



APPELS À PROJETS OPEN DATA 2020 ET 2021

SYNTHÈSE ET *LESSONS LEARNED* DES RAPPORTS
FINAUX

Auteure : Constance Delamadeleine
Date : Avril 2023





TABLE DES MATIERES

I/	INTRODUCTION	3
II/	UN MOMENT « DONNÉES »	4
III/	FAIRE « FAIR »	5
IV/	LES DÉPÔTS DE DONNÉES	8
V/	ASPECTS JURIDIQUES ET ETHIQUES	10
VI/	CONCLUSION	12
	BIBLIOGRAPHIE	13





I. INTRODUCTION

En 2020, le Rectorat a lancé un appel à projets Open Research Data (ORD), renouvelé en 2021, afin de financer l'ouverture des données de recherche issues des projets menés par les chercheurs.euses de la HES-SO. Cet appel à projets s'inscrit dans le cadre de la stratégie nationale Open science, stratégie définie par swissuniversities sur mandat du SEFRI en février 2017¹. Cet appel poursuit deux objectifs principaux :

1. Favoriser l'appropriation de la pratique ORD en menant des réflexions autour du partage des données de recherche, en mettant en commun des outils et des procédures ainsi qu'en dressant un état de l'art spécifique aux domaines de la HES-SO.
2. Disposer, grâce à la pratique, d'une base méthodologique pour la gestion des données propres aux spécificités disciplinaires de la HES-SO.

A l'issue de ces deux appels, le Rectorat a financé au total 13 projets à hauteur de CHF 195'400 en provenance de tous les domaines de la HES-SO :

Design & Arts visuels (1)
Economie et Services (2)
Ingénierie et Architecture (5)
Musique et Arts de la scène (1)
Travail social (2)
Santé (2)

Au terme de leur projet, et conformément aux exigences attendues, les chercheurs.euses ont rédigé un rapport scientifique incluant les sections suivantes :

- Résumé de la recherche
- Description des données (type de données, format, quantité, degré de sensibilité, etc.)
- Bref état de l'art des dépôts de données (Data repositories) existants et une argumentation justifiant le choix final
- Méthode d'implémentation des principes FAIR lors du dépôt
- Etat des lieux des difficultés juridiques et éthiques relatives au partage des données.

Ce document synthétise les réflexions, les problématiques ainsi que les stratégies empruntées par les chercheurs.euses afin de rendre leurs données conformes aux principes FAIR.

Ces rapports sont archivés sur ArODES et disponibles sur la [page Open Research Data du site internet de la HES-SO](#).

¹ swissuniversities a adopté la stratégie Open Data en juillet 2021 et diffusé son plan d'actions en décembre 2021. Plus d'information sur la stratégie et le plan d'action : <https://www.swissuniversities.ch/fr/themes/digitalisation/open-access-campagne/strategie-nationale-et-plan-daction>.





II. UN MOMENT « DONNÉES »

L'une des activités clés de l'ORD consiste à décrire ses données de recherche. Comme l'illustre le tableau ci-dessous, ces données peuvent prendre plusieurs formes telles que des notes d'observation, du code texte, ou encore des plans d'appareil électronique. Ces formes ne sont pas spécifiques à un domaine mais aux méthodes de production.

Domaine	Donnée de recherche
Design et Arts visuels	Images (numérique, Spectra, chemical map) Processus Texte
Economie et Services	Code (texte) Géospatiales Image médicale Texte
Ingénierie et Architecture	Code Code source Instructions Plans Scripts
Musique et Arts de la scène	Texte > entretiens, notes d'observation, questionnaires, transcriptions d'entretiens
Santé	Texte > entretiens, notes d'observation, questionnaires, transcriptions d'entretiens
Travail social	Texte > entretiens, notes d'observation, questionnaires, transcriptions d'entretiens

Tableau 1. Données de recherche collectées et/ou générées par les participants.es des appels à projets listées par domaine.

