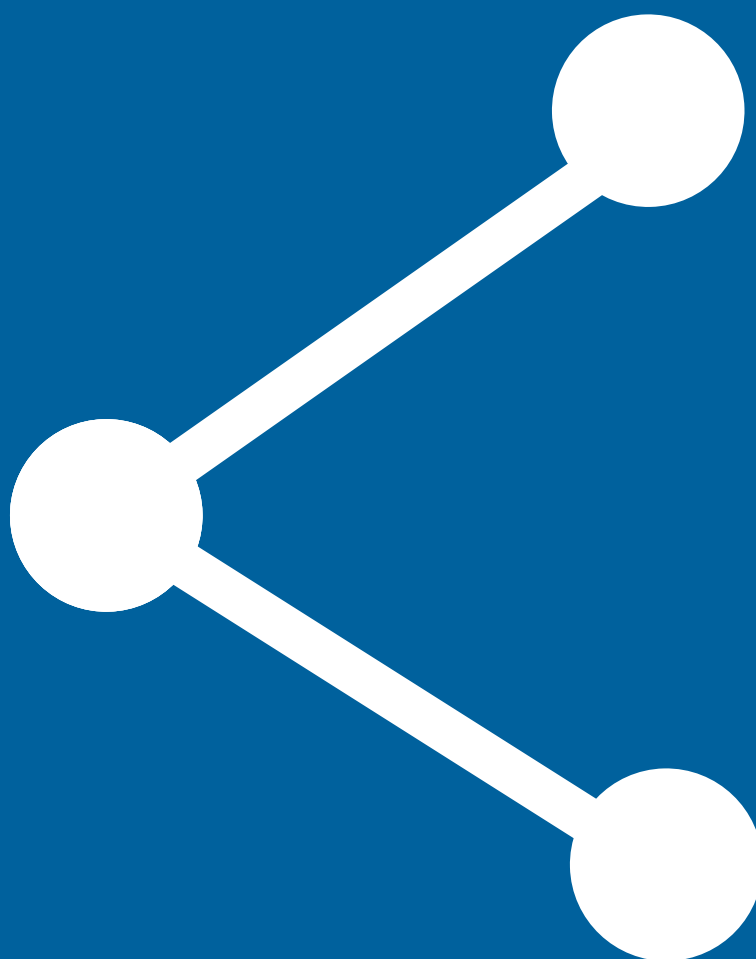
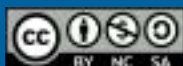


GUIDE RAPIDE SUR LA GESTION DES DONNÉES DE RECHERCHE



Constance Delamadeleine
Chargée de projet Open Research Data
Dicastère Recherche et Innovation



Selon le Fonds national suisse pour la recherche scientifique (FNS), «le libre accès aux données de la recherche (Open research Data) découle de l'idée que le partage des données de recherche apporte une contribution essentielle à la recherche scientifique en termes d'impact, de transparence et de reproductibilité»¹.

Si l'Open Research Data (ORD) représente un changement majeur et bouleverse les pratiques des chercheurs.euses, ses avantages sont multiples.

7 Bonnes raisons de partager ses données de recherche :

1. Augmente la visibilité de la recherche et par là-même vos résultats de recherche jusqu'à 25%²
2. Facilite l'accès aux données de recherche
3. Encourage les collaborations scientifiques
4. Améliore la transparence et la fiabilité de la recherche
5. Évite les pertes de données
6. Évite de faire ce qui a déjà été fait
7. Démocratise le savoir.

La HES-SO encourage la diffusion, la publication et le partage des données de la recherche et les codes logiciels en libre accès, selon le principe « aussi ouvert que possible, aussi fermé que nécessaire ».

Certaines données ne peuvent pas être partagées pour différentes raisons telles que la protection de la vie privée, la confidentialité commerciale, la sécurité, le caractère « sensible » des données, qui empêchent une ouverture complète de ces données.

¹ Fonds national suisse (2023). <https://www.snf.ch/fr/dMILj9t4LNk8NwyR/dossier/open-research-data>.

² Le partage des données est associé à une augmentation des citations des articles publiés allant jusqu'à 25%. Voir : Colavizza, G., Hrynaszkiewicz, L., Staden, L., Whitaker, K., McGillivray, B. (2020). The citation advantage of linking publications to research data. Plos One, 15(4): e0230416. <https://doi.org/10.1371/journal.pone.0230416>

DONNÉES DE RECHERCHE : DÉFINITION, TYPE ET EXEMPLES DE DONNÉES

La recommandation mise à jour en 2021 du Conseil concernant l'accès aux données de la recherche financée sur fonds publics de l'OCDE, définit les données de recherche comme suit: « documents factuels (tels que des notes numériques, des documents textuels, des images et des sons) résultant d'une recherche partiellement ou totalement financée par des fonds publics, utilisés comme sources primaires pour la recherche scientifique et communément acceptés dans la communauté scientifique comme nécessaires pour valider les résultats de la recherche»³.

