disposition en tant que package Python. Dans un autre cas, d'autres codes du groupe sont partagés publiquement sur GitHub. La gestion et l'ouverture des codes et logiciels présentent un double défi :

a) Aucune pratique établie

En général aucune pratique n'est établie concernant la documentation du code. Les chercheur·euses ont pour habitude de commenter leurs codes, mais plutôt à un usage interne (par exemple pour se souvenir de ce qu'ils ont fait, débbugger, reprendre leur code plus tard).

b) Valeur ajoutée des codes

Un·e chercheur·euse pointe ne pas voir d'inconvénient à partager des codes dans le cadre d'un projet financé par le FNS par exemple, mais ne le ferait pas dans le cadre d'un projet avec une entreprise (par exemple un financement InnoSuisse) : le code est une valeur ajoutée pour les chercheur·euses et le partager leur ferait perdre des opportunités futures.

2.7 Le partage des données

Tous les chercheur·euses interrogé·es connaissent de manière générale l'ORD. Cependant, les expériences de partage de données restent rares, avec seulement un·e chercheur·euse sur douze ayant déjà déposé des données ouvertes. Les dépôts de données sont relativement peu connus : les chercheur·euses citent principalement Zenodo. Les principes FAIR sont peu ou superficiellement connus des chercheur·euses. Certains chercheur·euses confondent également Open Access (OA) et ORD. De plus, plusieurs chercheur·euses doutent de la faisabilité du partage de leurs données. Les raisons évoquées sont décrites dans la section ci-dessous.

Problèmes liés à l'utilité perçue du partage des données

Beaucoup de chercheur·euses doutent de l'utilité de partager leurs données. Ces points ont été soulevés dans le cas de recherche qualitative.

a) Perte de dimension si on ne fait pas la collecte soi-même

Les données partagées sont lissées, ce qui empêche de saisir pleinement ce qui se passe réellement lors des entretiens ou observations. De nombreuses informations restent dans la tête des chercheur·euses et ne peuvent pas être retranscrites. La transcription est factuelle, mais ne rend pas toutes les subtilités dont le·la chercheur·euse initial a conscience : « la transcription des données est froide et plate ».





b) Spécificité des données à une question de recherche

Les données sont produites pour un projet bien particulier dans un contexte bien particulier. Il faudrait fournir beaucoup d'informations contextuelles pour une réutilisation secondaire potentielle. Un exemple : un·e chercheur·euse filme des sessions d'utilisation d'un logiciel propriétaire et à usage exclusif pour analyser les interactions entre les utilisateurs et ce logiciel. Il faut donc avoir des informations sur le logiciel (par exemple comprendre ce qu'il se passe lorsque l'utilisateur clique à un certain endroit de l'écran).

c) Nature du travail

Un·e chercheur·euse se demande si d'autres chercheur·euses trouveront une valeur à ses données partagées. Ce·tte chercheur·euse ne pourrait pas réutiliser de telles données pour son travail car il manque d'importants détails, qui ne peuvent être partagés. Cependant, il n'est pas exclu que d'autres chercheur·euses puissent avoir des questions de recherche non envisagées initialement.

Problèmes liés à l'anonymisation / protection des données

Certains jeux de données contiennent des informations personnelles et/ou sensibles. Il est donc indispensable d'effectuer une anonymisation des données avant de les déposer.

a) Connaissance en anonymisation

Les chercheur·euses ne reçoivent pas d'aide pour anonymiser les données. La plupart effectuent l'anonymisation de manière manuelle, en se basant sur leurs expériences passées. Certain·es chercheur·euses font part de leur inquiétude face à des problèmes potentiels qui pourraient survenir si l'anonymisation n'est pas suffisante, par exemple en cas de réidentification par croisement de jeux de données.

b) Risque de réidentification

Certaines populations étudiées sont spécifiques et de petite taille, ce qui augmente le risque de réidentification.

c) Utilité des données après anonymisation

Dans certains cas, les données identifiantes sont importantes pour répondre à la question de la recherche. La nature de la recherche implique de maintenir certaines caractéristiques au risque de voir les données potentiellement inexploitable.

Problèmes liés à la préparation des données :

La préparation des données en vue du partage est également un problème à prendre en compte.

a) Charge de travail pour la préparation des données

Certains chercheur·euses s'inquiètent de la charge de travail que représente la préparation des données : l'anonymisation, le nettoyage des données, etc. Par exemple, concernant l'anonymisation, un·e chercheur·euse n'a anonymisé qu'une partie de ses données par manque de ressources.

b) Documentation à préparer

L'écriture d'une documentation des données peut être problématique. La rédaction de documentations suffisamment complètes et claires pour des utilisations externes





ne semblent pas faire partie des pratiques courantes des chercheur·euses, probablement par manque de temps.