Pour certains.es chercheurs.euses, décrire ses données est un nouveau processus qui s'accompagne bien souvent de la question : qu'est-ce qu'une donnée de recherche ? Depuis l'apparition de l'ORD, les définitions abondent et fournissent de nombreux exemples de données de recherche². Toutefois, ces définitions ne tiennent pas toujours compte de la réalité de certaines communautés. Les sections suivantes s'intéressent aux réflexions menées par les chercheurs.euses autour de la définition d'une donnée de recherche à l'heure de l'ORD.

2.1 Le journal/carnet de terrain : une donnée de recherche ?

Comme le relève Patricia Perrenoud, la Société Suisse d'Ethnologie (SSE) met en lumière les défis et enjeux posés par l'ORD dans un rapport publié en 2021 (Hertz, 2021 ; Perrenoud, 2022). Le rapport précise que de nombreuses discussions ont été menées autour des carnets de terrain. Certains.es chercheurs.euses considèrent que ce type de données « ne devraient pas être classé comme des « données » du tout » pour différentes raisons (Hertz, p.10). D'une part, leurs contenus sont souvent subjectifs et incompréhensibles pour les autres personnes et d'autre part, ils sont « pratiquement impossibles à anonymiser, ce qui crée de réels risques de

² La définition de l'OCDE (2021) est couramment citée. Voir : <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>.





sécurité pour l'anthropologue et pour ses sujets de recherche » (Hertz, p.10). La SSE encourage « les anthropologues à maximiser les mesures de protection des données pour ce type de données et à ne pas (se sentir obligés de) les mettre à la disposition d'autres chercheurs.euses ou du public par le biais de dépôts de données ou d'autres mécanismes de partage de données » (Hertz, p.10). Ce rapport met en évidence l'impact que l'ORD entraîne sur certaines communautés de recherche.

2.2 La recherche-création : des données uniques et non reproductibles

Du côté des domaines artistiques, la description et la documentation des données produites dans le cadre d'une recherche basée sur la pratique artistique (arts de la scène, arts visuels, design et musique) soulèvent aussi certaines problématiques. Comme le souligne Angelika Gusewell, ces disciplines empruntent « des modes d'abstraction, de systématisation et de réflexivité spécifiques qui se distinguent des modes de théorisation les plus répandus dans le monde scientifique » (2022, p.10). Les données produites peuvent inclure des processus de travail, des œuvres d'art ou encore des idées qui sont des données originales et uniques et qui par conséquent sont difficilement réutilisables. De plus, les données produites « dans le cadre de la recherche « par l'art » », précise Gusewell, « sont souvent incorporées par l'artiste/designer lui-même et ne prennent pas toujours la forme de données archivables, mais de savoir-faire transmissibles uniquement dans des situations pratiques » (2022, p.10).

III. FAIRE « FAIR »

Les principes FAIR « permettent d'assurer qu'un jeu de données puisse être trouvé (« Findable »), qu'il soit accessible (« Accessible »), interopérable (« Interoperable ») et réutilisable (« Reusable »)³. Ces quatre principes directeurs s'accompagnent d'une série d'activités qui se déploie tout au long du projet. Si certains principes sont simples à appliquer, d'autres nécessitent un certain investissement de la part du chercheur.euse.

3.1 FAIRiser : du temps pour sa communauté

La réutilisation des données de recherche est l'un des objectifs de l'ORD dont le succès repose en grande partie sur une riche documentation des données. Il s'agit de fournir des explications détaillées sur le contexte de production des données, sur leur structure ainsi que leur contenu et d'indiquer les manipulations qui ont été effectuées sur les données. Idéalement, tout doit être décrit. En d'autres termes, « FAIRiser » implique de s'investir pour sa communauté scientifique.

³ Fonds national suisse, Open Research Data (ORD) : quels dépôts de données utiliser ?
<https://www.snf.ch/fr/7GhWDP8omTMLZ000/news/news-210122-open-research-data-quels-depots-de-donnees-utiliser>.





