

Gestion du cycle de vie des données de recherche à la Haute Ecole de santé Vaud : diagnostic et recommandations



**Travail de Master réalisé par :
Eléonore Crausaz**

Sous la direction de :
Pierre-Yves Burgi, Directeur SI adjoint – Université de Genève

Genève, le 15 août 2022

**Information documentaire
Haute École de Gestion de Genève (HEG-GE)**

Remerciements

Je remercie chaleureusement toutes les personnes qui m'ont apporté aide et soutien lors de la réalisation de ce travail :

Pierre-Yves Burgi, mon directeur de travail de Master, pour sa confiance, ses précieux conseils, son soutien et son implication ;

Eliane Blumer, mon experte externe, pour son enthousiasme à expertiser ce travail, sa bonne humeur et sa disponibilité ;

Laurent Amiotte-Suchet, mon mandant, pour ses réponses rapides, sa collaboration et sa bienveillance tout au long du projet ;

Anne Bréaud et Magali Serex, pour le temps qu'elles m'ont accordé pour répondre à mes questions et leur intérêt pour le domaine ;

Ainsi que les participant-e-s aux entretiens dirigés, pour la mise à disposition de leur temps, la qualité des informations transmises et leur intérêt pour les retombées de ce travail.

Enfin, mes sincères remerciements vont à ma famille et mes amis qui m'ont témoigné leur soutien sans faille tout au long de cette étape :

Edwine et Amandine, pour leur bienveillance, leurs précieux conseils et les nombreux cafés sur Teams ;

Mes parents adorés, pour leur présence et leur affection ;

Et Guillaume, pour son amour, sa patience et ses encouragements réconfortants lorsque le moral baissait.

Résumé

La promotion du mouvement *Open Science* par les institutions académiques suisses et plus particulièrement par les instances de financements de projets de recherche s'exprime par de nouvelles exigences, notamment par l'ouverture et le partage des données de recherche. Pour répondre à ces directives, les institutions et leurs chercheu-r-se-s doivent choisir un dépôt de données parmi un panel de dispositifs existants. Malgré les recommandations déjà mises en place par les bailleurs de fonds, les chercheu-r-se-s manquent parfois d'informations et de renseignements afin d'effectuer leur choix.

Au partage des données de recherche s'ajoute la problématique de leur pérennisation. Certaines institutions suisses ne possèdent pas l'infrastructure ou les dispositifs permettant la conservation de données à long terme. Il existe aujourd'hui plusieurs solutions permettant aux chercheu-r-se-s de partager leurs données et de les déposer pour assurer leur conservation.

Ce travail vise à fournir des recommandations pour la gestion du cycle de vie des données de recherche pour une unité de recherche d'une Haute école suisse traitant des données de santé. Essentiellement orienté sur l'analyse du cycle de vie des données et sa gestion, ce travail se concentre particulièrement sur une étape du cycle : la conservation/pérennisation des données de recherche.

Dans un premier temps, la méthodologie détaillée employée pour aboutir aux propositions est expliquée : la construction de la revue de littérature, la méthode de collecte ainsi que la construction des outils permettent de comprendre le fil intellectuel ayant amené aux solutions implémentables à l'institution.

Puis, la revue de littérature situe les données de recherche dans une optique évolutive en analysant plusieurs définitions, permet de comprendre le cycle de vie des données sur la base de différents modèles existants et d'obtenir un aperçu du cadre légal à respecter dans le domaine des données de recherche.

L'élaboration d'un inventaire des recherches menées au sein de l'institution a permis d'obtenir une visibilité des données traitées par les équipes de recherches. De plus, des entretiens menés auprès de chercheu-r-se-s de l'institution ont permis de comprendre et d'identifier les pratiques adoptées au sein des équipes de recherche afin d'effectuer un diagnostic des pratiques sur la base d'un modèle de maturité. De plus, chaque étape du cycle de vie a été analysée afin de fournir un second diagnostic.

Enfin, les recommandations disponibles en fin de travail fournissent des pistes de réflexions, des propositions de gestion ainsi que des propositions d'outils pour que les chercheu-r-se-s intègrent au mieux le cycle de vie des données, afin de gérer leurs données de la façon la plus efficiente possible.

Mots-clés : Données de recherche, cycle de vie des données, gestion du cycle de vie des données, dépôt de données, principes FAIR, données trouvables, données accessibles, données interopérables, données réutilisables, partage des données, conservation des données, pérennisation des données, archivage des données.

Table des matières

Remerciements.....	i
Résumé	ii
Liste des tableaux	v
Liste des figures.....	vi
Liste des abréviations.....	vii
1. Introduction.....	1
1.1 Problématique du mandat.....	1
1.2 Objectifs du mandat	1
1.3 Considérations contextuelles.....	3
1.3.1 L'Open Science à la HES-SO	3
1.3.2 HESAV : Open Science, Open Access et Open Data	4
2. Méthodologie	8
2.1 Revue de littérature	8
2.2 Analyse de l'existant	8
2.2.1 Inventaire	8
2.2.2 Entretiens	10
2.3 Analyse des besoins en formation.....	11
2.4 Présentation des résultats	12
2.5 Analyse des résultats.....	12
2.6 Propositions d'outils	13
2.6.1 Schéma du cycle de vie des données	13
2.6.2 Checklist pour la gestion des données	13
2.6.3 Sélection de dépôt d'après quatre critères de sélection	13
2.6.4 Vers une politique de gouvernance	14
3. Revue de littérature	15
3.1 Considérations terminologiques.....	15
3.1.1 Donnée de recherche et donnée de recherche ouverte	15
3.1.2 Donnée personnelle et donnée sensible	16
3.1.3 Dépôt de données	17
3.1.4 Cycle de vie des données de recherche	17
3.2 Cadre légal	20
3.2.1 Loi sur le droit d'auteur (CH)	20
3.2.2 Loi sur la protection des données personnelles (CH)	21
3.2.3 Loi fédérale relative à la recherche sur l'être humain (CH)	21
3.2.4 Règlement général sur la protection des données (UE)	22
4. Présentation des résultats	23
4.1 Inventaire	23
4.2 Pratiques de gestion	26

4.2.1	Connaissances et intérêts pour la gestion des données de recherche	26
4.2.2	Pratiques de gestion des données de recherche	27
5.	Diagnostic	32
5.1	Diagnostic des pratiques appuyé sur un modèle de maturité	32
5.2	Diagnostic du cycle de vie des données de recherche	38
5.3	Besoins en formation	41
5.3.1	Formation des chercheurs	41
5.3.2	Formation de l'équipe Bibliothèque	41
6.	Recommandations	43
6.1	Propositions de gestion de ressources humaines	43
6.1.1	Proposition 1	43
6.1.2	Proposition 2	44
6.1.3	Proposition 3	46
6.2	Aides à l'archivage et au partage des données	47
6.2.1	L'archivage des données	47
6.2.2	Le partage des données	49
6.3	Sélection de dépôts pour la HESAV	51
6.4	Propositions d'aides à la GDR	57
6.4.1	Schéma du cycle de vie des données à la HESAV	57
6.4.2	Checklist pour gérer ses données tout au long d'un projet de recherche	57
6.4.3	Vers une politique de gestion des données de recherche	59
7.	Conclusion	63
	Bibliographie	65
	Annexe 1 : Guide d'entretien vierge : Chercheur-se-s	75
	Annexe 2 : Guide d'entretien vierge : Bibliothèque	79
	Annexe 3 : Modèle de maturité	82
	Annexe 4 : Schéma du cycle vie des données pour la HESAV	83
	Annexe 5 : Checklist aux différentes étapes du cycle de vie	84

Liste des tableaux

Tableau 1 : Résumé de l'étape 1	2
Tableau 2 : Résumé de l'étape 2	2
Tableau 3 : Résumé de l'étape 3	3
Tableau 4 : Stratégie de l'institution	5
Tableau 5 : Modèle de maturité appliqué aux participant-e-s – Collecte	33
Tableau 6 : Modèle de maturité appliqué aux participant-e-s – Gestion	35
Tableau 7 : Modèle de maturité appliqué aux participant-e-s – Sécurité	37
Tableau 8 : Résumé du diagnostic.....	38
Tableau 9 : Analyse de la proposition 1	44
Tableau 10 : Analyse de la proposition 2	45
Tableau 11 : Analyse de la proposition 3	47
Tableau 12 : Formats adaptés à l'archivage.....	49
Tableau 13 : Rédaction d'un fichier README	50
Tableau 14 : Recommandations de dépôts par neuf institutions académiques suisses.....	52
Tableau 15 : Localisation des serveurs des dépôts sélectionnés	53
Tableau 16 : Niveaux de curation – Core Trust Seal	54
Tableau 17 : Niveau de curation des dépôts sélectionnés	56
Tableau 18 : Correspondance au modèle OAIS pour les dépôts sélectionnés	56

Liste des figures

Figure 2 : Modélisation de la stratégie <i>Open Data</i> à HESAV.....	6
Figure 3 : Population et échantillonnage	10
Figure 25 : Les six entités fonctionnelles e l'OAIS.....	14
Figure 4 : Donnée personnelle.....	16
Figure 5 : Cycle de vie des données de recherche (UNIGE)	18
Figure 9 : Cycle de vie des données de recherche (EPFL)	18
Figure 6 : Cycle de vie des données de recherche (UNIL)	19
Figure 8 : Cycle de vie des données (DLCM).....	19
Figure 10 : Type de données traitées à la HESAV	23
Figure 11 : Format des données	24
Figure 12 : Données relevant de la LRH	24
Figure 13 : Production et/ou réutilisation.....	25
Figure 14 : Partage sur un dépôt de données	25
Figure 15 : Visualisation des avantages de la gestion des données de recherche	26
Figure 16 : Visualisation des inconvénients de la gestion des données de recherche.....	27
Figure 17 : La sécurité des données à la HESAV	28
Figure 18 : Partage des données au sein des équipes.....	29
Figure 19 : Dépôt des données sur le serveur institutionnel par les équipes	30
Figure 20 : Réutilisation des données	31
Figure 21 : Arbre décisionnel pour le formatage des données	48
Figure 22 : Construction de métadonnées	51
Figure 23 : Niveau de curation nécessaire pour le maintien de l'accessibilité et la réutilisation des données	54
Figure 24 : Curation de la plateforme SWISSUbase	55
Figure 26 : Processus de création d'une politique de gouvernance.....	60

Liste des abréviations

aLPD	Ancienne Loi fédérale sur la protection des données
ARODES HES-SO	Archive ouverte des domaines de la HES-SO
CAS	Certificate of advanced studies
CCSD	Centre pour la communication scientifique directe
CER-VD	Commission cantonale d'éthique de la recherche sur l'être humain
CNIL	Commission nationale de l'informatique et des libertés
DAUnit	Data Acquisition Unit
DLCM	Data Life Cycle Management
DMM	Data Maturity Model
DMP	Data Management Plan
DR	Données de recherche
EPFL	Ecole polytechnique fédérale de Lausanne
FAIR	Findable, Accessible, Interoperable, Reusable
FNS	Fonds national suisse
GDR	Gestion des données de recherche
HEC	Haute école de commerce
HESAV	Haute Ecole de santé Vaud
HES-SO	Haute École Spécialisée de Suisse occidentale
LDA	Loi sur le droit d'auteur
LPD	Loi sur la protection des données
LRH	Loi fédérale relative à la recherche sur l'être humain
nLPD	Nouvelle Loi fédérale sur la protection des données
OAIS	Open Archival Information System
OFSP	Office fédérale de la santé publique
RDSN	Research Data Service Network
RGPD	Règlement général sur la protection des données
SIP	Submission Information Package
SONAR	Swiss <i>Open Access</i> Repository

1. Introduction

1.1 Problématique du mandat

La Haute Ecole de santé Vaud (abrégée pour la suite du travail : HESAV), située à Lausanne et membre de la Haute Ecole Spécialisée de Suisse occidentale, est une institution active dans les sciences appliquées. HESAV propose quatre filières de Bachelor ainsi que des formations continues et postgrades. Outre sa mission d'enseignement, HESAV occupe une place centrale dans la recherche au sein des Hautes écoles avec une unité de recherche au service de la santé, mélangeant les disciplines pour favoriser une recherche de haut niveau (HESAV 2022a).

L'unité de recherche, composée d'une cinquantaine de chercheu-r-se-s et professionnel-le-s de santé, s'intéresse à des enjeux sociaux-sanitaires et grâce à une interdisciplinarité faisant sa force, HESAV permet une compréhension globale des problématique liées à la santé. (HESAV 2022c ; HESAV 2022d)

Ces dernières années, les questions liées à l'*Open Science* sont devenues centrales pour l'unité de recherche, impliquant dans ce sillage plusieurs services de l'école tels que le service informatique, le service juridique ainsi que l'équipe Bibliothèque. Actuellement, les questions de publication des résultats en *Open Access* sont essentiellement prises en charge par les bibliothécaires qui offrent aux chercheu-r-se-s l'accompagnement nécessaire. Les pratiques de publication scientifiques en *Open Access* sont standardisées, connues et appliquées par toutes et tous. Reste aujourd'hui la question des données de recherche ouvertes et leur gestion.

Ces dernières ne font pas encore l'objet d'un traitement systématique et standardisé au sein de la HESAV. Le manque de connaissance ou de compétence des chercheu-r-se-s freinent la gestion du cycle de vie des données, ces dernier-ère-s peinant parfois à répondre aux exigences des bailleurs de fonds.

Le partage des données sur des dépôts est une pratique valorisée par le mouvement *Open Science* et devenue obligatoire dans le cadre de recherches subventionnées. Cette pratique est encore peu adoptée et pratiquée par les chercheu-r-se-s de la HESAV. De plus, le cycle de vie des données de recherche entre le lancement du projet et la publication des résultats, ne semble pas faire l'objet de pratiques systématiques et contrôlées. À terme, la volonté de la HESAV est la mise en place d'une pratique institutionnelle et standardisée de gestion de données de recherche tout au long de leur cycle de vie, ceci permettant la publication et la préservation des sets pour assurer leur conservation sécurisée, leur récupérabilité et leur partage.

1.2 Objectifs du mandat

Les principaux objectifs de ce travail sont de dresser un panorama des recherches actuelles et récentes menées à la HESAV, de comprendre les pratiques de gestion des données de la part des chercheu-r-se-s et d'en tirer un diagnostic. L'analyse du cadre juridique demeure un objectif important afin de définir sous quelle juridiction tombent les données au cours de la recherche. L'objectif final de ce travail réside dans l'établissement de recommandations de

gestion pour la HESAV ainsi que dans la proposition d'outils de gestion correspondant aux besoins actuels de l'institution.