- c) Qualité des données
Un·e chercheur·euse mentionne la peur de déposer des données de mauvaise qualité, par exemple qui contiendraient des erreurs.

- d) Travail jamais figé
En particulier pour les codes et logiciels, il s'agit souvent de travail en cours et non pas un produit abouti. Certain·es chercheur·euses estiment leur travail pas suffisamment mature pour être partagé.

Problèmes liés à la propriété des données :

La notion de propriété des données doit être prise en compte avant de pouvoir partager des jeux de données. Si la question de propriété des données est importante au début d'un projet, elle l'est encore plus dans le cadre de l'ouverture des données.

- a) Propriété des données
Le problème se pose particulièrement pour les projets impliquant plusieurs groupes de recherche ou les projets reposant sur des données réutilisées. Un·e chercheur·euse mentionne la peur de déposer des données qui ne lui appartiennent pas vraiment. Un·e autre chercheur·euse mentionne utiliser des données secondaires pour lesquels un engagement à utiliser les données uniquement pour sa recherche a été signé : les données ne peuvent donc pas être partagées.

- b) Gestion des licences lors du partage de jeux de données liés à des licences
La gestion des données dérivées à partir de plusieurs ensembles de données devient particulièrement délicate lorsque les licences des sources de données divergent. Dans de tels cas, il est essentiel de comprendre en détail les termes et les restrictions de chaque licence applicable. Dans certains cas, les licences peuvent être contradictoires (par exemple, certaines autorisent l'utilisation commerciale tandis que d'autres ne le permettent pas), nécessitant de trouver des solutions créatives ou de négocier des accords spécifiques avec les détenteurs de licence pour harmoniser les conditions d'utilisation et de partage des données dérivées.

- c) Données confidentielles
Bien que le cas ait été peu mentionné lors des entretiens, certains projets reposent sur des données confidentielles. C'est en particulier le cas des projets sur mandat. Dans de tel cas, le partage des données ne sera pas possible.

Autres problèmes :

- a) Peur de l'utilisation des données
Un·e chercheur·euse mentionne le fait de vouloir être sûr·e que l'utilisation de ses données se feront uniquement à but scientifique.

- b) Peur que ça complique la collecte
Un·e chercheur·euse craint que l'ouverture des données complique la collecte des données. Les participant·es, pour certain·es déjà réticent·es à participer aux entretiens, pourraient être effrayé·es par l'ouverture des données et être moins enclins à participer, ou ne pas divulguer certaines informations.





c) Impact environnemental

Certain·es chercheur·euses travaillent sur des jeux de données très volumineux (Big Data). L'ouverture de tels sets de données implique une consommation accrue de ressources énergétiques pour le stockage. Il est donc crucial d'évaluer les avantages et les coûts environnementaux de l'ouverture de ces jeux de données. Si les données peuvent bénéficier à de nombreux chercheur·euses ou applications, cela justifie mieux l'impact environnemental.

2.8 La réutilisation des données

La réutilisation de données est une pratique relativement courante pour les chercheur·euses interviewé·es travaillant avec des données quantitatives. Cependant, ces données ne sont généralement pas issues de la recherche, mais plutôt de sources extérieures aux chercheur·euses, telles que des statistiques fédérales ou cantonales, des données fournies par des entreprises privées ou des bases de données publiques. De plus ces expériences de réutilisations ne sont pas forcément liées aux activités actuelles des chercheur·euses au sein de la HES-SO.

Cas de réutilisation :

- Trois chercheur·euses ont mentionné l'utilisation de jeux de données dans une expérience professionnelle précédente, généralement liée à leur PhD.
- Trois chercheur·euses ont mentionné l'utilisation de jeux de données utilisés comme benchmark pour fine-tuner, évaluer ou comparer leurs systèmes (jeux de données trouvés sur Zenodo, Kaggle, TREC, CLEF notamment).
- Un·e chercheur·euse a mentionné l'utilisation de jeux de données FAIR, mais n'est pas au courant des détails : cette partie a été gérée par des partenaires du projet plus expérimentés à ce sujet.

Limitations :

- Des chercheur·euses mentionnent la difficulté à retrouver des jeux de données.
- Certains chercheur·euses aiment avoir le contrôle de la manière dont les données sont produites et n'envisagent ainsi pas de réutiliser des données : « j'aime rencontrer les gens, poser des questions, avoir les subtilités ».
- Plusieurs chercheur·euses mentionnent qu'il leur est difficile de voir des cas de réutilisation de données dans le domaine qualitatif.