3.3.1 Tout documenter : le cas des entretiens

Les données produites dans le cadre d'entretiens sont spécifiquement complexes à documenter. Afin de guider au mieux les futurs.es utilisateur.rices du jeu de données, il s'agit, comme indiqué plus haut, d'inclure un maximum d'informations relatives au contexte, aux processus de recherche, au contenu et à la structure des données. Les rapports fournissent des exemples d'éléments contextuels à intégrer tels que les courriels transmis aux enquêtés.es, les invitations à participer aux focus-groups, les formulaires de consentement, les consignes, le guide d'entretien ou encore les mind-maps. Il est conseillé d'inclure une liste des différentes manipulations et mesures appliquées pour rendre les données « FAIR » telles que la codification ou la suppression de certaines données. Combiné avec des indications sur la structure des fichiers de données et les relations entre les fichiers, cet ensemble d'informations renforce le degré de compréhension du jeu de données et favorise sa réutilisation.

3.3.2 Codebook versus spreadsheet

Le codebook ou livre de codes, fournit des informations et des instructions sur la manière de lire, d'analyser, d'interpréter et de vérifier les données de recherche. C'est un document destiné à faciliter l'interprétation des données à des fins de reproductibilité et de réutilisation. Il explique tous les concepts et/ou variables qui sont présents dans les fichiers de données. Il peut s'agir d'explications sur la signification, les valeurs numériques, l'étendue, les unités de mesure, les sources (des questionnaires par exemple), les schémas de classification utilisés et/ou les étiquettes des questions posées. La rédaction d'un livre de codes est une étape importante dans la gestion de tout projet de recherche, qui servira de référence à l'ensemble de l'équipe et aux futurs.es utilisateurs.rices du jeu de données. Travail chronophage, cette solution de documentation peut être remplacée par un spreadsheet qui est une version simplifiée du codebook (Savioz et al., 2023). Cependant, si le codebook permet d'encoder et de décoder des variables potentiellement identifiantes et donc d'assurer la confidentialité des données personnelles et sensibles, ce n'est pas le cas du spreadsheet. En d'autres termes, dans la perspective de respecter les principes FAIR, il est « cependant clair que le recours à un « spreadsheet » demeure relativement plus limité que celui d'un « codebook » » (Savioz et al., 2023, p.9).

3.3.3 Les standards de métadonnées

La description standardisée des données est également une composante essentielle voire une condition qui permet la réutilisation des données. Certains dépôts de données fournissent des guides d'utilisation pour décrire ses données ou des schémas de métadonnées spécifiques à une discipline. C'est le cas par exemple, de SWISSUbase qui fournit un schéma de métadonnées en sciences sociales. Certaines.es chercheurs.euses se sont basés.es sur des ressources en ligne telles que [DataCite Metadata Schema](#) et [schema.org](#). Pour la gestion des données issues de l'imagerie médicale, DICOM reste la norme de référence.





3.2 Accès immédiat ou restreint?

Certains dépôts de données proposent des paramètres de restriction et de contrôle sur l'accès aux jeux de données. Par exemple, les auteurs.es du jeu de données peuvent imposer un accord préalable au téléchargement de leur jeu de données. Les rapports montrent que ces modalités d'accès font l'objet d'avis divergents. En effet, pour certains.es, restreindre l'accès constitue un obstacle à l'accès immédiat aux données. Pour d'autres, l'accès aux données sur accord préalable « permet de s'assurer de l'existence d'un point de rencontre épistémologique, méthodologique et éthique entre l'équipe de recherche ayant produit les données primaires et celle qui a l'intention de réaliser une analyse secondaire » (Perrenoud, 2021, pp.10-11).

3.3 Open Data = Open format

Les formats conformes aux principes FAIR sont des formats dit « ouverts ». Dans de nombreux cas, il convient de procéder à la conversion d'un format propriétaire vers un format ouvert.