Ainsi, les grandes étapes du projet ont permis une subdivision des objectifs aboutissant à un livrable grâce à des outils spécifiques. Pour l'identification des pratiques dans le domaine des données de recherche, il a été question de réaliser deux objectifs :

- Réalisation d'un état de la littérature sur les données de recherche
- Identification des possibilités en matière d'archivage/partage

Ces deux objectifs ayant abouti aux résultats suivants :

- Une revue de la littérature
- Une liste préliminaire des dépôts recommandés pour la HESAV

Tableau 1 : Résumé de l'étape 1

Etape du projet	Objectif	Outil	Résultat
Identification des pratiques en matière de données de recherche	Réalisation d'un état de la littérature sur les données de recherche	Grille de lecture	Revue de la littérature
	Identification des possibilités en matière d'archivage/partage	Littérature + Analyse des dispositifs existants	Liste préliminaire de dépôts en vue des recommandations

Pour effectuer un état des lieux des types de données et une identification des pratiques de gestion à la HESAV, les objectifs ont été les suivants :

- Inventaire des recherches de 2018 à 2022
- Identification des connaissances et des pratiques des chercheur-se-s de la HESAV

Ces objectifs ont permis d'aboutir aux résultats qui suivent :

- Un inventaire des recherches
- Un diagnostic des pratiques
- Un diagnostic du cycle de vie des données de recherche

Tableau 2 : Résumé de l'étape 2

Etape du projet	Objectifs	Outil(s)	Résultat(s)
État des lieux de la recherche et des pratiques de gestion des données de recherche	Identification des connaissances, des pratiques et des dépôts utilisés par les chercheur-se-s	Entretiens dirigés	Diagnostic et recommandations
	Inventaire des recherches de 2018 à 2021	Fichier Excel	Inventaire des recherches

Enfin, pour parvenir à formuler des propositions de gestion et des recommandations liées à l'existant, les objectifs ont été les suivants :

- Analyse des besoins en formation de l'équipe bibliothèque à la HESAV
- Proposition d'un schéma du cycle de vie des données à la HESAV
- Proposition d'une checklist pour gérer les données tout au long d'un projet
- Proposition pour la mise en place d'une future politique de gestion des données
- Proposition de modèles de gestion des ressources humaines pour la gestion des données de recherche

Tableau 3 : Résumé de l'étape 3

Etape du projet	Objectifs	Outil(s)	Résultat(s)
Propositions et recommandations	Analyse des besoins de l'équipe bibliothèque à la HESAV	Entretiens dirigés	État des lieux des compétences, inventaire des besoins et propositions de gestion
	Proposition de modèles pour la gestion des données de recherche pour les chercheu-r-se-s	Modèles existants	Schéma du cycle de vie des données de recherche à la HESAV + Checklist
	Définition du cadre juridique pour l'archivage et le partage des données	Inventaires des normes et lois en matière de traitement des données	Vers une politique de gouvernance
	Guide pour le traitement, l'archivage et le partage de données	Analyse de modèles parallèlement aux pratiques en place	Checklist + Vers une politique de gouvernance

1.3 Considérations contextuelles

1.3.1 L'Open Science à la HES-SO

On retrouve sous le terme d'*Open Science* diverses significations selon le contexte dans lequel il est utilisé. Pour ce travail, nous conserverons les caractéristiques principales proposées en 2016 par la Commission européenne à savoir le travail collaboratif pour optimiser les processus scientifiques en s'appuyant sur de nouvelles technologies dans le but d'une diffusion de la connaissance. Outre ses objectifs initiaux de proposer gratuitement et

publiquement les résultats de recherche sous forme de publications scientifiques, le mouvement *Open Science* s'est élargi pour couvrir d'autres notions plus spécifiques telles que l'*Open Access* et l'*Open Data*, notions qui préoccupent particulièrement les bailleurs de fonds publics ainsi que les institutions publiques depuis quelques années. (Sauthier, Pirinoli, 2018)

Au niveau de la HES-SO, une stratégie *Open Science* a été adoptée en décembre 2018, appuyée sur deux documents : « *Open Science* : problématique et état de l'art » ainsi que « *Stratégie Open Science* ». La politique de publication a été modifiée afin de répondre aux exigences fixées au niveau national et a pour objectif d'ici à 2024 de publier 100% des publications scientifiques en *Open Access*. (Swissuniversities 2016)

Dans cette optique, plusieurs projets d'*Open Access* ont été mis en place au sein de la HES-SO notamment ARODES HES-SO ainsi que SONAR qui proposent un accès stable aux publications et valorisent les publications des chercheur-r-se-s.

Dans cette lignée, les écoles doivent inclure à leur vision la seconde notion spécifique de l'*Open Science*, à savoir les *Open Data*. Depuis l'exigence en 2017 par la Commission européenne de définir par défaut le libre accès aux données de recherche dans tous les nouveaux projets du programme Horizon 2020, parallèlement aux exigences du FNS de déposer les données sur une plateforme en vue de leur partage, les interrogations liées aux données de recherche sont nombreuses.

Ainsi, la HES-SO a participé à plusieurs projets, notamment au projet national DLCM (DLCM, 2020c) pour la gestion des données de recherche. De plus, un groupe interdisciplinaire a été constitué pour traiter les questions de protection des données. Enfin, deux autres projets liés aux *Open Data* peuvent être mentionnés : le projet BioMedIT, lié à la recherche biomédicale et des données y relatives ainsi que le projet DAUnit, mené par la HES-SO Valais-Wallis visant l'accompagnement des chercheur-r-se-s en santé. (Sauthier et Pirinoli 2018)

Vivement encouragés depuis le début des années 2000 en Angleterre (Huyghe et al. 2018), la question des *Open Data*, du partage de données et de l'analyse secondaire relevaient, au début des réflexions sur le sujet, de la volonté des financeurs d'archiver les données produites afin de les mettre à disposition. Notons de fait que l'*Open Data* induit des changements complexes de processus de recherche établis depuis une cinquantaine d'années. Les enjeux sont principalement la communication et la sensibilisation des chercheur-r-se-s mais aussi et surtout un accompagnement de cette communauté à travers ce changement. (Sauthier et Pirinoli 2018)

1.3.2 HESAV : *Open Science*, *Open Access* et *Open Data*

Le rapport d'activité *Open Science* de la HESAV datant de janvier 2022 établit la situation actuelle, les personnes compétentes et leur délégation ainsi que les formations internes prévues dans le but de développer une culture de l'*Open Science* auprès des chercheur-r-se-s. Les personnes compétentes sont référencées et se comptent au nombre de sept :

- Un-e adjoint-e scientifique répondant-e *Open Science*, 20%
- Trois personnes pour l'unité juridique en santé pour les aspects juridiques et éthiques
- Un-e membre de la CER-VD, à contacter pour les démarches spécifiques auprès de la commission

- Un-e responsable de bibliothèque pour les questions liées aux publications scientifiques en *Open Access*
- Un-e membre du service informatique pour la sécurisation des données de recherche

Depuis la fin de l'année 2020, plusieurs membres de la HESAV se sont impliqués dans les groupes de travail de la communauté *Open Science* de la HES-SO. Parallèlement aux réunions plénières, le répondant *Open Science* s'investit dans plusieurs groupes de travail : « *Open Data* », « *Open Science Event* » et « *Open Science Guidelines* » collaborant aux séances liées à la sécurité des données de recherche, à leur partage ou encore en préparant « Les journées *Open Science* de la HES-SO ». De plus, le groupe de travail « *Guidelines* » se donne pour objectif la rédaction de fiches techniques concernant tous les aspects de l'*Open Science*, à partir d'un glossaire exhaustif et de mettre à terme en ligne ces documents pour toutes les écoles de la HES-SO. Le tableau suivant résume les éléments mis en place par le groupe de travail :

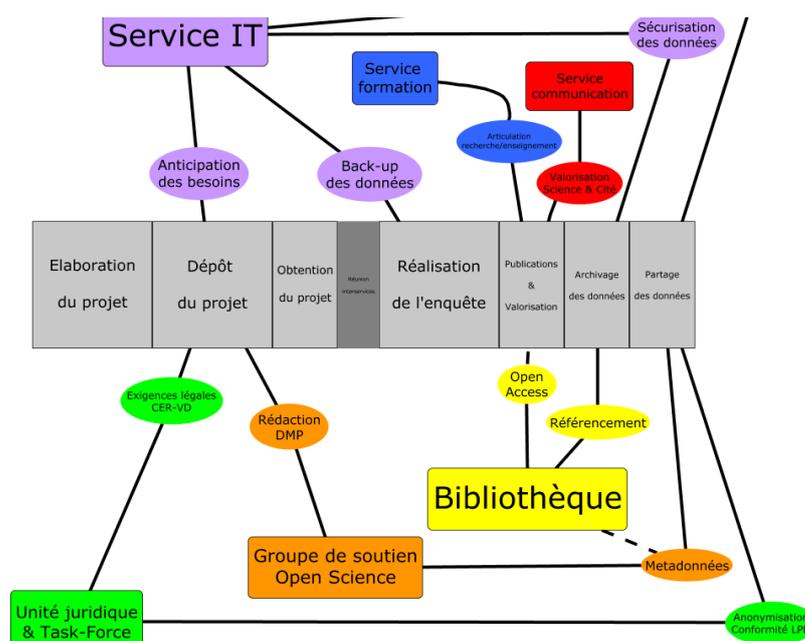
Tableau 4 : Stratégie de l'institution

Mise à disposition de documents et d'outils	Développement d'une « culture » de l'Open Science à HESAV	Contribution à la réflexion au niveau HES-SO
<ul style="list-style-type: none"> - Sécurisation des données (annexe 1) - Aide-mémoire juridique (annexe 2) - Guide pour la rédaction du DMP (annexe 3) - Modèle de sites internet pour les recherches - Présentation de la diversité des bases de données (site HES-SO) - Timeline pour la conduite de projets (à venir) 	<ul style="list-style-type: none"> - Groupe de réflexion interdisciplinaire sur les évolutions de l'Open Science - Désignation d'un délégué à la protection des données par projet - Réunions régulières direction de la recherche, services juridiques et informatiques, chercheur-e-s - Rédaction d'une prise de position des chercheur-e-s en santé 	<ul style="list-style-type: none"> - Participation à une communauté Open Science - Contribution au site internet HES-SO - Diffusion de nos rapports et travaux

(Source : Amiotte-Suchet 2019, p.19)

Enfin, la vision stratégique de la HESAV à moyen terme est d'obtenir une collaboration entre tous les services de l'institution, ceci afin de penser l'entier du cycle de vie des données de recherche depuis l'élaboration du projet jusqu'à l'archivage et au partage de celles-ci.

Figure 1 : Modélisation de la stratégie *Open Data* à HESAV



(Source : Amiotte-Suchet 2021, p. 4)

En début d'année 2022, la HESAV a la volonté de développer des outils et ressources pour l'encadrement et le soutien des chercheur-r-se-s. Pour ce faire, l'Intranet de l'institution a été augmenté mettant à disposition des documents spécifiques tels qu'un guide de rédaction d'un Data Management Plan ou des questions spécifiques à la sécurité des données, ainsi que des documents et ressources encadrant l'aspect légal et éthique explicitant notamment les démarches à entreprendre auprès de la commission d'éthique.

À l'heure de ce mandat, les besoins des chercheur-r-se-s de la HESAV s'établissent en deux points (Amiotte-Suchet, 2022) :

- Besoin d'une meilleure compréhension des enjeux liés à l'*Open Science* et l'*Open Data*
- Besoin de soutien dans l'ensemble des démarches à accomplir pour s'inscrire dans la stratégie *Open Science*, tant dans l'*Open Access* que dans l'*Open Data*.

Les données sensibles et les sujets délicats abordés dans les projets de recherche soulèvent de nombreux questionnements sur les données ouvertes et leur partage. En effet, le groupe de recherche traitant par exemple des « Usages, expériences et besoins des femmes immigrantes en regard des TIC autour de la période de la naissance » ont rencontré un refus de la part du FNS dans leur volonté de ne pas partager leurs données au vu de leur grande sensibilité. À la suite de ce refus, une refonte du DMP a été établie pour ne partager que les retranscriptions d'entretiens des professionnel-le-s ainsi que des interprètes. Toutefois, l'équipe a émis un doute sur le sens de partager uniquement une partie des données. (Pfund et al. 2022)

Le nombre de chercheur-r-se-s concerné-e-s par la diffusion, la mutualisation et la réutilisation des données qualitatives augmente, mais les pratiques ne sont pas encore systématiques. Encore souvent présentées comme un idéal à atteindre, ces pratiques visent à éviter, d'une

part, des travaux contre-productifs coûteux et laborieux (Amiotte-Suchet, 2022), et d'autre part, à atteindre un niveau optimal de capitalisation, de transparence de la démarche et de qualité de la preuve scientifique. (Huyghe et al. 2018) Le partage des données revêt d'importants avantages scientifiques tels que la possibilité de réutilisation pour des recherches comparatives (Corti 2011), mais par manque de temps souvent, rare est l'exploitation complète des données qualitatives. (Huyghe et al. 2018).

À la HESAV, la grande majorité des équipes de recherches collectent et produisent des données qualitatives. Fondées sur des entretiens, des documents personnels, confidentiels ou encore sur des observations, ces recherches nécessitent un traitement spécifique des données à savoir l'anonymisation et la contextualisation de la récolte. Toutefois, quelques craintes peuvent s'immiscer entre *Open Data* et recherche qualitative ; ces deux aspects sont difficilement conciliables (Amiotte-Suchet, 2022). L'anonymisation, qui ne garantit parfois que partiellement l'anonymat des participants-es, est en effet un exercice chronophage et compliqué. (Amiotte-Suchet, 2022)

Sur la base de ces considérations, nous allons mettre en lumière les éléments suivants :

- Les difficultés et craintes rencontrées pour le partage des données sensibles
- Le décalage entre recherche qualitative et exigences de partage des données
- Les besoins de l'institution face à ces nouvelles directives

Pour comprendre les éléments mis en exergue dans ce travail, la méthodologie, explicitée au point suivant, détaille le processus de recherche.

2. Méthodologie

La collecte s'est déroulée en trois grandes phases. La première a consisté en l'élaboration d'une revue de littérature sur la question de la gestion des données de recherche pour répondre principalement aux deux premiers objectifs. La deuxième phase constitue le volet quantitatif, avec la construction d'un inventaire des projets de recherche à la HESAV de 2018 à 2022. La dernière phase est constituée du volet qualitatif, avec le recueil de données issues d'entretiens dirigés, pour permettre une visibilité des pratiques de gestion de données au sein de l'institution.

2.1 Revue de littérature

La revue de littérature nous a permis d'atteindre nos deux premiers objectifs, à savoir la compréhension des enjeux autour de la gestion des données de recherche et l'élaboration d'une liste préliminaire de *Data repositories* adaptés aux enjeux et aux pratiques de l'institution. Quatre grands axes ont été définis pour construire une grille de lecture et faciliter la recherche d'informations :

- Cadre général : définitions et historique
- Cadre légal et normatif
- Dispositifs existants : modèles et politiques
- Outils et compétences

Pour chaque axe, un regroupement de littérature tant académique que professionnelle a été effectué. La grille de lecture a permis d'orienter la recherche et a facilité l'analyse et le recoupement/confrontation des sources grâce aux éléments clés de chaque référence reportés dans la grille.