Exemples de données de recherche

Ces données peuvent-être brutes ou traitées

- Base de données
- Code, algorithme
- Croquis (dessins, notes dessinées)
- Data mining (résultats)
- Document textuel
- Enregistrement audio
- Enregistrement vidéo
- Géospatiale
- Mesures
- Objet physique (e.g. œuvres d'art, prototype, textile)
- Photographie
- Protocole
- Questionnaire
- Rapport
- Sondage
- Statistiques
- Transcription

N'est pas considéré comme données de recherche⁴

- Analyses préliminaires
- Annexe de publication (ex: graphique ou image)
- Communications personnelles avec des collègues
- Examens par les pairs
- Fichier lié à l'administration du projet
- Programmes de travaux futurs
- Projets de documents scientifiques
- Texte issu de publication

Types de données par méthode de production⁵

- **Observationnelles :**
Données capturées en temps réel, elles sont uniques et irremplaçables.
Ex: capture vidéo, neuro-imagerie, photographie, sondage.
- **Expérimentales :**
Ces données sont produites avec des instruments de laboratoire ou des méthodes standardisées.
Ex : séquençage génomique.
- **Simulation/modélisation :**
Ces données sont produites par des modèles expérimentaux.
Ex: modèles économiques ou climatologiques, tableau peint par un robot.
- **Dérivées/compilées :**
Ces données sont issues d'un travail effectué sur des données primaires.
Ex : base de données compilées.
- **De référence ou canoniques :**
Ces données prennent la forme de corpus, généralement publiés et édités, de contenus faisant référence dans un domaine.
Ex : bases de données de gènes, bases de données d'images anciennes, collections d'archives.
- **Code informatique**

³ OCDE (2021). Recommandation du Conseil concernant l'accès aux données de la recherche financée sur fonds publics. www.legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0347

⁴ Ibid.

⁵ Université de Genève. www.unige.ch/researchdata/fr/planifier/identifier-donnees-de-recherche/.

GÉRER SES DONNÉES DE RECHERCHE

La gestion des données de recherche ou Research Data Management (RDM), concerne l'organisation, la documentation, le stockage, l'archivage et le partage des données. La gestion des données s'applique tout au long du projet de recherche, elle fait partie du processus de recherche.

Ce guide propose une liste (non exhaustive) d'activités clés détaillées ci-dessous en fonction des phases du projet de recherche.

Préparation de la requête

1. PLANIFIER



- Établir le DMP
- Prévoir les coûts (gestion des données)
- S'accorder sur la propriété et le partage des données (si partenariat)
- Prendre contact avec la commission d'éthique du canton concerné (si projet de recherche entrant dans le champ d'application de la LRH)
- Explorer les jeux de données existants



Pendant le projet

2. COLLECTER ET ANALYSER



- Organiser et nommer vos fichiers
- Documenter et décrire ses données
- Créer un Readme File
- Partager, sauvegarder, et stocker de façon sécurisée
- Évaluer les aspects, les enjeux juridiques, et éthiques

3. ÉVALUER, TRAITER, ET STOCKER



- Conserver/Détruire
- Anonymiser/Pseudonymiser
- Formater les fichiers (formats ouverts)
- Stocker les données qui ne seront pas partagées



Partage des données

4. PARTAGER, PRÉSERVER, ET VALORISER



- Choisir un Data Repository FAIR
- Ajouter un identifiant unique et durable (ex : DOI)
- Définir les droits d'accès
- Définir le délai d'embargo
- Choisir la licence
- Valoriser votre jeu de données avec un data paper



1. PLANIFIER

LES PRINCIPES FAIR⁶

Les principes FAIR « permettent d'assurer qu'un set de données puisse être trouvé (« Findable »), qu'il soit accessible (« Accessible »), interopérable (« Interoperable ») et réutilisable (« Re-usable »)⁷. Il est nécessaire de prendre connaissance de ces différents principes dès la planification du projet.

FINDABLE	ACCESSIBLE	INTEROPERABLE	REUSABLE
<ul style="list-style-type: none"> Les (méta)données doivent être liées à un identifiant unique (ex : DOI) Les données doivent être accompagnées de métadonnées complètes Le(s) jeu(x) de données est déposé sur une plateforme indexée Les métadonnées sont utilisées pour construire des index d'ensembles de données facilement consultables. 	<ul style="list-style-type: none"> Les (méta)données peuvent être retrouvées par leur identifiant en utilisant un protocole de communication standardisé, ouvert et gratuit (ex : HTTP, FTP). Le protocole prévoit une procédure d'authentification et d'autorisation, lorsque c'est nécessaire. Les métadonnées sont accessibles, même lorsque les données ne sont pas ou plus disponibles. 	<ul style="list-style-type: none"> Les (méta)données utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances. Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR. 	<ul style="list-style-type: none"> Les (méta)données sont richement décrites avec une pluralité d'attributs précis et pertinents. Les (méta)données sont publiées avec une licence (e.g. Creative Commons). Les (méta)données sont associées à leur provenance. Les (méta)données répondent à des normes communautaires pertinentes pour le domaine.