3.3.1 Logiciels commerciaux courants, que faire ?

Le logiciel SPSS est largement utilisé en sciences sociales et en psychologie. Toutefois ce logiciel est payant. Il est conseillé d'enregistrer les données quantitatives issues de questionnaires, importées dans SPSS au format .sav puis de les importer soit dans le logiciel gratuit R ou GNU PSPP, très similaire à SPSS.

Largement utilisés, les logiciels Microsoft sont des formats commerciaux qui ne sont pas compatibles avec les principes FAIR. Deux alternatives pour le traitement de texte sont suggérées :

1. convertir tous les fichiers .docx en .pdf > Adobe Acrobat Reader DC est disponible gratuitement
2. utiliser des logiciels non commerciaux et/ou open source, tels que Notepad++ pour la création de fichiers texte .csv, librement accessibles et modifiables.

Quant aux fichiers .xlsx (Excel), il s'agit de les convertir au format .pdf ou .csv.

3.3.2 Imagerie médicale : conversion au format standard DICOM

Pour les images médicales, DICOM est un format standard qui assure l'interopérabilité des données. Toutefois, si la recherche nécessite de segmenter et de générer des ROIs (Region of interest), cela implique la conversion des images DICOM au format NiftI (les fichiers ROIs sont créés au format NiftI). Afin d'harmoniser les formats de données, les ROI NiftI devaient être reconvertis au format de segmentation DICOM. Une solution proposée est la suivante: l'utilisation d'un package Python appelé pydicom-seg (Depeursinge, 2022, p. 9).





IV. LES DÉPÔTS DE DONNÉES

Les chercheurs.euses ont dressé un état de l'art des dépôts de données conformes aux principes FAIR. Pour ce faire, ils.elles se sont appuyés.es sur différentes sources dont le rapport de monitoring du Fonds national suisse pour la recherche scientifique (FNS) sur l'Open Research Data publié en 2020 ou encore le travail de Master de Marielle Guirlet (2020)⁴. D'autres ont consulté des registres de dépôts de données tels que CatOPIDoR, FAIRsharing ou encore re3data⁵. Recommandé par le programme Horizon Europe de la Commission Européenne, re3data recense plus de 2500 dépôts.

4.1 Etat de l'art: SWISSUbase et Zenodo en tête de liste

Au total, une vingtaine de dépôts émergent de ces états de l'art. Sont listés des dépôts institutionnels tels que l'ETH Data archive et ETH Zürich Research Collection. Ces derniers sont d'emblée exclus de la sélection car ils sont réservés exclusivement aux chercheurs.euses de l'institution tout comme Yareta de l'Université de Genève qui a la particularité d'être destiné aux hautes écoles genevoises. Parmi les autres dépôts cités, certains sont à orientation disciplinaires :

- Domaines artistiques : Global Performing Arts database, Research catalogue
- Sciences de la vie/sciences naturelles : AHEAD, CARIBIC, Cancer imaging archive, EBRAINS, ELI ERIC/Open, Health on the Net Foundation.
- Sciences humaines : DaSCH, Sammlung Schweizerischer Rechtsquellen online
- Sciences sociales : FORSbase devenu SWISSUbase.

Enfin, des solutions généralistes sont également mentionnées telles que Dryad, Figshare, Harvard Dataverse, Open Science Framework, Olos, SWISSUbase ou encore Zenodo.

Sur l'ensemble de ces dépôts seulement 4 ont été finalement sélectionnés par les chercheurs.euses, avec une large majorité en faveur de solutions généralistes : SWISSUbase (4) et Zenodo (7)⁶. Le tableau ci-dessous répertorie les critères sur lesquels les chercheurs.euses se sont basés.es pour sélectionner ces dépôts. La gratuité ainsi que localisation nationale du serveur figurent parmi les critères les plus fréquemment cités.

⁴ <https://www.snf.ch/fr/W/tezJ6qxuTRnSYgF/dossier/open-research-data-quels-depots-de-donnees-peuvent-etre-utilises>.