La recherche documentaire a été effectuée de façon itérative : chaque recherche a permis l'enrichissement ou la réorientation des termes de recherche employés. Les recherches faites en amont du volet qualitatif faites avec des termes de recherche généraux tels que : « données de (la) recherche », « données de (la) recherche ouvertes », « gestion des données de (la) recherche », « GDR », « management des données de (la) recherche » ainsi que leur traduction anglophone ont permis d'obtenir un cadre historique ainsi que les définitions et enjeux autour des données de recherche. En affinant la recherche grâce à des termes spécifiques au domaine tels que : « dépôts de données », « dépôts de données ouvertes », « politique de gestion des données de (la) recherche », la liste préliminaire des dépôts de données compatibles aux besoins à l'institution ainsi que celle des dispositifs existants ont pu être établies.

2.2 Analyse de l'existant

Deux volets de collecte ont permis d'obtenir une visibilité de l'existant : un volet quantitatif constitué d'un inventaire des projets de recherche à la HESAV ainsi qu'un volet qualitatif comportant des données issues d'entretiens dirigés.

2.2.1 Inventaire

Nous avons choisi d'établir l'inventaire des recherches en format Excel. En premier lieu, ce format est le format initial de la première extraction AGP. Ensuite, cet outil est facile d'utilisation, connu de par répondant Open Science, et permettait l'intervention et

l'augmentation du fichier directement par les requérants principaux. De plus, il permet une première analyse statistique facilitée grâce aux fonctionnalités de l'outil.

La première étape de l'inventaire des recherches a constitué en une extraction AGP (accelerated graphics port) effectuée par un membre du service informatique. Les données sont exportées sous forme de fichier Excel et représente un avantage car elles sont directement téléchargeables et ainsi modifiables. (SageX Wiki, sans date)

Deux personnes internes à la HESAV ont ensuite nettoyé le fichier par la suppression de données inutiles à l'inventaire telles que les journées scientifiques, les forfaits de réparation etc. Après ce nettoyage, les onglets extraits étaient les suivants :

- Institut
- Chef de projet
- Numéro de recherche
- Acronyme
- Titre
- Ecole
- État de la recherche
- Date de dépôt
- Date de fin
- Filière
- Pilier

Nous avons augmenté le fichier en insérant les résumés de recherches. Puis, en collaboration avec le mandant, les colonnes suivantes ont été ajoutées pour obtenir le plus d'informations possibles :

- Autres institutions partenaires
- Type de données
- Langue des données
- Analyse des données
- Technologies utilisées
- Données produites / réutilisées
- Lieu de stockage
- Format
- Volume
- Partage des données
- URL du set
- Données relevant de la LRH

Une fois le fichier Excel construit, il a été déposé dans un espace Switch afin que chaque requérant-e principal-e de recherche puisse compléter les colonnes et fournir une base solide à l'inventaire.

2.2.2 Entretiens

L'unité de recherche de la HESAV compte plus d'une cinquantaine de chercheur-r-se-s. Afin d'obtenir un panorama des pratiques de gestion à la HESAV, des entretiens dirigés sur un échantillon de 7 chercheur-r-se-s ont été effectués.

2.2.2.1 Population et échantillonnage

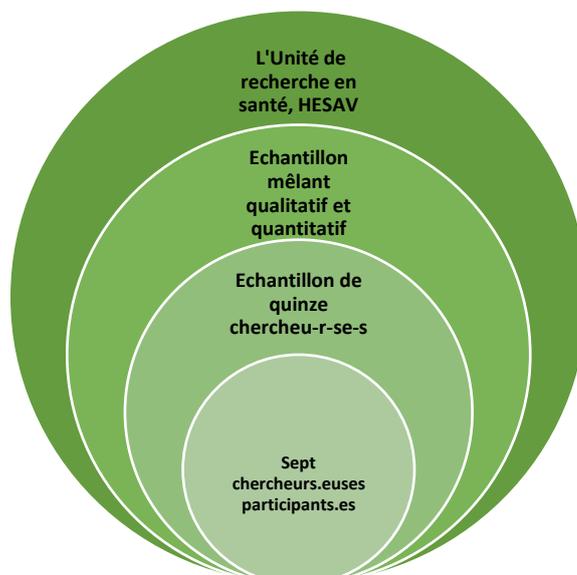
Sur la base d'une première extraction effectuée par le service informatique, la directrice de l'unité de recherche a effectué une première sélection des études à faire figurer en priorité dans ce mandat. D'une part pour obtenir une proportion adéquate entre les différents types de données et les thématiques traitées, d'autre part pour optimiser la coopération avec un échantillon réactif dont les intérêts et les connaissances pour la thématique de la gestion des données de recherches sont différentes et variées.

Une première sélection entre les recherches produisant des données qualitatives ou quantitatives a été effectuée. Les notions de volume et de support ont été rapidement examinées. Enfin, la représentativité de l'échantillon a été assurée en mêlant recherches en cours et abouties, financées par différents bailleurs de fonds, mélangeant aussi l'expérience des chercheur-r-se-s allant d'une recherche FNS à plus de trois recherches.

Ainsi, pour quinze chercheur-r-se-s contacté-es, sept se sont annoncé-e-s comme participant-e-s au projet, alors que deux ont refusé la participation. Les six dernières personnes contactées n'ont malheureusement pas donné réponse et ne figurent pas dans l'échantillon présenté.

Ainsi si l'on reprend le modèle établi par Fortin et Gagnon (2016) qui détermine la population concernée, la population cible et la population accessible, nous obtenons le schéma suivant :

Figure 2 : Population et échantillonnage



2.2.2.2 Méthode de collecte

Au vu de la population de l'échantillon (n=7) et de l'aspect exploratoire de ce mandat, l'entretien dirigé a été choisi afin d'optimiser un maximum de réponses, tout en permettant argumentation et digressions et permettant aussi à l'étudiante de compléter un besoin

d'informations. Le dernier avantage de cette méthode est la découverte de nouveaux éléments de réponses susceptibles de faire évoluer la recherche (Makhlouf Shabou 2020).

Dans cette optique, le guide d'entretien (Annexe 1) a été construit en cinq parties.

La première partie du guide constitue une base introductive. La deuxième partie vise à identifier les connaissances et intérêts des chercheur-r-se-s sur la gestion des données de recherche. La troisième partie, la plus conséquente d'un point de vue de l'analyse générale, vise à identifier et comprendre des mécanismes de gestion de données communes à tous les projets menés par la personne interrogée. Cette partie permet de dégager des éléments essentiels sur les pratiques générales des chercheur-r-se-s de la HESAV. La quatrième partie est accentuée sur les pratiques détaillées et spécifiques à un projet de recherche. Cette section permet à l'étudiante de voir, d'identifier et de comprendre directement les méthodes de gestion telles que la rédaction du DMP, l'élaboration d'un readme etc. d'une recherche spécifique en consultant avec l'accord de la personne interrogée les documents relatifs à un projet de recherche. Ces deux parties, construites en suivant le cycle de vie des données, permettent une visibilité de l'ensemble du cycle et de comprendre dans quelle(s) étape(s) s'opèrent les difficultés liées aux connaissances ou compétences techniques. Elles permettent en outre de déterminer quels techniques/outils sont utilisés pour gérer le cycle de vie des données de recherche.

Finalement, la cinquième partie permettra de comprendre s'il existe des besoins en formation, des sentiments spécifiques face à la gestion des données de recherches ou des attentes particulières liées au mandat.

Cet instrument de collecte a été construit afin de garantir sa réutilisation. En effet, les questions ont été élaborées pour assurer leur reproductibilité auprès de chercheur-r-se-s, professionnel-le-s ou étudiant-e-s francophone-s susceptible-s de gérer des données de recherche. Axées sur les connaissances des chercheur-r-se-s, sur les pratiques de gestion, les questions du guide sont implémentables pour toute institution souhaitant se confronter à la question de la gestion du cycle de vie des données de recherche. Enfin, construit selon le cycle de vie des données, le guide couvre l'ensemble des questions liées aux données de leur création à leur archivage et publication permettant ainsi à tout-e utilisateur-trice d'accentuer l'une ou l'autre des étapes du cycle de vie en laissant de côté d'autres étapes en fonction des questions de recherche.

2.3 Analyse des besoins en formation

Afin de comprendre et d'identifier les besoins en formation pour l'équipe Bibliothèque, un entretien dirigé a été effectué avec la responsable de la bibliothèque ainsi qu'avec une deuxième spécialiste en information documentaire, membre du groupe de travail *Open Science*.

Cet entretien a été effectué à partir d'un guide en cinq parties construit sur la même base que celui établi pour les entretiens avec les chercheur-r-se-s. La partie 3 cible les besoins en formation alors que la partie 4 s'oriente sur des propositions de gestion implémentables ou non à l'institution.

Cet outil n'est malheureusement pas réutilisable pour l'ensemble des institutions suisses gérant des données de recherche, car les propositions de gestion sont spécifiquement

construites sur la base du fonctionnement de l'institution et de la collaboration potentielle entre divers établissements du canton de Vaud.

2.4 Présentation des résultats

Afin de présenter les résultats de façon visuelle, nous avons choisi de les présenter sous forme de graphiques compartimentés pour la partie quantitative liée à l'inventaire des recherches.

Pour présenter les résultats liés aux avantages et aux inconvénients de la gestion des données de recherche, nous avons élaboré des nuages de mots. Pour ce faire, une grille d'analyse a été construite comme suit :

Une entrée par participant, puis deux entrées pour les points que l'on qualifiera ici par mesure de simplification de « négatifs » et de « positifs ». Grâce aux retranscriptions partielles et à l'aide des enregistrements, les occurrences des termes employés ont été comptabilisés et ainsi transposés dans un outil de nuage de mots en ligne gratuit. Trois termes ont nécessité un regroupement intellectuel : « efficacité », « compétences » et « chronophage ».

Dans les avantages, les éléments liés à la rapidité de recherche ainsi qu'à la découverte des fichiers/dossiers ont été transposés sous le terme « efficacité ».

Ceux liés aux points négatifs relevant du manque de compétences informatiques, juridiques, techniques ou un manque de compétence lié au fait de n'avoir encore jamais rédigé de Data management Plan ont été comptabilisés sous le terme de « compétences ». Le terme « chronophage » regroupe quant à lui tous les aspects liés au temps que demandent les tâches liées à la gestion des données allant de la réflexion liée à la rédaction liée au DMP jusqu'à l'élaboration de certains outils de gestion tels que le nommage des fichiers ou le choix d'un dépôt de données.

2.5 Analyse des résultats

Pour l'analyse des résultats, le modèle de maturité a été choisi comme outil d'analyse afin de classer les différentes pratiques selon leur niveau de gestion.

À l'origine, cet outil a pour but la gestion et le maintien de la qualité d'un domaine dans une institution (Brookes et Clark 2009 ; Makhlouf Shabou 2021c). Il représente un outil adéquat dans la perspective d'une amélioration. Ces outils souvent représentés sous forme de tableaux proposent des paliers progressifs qui permettent une amélioration continue de la performance, tout en donnant des pistes de développements pour aboutir à des processus plus sophistiqués (Langston et Ghanbaripour 2016, pp. 69-70 ; Makhlouf Shabou 2021c).

Le modèle de maturité choisi est le Data Maturity Model, abrégé DMM, de Baolong, Hong et Haodong (2018), issu d'une comparaison entre différents modèles de maturité orientés gestion de données tels que le DCAM, DMM, DCMM ou le modèle d'IBM. Incluant cinq niveaux de maturité partant du niveau initial jusqu'au niveau optimal, il comprend quatre secteurs d'analyse : la collecte de données, le management de données, la sécurité et les services.

Afin de déterminer les différents niveaux de maturité entre les participants de l'échantillon, ce modèle a été utilisé comme système de classification et permet une vue non-exhaustive des pratiques de gestion des données de recherche à la HESAV.

2.6 Propositions d'outils

Nous exposons ici la méthode utilisée pour la construction des trois propositions d'outils : le schéma du cycle de vie, la checklist et l'accompagnement pour une future politique de gouvernance.

2.6.1 Schéma du cycle de vie des données

Bon nombre de modèles de cycles de vie existent et seront explicités au point 3.4.1 (Université de Genève [sans date]^b ; Université de Lausanne [sans date]^b ; Van Gulick 2016 ; EPFL 2022g). Sur la base de ces différents modèles existants, une analyse comparative a été effectuée afin de décider le mode de présentation du modèle adapté à la HESAV. Puis, sur la base du fonctionnement actuel et des services proposés à la HESAV, ainsi que sur la base des pratiques de gestion des chercheu-r-se-s, une proposition de modèle a été établie (Annexe 4). Ainsi, le nombre d'étapes et leur composition ont été décidés, les opérations et traitements à effectuer ont été énumérés et l'élément central dédié au stockage et à l'archivage a été placé au centre du modèle, afin de connaître les opérations liées à la préservation des données en fonction des étapes du cycle et de la recherche.

La HESAV ne fournissant pas véritablement d'aide à l'archivage, mais une solution de soutien au stockage sécurisé durant la phase active des projets, ainsi qu'une proposition de stockage à long terme, nous proposons d'archiver les données sur un dépôt en parallèle de leur partage.

2.6.2 Checklist pour la gestion des données

Les besoins des chercheu-r-se-s, le fonctionnement de l'institution ainsi que les pratiques de gestion en place ont été la base de la construction de cette checklist. Ces besoins, mis en parallèle avec les différents processus existants dans d'autres institutions suisses, ont permis de dresser une liste des opérations à ne pas oublier et à effectuer durant le cycle de vie. Afin de familiariser et d'inscrire un process systématisé dans la pratique des chercheu-r-se-s, la checklist a été construite selon le cycle de vie des données proposé dans le schéma du cycle de vie établi pour la HESAV. Ainsi construit en six parties, il définit les différents processus à effectuer et les éléments indispensables à prendre en compte pour gérer ses données de recherches tout au long d'une enquête.

2.6.3 Sélection de dépôt d'après quatre critères de sélection

Pour proposer une solution de dépôt correspondant aux besoins de l'institution, nous avons défini quatre critères de sélection :

- Correspondance aux principes FAIR
- Localisation des serveurs
- Niveau de curation
- Correspondance au modèle de référence OAIS

Deux critères nécessitent d'être explicités ici ; les principes FAIR ainsi que le modèle de référence OAIS.

Les principes FAIR

Les principes FAIR décrivent l'organisation des données pour que celles-ci soient trouvables, accessibles, interopérables et réutilisables. (CCSD, sans date) Ces principes s'appliquent aux données et assurent que l'ensemble des données répondent à des caractéristiques

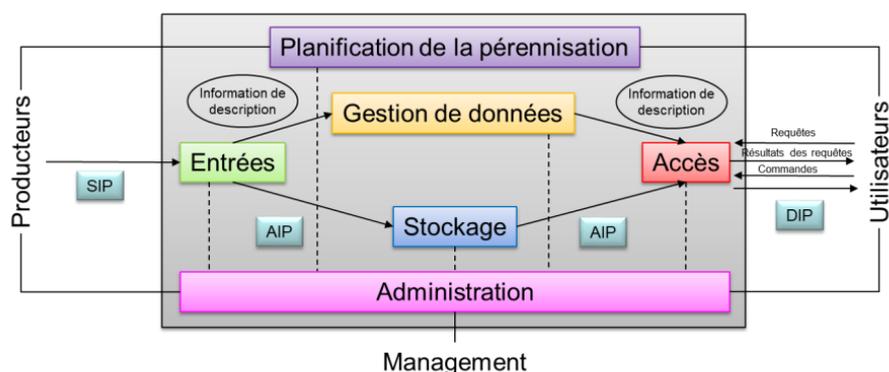
d'accessibilité et de réutilisation. (FNS, 2022) Le FNS renvoie d'ailleurs au document de Force11 [The FAIR Data Principles](#) qui détaille pour chaque principe les éléments à prendre en considération et a élaboré une [liste](#) utile à tous-tes les chercheu-r-se-s pour analyser les possibilités de dépôts correspondants aux principes FAIR dans la pratique, Enfin, une [checklist](#) permet d'identifier les dépôts qui satisfont aux principes de trouvabilité, d'accessibilité, d'interopérabilité et de réutilisation.