OUTILS :

- [Guide FAIR⁸](#)
- [Outil d'évaluation FAIR de mes données⁹](#)
- [Présentation interactive et ludique des principes FAIR de DoRANum¹⁰](#)

⁶ <https://force11.org/info/the-fair-data-principles/>; https://www.snf.ch/media/fr/s1eRHVxx1Y41S5le/FAIR_principles_translation_SNSF_logo.pdf.

⁷ <https://www.snf.ch/fr/7GhWDP8omTMLZ000/news/news-210122-open-research-data-quels-depots-de-donnees-utiliser>.

⁸ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. (2016). Sci Data 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

⁹ Australian Research Data Commons. <https://ardc.edu.au/resource/fair-data-self-assessment-tool/>.

¹⁰ DoRANum. view.genial.ly/5d64fbbd83352350fa3d22603.

ETABLIR UN DATA MANAGEMENT PLAN (DMP)

Le DMP est un document écrit qui décrit la manière dont les données de recherche seront gérées avant, pendant et après le projet. Ce document est une Roadmap de recherche, un instrument qui garantit une bonne gestion des données.

Il permet de gagner du temps et d'anticiper certaines problématiques (ex. coûts, destruction ou perte des données, infrastructure, etc.)

QUE CONTIENT LE DMP ?

La description des données de recherche conformément aux principes FAIR fait partie intégrante d'un DMP. Le DMP doit notamment contenir des informations sur les points suivants :

- Collecte et documentation des données
- Questions éthiques, juridiques et de sécurité
- Stockage et conservation des données
- Échange et réutilisation des données.

QUAND ÉTABLIR UN DMP ?

De plus en plus de bailleurs de fonds exigent de remplir un DMP lors du dépôt de la requête tels que le FNS, HORIZON Europe, AXA Research Fund, U.S Federal Grants, Ligue genevoise, vaudoise contre le cancer, etc.

OUTILS :

- Check-list du [DLCM](#)¹¹
- Modèles de DMP [FNS](#) ou [Horizon Europe \(ERC\)](#)¹²

Le DMP est un document vivant, il est recommandé de régulièrement le mettre à jour tout au long du projet. Le DMP est un outil qui permet d'assurer l'implémentation des principes FAIR

DMP et FNS

Le DMP est maintenant requis seulement pour les projets approuvés.
Par ailleurs le FNS invite les chercheurs·euses à actualiser le DMP.
Une version finale sera demandée à la fin du subside.

PRÉVOIR LES COÛTS RDM

Fonds national suisse (FNS) : jusqu'à 10'000.- pour la préparation et l'archivage des données sous condition que le Data Repository (DR) soit géré par une organisation à but non commercial.

Horizon Europe : l'éligibilité au remboursement est définie dans la convention de subvention type¹³.

OUTIL :

- [Calculer les coûts RDM](#)¹⁴

¹¹ La HES-SO est partenaire du projet DLCM (Data Life Cycle Management). Ce projet vise en partie, à conseiller les chercheurs·euses (affiliés.es aux institutions membres) en matière de DMP, voir : <https://www.dlcm.ch/resources/dlcm-dmp>.

¹² Modèle du FNS : <https://www.dlcm.ch/blog/dlcm-template-snsf-data-management-plan>;

Modèle du programme Horizon Europe : https://www.hes-so.ch/fileadmin/documents/HES-SO/Documents_HES-SO/pdf/open-science/data-management-plan_he_en.docx.

¹³ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2023-2024/wp-13-general-annexes_horizon-2023-2024_en.pdf

¹⁴ <https://dam.ukdataservice.ac.uk/media/622368/costingtool.pdf>.

S' ACCORDER SUR LA PROPRIÉTÉ ET LE PARTAGE DES DONNÉES

Selon la convention intercantonale sur la haute école spécialisée de suisse occidentale (HES-SO), les hautes écoles « sont titulaires des droits de propriété intellectuelle portant sur toutes les créations intellectuelles ainsi que les résultats de recherches obtenus dans l'exercice de leurs fonctions par les personnes ayant une relation contractuelle de travail avec ces dernières »¹⁵. Toutefois, dans le cas d'un partenariat avec une autre institution suisse ou étrangère, ou une entreprise, il est essentiel de s'accorder dès la planification du projet, sur la propriété et le partage des données.