⁵ Liens vers les sites internet : https://cat.opidor.fr/index.php/Entrepôt_de_données; <https://fairsharing.org/>;
<https://www.re3data.org/>.

⁶ En 2021 FORSbase devient SWISSUbase. Initialement dédié aux sciences sociales, SWISSUbase est aujourd'hui un dépôt généraliste.





Nom du dépôt	Domaine des chercheurs.euses ayant choisi ce dépôt	+	-
Cancer imaging archive	<ul style="list-style-type: none"> Economie et Services 	<ul style="list-style-type: none"> Dédié à l'imagerie médicale Prend en charge les fichiers DICOM 	
FORSbase (devenu SWISSUbase en 2021)	<ul style="list-style-type: none"> Musique et Arts de la scène Santé Travail social 	<ul style="list-style-type: none"> Assistance pour la curation⁷ Gratuit (pour les chercheurs.euses en sciences sociales) Expertise en sciences sociales Interface centrée-utilisateur.rice Sécurité optimale Serveur situé en Suisse 	<ul style="list-style-type: none"> Peu de visibilité internationale
Materials Cloud Archive	<ul style="list-style-type: none"> Ingénierie et Architecture 	<ul style="list-style-type: none"> Bonne visibilité dans le paysage suisse Dédié au Computational Materials Science Dépôt certifié (Seal of approval) Serveur situé en Suisse 	
Zenodo	<ul style="list-style-type: none"> Design et Arts visuels Economie et Services Ingénierie et Architecture Travail social 	<ul style="list-style-type: none"> Contient une catégorie « physical object » Gratuit Interface ergonomique Lien avec Github (code) Permet de créer ou rejoindre une communauté Plateforme recommandée par le FNS Possibilité de déposer plusieurs jeux de données Serveur situé en Suisse 	<ul style="list-style-type: none"> Pas de manuel d'utilisation

Tableau 2. Dépôts de données sélectionnés par les participants.es de l'appel à projet (2020/2021).

⁷ Certains dépôts de données proposent un service de curation. Par exemple FORS, le centre d'expertise en sciences sociales fournit une aide à la communauté en sciences sociales pour le dépôt des données sur SWISSUbase. Le service procède entre autre, à une vérification du jeu de données et communique les points à modifier, corriger si nécessaire certains aspects non compatibles ou conformes aux principes FAIR. [Voir le rapport d'Angelika Güsewell qui inclut une liste des commentaires communiqués par FORS à la suite du dépôt de son jeu de données \(2022, pp. 15-16\).](#)





4.2 Les plateformes Open Source et Github

Deux projets du domaine Ingénierie et Architecture ont déposé une partie de leur jeu de données sur des plateformes collaboratives (Open Source)⁸. Cette stratégie permet de mettre à disposition certaines (meta)données telles que les plans de constructions et code source de logiciel relatifs à la création d'objets physiques à des communautés spécifiques (les fabricants/designers) soit à un plus large public selon les principes de la science participative.

Une pratique également récurrente lorsque le jeu de données contient du code, consiste à déposer le code sur Github et le reste du jeu de données sur Zenodo⁹.

V. ASPECTS JURIDIQUES ET ETHIQUES

Les pratiques de l'ORD sont soumises aux contraintes juridiques et éthiques. Les domaines de la santé et des sciences sociales sont majoritairement concernés par ces aspects. Le traitement des données sensibles et personnelles collectées lors d'entretiens est soumis à la Loi fédérale sur la Protection des données (LPD) ainsi qu'aux lois cantonales. Afin de se conformer aux exigences légales et éthiques, il s'agit de mettre en place des mesures techniques et organisationnelles de protection telles que l'anonymisation, la pseudonymisation ou le consentement éclairé. Ces mesures peuvent-être combinées.