Le modèle OAIS

L'Open Archival Information System (OAIS), traduit en français par Système ouvert d'archivage d'information, a été développé au milieu des années 90. L'OAIS est un référentiel qui saisit les spécificités, les considérations et les acteurs de l'archivage numérique à long terme. On peut aussi le considérer comme un modèle conceptuel de gestion et de conservation à long terme pour les documents numériques (Makhlouf Shabou2021b).

Composé de six entités, ce modèle peut être représenté comme suit :

Figure 3 : Les six entités fonctionnelles e l'OAIS



(adapté de CCSDS 2017, p. 4-1)

2.6.4 Vers une politique de gouvernance

Une politique de gouvernance ne s'écrivant pas en cavalier seul, une liste des éléments à faire figurer dans une politique future a été établie en fonction

- Des objectifs d'une politique
- De la structure d'une politique
- Du contexte organisationnel de l'institution

En effet, sur la base de la littérature (Makhlouf Shabou et al. 2019 ; Makhlouf Shabou 2021e ; Jules et al. 2013 ; DLCM, sans date) les éléments principaux à faire figurer au sein d'une éventuelle politique ont été énumérés. Ainsi, les objectifs à court, moyen et long terme ont été analysés pour répondre au mieux aux besoins de l'institution en matière de gestion des données de recherche et d'établir des propositions dans l'axe voulu par la stratégie de la HESAV.

3. Revue de littérature

3.1 Considérations terminologiques

3.1.1 Donnée de recherche et donnée de recherche ouverte

Depuis quelques années, plusieurs définitions encadrent les données de recherche et la fameuse phrase qui qualifie désormais leur ouverture : des données aussi ouvertes que possibles, aussi fermées que nécessaires.

L'Université de Stanford définissait en 1997 les données de recherche selon leur forme et leur utilité dans la présentation et l'analyse des résultats sous les termes suivants :

“Research data include laboratory notebooks, as well as any other records that are necessary for the reconstruction and evaluation of reported results of research and the events and processes leading to those results, regardless of the form or the media on which they may be recorded.” (Stanford University 1997)

S'ajoute à cet aspect formel l'affirmation des données de recherche comme « sources principales pour la recherche scientifique » reconnue par la communauté scientifique comme nécessaire pour la validation des résultats. (OCDE 2007, p. 18)

Les définitions s'affinent au cours des années et l'Université d'Edinburgh en 2020 ainsi que le projet DLCM s'accordent sur les différents modes de production des données de recherche en abordant notamment la forme digitale :

« Research data : That which is collected, observed, or created in a digital form, for purposes of analysing to produce original research results. »

Souvent retenues par les chercheur-s participant-e-s à ce travail comme les informations collectées au cours d'entretiens ou de questionnaires dans le but de répondre à des questions de recherche scientifiques propres à chaque projet, la définition de l'AFF en 2014, regroupe les principaux éléments que constituent les données de recherche pour l'unité de recherche en santé de la HESAV :

« Les données de la recherche sont l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. » (AAF Section Aurore, 2014, cité par Jacquemot-Perbal, Cosserat et Ciolek-Figiel 2016)

Le Geneva Graduate Institut confirme en ce sens la production de résultats avec la définition suivante :

« On désigne par "données de la recherche" ou "données de recherche", les données issues de l'observation, de l'expérimentation ou issues de sources existantes et qui sont utilisées pour produire des résultats dans le domaine de la recherche. » (Geneva Graduate Institut 2022)

Ajoutons à ces éléments la contextualisation de la collecte, notion prise en compte par le Data Curation Network en 2016 qui définit les données de recherches comme suit :

*« Facts, measurements, recordings, records, or observations about the world collected by scientists and others, **with a minimum of contextual interpretation**. Data may be any format or medium (e.g., numbers, symbols, text, images, films, video, sound recordings, drawings, designs or other graphical representations, procedural manuals, forms, data processing algorithms, or statistical records.) »*

Les principaux aspects que nous considérons dans ce travail définissent les données de recherche comme des éléments indispensables à tout projet, que leur mode de production varie en fonction des objectifs et des questions de recherche. De plus, selon les méthodes de collecte, ces données peuvent exister sous différents supports et dans des formats très variés. Dans la mesure du possible, ces données sont gérées afin que tous les membres des équipes de recherche soient efficaces, que les données restent intègres, sécurisées et de qualité tout au long du projet. Finalement, elles doivent être encadrées d'un minimum d'éléments contextuels.

Gardons enfin à l'esprit que l'intérêt des données de recherche revêt aussi, dans l'optique du mouvement *Open Science*, des intérêts supplémentaires à celui de la validation d'un résultat scientifique, intérêts que l'on observe dans la dernière phase du cycle de vie, à savoir le partage des données. Ces intérêts sont variés et servent la communauté scientifique grâce à l'ouverture et au partage des données de recherche comme l'annonce l'Université de Lausanne ([sans date]c) :

*« Facilite la reproduction et la vérification des expériences et recherches menées ;
Rend davantage visible les travaux et résultats de recherche ;
Augmente le nombre de citation des articles scientifiques pour lesquels les données de recherche sont également publiées ;
Encourage les nouvelles collaborations et nouvelles pistes de recherche ;
Soutient le principe d'intégrité scientifique ;
Répond aux exigences de certains bailleurs de fonds et éditeurs scientifiques »*

Pour répondre à ces intérêts supplémentaires, ces données de recherche doivent être ouvertes. On qualifie des données de recherche comme ouvertes si leur accès est libre et que leur utilisation et leur modification le sont aussi, sous condition de reconnaître leurs auteurs (Hodson, Jones et al. 2018, cité dans Guirlet 2020a). Les données de recherche ouvertes sont des données gratuitement accessibles sur des plateformes en ligne (HES-SO 2021). Pour être qualifiées d'ouvertes, ces données doivent être déposées sur un dépôt généraliste ou spécialisé afin de pouvoir les trouver, procéder à leur extraction en vue de leur réutilisation, ceci tout en protégeant l'intégrité et l'authenticité des données (Johnson 2017).

3.1.2 Donnée personnelle et donnée sensible

Figure 4 : Donnée personnelle



(Source : Vernaz 2021)

Comme le mentionne le tableau ci-dessus, les données personnelles regroupent toutes les informations qui se rapportent à une personne physique identifiée ou identifiable. Ainsi, le nom et le prénom permettent d'identifier aisément une personne. En parallèle, une personne peut être identifiable à partir d'une seule donnée ou d'un croisement de jeux de données.

Les données sensibles composent une catégorie particulière des données personnelles (CNIL [sans date]d). Elles relèvent en effet de l'origine raciale, des opinions politiques, de la confession mais aussi des données liées à la santé telles que données médicales, génétiques, biométriques ou encore l'orientation ou la vie sexuelle d'un individu. (CNIL [sans date]d ; Bayle 2020 ; Vernaz 2021)

3.1.3 Dépôt de données

La première définition que nous pouvons donner pour un dépôt de donnée est la suivante : « A digital archive that provides services for the storage and retrieval of digital content. » (Johnston et al., 2016).

Simple dans sa présentation des fonctions d'un dépôt, cette définition peut être complétée par celle proposée par l'Université de Genève qui apporte la notion de transparence et de réutilisation par la communauté scientifique :

« Un data repository, en français un dépôt de données, est un espace dédié au téléversement de données de recherche dans l'objectif de leur archivage et/ou de leur mise à disposition dans une perspective de transparence ou de réutilisation par des pairs. » (Université de Genève [sans date]a)

Afin d'optimiser la qualité d'un dépôt et de promouvoir cette réutilisation par les pairs, les caractéristiques de ces dépôts sont explicitées dans un article lié à la certification CoreTrustSeal :

« First, a digital repository's mission should be to give reliable long-term access to the digital data under their care, now and in the future. Second, there should be permanent monitoring, planning and maintenance. The threats and risks within their systems must be understood. Finally, there should be a regular audit and certification cycle in place. » (Dillo et De Leuw 2018)

Il existe divers outils pour le choix d'un dépôt notamment la liste de recommandations du FNS mais aussi l'utilisation de registre tel que Re3data comptant plus de 2'500 dépôts. (Re3data.org, sans date)

3.1.4 Cycle de vie des données de recherche

Selon le glossaire dédié aux données de recherche de l'Université de Genève, le cycle de vie des données de recherche « décrit les différentes étapes de vie à travers lesquelles évoluent les données de recherche, de leur création à leur archivage. » (Université de Genève [sans date]b)

Il peut être constitué de différentes étapes selon les modèles, ces dernières se situant entre six et sept, selon la prise en compte de la phase d'élaboration ou non. Le UK Data Service divise ce cycle en 6 étapes :

1. Planification de la création de données
2. Collecte des données

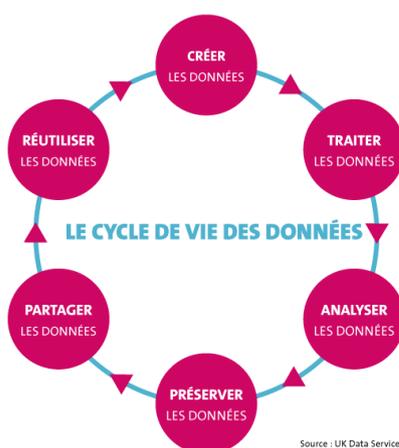
3. Préparation et analyse des données
4. Publication et partage des données
5. Préparation les données pour la préservation
6. Réutilisation des données.

Plusieurs institutions en Suisse romande proposent à leurs chercheu-r-ses-s des cycles de vies afin de favoriser la connaissance de la GDR et une meilleure gestion du cycle de vie de leurs données. De plus, le projet national DLCM propose aussi un cyclede vie des données de recherche pour donner des clés de gestion aux dépositaires.

Sur la base de la comparaison entre ces modèles existants, nous pourrons établir une proposition de cycle de vie pour l'institution HESAV.

L'Université de Genève propose un modèle en six étapes, basé sur celui proposé par le UK Data Service sans y intégrer de phase de planification. Le schéma est composé uniquement de verbes infinitifs qui permettent aux utilisateur-trice-s de visualiser rapidement les principales étapes et actions à organiser durant le cycle de vie.

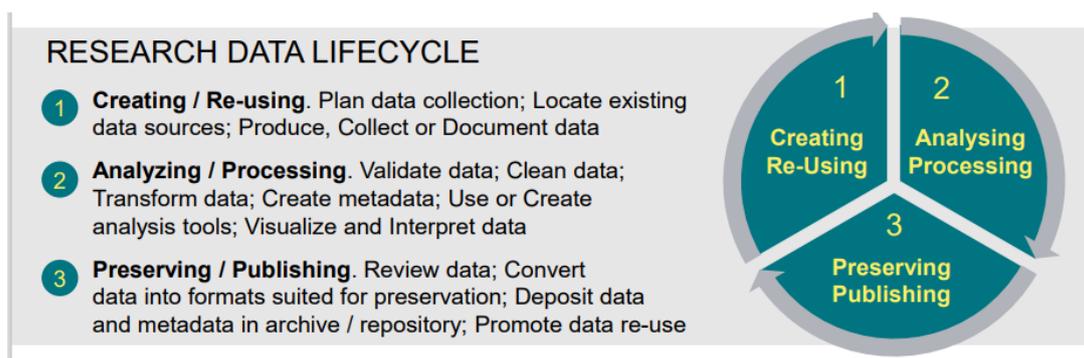
Figure 5 : Cycle de vie des données de recherche (UNIGE)



(Source : Université de Genève [sans date]b)

Le modèle proposé par l'EPFL dans les fiches techniques qui soutiennent les chercheu-r-se-s, est construit en trois temps mais comprend deux processus de traitement sur les données :

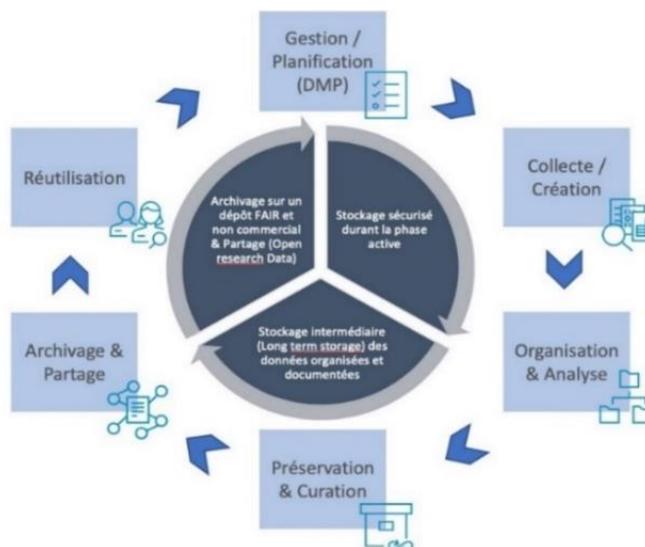
Figure 6 : Cycle de vie des données de recherche (EPFL)



(Source : EPFL 2022g)

Le schéma proposé par l'Université de Lausanne est aussi composé de six parties mais compte plusieurs traitements/processus dans chacune des étapes. Ainsi, nous pouvons noter des combinaisons de type préservation et curation. En plus du modèle proposé par le UK Data Service, une phase d'archivage et de partage simultanés avant la réutilisation des données.

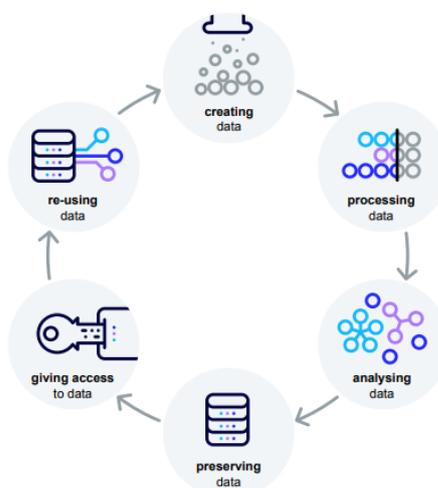
Figure 7 : Cycle de vie des données de recherche (UNIL)



(Source : Université de Lausanne, [sans date]b)

Le projet DLCM ajoute à l'importance du suivi de ce cycle, la validité à long terme des données ainsi que la prévention à l'obsolescence des données. (DLCM, 2018) Ce projet propose un cycle de vie en six étapes afin de promouvoir une bonne pratique scientifique assurant la reproductibilité des résultats :

Figure 8 : Cycle de vie des données (DLCM)



(Source : DLCM 2018)

Enfin, l'importance de la connaissance du cycle de vie des données de recherche est le maintien de la qualité, de l'intégralité et de la sécurité des données de recherche grâce à différents processus et actions qui peuvent être mis en place (Université de Genève, sans date).