OUTIL :

- Pour plus d'information sur la propriété intellectuelle des données, voir le guide « [Ouverture des données de la recherche. Les grands principes juridiques et les recommandations de la HES-SO](#) ».

PROJET DE RECHERCHE ET LRH

« Dans le cas où la recherche tombe sous le coup de la Loi fédérale relative à la recherche sur l'être humain (LRH), généralement des études menées dans un contexte biomédical, les chercheur·e·s doivent obligatoirement prendre contact avec la commission d'éthique pour soumettre leur protocole et s'assurer de la conformité légale de leur formulaire de consentement »¹⁶.

EXPLORER LES JEUX DE DONNÉES EXISTANTS: VERS LA RÉUTILISATION DES DONNÉES

La réutilisation des données de recherche (jeux de données) est l'un des principes FAIR. Lors de l'écriture du projet, il est recommandé d'établir un état des lieux des jeux de données en lien avec votre discipline pour les raisons suivantes :

- Les données existantes peuvent faire partie de la base sur laquelle votre recherche s'appuiera
- Vous pouvez examiner le besoin de nouvelles données
- Si vous envisagez de collecter ou de générer des données, vous devez vous assurer à un stade précoce que les mêmes données n'ont pas déjà été collectées
- Vous pouvez découvrir s'il existe des normes et des standards à suivre pour la structure, l'étiquetage et la documentation.

Les chercheurs.euses ont l'entière responsabilité de la gestion des données.

Dans le cas d'une équipe de recherche, il est recommandé de définir lors de la planification la ou les personnes qui sera(ont) en charge de la gestion des données.

¹⁵ Convention intercantonale sur la Haute école spécialisée de suisse occidentale (HES-SO) du 26 mai 2011. Voir : <https://www.hes-so.ch/la-hes-so/a-propos/reglements-et-procedures>.
¹⁶ Cruchon, B.; Lucas, I. (2023). Ouverture des données de la recherche. Les grands principes juridiques et les recommandations de la HES-SO.



2. COLLECTER / ANALYSER

ORGANISER ET NOMMER VOS FICHIERS

Une bonne organisation des données vous permet de rapidement identifier et localiser vos données. Afin de garantir l'accessibilité des documents, éviter les problèmes techniques, harmoniser les pratiques et faciliter la recherche, il est conseillé de fixer par écrit l'ensemble des règles de nommage des fichiers ainsi que les modalités d'accès aux fichiers.

OUTIL :

- Guide pour nommer les fichiers [File naming guidelines](#)¹⁷

DOCUMENTER ET DÉCRIRE SES DONNÉES

La documentation et les métadonnées sont deux choses différentes. Si la documentation est destinée à être lue par des humains, les métadonnées sont conçues davantage pour le traitement par machine que pour la lisibilité humaine. Toutefois, les métadonnées servent de documentation.

DOCUMENTATION

La documentation consiste à expliquer comment vos données ont été créées, quel est le contexte dans lequel elles s'inscrivent, leur structure et leur contenu, ainsi que toutes les manipulations qui ont été effectuées. Toutes les données collectées ou générées doivent être documentées lors de votre projet. Il est également recommandé de documenter tous les formats de fichiers utilisés. La documentation des données fait partie du processus de conservation des données.

LES MÉTADONNÉES

Il s'agit d'un ensemble d'informations structurées qui permet de décrire les caractéristiques d'une donnée ou d'un jeu de données. Leur rôle vise à faciliter la recherche, la gestion et la réutilisation des données. Ces métadonnées peuvent être descriptives (ex : titre, auteur, date du projet, etc.), techniques (ex : formats utilisés, etc.) ou de droits d'usage et d'accès (ex : titulaires des droits). Il est recommandé de créer les métadonnées le plus tôt possible.

OUTILS :

- [DataCite](#) fournit une liste d'exemples des schémas de métadonnées¹⁸
- [Digital Curation Center](#) fournit des standards de métadonnées par disciplines¹⁹
- [Massive File Renamer](#) vous permet de renommer vos fichiers²⁰

Les métadonnées sont indispensables à l'implémentation des principes FAIR.

¹⁷ University of British Columbia, <https://researchdata.library.ubc.ca/plan/organize-your-data/>.

¹⁸ <https://schema.datacite.org/>.

¹⁹ <https://www.dcc.ac.uk/guidance/standards/metadata>.

²⁰ <http://ivanrf.com/en/portfolio-posts/massive-file-renamer/>.