III. BESOINS ET SOLUTIONS

3.1 Les besoins

Dans cette section sont listés les besoins exprimés par les chercheur·euses lors des entretiens. Peu de chercheur·euses ont obtenus de l'aide pour gérer leurs données, cependant peu d'entre eux en ont sollicité.

Besoins liés au temps :

Les chercheur·euses ont fait mention de leur préoccupation en regard du peu de temps qu'ils ont à disposition pour l'ORD.

- Besoin d'un processus simple et clair : les chercheur·euses partageraient plus volontiers leurs données s'il était possible de le faire facilement.





- Besoin d'informations ciblées : les chercheur·euses n'ont pas le temps pour de formations généralistes.
- Besoin de reconnaissance et valorisation : les chercheur·euses sont sous pression pour leur recherche et ressentent l'ORD comme une charge de travail supplémentaire et non valorisée.

Besoins liés aux connaissances :

Les chercheur·euses font souvent mention d'un manque de connaissance et d'un besoin d'informations. Il a été mentionné le fait de ne pas savoir où chercher l'information ou l'existence de procédures à suivre au niveau de l'école/institution. En particulier, les sujets suivants ont été évoqués :

- Besoin d'une feuille de route claire pour savoir quoi faire, quand le faire et comment le faire ; d'être informé sur les procédures (si existantes) et recommandations.
- Besoin de connaître l'identité des personnes de contact.
- Besoin d'informations sur les logiciels à utiliser, par exemple quel logiciel est recommandé pour la transcription, quel logiciel est recommandé pour les questionnaires, etc.
- Besoin d'informations sur le stockage et les sauvegardes de leurs données, notamment sur le stockage à long terme.
- Besoin d'aide au sujet des métadonnées.
- Besoin d'aide juridique.
- Besoin d'aide pour les questions éthiques pour les projets non liés à la santé (guidelines, workshop).
- Besoin d'aide pour la préparation des données.
- Besoin d'un service à l'école/institution pour pouvoir clarifier de manière efficace et homogène les questions des chercheur·euses.
- Besoin d'un accompagnement personnalisé : que quelqu'un se mette dans leurs problématiques et leur disent ce qui est utile pour eux. Des chercheur·euses mentionnent qu'ils auraient besoin d'être accompagnés quelques fois jusqu'à ce que ces nouvelles pratiques deviennent des habitudes.

Besoins liés à l'utilité des données :

- Beaucoup de chercheur·euses doutent de l'utilité du partage de leurs données. Il y a un besoin de savoir que l'effort fourni est utile : les chercheur·euses ont besoin de savoir que leurs données sont effectivement réutilisées.
- Des chercheur·euses ont également besoin d'aide pour réutiliser les jeux de données. Il leur est parfois difficile de trouver des jeux de données intéressants.

3.2 Les solutions

Quelques solutions ont été mentionnées par les chercheur·euses pour faire face à leurs problèmes et besoins. Mis à part l'entraide entre collègues, à laquelle la plupart des chercheur·euses ont recours, les autres solutions évoquées restent au stade de suggestion et ne sont pas mises en pratique.

Entraide entre collègues

Plusieurs chercheur·euses ont mentionné faire appel à leurs collègues lorsque des difficultés sont rencontrées concernant la gestion des données de la recherche, telles que l'écriture d'un DMP, trouver le bon dépôt de données pour leurs données, etc. Cette solution est couramment utilisée par les chercheur·euses.





Répondant ORD au niveau de la HES-SO ou des écoles

Plusieurs chercheur·euses ont indiqué ne pas avoir d'aide en matière d'ORD au niveau de leur école. Dans ces cas, les bibliothécaires les soutiennent plutôt sur les questions liées à l'OA, mais pas réellement pour l'ORD. Plusieurs personnes insistent sur la nécessité d'avoir un·e répondant·e avec un profil « recherche ». Généralement, les chercheur·euses n'accordent pas d'importance à la provenance de l'aide (école ou rectorat par exemple). Cependant, une autre personne pointe également du doigt qu'il est parfois plus simple de se débrouiller que de devoir expliquer ses problématiques à une autre personne.

Obligations

Certains chercheur·euses pensent que la solution est de rendre l'ORD obligatoire, que ce soit par les bailleurs de fonds, les exigences des journaux ou des écoles. Cette opinion n'est pas partagée par tous : d'autres estiment que ce serait contre-productif.