5.1 Les entretiens

5.1.1 Les identifiants sous microscope

Certaines disciplines ou domaines mobilisent des méthodes (enquêtes, interviews, observations etc.) qui génèrent des données personnelles et/ou sensibles. Afin de minimiser les risques d'identification des personnes, ces données sont soumises à des traitements particuliers. Les identifiants directs et indirects sont la cible de ces traitements. Il s'agit dans un premier temps de les repérer, puis de les coder ou de les supprimer en utilisant les techniques d'anonymisation ou de pseudonymisation ou les deux. Le protocole de traitement dépend de plusieurs facteurs tels que le nombre et la précision des variables, l'hétérogénéité du corpus traité (plusieurs documents, plusieurs grilles de questionnement) ou encore l'hétérogénéité des supports (papier, numérique). Le codage ou la suppression des identifiants peuvent dans certains cas, devenir un processus extrêmement complexe et chronophage à travers lequel il s'agit de rechercher un équilibre entre partager autant de données que possible et respecter la confidentialité des enquêtés.es.

⁸ Pour plus d'information, voir les rapports suivants : [Roland, J. & Poulin, M. \(2022\). AXO – An immediate post-operative prosthesis for transtibial amputees](#) ; [Vorlet, O. \(2021\). Développement d'un dispositif performant, écologique et économique de contrôle de la qualité des médicaments dans les pays émergents.](#)

⁹ Le FNS suggère de créer « une connexion entre Zenodo et Github pour assurer un enregistrement permanent de votre code et le rendre accessible ». Quels dépôts de données peuvent être utilisés ? <https://www.snf.ch/fr/WtezJ6qxuTRnSYgF/dossier/open-research-data-quels-depots-de-donnees-peuvent-etre-utilises>.





L'anonymisation est une opération irréversible qui élimine tout identifiant direct d'un ensemble de données. Il existe plusieurs stratégies d'anonymisation telles que l'anonymisation par généralisation qui consiste à modifier 'l'échelle des attributs des jeux de données, ou leur ordre de grandeur, afin de s'assurer qu'ils soient communs à un ensemble de personnes' (CNIL, 2020) (Benz & Thiévent, 2021, p.12). Par exemple, dans le cas d'une enquête incluant comme variable le lieu de résidence, cette stratégie vise à « diminuer la précision du lieu de résidence et de la nationalité par un recodage du lieu de résidence en districts de résidence (Porrentruy, Delémont, Franches-Montagnes) et de la nationalité en trois zones géographiques (Suisse, pays de l'Union européenne, pays hors Union européenne) » (p. 12). Toutefois, l'anonymisation par généralisation peut se être insuffisante et doit être accompagnée de mesures complémentaires notamment lorsque l'enquête contient des questions ouvertes¹⁰. Afin de diminuer les risques d'identification, plusieurs stratégies « ont été mise en œuvre en fonction des variables concernées : (1) la suppression de toutes les valeurs d'une variable string, (2) le recodage des variables string de type « autre » en variables binaires et (3) la suppression sélective de certaines valeurs des variables string permettant un long développement » (Benz & Thiévent, 2021, pp.12-13)¹¹. Une des mesures de sécurité additionnelle est la relecture et le contrôle de toutes les valeurs des questions ouvertes.

La pseudonymisation est appliquée dans certains cas. Cette option est largement conseillée en cas d'absence de variables directement identifiantes (nom et prénom, adresse, etc.) et peut-être combinée avec l'anonymisation afin de renforcer le degré de sécurité. Une autre stratégie de sécurité est celle du consentement éclairé. Procédure parfois fastidieuse à mettre en place, elle permet cependant, d'obtenir l'autorisation des enquêtés.es de partager leurs données personnelles lorsque l'anonymisation n'est pas possible.