README FILE

Ce document est une forme de documentation qui contient des métadonnées. Il explique le contenu et la structure de votre ensemble de données, et donne suffisamment d'informations pour qu'un utilisateur potentiel puisse déterminer si les données l'intéressent ou non. Si votre ensemble de données nécessite un livre de codes (Codebook), celui-ci peut être inclus dans le fichier Readme. Vous pouvez également créer des fichiers Readme secondaires dans des sous-dossiers pour documenter des parties spécifiques de vos données.

BONNES PRATIQUES :

- Placer ce fichier idéalement au tout premier niveau de l'arborescence avec un titre commençant par « ReadmeFile »
- Choisir un format .txt brut ou .pdf/a
- Inclure les éléments qui sont utiles et/ou nécessaires pour interpréter, évaluer et réutiliser correctement votre ensemble de données.

OUTILS :

- [Lien vers un modèle de Readme File²¹](#)
- [Lien vers un guide pratique²²](#)

CODEBOOK

Le Codebook ou livre de codes est un document (généralement un tableau) décrivant les variables présentes dans un ensemble de données. Son objectif est d'enregistrer des informations détaillées sur chaque variable.

Le Codebook contient généralement des codes, symboles et abréviations utilisés dans les fichiers, une liste des variables avec leur nom complet et leur définition, des unités de mesure et formats de données (par exemple AAAA-MM-JJ), le traitement des données manquantes (code, etc.).

PARTAGER, SAUVEGARDER, STOCKER VOS DONNÉES DE FAÇON SÉCURISÉE

Afin de sauvegarder et d'utiliser vos données tout au long du projet, il est conseillé de stocker vos données sur les serveurs de votre institution.

Afin d'optimiser la sécurité et d'éviter la perte de des données, il est fortement conseillé de sauvegarder l'ensemble de vos données à intervalles réguliers, sur 3 supports différents : serveur de l'institution, disque dur externe et votre ordinateur.

Dans le cas d'un partenariat avec une équipe basée à l'étranger et dont les membres n'ont pas accès à SWITCHdrive, Switch peut fournir un « voucher », un accès sécurisé à chacun de ces membres afin de permettre le partage de données de façon sécurisée²³.

OUTIL :

- [Voir le Fast guide #10 de l'EPFL « Store, publish, preserve data »²⁴](#)

Il est déconseillé d'utiliser une clé USB ou Dropbox pour partager, stocker et sauvegarder ses données.

Pour plus d'information sur les solutions de stockage à court ou long terme, veuillez prendre contact avec le service informatique de votre institution.

²¹ <https://cornell.app.box.com/v/ReadmeTemplate>, Université de Cornell.

²¹ <https://data.research.cornell.edu/content/readme>, Université de Cornell.

²³ Pour plus d'information sur cette solution, consulter le lien suivant : <https://help.switch.ch/drive/faq/filesandsharing/#collapse-533e115d-dc49-11e8-b90c-5254009dc73c-3>.

²⁴ <https://go.epfl.ch/rdm-fastguide>.

EVALUER LES ASPECTS, LES ENJEUX JURIDIQUES ET ETHIQUES

Si le projet de recherche nécessite de recueillir des informations personnelles et/ou sensibles auprès d'individus, il est obligatoire d'assurer la confidentialité des données collectées. Lors de cette étape du projet, les chercheurs·euses ont la responsabilité de réfléchir et d'anticiper les risques d'identification des participant·es à la recherche.

QU'EST-CE QU' UNE DONNÉE PERSONNELLE ?

En Suisse les données personnelles sont définies par la Loi fédérale sur la protection des données (LPD) et les lois cantonales comme étant : « Toutes informations qui se rapportent à une personne physique identifiées ou identifiable »²⁵.

Exemples : une date de naissance, un nom, un numéro de téléphone, une photographie, une vidéo, une adresse IP, etc.

Sont qualifiées de « sensibles » les données personnelles sur :

- Les données sur les opinions ou les activités religieuses, philosophiques, politiques ou syndicales
- Les données sur la santé, la sphère intime ou l'origine raciale ou ethnique,
- Les données génétiques
- Les données biométriques identifiant une personne physique de manière univoque
- Les données sur des poursuites ou sanctions pénales et administratives
- Les données sur des mesures d'aide sociale.

LA LOI FÉDÉRALE SUR LA PROTECTION DES DONNÉES (LPD)

Elle s'applique aux projets menés en Suisse. Le consentement préalable et éclairé de la personne concernée est obligatoire pour collecter des données personnelles et sensibles.

Le consentement : la collecte des données personnelles et sensibles n'est possible qu' avec le consentement préalable et éclairé de la personne concernée. Ces consentements concernent à la fois la publication issue du projet de recherche ainsi que le dépôt du jeu de données.