Valoriser l'ouverture des données de la recherche

Plusieurs chercheur·euses souhaiteraient que l'effort soit valorisé, par exemple via des incitatifs financiers, en les répertoriant dans ArODES ou en comptabilisant les heures passées à ouvrir les données de la recherche dans les feuilles de charge des chercheur·euses.

Mettre en place de l'aide financière pour l'ouverture des données

Hormis dans le cadre des FNS où il est possible de demander un support financier pour l'ouverture des données, les chercheur·euses ne bénéficient d'aucune aide financière. Certains chercheur·euses souhaiteraient une aide financière de l'école/institution.

Mettre en avant l'intégrité académique

En ouvrant leurs données de recherche, les chercheur·euses permettent à leurs pairs de valider l'intégrité de leurs travaux en reproduisant les expériences. De plus, le partage des données offre aux jeunes chercheur·euses des opportunités d'apprentissage et de progression. Cela permet également aux chercheur·euses des pays en développement d'accéder à des données qu'il ne leur serait pas possible d'obtenir. En partageant leurs données, les chercheur·euses contribuent ainsi à une plus grande équité dans le domaine scientifique.

Installer de nouvelles habitudes

Mettre en place de nouvelles habitudes est souvent compliqué. Les chercheur·euses peuvent ressentir une injustice à devoir consacrer du temps au partage de leurs données sans bénéficier des mêmes avantages de la part de leurs pairs. Il est essentiel de changer cette mentalité pour que le partage de données devienne une pratique naturelle. Les recommandations et politiques des revues scientifiques peuvent jouer un rôle crucial dans ce changement en incitant fortement au partage des données.

IV. CONCLUSION ET RECOMMANDATIONS

4.1 Conclusion

Ce rapport a mis en lumière plusieurs défis et besoins liés à l'ORD chez les chercheur·euses du domaine Économie et Services, listés ci-dessous :





1. Connaissance de l'Open Research Data (ORD) :

La majorité des chercheur·euses connaissent l'ORD, mais très peu ont déjà déposé des données ouvertes.

2. Défis liés à la propriété des données

- Confusion entre propriété intellectuelle et propriété des données : des chercheur·euses ne sont pas au clair sur les notions de propriété des données.
- Propriété partagée des données : l'établissement de contrats entre institutions académiques est peu courant.
- Droits des participant·es : la gestion des droits des participant·es concernant leurs données personnelles reste floue.
- Propriété de tiers : l'utilisation de données de tiers nécessite de respecter des conditions spécifiques et peut être un frein à l'ouverture des données.

3. Défis liés aux types et formats de données collectées/produites

Les formats utilisés sont pour la plupart ouverts et ne présentent donc pas de défis majeurs.

4. Défis liés au stockage des données

- Stockage de données volumineuses : des solutions individuelles sont mises en place par les chercheur·euses, telles que l'utilisation de disques durs externes ou de serveurs internes.
- Stockage des données partagées entre partenaires : des solutions individuelles sont mises en place par le service informatique, les chercheur·euses ou une autre institution.
- Stockage des données confidentielles : des solutions individuelles sont mises en place par le service informatique ou les chercheur·euses.

5. Défis liés aux aspects juridiques

- Spécificité des besoins : certains besoins sortent du domaine de compétence des juristes de l'école.

6. Défis liés aux données personnelles et sensibles

- Anonymisation : l'anonymisation est un processus complexe et les chercheur·euses ne bénéficient pas ou peu d'aide.
- Dépersonnalisation : dans certains cas, seule une dépersonnalisation est réalisable. Les possibilités de partages de données dépersonnalisées soulèvent des questionnements.

7. Défis liés au développement de codes et logiciels

- Bonnes pratiques : l'établissement de bonnes pratiques (documentations, etc.) est peu courant.
- Maturité du travail : des chercheur·euses considèrent dans certains cas leurs codes comme pas suffisamment mature pour être partagés.

8. Défis liés au partage des données

- Utilité perçue du partage des données : en particulier pour les données qualitatives, des chercheur·euses mentionnent des problèmes tels que la perte de dimensions si on ne fait pas la collecte soi-même, la spécificité de la question de recherche et la valeur du travail pour d'autres chercheur·euses.





- Protection des données personnelles et sensibles : des chercheur·euses s'inquiètent des risques de réidentification des données anonymisées, ainsi que de l'utilité des données anonymisées.
- Charge de travail pour la préparation des données : la préparation des données représente une charge de travail conséquente, avec notamment une documentation à préparer.
- Propriété des données : les chercheur·euses doivent s'assurer de respecter les droits d'auteurs, licences ou autres contrats.
- Impact environnemental : dans la cadre de l'ouverture de jeux de données très volumineux, les coûts environnementaux devraient être pris en compte.