5.1.2 Le consentement

Généralement, les modèles de formulaire de consentement ne contiennent pas d'item « partage des données ». Il est conseillé d'ajouter cet item dans le formulaire de consentement afin d'informer les enquêtés.es de cette nouvelle procédure et ainsi les familiariser le plus tôt possible avec cette culture de la science ouverte (Perrenoud, 2021, p.7). Des chercheurs.euses ont notamment souligné la difficulté de fournir une explication claire et intelligible quant aux usages qui pourraient être faits de ces données lors d'analyses secondaires. Par ailleurs, il est recommandé de procéder à la codification et au nettoyage des données le plus tôt possible après la transcription des entretiens. En effet, à l'issue de ce processus, les enquêtés.es sont recontactés.es en vue de vérifier le consentement. Une charge de travail supplémentaire s'imposera si les enquêtés.es changent de coordonnées entre la période d'entretien et le traitement des données personnelles et sensibles.

Le domaine Santé dont « la recherche qualitative est soumise à des procédures éthiques qui sont dérivées de celles utilisées dans la recherche médicale et la recherche quantitative » est

¹⁰ Les contrôles assurés par le service de curation de FORS lors des dépôts des jeux de données ont permis dans certains cas, de révéler des identifiants malgré l'anonymisation effectuée par les chercheurs.euses.

¹¹ Une « variable string » contient des valeurs qui sont traitées comme du texte. Cela signifie que ces variables peuvent inclure des chiffres, des lettres ou des symboles.





confronté au problème du consentement unique (Perrenoud, 2021, p. 13). Si ce modèle fonctionne dans certains cas et pour certaines méthodes, pour d'autres, « [l]e consentement est recherché de manière itérative lors de la planification de nouveaux entretiens ou observations ou lorsque l'évolution de la recherche change les conditions qui avaient été convenues » (p. 13). En d'autres termes, le consentement nécessite d'être vérifié ou modifié ultérieurement. Le modèle unique n'est donc pas approprié.

5.2 Les images

Si un jeu de données contient des images de personnes physiques, il s'agit de retirer toutes les métadonnées susceptibles de contenir des informations personnelles ou d'identification. L'image peut-être soit retirée du jeu de données ou alors éditée à l'aide d'un logiciel de traitement afin de supprimer tout identifiant. Par exemple, l'image peut-être tronquée ou floutée.

VI. CONCLUSION

Le FNS souligne la contribution essentielle du partage des données de recherche « à la recherche scientifique en termes d'impact, de transparence et de reproductibilité »¹². Si l'ORD présente de nombreux avantages, l'analyse des rapports montre que ce nouveau paradigme s'accompagne de transformations notoires dans la pratique du chercheur.euse et soulève de nombreux enjeux d'ordres scientifiques, éthiques, juridiques, techniques ou encore économiques.

Tout d'abord, rendre les données de recherche conformes aux principes FAIR nécessite des expertises multiples notamment juridiques et informatiques. Ensuite, il s'agit d'un processus chronophage qui a un fort impact sur le coût du projet. Enfin, et comme le souligne Patricia Perrenoud, le partage des données pose des défis d'ordre épistémologiques et méthodologiques car ces processus ont été « [p]ensés au préalable pour des champs de recherche travaillant à partir d'épistémologies et de méthodes quantitatives » (p. 11). Si les principes « FAIR » s'appliquent aux données issues d'un large éventail de méthodologies de recherche, ils mettent l'accent sur la standardisation et la reproductibilité, deux principes qui sont au cœur de la recherche quantitative. La manière dont les connaissances sont produites dans la recherche qualitative et les conditions dans lesquelles les données sont créées ne sont pas toujours compatibles avec les principes « FAIR ». Par exemple, les données qualitatives issues d'enquêtes ou d'observations sont la plupart du temps, liées au contexte dans lequel elles ont été obtenues. Ces informations contextuelles qui contiennent bien souvent des valeurs subjectives mais aussi des données personnelles sont difficilement partageables. Leur suppression affecterait l'interprétation du jeu de données, et le rendrait potentiellement inutilisable. Une autre crainte formulée par les chercheur.euses qui ont recours aux entretiens, est celle du risque non négligeable de perdre l'accès au terrain. En effet, les personnes enquêtées pourraient être moins disposées à participer à des projets dont les données seront ouvertes et réutilisées par d'autres.