OUTIL :

- [Modèle de consentement de l'Association suisse des Commissions d'éthique de la recherche](#)²⁶

LA LOI FÉDÉRALE RELATIVE À LA RECHERCHE SUR L'ÊTRE HUMAIN (LRH)

Cette loi s'applique à la recherche sur les maladies humaines et sur la structure et le fonctionnement du corps humain, pratiquée sur des personnes ; sur des personnes décédées ; sur des embryons et des fœtus ; sur du matériel biologique ; sur des données personnelles liées à la santé. Dans le cas d'une recherche couvrant ces sujets, les chercheurs·euses doivent soumettre leur projet à la Commission d'éthique cantonale liée à leur affiliation et cela avant même de commencer à collecter leurs données.

LA LOI SUR LA PROTECTION DES DONNÉES DE L'UNION EUROPÉENNE

Au niveau européen, le Règlement Général de Protection des Données (RGDP) ajoute à la liste des exemples de « données à caractère personnel » : « [...] un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle, ou sociale »²⁷.

OUTILS:

- Voir le guide « [Ouverture des données de la recherche. Les grands principes juridiques et les recommandations de la HES-SO](#) ».
- [DMLawTool](#) un outil en ligne qui propose des solutions possibles aux problèmes de droits d'auteur et de protection des données²⁸.

²⁵ https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/fr#art_3.

²⁶ <https://swissethics.ch/fr/templates>.

²⁷ <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679>.

²⁸ <https://dmlawtool.ccdigitalaw.ch/>.



3. ÉVALUER, TRAITER, ET STOCKER

CONSERVER/DETRUIRE

A ce stade du projet, une réflexion doit être menée sur les données qui seront archivées et celles qui seront partagées sur un Data Repository (DR). Afin de déterminer quelles données seront conservées en vue du partage, le Digital Curation Centre (DCC) propose une série de questions dont les réponses permettront d'identifier les données à partager :

- Quelles sont les exigences de mon bailleur de fonds ?
- Mes données sont-elles suffisamment documentées et conformes aux principes FAIR ?
- Mes (meta)données respectent-elles la confidentialité, la sécurité et les exigences légales en ce qui concerne les données personnelles et sensibles ?

ANONYMISER/PSEUDONYMISER

La HES-SO recommande d'anonymiser toutes les données personnelles et sensibles aussi tôt que possible dans le processus de recherche, durant toute la durée du projet, ou au minimum et dans tous les cas avant la valorisation et le dépôt des données de recherche dans une base de données. Si l'anonymisation est impossible, dans ce cas, la pseudonymisation est recommandée. Pour des projets soumis aux Commissions cantonales d'éthique de la recherche sur l'être humain, des conditions plus strictes ou spécifiques peuvent être imposées et doivent être respectées.

ANONYMISATION > UN PROCESSUS IRRÉVERSIBLE

« L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et ce de manière irréversible »³⁰.

LA PSEUDONYMISATION

Elle consiste elle aussi à empêcher l'identification d'une personne sur la base de ses données personnelles. Mais contrairement à l'anonymisation, la pseudonymisation est réversible et les personnes autorisées sont capables de relier une personne à ses données personnelles.

OUTIL :

- Guide du [DCC](#)²⁹

TECHNIQUES D' ANONYMISATION :³¹

- Mélanger : permuter les données entre elles afin d'altérer la véracité des informations tout en conservant les données elles-mêmes.
- Simuler : ajouter des fausses données afin de prévenir l'identification de données spécifiques
- Retirer : supprimer certaines données

TECHNIQUES DE PSEUDONYMISATION

- Remplacer les données identifiantes (tel que le nom, prénom, etc.) par un code ou un numéro de référence, de telle sorte qu'il ne soit plus possible de relier telles données à telle personne sans devoir passer par ce numéro de référence et par la table de correspondance. Contrairement à l'anonymisation, la pseudonymisation est réversible.

OUTILS :

- [ARX, outil d'anonymisation des données](#)
- [Amnesia, anonymise vos bases de données](#)
- Guide de FORS : Data anonymisation : legal, ethical, and strategic consideration, 2020³².

²⁹ <https://www.dcc.ac.uk/sites/default/files/documents/publications/Five%20Steps%20to%20decide%20what%20data%20to%20keep.pdf>.

³⁰ Commission Nationale de l'informatique et des libertés (CNIL, Autorité administrative française), <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>.

³¹ Techniques suggérées par l'EPFL Library « Research Data Management Fast guide », <https://go.epfl.ch/rdm-fastguide>.

³² https://forscenter.ch/wp-content/uploads/2022/02/kleinerstam_fg1anonymisation1v1.1_final_2022.pdf.

FAIR DATA = OPEN FORMAT

Pour assurer l'accessibilité et la réutilisabilité de vos données, le choix du format de fichier approprié est important. Un format ouvert est un format de fichier qui peut être utilisé par n'importe qui. Au contraire, un format propriétaire (fermé) est un format lié à un logiciel particulier d'un fournisseur. Si le logiciel n'est plus disponible, les fichiers deviennent généralement inutilisables et/ou illisibles. Vos données sont donc inaccessibles.