9. Défis liés à la réutilisation des données FAIR

- Limitations pour la réutilisation de données qualitatives : la réutilisation est plus courante pour les données quantitatives provenant de sources externes, mais limitée dans la recherche qualitative.
- Difficulté à trouver des jeux de données : les chercheurs réutilisent peu de jeux de données FAIR.

10. Besoins exprimés par les chercheur·euses :

- Simplicité et clarté du processus de partage.
- Formation et informations ciblées.
- Reconnaissance et valorisation de l'effort consacré à l'ouverture des données.
- Aide financière et technique pour la préparation des données (par exemple pour l'anonymisation).
- Accompagnement personnalisé et soutenu pour adopter de nouvelles pratiques.
- Besoin de voir l'utilité du partage des données

Les principaux résultats montrent que bien que la connaissance de l'ORD soit généralement répandue, les expériences de partage et de réutilisation des données restent limitées. Les chercheur·euses rencontrent des obstacles variés, allant de l'utilité perçue du partage de données, à la complexité de l'anonymisation et à la charge de travail liée à la préparation des données.

4.2 Recommandations

Plusieurs actions ont déjà été menées ou sont en cours pour répondre à certains de ces défis, tels que :

- Répertorier les actrices et acteurs de l'aide à l'ORD de manière claire sur le site de la HES-SO.
- Répertorier les guides et formations existants, de la même école ou d'autres écoles, pour répondre aux besoins spécifiques des chercheur·euses.
- Proposer des documents types, tels que des fichiers README, des formulaires de consentements, des Data Transfer Agreements (DTA), des modèles de DMP, etc.
- Intégrer la contribution à l'ORD dans les critères d'évaluation des chercheur·euses, en valorisant cet effort, par exemple dans les subventions ou en comptabilisant le temps consacré à l'ouverture des données dans les feuilles de charge.

Toutefois, ces efforts doivent être poursuivis afin de répondre aux nombreux besoins disciplinaires. Les recommandations suivantes sont donc proposées :

Renforcement du soutien :





- Engager un répondant·e ORD pour chaque Haute école afin de fournir un soutien personnalisé et continu. Ce service devrait offrir des consultations individuelles et des ateliers pratiques.
- Développer de nouveaux guides et formations, selon les besoins des chercheur·euses : par exemple sur l'écriture des data papers, sur la coopération pour la valorisation des jeux de données, sur la recherche de jeux de données existants, sur l'ouverture de jeux de données dérivées, etc.
- Créer des feuilles de route claires pour le processus de gestion et d'ouverture des données, incluant des étapes spécifiques et des ressources disponibles.
- Sensibiliser les chercheurs aux solutions de stockage, de sauvegardes, d'archivages et de transferts de leurs données.
- Assister l'écriture du DMP en utilisant des logiciels, tels que le Data Stewardship Wizard (DSW).
- Développer un réseau de partage d'expérience entre chercheur·euses (Data Champions) permettant de trouver rapidement des solutions aux problèmes courants et de bénéficier des conseils de collègues (de la même école ou d'autres écoles)
- Sensibiliser les chercheur·euses aux aspects de propriétés des données, en particulier concernant la propriété partagée et les cas de données dérivées.
- Former les nouveaux arrivant·es.

Valorisation et reconnaissance :

- Valoriser les données partagées des chercheur·euses de la HES-SO, par exemple en référençant les DOI des jeux de données dans ArODES ou en créant une communauté dans Zenodo. Le référencement dans ArODES pourrait constituer une approche plus globale, permettant ainsi aux chercheur·euses de déposer dans le dépôt de leur choix. La création d'une communauté sur Zenodo nécessiterait des ressources pour assurer la curation des jeux de données soumis.
- Sensibiliser les chercheur·euses à la possibilité de publier des « data papers » pour valoriser leurs jeux de données déposés.
- Sensibiliser les chercheur·euses aux avantages du partage des données pour la validation de l'intégrité de leurs travaux et l'équité dans la recherche, via des campagnes de communications et des témoignages de pairs.

Promotion de l'utilité du partage des données :

- Avoir une métrique de la réutilisation des données des chercheur·euses de la HES-SO. Une telle métrique pourrait être obtenue par fouille de données dans les publications pour identifier la réutilisation de jeux de données. Toutefois, la nouveauté du partage des données peut rendre cette analyse complexe.
- Aider les chercheur·euses à identifier des jeux de données pertinents pour leurs recherches, par exemple via des outils tels que Datacite ou OpenAlex.