Les modèles et schémas sont donc nombreux entre les différentes institutions, notamment en Suisse. Il reste tout de même important de mentionner que tous ont des éléments communs ; ceux de la préservation/archivage et de la publication/partage des données à des fins de réutilisation. Pour mener à bien cette réutilisation, il est nécessaire d'avoir une connaissance des étapes du cycle de vie des données pour garantir leur qualité, leur transparence, leur fiabilité, leur intégrité ainsi que leur authenticité (Makhlouf Shabou 2021a ; Makhlouf Shabou 2021d).

Notons finalement que bien que ces modèles soient nombreux, il est important d'adapter le schéma du cycle de vie en fonction des besoins des chercheur-r-se-s et du fonctionnement de son institution.

3.2 Cadre légal

3.2.1 Loi sur le droit d'auteur (CH)

La création, le traitement ainsi que l'analyse et le stockage de données dans un cadre de recherche scientifique ne posent pas de question d'ordre juridique. Toutefois, les exigences des bailleurs de fonds imposent le partage de ces données et leur accessibilité aussi ample que possible. Cela soulève quelques interrogations notamment sous l'angle de la recherche scientifique. L'article 24d de la LDA stipule notamment que : « La reproduction d'une œuvre à des fins de recherche scientifique est autorisée lorsqu'elle est nécessaire à l'application d'un procédé technique et que l'accès à l'œuvre reproduite est licite ». Le second alinéa du même article aborde la reproduction pour être conservée à des fins d'archivage et de sauvegarde une fois la recherche terminée.

Ainsi, le Geneva Graduate Institut tranche la soumission à la LDA en différenciant les données brutes des données organisées : « En termes de droit d'auteur, les données brutes et les métadonnées ne sont pas protégées par le droit d'auteur mais les bases de données ou les collections de données organisées le sont sous réserve qu'elles présentent un caractère individuel » (Geneva Graduate Institut [sans date]). Néanmoins, si l'on considère un abstract ou un titre comme une métadonnée, ces éléments sont alors protégés par la LDA.

Toutefois, la prudence est de rigueur lorsque l'on aborde la protection de contenu type « données » par la LDA. Par précaution, la réponse est semble-t-il affirmative au vu de la définition très large de la notion d'œuvre dans la loi (Santos 2020). De plus, en droit suisse, contrairement à l'Europe, aucune protection n'existe pour les bases de données. En Europe ce droit garantit un droit exclusif au créateur pendant 15 ans et interdit la réutilisation ou l'extraction de partie de ces bases de données (Aliprandi 2012). Alors qu'en Suisse les bases de données sont considérées comme des recueils au sens de la loi sur le droit d'auteur (Baggi, sans date, cité dans Santos 2020). Les données conservées dans ces bases peuvent être aussi protégées si elles sont des créations de l'esprit et possèdent un caractère individuel (cf. définition de l'œuvre dans la LDA, art. 2 al. 1). Quant aux métadonnées, elles ne sont pas protégées dans la majorité des cas car dépourvues d'individualité (Santos 2020).

3.2.2 Loi sur la protection des données personnelles (CH)

Rappelons en premier lieu la distinction entre données personnelles et données sensibles : les premières sont des informations se rapportant à une personne identifiée ou identifiable directement ou indirectement. Alors que les secondes relèvent de l'origine raciale, des opinions politiques, des confessions et regroupe aussi les données médicales, les données génétiques, biométriques ou encore des éléments d'orientation sexuelle (Bayle 2020). La LPD intervient lorsque ce type de données est traité et édicte les principes à respecter au tel cas. Le traitement des données personnelles sans le consentement de ladite personne, est considéré comme une atteinte à la personnalité (Métille 2014). Le but de cette loi est la protection de la personnalité et des droits fondamentaux des personnes qui font l'objet de traitement de données (LPD, art.1). Attention toutefois au fait que la LPD ne s'applique pas aux données collectées anonymement, ou anonymisées (Meier, 2011 cité dans Santos 2020).

Qu'en est-il des données sensibles, majoritairement au cœur des projets de recherche de la HESAV ? Définies aussi dans l'article 3, lettre c de la LPD, ces données personnelles particulières sont protégées par la loi, du fait que les profils de personnalité profitent d'une protection identique à celle des données sensibles. (Santos 2020)

Adoptée le 25 septembre 2020 par le Parlement, la nouvelle Loi fédérale sur la protection des données s'aligne au RGPD mais s'en écarte sur quelques points pour aller parfois même au-delà du texte européen. (Rosenthal et al. 2020). Si les grands principes d'application de la LPD restent identiques, la nLPD a élargi plusieurs définitions notamment celle des « données personnelles » reprise du RGPD. La définition des données personnelles sensibles n'est quant à elle pas l'exact reflet du texte européen, puisque le texte suisse définit de manière plus large les données personnelles sensibles et comprend aussi les données relatives aux poursuites ou sanctions pénales et administratives, les données relatives à l'aide sociale ou encore relevant de la sphère intime (Rosenthal et al. 2020).

3.2.3 Loi fédérale relative à la recherche sur l'être humain (CH)

La LRH encadre les activités de recherches scientifiques sur les maladies humaines ainsi que le fonctionnement du corps humain. Elle s'applique lorsque les recherches sont effectuées sur des personnes ou qu'elles portent sur des données de patient-e-s liées à leur état de santé (LRH art. 2, al. 1 ; Erard 2021). Toutefois, lorsque les données ont été collectées et anonymisées, ces données ne sont alors pas soumises à la LRH. (art. 2 al. 2 let. C LRH ; Erard 2021). Qu'en est-il du partage des données récoltées lors de ces projets de recherche ? L'article 17 de la LRH stipule que :

« si, lors [...] de la collecte de données personnelles liées à la santé, leur réutilisation est envisagée à des fins de recherche, le consentement de la personne concernée doit être recueilli dès ce moment-là et la personne concernée doit être informée qu'elle a le droit de s'y opposer ».

Ainsi, les projets de recherche soumis à la LRH doivent être encadrés de formulaires de consentement spécifiques et doivent être construits dès l'initiation du projet, en vue d'un éventuel partage des données. De plus, les données personnelles non génétiques liées à la santé peuvent être réutilisées à des fins de recherche, ceci sous une forme non codée lorsque la personne concernée ou ses représentants légaux ont donné leur consentement (LRH, art. 33, al. 1). De plus, la forme non codée des données pour une réutilisation est possible, notamment si la personne concernée ou ses représentants légaux ne s'y sont pas opposés après avoir été informés (LRH, art. 33, al. 2).

3.2.4 Règlement général sur la protection des données (UE)

Le RGPD ou Règlement général sur la protection a pour but la création d'un socle commun de protection des données sur le territoire européen, auparavant grandement hétérogène. Il a aussi pour but le renforcement de la protection des personnes, de renforcer les sanctions, de faciliter la circulation des données et de répartir les responsabilités et obligations. (Bayle 2020) Bénéficiant de plusieurs principes tels que la loyauté, la transparence, la limitation des finalités, la proportionnalité, l'exactitude, la conservation limitée ainsi que la sécurisation et la confidentialité (Bayle 2020 ; Vernaz 2021 ; Santos 2020), le RGPD couvre aussi des principes encore méconnus en Suisse tel que l'*accountability*, qui encadre les responsabilités et devoirs du responsable de traitement, ainsi que les principes de *privacy by design* et *privacy by default* qui implique pour le premier, l'étude du traitement et l'impact qui en découle sur la vie privée des personnes et pour le second, qui garantit le plus haut niveau de protection aux personnes dont les données personnelles sont traitées (Bayle 2020 ; Santos 2020). Les projets de recherche de la HESAV étant parfois inter-institutionnels et s'étalant sur d'autres territoires que le sol helvétique, de telles notions juridiques européennes au niveau de la protection des données et du traitement de celles-ci semblent essentielles.

4. Présentation des résultats

4.1 Inventaire

L'inventaire des données de recherche à la HESAV entre 2018 et 2022 réalisé sur la base de la checklist « How to create a data inventory » du Data Sharing Toolkit, a permis de mettre en lumière des éléments sur le type de données traitées et leurs formats. Le pourcentage de données partagées, les tendances de production et/ou réutilisation, ainsi que la soumission à la LRH ou non ont pu être mis en avant.

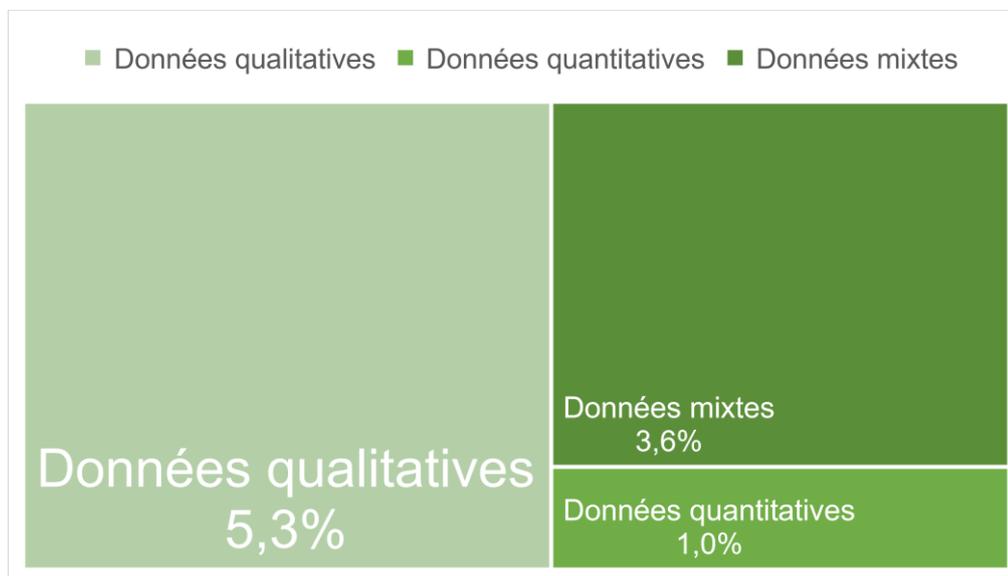
Type de données et de formats

Nous pouvons soulever que la plus grande partie des données de recherche de la Haute école sont de type qualitatif, un petit pourcentage de données relève du quantitatif et plus d'un tiers des projets comprennent des données mixtes. La collecte des données qualitatives se fait selon plusieurs méthodes : entretiens, observations, journaux de terrain, littérature primaire et secondaire, enregistrements vidéo, photographies ou encore recoupement de documents administratifs.

Le volet quantitatif est généralement obtenu grâce à une collecte via des questionnaires ou des données statistiques.

La méthode mixte, représentant la seconde plus grande proportion de type de données à la HESAV, regroupe en grande partie la combinaison entretiens/questionnaires.

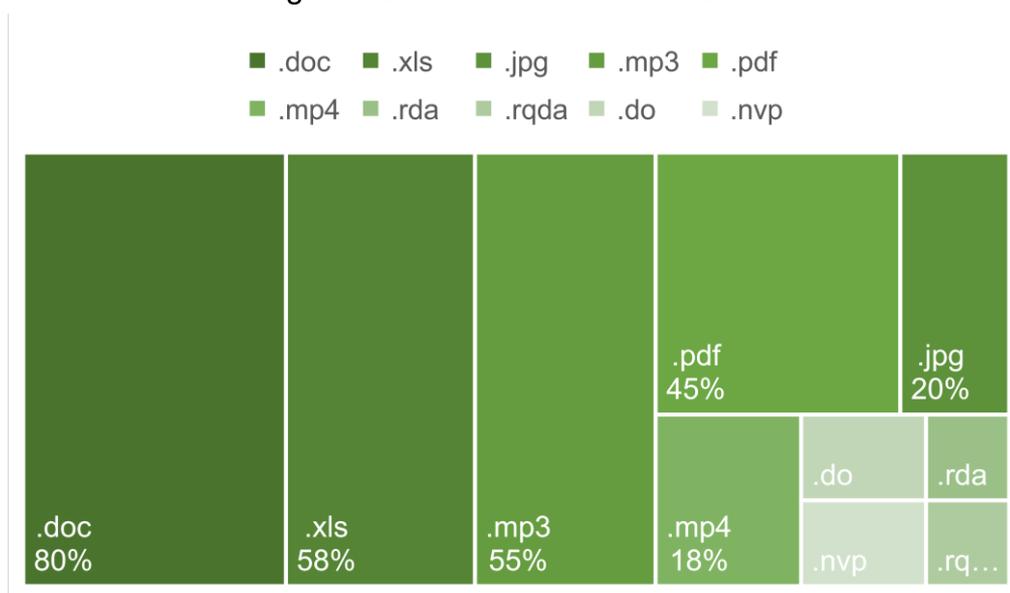
Figure 9 : Type de données traitées à la HESAV



Les formats des données sont très variés au sein de l'institution. Une grande proportion des données sont des formats de logiciels de traitement de texte ou transformés en PDF.

Les données sont généralement traitées dans des documents Excel, représentant un autre grand pourcentage des formats. Dans le volet qualitatif, les données d'entretiens sont généralement enregistrées grâce à des outils d'enregistrements plus ou moins sécurisés, dont les formats sont majoritairement en mp3, et dans une moindre mesure en mp4. Les données traitées via les logiciels de traitements tels que Nvivo, STATA, MAXQDA ou RedCap sont conservées dans les formats liés au logiciel.

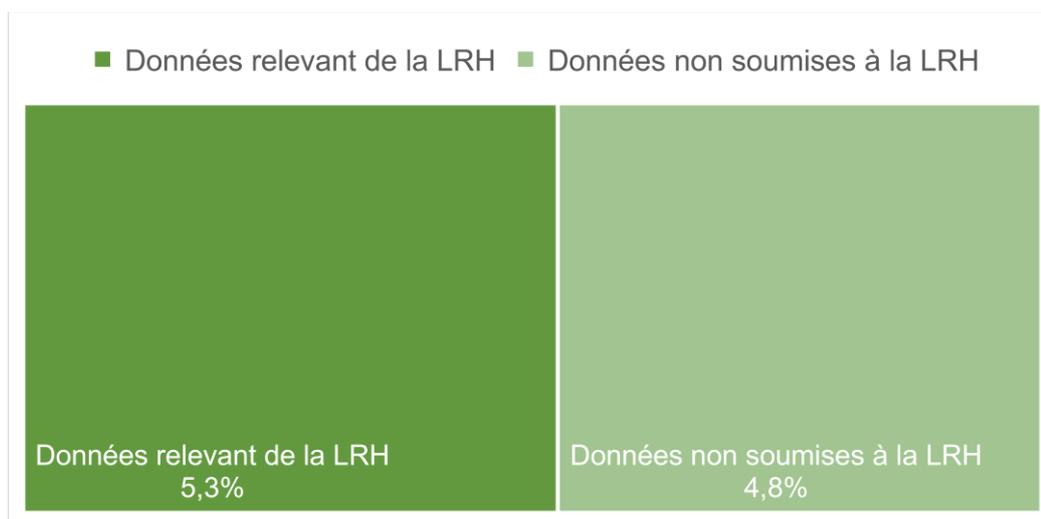
Figure 10 : Format des données



Cadre légal

Dans le cadre légal, en moyenne, un projet de recherche sur deux relève de la loi fédérale sur la recherche sur l'être humain. Cette proportion nécessite une attention particulière sur le cadre légal afin de connaître les dispositions exactes en cas de partage sur un dépôt.