Le tableau suivant indique les formats couramment utilisées, les formats recommandés et les formats propriétaire (non compatibles pour un partage FAIR) pour divers types de données³³.

TYPE DE DONNÉE	SOUS-TYPE	FORMAT COURANT format ouvert ou propriétaire	FORMAT POUR LE PARTAGE ouvert et non propriétaire	FORMAT PROPRIÉTAIRE : non compatible pour un partage FAIR
Code/Computation	.aiml, .iypnb, netcdf, .rmd, .rstudio	.aiml, .iypnb, netcdf, .rmd, .rstudio	.aiml, .iypnb, .netcdf, .rmd, .rstudio	mat, .rdata
Géospatiales		.dbf, .prj, .sbx, .sbn, .shp, .shx, .tfw, .tif	dbf, .sbx, .sbn, .shp, .shx, .tfw	
Multimédias	Audio	.aac, .flac, .mp3, .wav	.aif, .flac, .mp3, .wav	
	Image	.gif, .jpeg, .png, .raw, .svg, .tiff	.jpeg, .png, .svg,	Adobe InDesign (.indd), Adobe Illustrator (.ai), Adobe Photoshop (.psd, .tiff)
	Image 2D/3D	.dxf, .jpeg, .png, .svg, .step, .tiff	.dxf, .jpeg, .step, .png, .svg	.dwg
	Image médicale	DICOM, NifTI	DICOM, NifTI	
	Image vectorielle	.eps, .pdf, .x3d, .x3dv	.dwg, .dxf, .pdf, .svg	Adobe Illustrator (.ai)
	Vidéo	.avi, .mj2, .mkv, .mov, .mpeg, .wmv	.avi, .mkv, .mp4	.mov, .wmv
Partition	Partition	.mei, .MusicXML	.mei, .MusicXML, .pdf	
Site/page web	Web	.css., .html, .warc	.pdf./a, .warc	
Tableau de données	BDD, feuille de calcul	.csv, .html, .txt, .xlsx	.csv, .html, .txt	.xlsx
Texte	Données textuelles	.docx, .json, .htm, .html, .md, .pdf, .odm, .odt, .rtf, .tex, .txt, .xml	.html, .pdf/a, .rtf, .txt, .xml	.docx, .key, .pptx, .xlsx.

³³ The Library of Congress, <https://www.loc.gov/preservation/resources/rfs/index.html> et l'EPFL library, <https://go.epfl.ch/rdm-fastguide>.

4. PARTAGER/PRÉSERVER ET VALORISER

Parallèlement à la publication des résultats de votre recherche dans une revue spécialisée, la HES-SO recommande de partager les données qui étayent vos résultats dans un dépôt de données (souvent nommé Data Repository – DR³⁴). Cela permet aux chercheurs·euses de reproduire vos résultats et de les réutiliser dans des recherches ultérieures.

QUE DÉPOSER ?³⁵

DÉPÔT DÉFINITIF	DÉPÔT ÉVENTUEL	DÉPÔT NON NÉCESSAIRE	PAS DE DÉPÔT
<ul style="list-style-type: none"> Données brutes obtenues à partir de l'analyse d'échantillons physiques Données d'observation qui ne peuvent être régénérées Jeux de données originaux Jeux de données non originaux mais difficilement disponibles en ligne Codebook Code logiciel original 	<ul style="list-style-type: none"> Versions intermédiaires d'analyses ou de code si elles sont potentiellement utiles pour d'autres personnes ou si elles ont été utilisées dans des publications ou des thèses. 	<ul style="list-style-type: none"> Version incomplète Jeu de données déjà disponible en ligne 	<ul style="list-style-type: none"> N'importe quelle donnée qui contient des informations d'identification personnelle pour les sujets humains.

CHOISIR VOTRE DR

Les DR sont des plateformes qui permettent de stocker, partager et préserver à long terme son ou ses jeux de données. Un dépôt contient des sets de données et leur description (métadonnées). Ainsi humains ou machines peuvent les retrouver et les réutiliser. La taille des sets de données est mesurée en kilo-, mega-, giga- ou tetrabytes. Il existe 3 types de dépôt de données : disciplinaires, institutionnels et généralistes (multidisciplinaires). Lors de la sélection du dépôt de données, il est recommandé de :

- Choisir un DR conforme aux principes FAIR
- Choisir un DR non-commercial pour assurer la préservation à long terme
- Vérifier si votre bailleur de fonds impose ou recommande un dépôt en particulier
- Vérifier où se situe le serveur du DR. Il est généralement déconseillé de déposer sur un serveur américain ou aux USA³⁶
- Choisir un dépôt pertinent pour votre communauté de recherche.