¹² <https://www.snf.ch/fr/dMILj9t4LNk8NwyR/dossier/open-research-data>.





Des réflexions sont à poursuivre afin de trouver un terrain d'entente entre les exigences de l'ouverture des données et celles des chercheurs.euses « qualitatifs.ves ».

Malgré les nombreux obstacles rencontrés, les chercheurs.euses ont répondu avec succès aux exigences relatives à l'ouverture des données.

BIBLIOGRAPHIE

Benz, P. & Thiévent, R. (2021). Mise en Open Data des données de l'enquête auprès des enfants, des jeunes et des acteurs jeunesse : rapport scientifique à la HES-SO. Lausanne, Haute école de travail social et de la santé Lausanne.

Cotting, A. & Calbimonte, J-P. (2021). OpenDataSyris = Open Data pour Syris : système de recommandation pour itinéraires pédestres de santé. Sierre, HES-SO Valais-Wallis.

Degrigny, C., Joseph, E., Brambilla, L., Siatou, A. & Russo, S. (2021). Conditions d'un dépôt optimal en libre accès des données de deux thèses de doctorat du projet ITN-CHANGE. Neuchâtel, HE-Arc Conservation-Restauration.

Depeursinge, A. & Schaer, R. (2022). Quality assessment for interoperable quantitative CT imaging (QA4IQI) - Open access to standardized quantitative imaging. Rapport final – Appel à projets Open Data HES-SO 2021. Sierre, HES-SO Valais-Wallis.

Guirlet, M. (2020). Ouverture des données de recherche dans le domaine académique suisse: outils pour le choix d'une stratégie institutionnelle en matière de dépôt de données. *Ressi*, 21. <http://www.ressi.ch/num21/article182>.

Güsewell, A. (2022). Musicdrops@work. Impact of short live music intervention in the working context. Rapport final - Appel à Projets Open Data HES-SO 2021. Lausanne, Haute école de musique.

Guzman Villegas-Frei, M., Jubin, J. & Oulevey Bachmann, A. (2021). Partage des données de recherche: Un enjeu capital pour une poursuite durable du développement des connaissances. Lausanne, Institut et Haute Ecole de la Santé La Source.

Hertz, E. (2021). Open Science and Data Management in Anthropological Research. Executive Summary of the Swiss Anthropological Association.

Lempesis, N. (2021). Bio-inspired omniphobic surfaces (BIOS) : process development for producing at scale omniphobic and drag free plastic surfaces. Fribourg, Haute école d'ingénierie et d'architecture.

Marti, R. (2022). NCCR CATALYSIS. Final report – Appel à projets Open Data HES-SO 2021. Fribourg, Haute école d'ingénierie et d'architecture.





Meteier, Q., Mugellini, E. & Carrino S. (2022). Advitam – Adaptive driver vehicle Interaction to make Future driving safer. Rapport final – Appel à projets Open Data HES-SO 2021. Fribourg, Haute école d'ingénierie et d'architecture.

Perrenoud, P. (2021). Rapport concernant les procédures et processus utilisés pour le partage de données qualitatives sur un DATA repository : Projet Open Data 2020 HES-SO. Lausanne, Haute Ecole de Santé Vaud.

Roland, J. & Poulin, M. (2022). AXO – An immediate post-operative prosthesis for transtibial amputees. Rapport final – Appel à projets Open Data HES-SO 2021. Fribourg, Haute école d'ingénierie et d'architecture.

Savioz, A., Kurt, S. & Borrelli, L. M. (2023). La création d'un "spreadsheet", le traitement d'une grande quantité des données empiriques et diversifiées : rapport final - appel à projets Open Data HES-SO 2021. Sierre, HES-SO Valais-Wallis, Haute École et École supérieure de travail social.

Vorlet, O. (2021). Développement d'un dispositif performant, écologique et économique de contrôle de la qualité des médicaments dans les pays émergents. Fribourg, Haute école d'ingénierie et d'architecture.