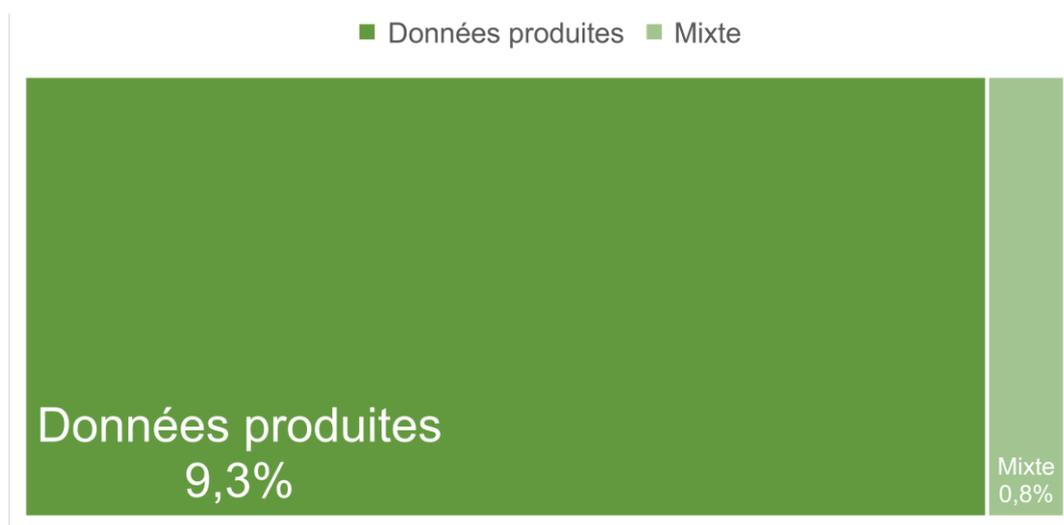
Figure 11 : Données relevant de la LRH



Gestion des données : production et partage

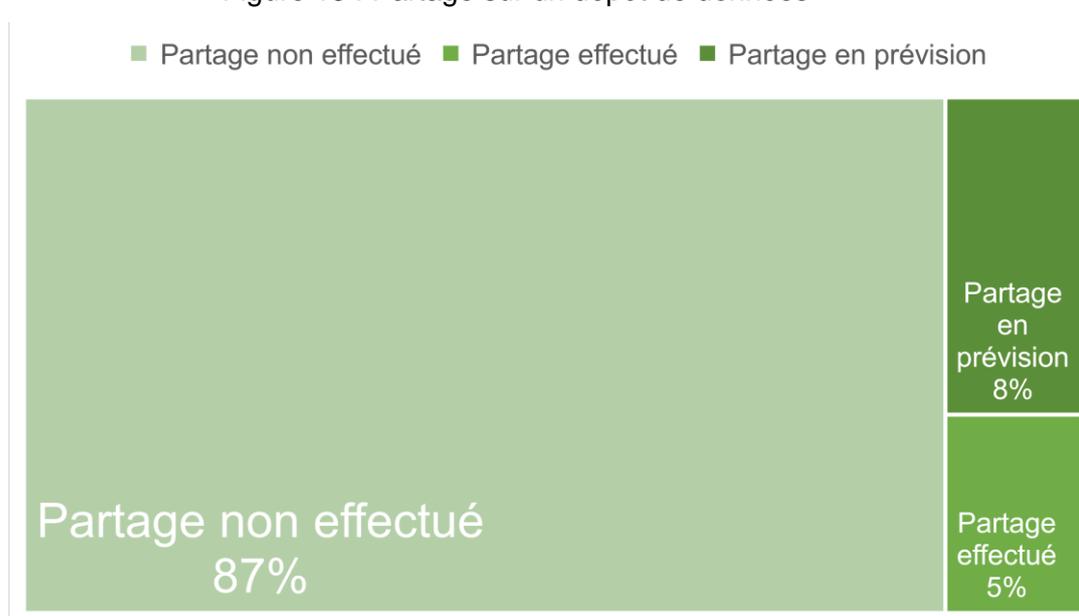
Quant à la gestion des données de recherche, nous avons choisi de mettre en lumière plusieurs tendances représentatives de la HESAV ; le partage ou non sur des dépôts ainsi que la production et/ou réutilisation des données. La très grande majorité des données sont produites. Un léger pourcentage des projets (7,5%) comprennent des données à la fois produites et réutilisées. Aucune recherche ne dispose que de données intégralement réutilisées.

Figure 12 : Production et/ou réutilisation



Le « non-partage » des données sur des dépôts est majoritaire. En effet, les exigences des bailleurs de fonds principaux tels que le FNS datant d'octobre 2017 (FNS 2017), la proportion est justifiée par la chronologie des événements. De plus, la prévision des partages est effectivement liée aux exigences des bailleurs de fonds et concorde avec la chronologie des événements.

Figure 13 : Partage sur un dépôt de données



4.2 Pratiques de gestion

Tout d'abord, au cours des entretiens il a été question de définir l'intérêt et les compétences des chercheur-r-se-s participant-e-s à l'étude. Nous ne présenterons ici que trois aspects de cette partie du guide pour des raisons de lisibilité : l'intérêt pour la gestion des données de recherche, une définition de celle-ci ainsi que les avantages et inconvénient liés à la gestion des données de recherche.

4.2.1 Connaissances et intérêts pour la gestion des données de recherche

L'intérêt pour la gestion des données de recherche relève chez certain-e-s chercheur-r-se-s d'une obligation à la suite des exigences des bailleurs de fonds, et pour d'autres, d'un besoin d'acquérir des connaissances et compétences techniques. À la suite de l'obligations du FNS de soumettre un DMP ainsi que de partager ses données en fin de projet, les chercheur-r-se-s ont été « contraint-e-s » de répondre à ces obligations, dont bon nombre de questionnements ont découlé : comment faire ? pourquoi le faire ? sur quel modèle ? Le flou autour des processus a semble-t-il été un frein dans la construction d'un intérêt ou d'une curiosité autour de la gestion des données de recherche.

Une définition ou des termes spécifiques spontanés ont été demandés aux participant-e-s afin de déterminer leurs connaissances terminologiques des données de recherche : la production est en grande majorité le point de départ de beaucoup de réponses. S'en suit de près la récolte de données via des entretiens ou des questionnaires puis les compléments de littérature, de documents administratifs ou d'archives. Un-e participant-e sur sept a alimenté sa réponse en ajoutant la reproductibilité des résultats et un-e participant-e sur sept a donné une définition des données de recherche en ce sens : elles regroupent l'ensemble des données dont a besoin une équipe de recherche pour reproduire un résultat.

Les réponses quant aux points positifs d'une gestion des données de recherche durant l'intégralité de leur cycle de vie ont été en grande majorité orientées sur la qualité des données. L'efficacité d'une telle gestion liée au gain de temps dans la recherche d'un document ou de l'auteur de celui-ci ont été relevés, l'importance du nommage des fichiers pour minimiser les biais d'erreur (créateur des fichiers, intervenants, transcription etc.) a aussi été soulevée comme point positif. Enfin, dans une moindre proportion mais tout de même présentes, les notions de structure, de traçabilité et de sécurité ont été mentionnées comme avantages à un tel type de gestion.

Figure 14 : Visualisation des avantages de la gestion des données de recherche



Les réponses quant aux inconvénients ont été principalement orientées vers le manque de compétences liées à la gestion de données ; qu'elles soient techniques ou intellectuelles, ces compétences représentent un frein pour beaucoup de chercheu-r-se-s. L'aspect chronophage découle de la réflexion liée à la rédaction du DMP qui pourrait être utile à l'argumentation d'autres éléments, le manque de connaissances informatiques ou juridiques liées à la sécurité et à la protection des données, le manque de pratique pour déposer des données sur un dépôt sont, entre autres, des éléments chronophages relevés comme étant des « obstacles » à la gestion des données.

Les réponses moins représentées relèvent de la sécurité des données, dont les lieux les plus sûrs sont difficiles d'accès (institutions partenaires n'ayant pas tous les accès, accès à Internet sécurisé, échange de documents parfois plus facile grâce aux messageries etc.).

Un décalage, parfois important, entre les demandes lors de la rédaction du DMP et la réalité a été soulevé. En découlent de fait des interrogations chez quelques chercheu-r-se-s, notamment sur la véritable utilité de la rédaction d'un tel outil.

Figure 15 : Visualisation des inconvénients de la gestion des données de recherche



4.2.2 Pratiques de gestion des données de recherche

Pour une présentation des résultats claire et suivie, nous avons décidé de les présenter en suivant le cycle de vie des données en six étapes, basé sur le guide d'entretien proposé aux participant-e-s.

Planification

Les équipes de recherche n'ont pas encore toutes été confrontées à la rédaction d'un DMP. Ainsi, certains y voient un passage obligé mais encore inconnu pour les prochains projets, alors que d'autres ont trouvé l'utilité de cette injonction qui structure les données de recherche en amont.

Les systèmes de nommage sont pour la totalité explicités oralement, mais aucun-e participant-e n'a mentionné l'existence d'un fichier officiel au sein de l'équipe pour le nommage des documents. Deux équipes ont délégué ces aspects de gestion des fichiers à des doctorants intégrés aux équipes de recherche afin de favoriser une gestion systématique.

Quant à l'organisation des fichiers, la majorité des équipes de recherche (4 sur 7) prévoient une organisation par nature de documents, puis chronologiquement ou alphabétiquement selon le système de nommage.

Le consentement est prévu dans tous les cas. En effet, devant soumettre les projets à la commission d'éthique, les projets de consentement des personnes participantes sont analysés en amont.

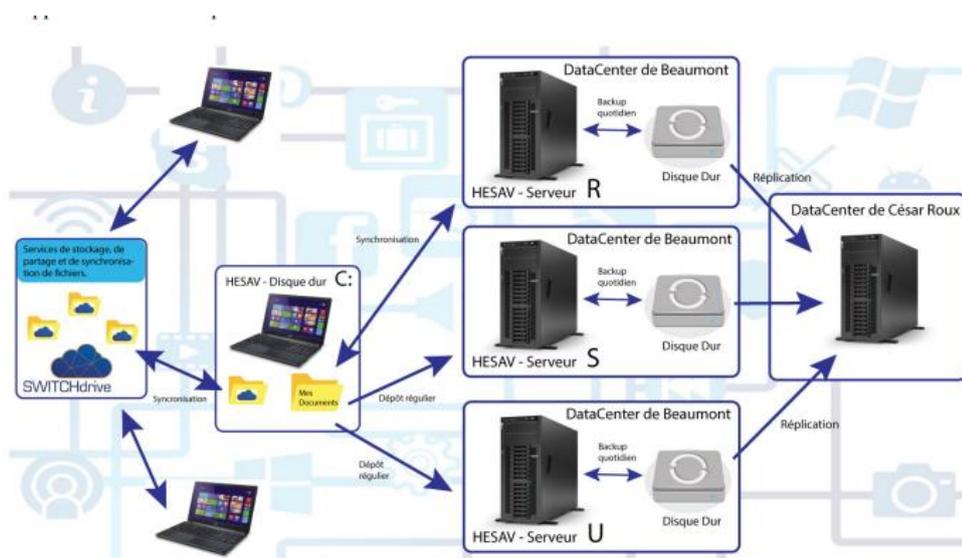
Trois participant-e-s sur sept ont admis des contrôles sur les données lors des différentes étapes du projet de recherche par souci de rigueur et de qualité, mais ce, sans planification officielle au sein du groupe de recherche. Les contrôles des données interviennent à des points stratégiques des projets comme le passage d'une étape du cycle à une autre (de la collecte à l'analyse) ou d'un traitement à un autre (de transcription à anonymisation par exemple).

Création et collecte

La gestion de la collecte est encadrée selon différents niveaux d'intensité en fonction les équipes. La totalité des recherches qualitatives possède des fichiers de collecte de données des participants séparés, anonymisés, sécurisés et dont les restrictions d'accès sont contrôlées.

Tout au long de chaque projet, la sécurité des données est assurée informatiquement par le service informatique selon le schéma ci-joint :

Figure 16 : La sécurité des données à la HESAV



(Source : Groupe de travail *Open Science* 2022)

Les chercheur-se-s interrogé-e-s suivent, pour la majorité, le même processus de stockage tout au long de la phase active des données, en utilisant le SWITCHdrive institutionnel pour un accès et un partage rapide et sécurisé des données. Les chercheur-se-s ont admis cette solution comme la plus confortable et la plus sûre.

Traitement et analyse

La grande majorité des données, nous l'avons vu au point précédent, relève du qualitatif. Les données sont collectées au travers d'entretiens enregistrés, retranscrits, anonymisés et encadrés pour trois chercheur-se-s sur sept par un contexte de collecte soit dans l'atmosphère de la récolte (sentiments, attitudes, langage corporel etc.) en vue de l'analyse, soit dans la

cadre stricte de collecte (nombre de personnes présentes, temps de l'entretien, etc.) pour encadrer la collecte de façon objective et purement descriptive.

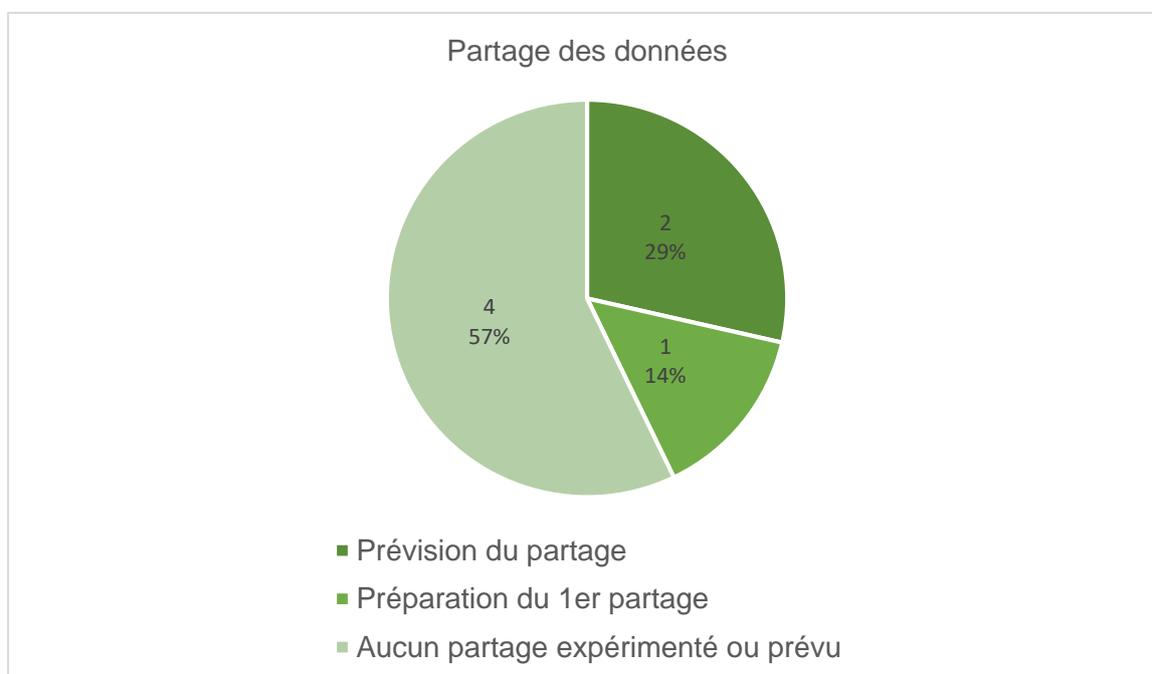
La majorité des groupes de recherche conservent l'intégralité des données et des versions tout au long du projet et ce, même après la publication des résultats. Plusieurs raisons ont été avancées pour justifier cette conservation : accident informatique, erreur humaine, volonté personnelle de conservation ou encore peur de la suppression.