³⁴ Cette recommandation s'inscrit dans le cadre de la stratégie nationale ORD.

Voir : https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/Swiss_National_ORD_Strategy_fr.pdf

³⁵ Université de Genève <https://www.unige.ch/researchdata/fr/partager/comment-partager/>

³⁶ Le préposé fédéral à la protection des données et la transparence liste les USA comme pays n'offrant pas un niveau adéquat de protection des données, <https://www.edoeb.admin.ch/edoeb/fr/home/protection-des-donnees/handel-und-wirtschaft/uebermittlung-into-usa/transmission-des-donnees-aux-etats-unis.html>

OUTILS POUR CHOISIR VOTRE DR:

- Le site de [re3data.org](https://www.re3data.org/) vous permet de filtrer les DR. Les filtres ci-dessous facilitent l'identification des DR conformes aux principes FAIR³⁷ :
 - AID systems (author identifier) = les auteurs sont identifiables de façon unique
 - Certificates = le DR est certifié FAIR via par exemple CoreTrustSeal³⁸
 - Data licenses = les données sont sous licence de réutilisation
 - Data access = les données sont librement accessibles
 - Metada standards = les standards de métadonnées établis sont utilisés
 - PID systems (persistent identifier) = le dépôt attribue des identifiants persistants à ses objets, par exemple des DOI ou ORCID.
- [Une liste de dépôts FAIR est disponible sur le site de la HES-SO.](#)

AJOUTER UN IDENTIFIANT UNIQUE ET DURABLE

Le DOI est une chaîne de chiffres, de lettres et de symboles unique utilisée pour identifier et localiser un contenu, et lui fournir une adresse web permanente (URL).

Les DOI sont généralement attribués lors du dépôt de vos données dans un DR ou via le Centre pour l'information scientifique de la HES-SO (CISO).

DEFINIR LE DELAI D'EMBARGO SI NECESSAIRE

Avant de rendre vos données accessibles, vous pouvez définir une période d'embargo le temps de publier vos résultats de recherche. Toutefois, les métadonnées associées doivent être accessibles pour signaler l'existence de ces données tout en préservant leur protection. Certains dépôts de données n'incluent pas l'option « embargo ».

DEFINIR LES DROITS D'ACCÈS SI NECESSAIRE

Lors du dépôt de vos données dans un DR, et selon les DR, vous avez la possibilité de restreindre ou encore de fermer l'accès à vos données. L'accès « restreint » permet à certaines personnes d'accéder à vos données après autorisation. Ces deux options concernent uniquement vos données et non pas vos métadonnées : dans les deux cas, vos métadonnées restent accessibles au public.

VALORISER VOTRE JEU DE DONNÉES AVEC LE DATA PAPER

Le Data paper est un article qui décrit un jeu ou un ensemble de jeux de données afin de faciliter et permettre l'éventuelle réutilisation de ces données. Cet article est soumis à la validation des pairs.

³⁷ <https://www.re3data.org/>.

³⁸ Si la certification est un critère important, cela ne signifie pas que les DR qui n'ont pas de certification ne sont pas conformes aux principes FAIR.

SERVICES GRATUITS POUR LES CHERCHEURS·EUSES EN SUISSE :

[DaSCH](#) est le Centre national suisse de données (DR) et de services pour les sciences humaines.

[DLCM](#) (Data Life Cycle Management) conseille les chercheurs·euses de la HES-SO en matière de DMP.

[FORS](#) est un centre de compétences qui offre un service de consultation aux chercheurs·euses dans le domaine des Sciences sociales. FORS est partenaire de SwissUbase.

CONTACTS ET INFORMATIONS SUR L'OPEN RESEARCH DATA À LA HES-SO

Constance Delamadeleine

Chargée de projet Open Research data
Dicastère Recherche et Innovation

Emails:

constance.delamadeleine@hes-so.ch
ou open@hes-so.ch

Site internet, page ORD:

[www.hes-so.ch/recherche-innovation/
open-science/open-data](http://www.hes-so.ch/recherche-innovation/open-science/open-data)

Pour vous tenir régulièrement informé.e sur l'actualité de l'Open Science, rejoignez la Communauté Open Science de la HES-SO (COS):

[Communauté Open Science HES-SO \(COS\)](#)

ABRÉVIATIONS ET VOCABULAIRES ANGLOPHONES COURAMMENT UTILISÉS DANS LA LITTÉRATURE

DMP: Data Management Plan

DR: Data repository

DOI: Digital Object Identifier

FAIR: Findability, Accessibility, Interoperability, Reusability

ORD: Open Research Data

OS: Open Science

RDM: Research Data Management