Malgré les motifs avancés, la conservation non contrôlée d'éléments non structurés demeure dysfonctionnelle en termes de gestion de données de recherche.

Partage

Deux équipes sur sept prévoient le partage des données et une équipe sur sept partagera pour la première fois ses données. Quatre participant-e-s n'ont pas encore expérimenté le partage et montrent quelques appréhensions quant à cette pratique.

Figure 17 : Partage des données au sein des équipes



Plusieurs motifs sont avancés pour justifier cette proportion, mais la majorité des personnes interrogées avancent un doute quant à l'utilité dudit partage. En effet, trois participant-e-s sur sept ont émis des doutes quant à l'utilité d'un partage car chaque collecte répond à une question de recherche spécifique et il serait difficile de reprendre un set de données récolté spécifiquement pour telle question dans le cadre d'une recherche dont les objectifs sont différents. La question du partage partiel est aussi avancée et suscite des interrogations sur la validité, l'utilité et la nécessité de proposer un set de données qui ne serait pas intégral. Des processus obligatoires peuvent inciter un-e chercheur-se sur sept à partager les données, toutefois avec la volonté de n'en partager que le strict nécessaire, ceci pour les raisons avancées plus haut. Deux participant-e-s sur sept émettent un doute quant au partage, du fait de la méconnaissance des dépôts et l'éventuelle charge de travail importante y relative.

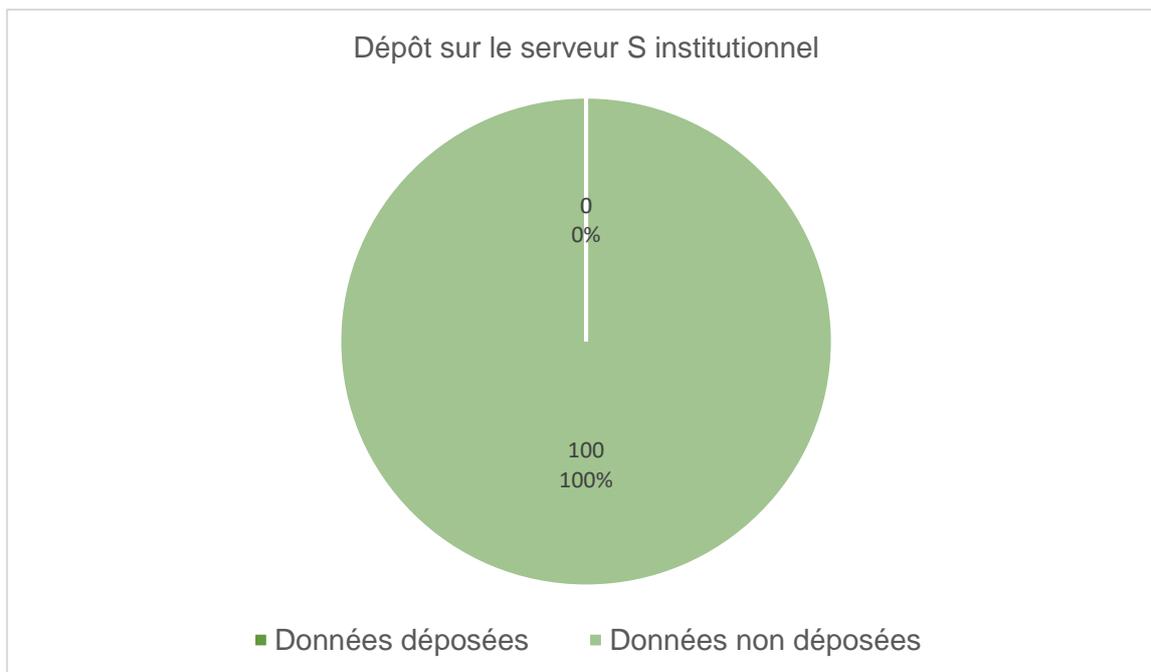
Trois participant-e-s sur sept témoignent de l'utilité de cette pratique et y voient une opportunité de ne pas laisser les données se perdre. Certain-e-s y voient une occasion de recherche

comparative. Un-e participant-e sur sept y voit aussi l'occasion de valoriser les données sous différentes formes ; l'enseignement ou encore la réutilisation sous un autre format tel que des films ou des rétrospectives de projet de recherche par exemple.

Archivage et pérennisation

Aucun-e chercheu-r-se n'a pour habitude systématique de déposer les données de fin de recherche sur le serveur S institutionnel.

Figure 18 : Dépôt des données sur le serveur institutionnel par les équipes



La pérennisation des données ne semble pas être au cœur des préoccupations notamment au niveau des formats adéquats méconnus pour six chercheu-r-se-s interrogé-e-s sur sept. Les habitudes de conservation à long terme se composent majoritairement de conservation sur des machines personnelles, aux formats inchangés et parfois en plusieurs versions pour notamment faciliter l'accès aux documents après la publication des résultats. Ainsi, les habitudes déjà observées lors du rapport d'activités de 2019 ont été confirmées lors des entretiens :

« À ce stade, peu de chercheur-e-s consacrent encore de l'énergie à nettoyer et classer les données de leur recherche afin de les archiver et de les rendre accessibles et compréhensibles à plus ou moins long terme. De fait, les données de recherches antérieures se trouvent bien souvent « éparpillées » sur différentes machines, dupliquées dans des versions différentes, souvent non-anonymisées et rarement accessibles/compréhensibles sans l'intervention de celui-celle qui a procédé à leur classification selon sa propre logique. » (Rapport d'activité 2019, p. 13)

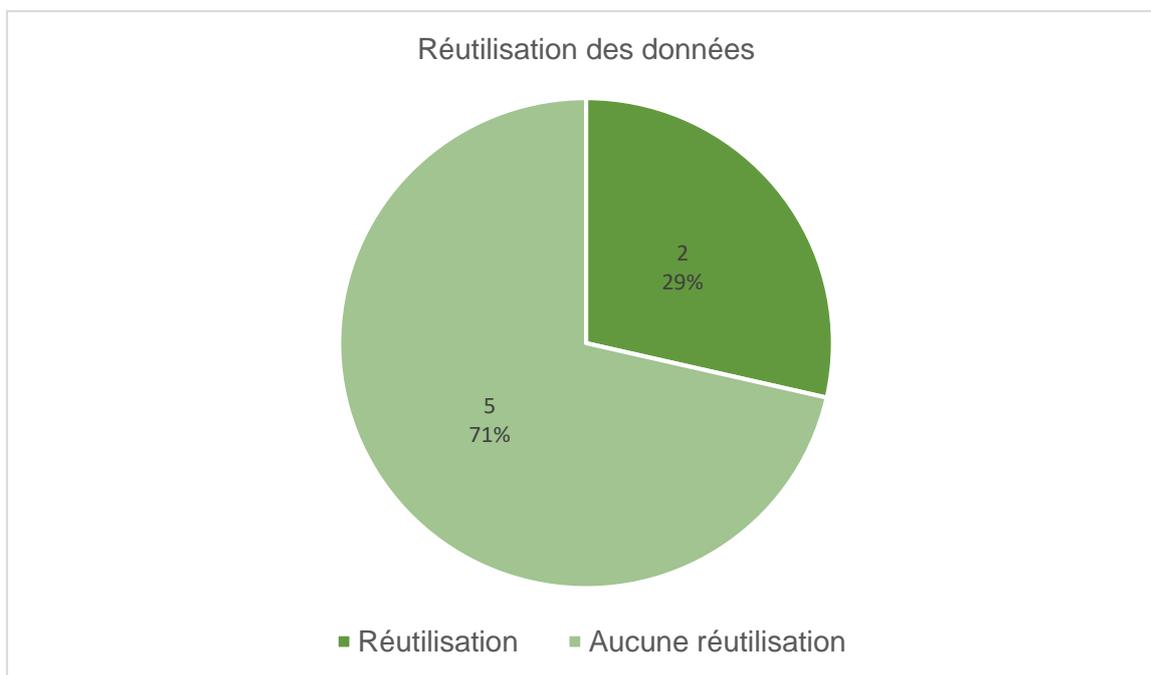
Réutilisation

Deux chercheu-r-se-s sur sept ont déjà réutilisé des données accessibles via des bases de données telles que les données issues de l'OFSP, notamment dans l'enseignement.

Cinq chercheu-r-se-s sur sept n'ont pas encore réutilisé ou eu l'occasion de réutiliser des données issues de dépôts.

La réutilisation se fait d'une autre manière ; deux chercheurs sur sept ont réutilisé leurs propres données issues d'anciens projets.

Figure 19 : Réutilisation des données



5. Diagnostic

Les diagnostics suivants permettent le listage d'éléments-clés à faire figurer dans une future politique et de construire une liste préliminaire de dépôts en intégrant notamment le niveau de curation et les solutions de pérennisation.

Appuyé sur un modèle de maturité spécifique au management des données (Baolong 2018, cf. Annexe 3), le diagnostic des pratiques est séparé selon trois des quatre catégories présentées dans le dit-modèle à savoir : la collecte (planification en amont, collecte, réutilisation etc.), la gestion des données (analyse, traitement, systèmes, etc.), la sécurité des données (systèmes utilisés, stockage, backup, etc.)

Ce modèle est construit en cinq niveaux que l'on peut traduire comme suit : niveau faible, niveau basique, niveau moyen, niveau élevé et niveau optimal.

5.1 Diagnostic des pratiques appuyé sur un modèle de maturité

Collecte

Pour la collecte des données, les niveaux de maturité sont établis comme suit :

- Niveau faible : collecte spontanée, non standardisée, sans différenciation des supports et des types de données
- Niveau basique : Système de collecte développé, sans processus standardisés
- Niveau moyen : Différents outils et processus de qualité standardisés au sein de l'équipe de recherche
- Niveau élevé : Différents outils et processus de qualité standardisés au sein de l'équipe de recherche et établissement d'une documentation de collecte
- Niveau optimal : Ajuster les mécanismes au cours du projet

Cinq participant-e-s se trouvent au niveau moyen pour la collecte des données. Ils-elles utilisent différents outils et processus de qualité standardisés au sein de l'équipe de recherche. Pour accéder au niveau supérieur, il manque une documentation de collecte stricte avec répartition des rôles, responsabilités et une planification stricte. La rédaction d'un DMP et la production d'une telle documentation à chaque projet permettrait de passer un niveau supérieur.

Deux participant-e-s se situent à un niveau de collecte élevé. En effet, ils fournissent une documentation de collecte qui confèrent un niveau supérieur de qualité aux données selon le modèle de maturité. Afin de passer au niveau supérieur, des contrôles réguliers et des ajustements peuvent être prévus dans la planification.

Le tableau suivant détaille le niveau de maturité de collecte de chaque participant-e.

Tableau 5 : Modèle de maturité appliqué aux participant-e-s – Collecte

	Collecte
1	<p>Rédaction rapide et dans l'urgence - Métadonnées gérées par une tierce personne – Pas de contrôles prévus – DMP = nouveauté – Pas d'acquisition - Documentation de collecte dans le devis</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Système de collecte développé avec procédures et documentation</p>
2	<p>Pas de rédaction de DMP – Pas de réel intérêt pour les métadonnées – Contrôles ponctuels – Pas d'acquisition – Peu de documentation de collecte</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Système de collecte développé avec procédures et documentation</p>
3	<p>Pas de rédaction de DMP - Pas de réel intérêt pour les métadonnées - Contrôles ponctuels – Pas d'acquisition – Aucune documentation de collecte</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Système de collecte développé avec procédures et documentation</p>
4	<p>Planification adaptée à l'équipe – Pas de rédaction de DMP – Organisation au cas par cas – Réel intérêt pour les métadonnées – Documentation de collecte</p> <p style="text-align: center;">Niveau 4 : Elevé</p> <p style="text-align: center;">Différents outils et processus de qualité standardisés au sein de l'équipe de recherche et établissement d'une documentation de collecte</p>
5	<p>Connaissance du DMP et utilisation - Planification stricte - Réutilisation des données propres – Intérêt pour les métadonnées – Pas de documentation de collecte</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Système de collecte développé avec procédures et documentation</p>
6	<p>Rédaction de DMP – Intérêt pour les métadonnées – Peu de documentation de collecte – Pas de réutilisation</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Système de collecte développé avec procédures et documentation</p>
7	<p>Rédaction de DMP – Partage prévu – Documentation de collecte – Réel intérêt pour les métadonnées – Planification de monitoring</p> <p style="text-align: center;">Niveau 4 : Elevé</p> <p style="text-align: center;">Différents outils et processus de qualité standardisés au sein de l'équipe de recherche et établissement d'une documentation de collecte</p>

Gestion

Les systèmes de niveau sont établis comme suit :

- Niveau faible : Gestion minimum sans procédures
- Niveau basique : Règles préliminaires de collecte, de gestion et d'utilisation respectées
- Niveau moyen : Différents outils et processus de qualité standardisés au sein de l'équipe de recherche et établissement d'une documentation de gestion pour assurer la qualité des données
- Niveau élevé : Utilisation d'un modèle standardisé pour les métadonnées et des standards d'acquisition, de nettoyage et de gestion
- Niveau optimal : Contrôle et maîtrise du cycle de vie des données dans son intégralité pour optimiser et maintenir la qualité des données

Cinq participant-e-s se situent dans le niveau basique. Ils respectent les règles préliminaires de collecte, de gestion et d'utilisation. Afin de passer au niveau de maturité supérieur, il serait bon d'encadrer la collecte d'une documentation de collecte, de prévoir des contrôles de la qualité et de l'intégrité des données au fil du projet, mais aussi de formater les données en fonction des phases du projets. La conservation en l'état des données collectées engendre un risque de perte de qualité de lecture, voire d'impossibilité de lecture en fonction des machines.

Un-e participant-e se situe dans le niveau moyen de gestion. Avec un formatage partiel des données, des contrôles répétés et prévus ainsi qu'un système de nommage et d'organisation pris en charge par une unique personne, la gestion atteint un niveau moyen et assure une bonne qualité des données. Toutefois, pour passer au niveau de maturité supérieur, il mériterait d'intégrer une documentation de gestion, un formatage contrôlé des données et un nombre plus élevé de contrôles.

Un-e participant-e se situe dans le niveau élevé. Avec plusieurs outils de contrôles prévus et fréquents, un système de nettoyage élevé et contrôlé, une documentation de gestion ainsi qu'une nomenclature documentée, la qualité des données est garantie. Toutefois, une réelle connaissance du cycle de vie des données et des processus y a relatifs permettrait d'atteindre le niveau de maturité optimal en matière de gestion des données.

Le tableau suivant détaille le niveau de maturité de gestion de chaque participant-e.

Tableau 6 : Modèle de maturité appliqué aux participant-e-s – Gestion

	Gestion
1	<p>Système basique de nommage non documenté – Traitements sur logiciels - Anonymisation faite par une tierce personne - Peu de contrôles en cours de projet – Pas de formatage - Partage envisagé</p> <p style="text-align: center;">Niveau 2 : Basique</p> <p style="text-align: center;">Règles préliminaires de collecte, nettoyage et gestion respectées</p>
2	<p>Organisation des fichiers stricte mais pas documentée – Nommage strict des fichiers mais pas documenté - Pas de formatage – Anonymisation garantie – Peu de contrôles</p> <p style="text-align: center;">Niveau 2 : Basique</p> <p style="text-align: center;">Règles préliminaires de collecte, nettoyage et gestion respectées</p>
3	<p>Documentation du traitement - Pas de formatage – Aucune documentation d'organisation des fichiers – Aucune documentation de nommage</p> <p style="text-align: center;">Niveau 2 : Basique</p> <p style="text-align: center;">Règles préliminaires de collecte, nettoyage et gestion respectées</p>
4	<p>Utilisation de logiciels sécurisés – Formatage standard – Contrôles répétés et prévus – Pas de partage réalisé ou prévu – Personne responsable de la gestion des fichiers et du nommage</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Utilisation de plusieurs outils de gestion</p>
5	<p>Système d'organisation et de nommage documenté – Pas de contrôles prévus, ponctuels - Pas de formatage</p> <p style="text-align: center;">Niveau 2 : Basique</p> <p style="text-align: center;">Règles préliminaires de collecte, nettoyage et gestion respectées</p>
6	<p>Personne responsable de la gestion des fichiers et du nommage - Pas de formatage – Anonymisation considérée comme basique - Contrôles aléatoires et imprévus</p> <p style="text-align: center;">Niveau 2 : Basique</p> <p style="text-align: center;">Règles préliminaires de collecte, nettoyage et gestion respectées</p>
7	<p>Nomenclature documentée – Gestion documentée - Formatage standard – Contrôles répétés et prévus – Contrôles des encodages – Nettoyage contrôlé et documenté</p> <p style="text-align: center;">Niveau 4 : Elevé</p> <p style="text-align: center;">Système de gestion développé avec documentations, classification, nombre de contrôles et nettoyage élevés</p>

Sécurité des données

Les systèmes de niveau sont établis comme suit :

- Niveau faible : Stockage des données sur disque dur personnel, faible niveau de sécurité
- Niveau basique : Données centralisées, backup de sécurité et stratégie de contrôle
- Niveau moyen : Mise à jour régulière, utilisation de systèmes sécurisés
- Niveau élevé : Systèmes de sécurités développés, politique de stockage et politique d'accès
- Niveau optimal : Gestion anticipée d'un sinistre pour consolider la sécurité des données

La majorité des participant-e-s se situent dans le niveau moyen (4/7), deux chercheu-r-se-s se situent dans le niveau basique, considérant l'utilisation de systèmes peu sécurisés tels que des clés ou la modification de documents en local sur des machines personnelles et un-e chercheu-r-se sur sept se situe dans le niveau élevé proposant un processus de stockage stricte et validé au sein de l'équipe ainsi qu'une documentation de conservation. Nous considérons le niveau de maturité de ces participants sur l'ensemble des dispositifs utilisés à savoir les solutions sécurisées proposées par l'institution en parallèle de leurs pratiques de gestion de la sécurité.

Chaque chercheu-r-se a démontré une réelle attention pour la sécurité des données avec toutefois l'utilisation de systèmes employés peu sécurisés. En effet, le stockage sur clé est encore utilisé pour deux chercheu-r-se-s sur sept, mais la majorité des chercheu-r-se-s se rendent compte du peu de sécurité de cette méthode. Toutefois, pour des raisons de confort et de rapidité, deux chercheu-r-se-s sur sept envoient des versions de documents sur une adresse électronique afin de les retravailler en local. De plus, deux chercheu-r-se-s sur sept utilisent leurs téléphones portables personnels pour enregistrer des entretiens, la sécurité des enregistrements n'étant pas assurée.

Le partage, le stockage et de synchronisation des documents sur les systèmes institutionnels sont encouragés par le service informatique et annoncé comme sécurisé. Cette pratique est suivie par la totalité des chercheu-r-se-s interrogé-e-s. De plus, la documentation relative à la sécurité des données et la sensibilisation effectuée par le groupe *Open Science* contribuent à cette grande attention à la sécurité des données.

Le tableau suivant détaille le niveau de maturité de sécurité de chaque participant-e.

Tableau 7 : Modèle de maturité appliqué aux participant-e-s – Sécurité

	Sécurité
1	<p>Clé – Disquette – Backup régulier - Utilisation du matériel institutionnel existant - Restrictions d'accès - Pas de documentation de conservation</p> <p style="text-align: center;">Niveau 2 : Basique</p> <p style="text-align: center;">Données centralisées, backup de sécurité et stratégie de contrôle</p>
2	<p>Attention portée au transfert des documents – Back up sur machines personnelles – Clé – Pas de courriels – Utilisation du matériel institutionnel existant - Restrictions d'accès - Pas de documentation de conservation</p> <p style="text-align: center;">Niveau 2 : Basique</p> <p style="text-align: center;">Données centralisées, backup de sécurité et stratégie de contrôle</p>
3	<p>Incertitude quant à la responsabilité des données -Utilisation du matériel institutionnel existant - Cut dans les fichiers audio – Restrictions d'accès - Pas de documentation de conservation</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Mise à jour régulière, utilisation de systèmes sécurisés</p>
4	<p>Restrictions d'accès - Utilisation du matériel institutionnel existant – Codage double – Back up régulier – Archivage sur machines personnelles - Pas de documentation de conservation</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Mise à jour régulière, utilisation de systèmes sécurisés</p>
5	<p>Restrictions d'accès - Utilisation du matériel institutionnel existant – Logiciels sécurisés - Pas de documentation de conservation</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Mise à jour régulière, utilisation de systèmes sécurisés</p>
6	<p>Restrictions d'accès - Utilisation du matériel institutionnel existant – Logiciels sécurisés – backup – Archivage standard – Pas de documentation de conservation</p> <p style="text-align: center;">Niveau 3 : Moyen</p> <p style="text-align: center;">Mise à jour régulière, utilisation de systèmes sécurisés</p>
7	<p>Restrictions d'accès - Utilisation du matériel institutionnel existant – Logiciels sécurisés – backup – choix stricte de processus de stockage – documentation de conservation</p> <p style="text-align: center;">Niveau 4 : Elevé</p> <p style="text-align: center;">Systemes de sécurités développés, politique de stockage et politique d'accès</p>

Résumé

En résumé, tous-tes les chercheu-r-se-s interrogé-e-s se situent entre le niveau de maturité basique et le niveau de maturité élevé pour les trois domaines confondus comme le met en évidence la zone grisée du tableau ci-dessous.

Tableau 8 : Résumé du diagnostic

	Collecte	Gestion	Sécurité
Niveau faible	X	X	X
Niveau basique	X		
Niveau moyen			
Niveau élevé			
Niveau optimal	X	X	X

La collecte demeure en moyenne de bonne qualité avec un ensemble de processus standardisés de qualités avec dans quelques équipes une volonté de documentation de collecte.

Les systèmes proposés par l'institution sont sécurisés, localisés en Suisse, proposent des restrictions d'accès et des sauvegardes régulières. L'utilisation de ces dispositifs est encouragée par le service informatique et demeure une solution sécurisée et confortable. Toutefois, certaines habitudes des chercheu-r-se-s mériteraient quelques ajustements. L'envoi de versions non définitives sur des messageries électroniques afin de retravailler les textes en local est à éviter. En effet, ce comportement de gestion augmente les risques d'erreur et aucun listage de modification ou suppression ne peut être assuré.

La gestion mériterait de passer au niveau supérieur pour la majorité des équipes. En effet, le but est d'atteindre le niveau optimal en contrôlant les données tout au long de leur cycle de vie. Pour ce faire, plusieurs outils sont proposés au sein de ce travail afin d'arriver à ce niveau : le schéma du cycle de vie des données permettrait une meilleure visibilité des étapes du cycle et la checklist permet de systématiser les processus aux différentes phases du cycle.

5.2 Diagnostic du cycle de vie des données de recherche

Planification

À l'heure de ce mandat, la rédaction d'un DMP ne s'effectue que si le bailleur de fonds l'exige. Remplir un DMP pour tous les projets, financés ou non par des fonds publics, devrait être une opération systématique car ce document permet une gestion anticipée des données de recherche et propose à l'équipe de recherche une vue d'ensemble des processus à entreprendre. De plus, l'utilité d'un tel outil a été confirmée par une majorité de chercheu-r-se-s interrogé-e-s, cette pratique devrait être donc encouragée de façon systématique.

De plus, nous souhaitons insister sur la prévision des contrôles qualité sur les données collectées : peu d'équipes le mettent en pratique de façon organisée et systématique. Cette pratique mériterait d'être encouragée afin de prévoir des contrôles aux différentes étapes de la recherche en vue d'optimiser la qualité des données de recherche.

Création et collecte

La collecte est dans l'ensemble bien menée, encadrée de règles strictes, notamment au niveau éthique. Toutefois, un guide de bonnes pratiques ou la checklist proposée dans les outils, recommande d'encadrer la collecte d'une documentation, d'un système de gestion des fichiers ainsi que d'un système de nommage des fichiers afin d'organiser efficacement la collecte au sein des équipes de recherche.

Traitement et analyse

Le rapport d'activité de 2019 relève que le nettoyage n'est pas une opération systématique (Amiotte-Suchet 2019) et ceci a été confirmé dans les entretiens.

Les processus d'anonymisation et les doutes des chercheurs interrogés ont mis en lumière un besoin d'encadrement et la nécessité de mettre en place un système d'anonymisation propre et structuré.

En premier lieu, il semble nécessaire de rappeler la distinction entre pseudonymisation et anonymisation. Le premier est un processus de transformation des données personnelles afin qu'elles ne puissent plus être attribuées à une personne « sans avoir recours à des informations supplémentaires ». (RGPD 2017 ; Bayle 2020 ; Maklhof-Shabou 2021a). Le second est un processus de transformation des données personnelles de manière à ce qu'elles ne puissent plus être ré-identifiées après traitement et demeure généralement utilisé pour les données qui seront publiées (RGPD 2017 ; Bayle 2020 ; Maklhof-Shabou 2021a ; EPFL, 2022b).

Le rappel des systèmes d'anonymisation pourrait être utile afin d'optimiser ce traitement sur les données (Makhlouf Shabou2021a) :

- Généralisation : diminution de la granularité en généralisant les variables
- Suppression : élimination des données d'identification ou de parties erronées
- Ajout de fausses données : empêche l'identification d'enregistrements spécifiques
- Mélange : consiste à mélanger plusieurs colonnes de données sans pour autant compromettre leur intégrité

La pseudonymisation peut être effectuée selon deux processus :

- Générer une clé d'identification, stockée de façon sécurisée avec des restrictions d'accès
- Méthode algorithmique pour un chiffrement des données également propice pour la conservation à long terme

Partage

Le partage des données n'est pas un processus habituel au sein des groupes de recherche à l'heure actuelle.

Si réutiliser les données s'avèrent un système moins chronophage et moins coûteux dans une recherche comparative par exemple, le partage est une pratique à promouvoir pour éviter les travaux doubles.

De plus, la rédaction d'un fichier README n'est pas une priorité pour les chercheu-r-se-s. En effet, aucun-e participant-e n'a montré d'intérêt ou de connaissance spécifique pour un tel fichier dans le cadre d'un partage. Pour cette raison, nous établissons une proposition de fichier README dans les recommandations de ce travail.

Archivage et pérennisation

Cette étape du cycle suscite le plus de difficultés et d'interrogations et demeure celle sur laquelle l'institution doit porter son attention.

Les réflexions actuelles sur la gestion des données de recherches *a fortiori* en sciences humaines et sociales portent sur la pérennisation des données. (Huyghe et al. 2018); Pérennisation, conservation, archivage, stockage à court ou long terme (Université de Lausanne, [sans date]a ; Université de Genève [sans date]b ; EPFL 2022h, Huyghe 2018 ; Van Gulick 2016) : les termes sont nombreux dans les cycles de vie, souvent méconnus et de ce fait intervertis ou utilisés l'un pour l'autre par les chercheu-r-se-s. Tout comme le souligne Marie Hyughe dans son rapport, « le stockage et l'archivage sont des pratiques de recherche à systématiser » aussi au sein de la HESAV. Ce manque de systématique a été relevé au cours d'une enquête (Amiotte-Suchet 2019) mais aussi lors des entretiens. Dans la pratique, il est essentiel de différencier le stockage de l'archivage ; le premier concerne la conservation pour une durée donnée, alors que le second process encadre une conservation définitive ou à long terme. (Hyughe 2018)

La possibilité de stocker les données en fin de projet sur le disque réseau S de l'institution est proposée, toutefois cette pratique n'est pas considérée comme de l'archivage et surtout n'est pas systématisée (Amiotte-Suchet 2019). Cet élément a été confirmé par les réponses aux entretiens dirigés. Il est donc important de mettre l'accent sur cette partie du cycle en adoptant, dans une politique future, l'archivage systématique des données en fin de recherche, non seulement sur les disques réseaux S institutionnels, mais aussi sur un dépôt de données proposant ce service d'archivage, et ceci aux formats adaptés.

Il existe un réel besoin de soutien et d'encadrement dans le formatage, la contextualisation ainsi que dans la création des métadonnées. Ces processus font partie intégrante de l'archivage et de la pérennisation et méritent d'être encouragés et systématisés. Pour ce faire, quelques recommandations dans la checklist ainsi qu'une aide à l'archivage figurent au point 6 de ce travail.

Réutilisation

Cette étape du cycle de vie des données de recherche reste encore très peu mise en avant au sein de la HESAV.

Cette pratique pourrait être utile pour des recherches comparatives ou pour l'enseignement. De plus, la réutilisation de ses propres données, pratique déjà mise en place par quelques chercheu-r-se-s, gagne à être encouragée.

5.3 Besoins en formation

5.3.1 Formation des chercheurs

Le groupe *Open Science* de la HESAV met à disposition sur l'intranet bon nombre de documents : « Aide-mémoire juridique », « Sécurité des données » ou « Aide à la rédaction d'un DMP ». Différents rendez-vous allant de la formation interne pour la rédaction d'un DMP à des ateliers participatifs autour de l'*Open Science* sont proposés.

De plus, les présentations pour SWISSUBase ainsi que les journées *Open Science* de la HES-SO sont rappelées aux chercheu-r-se-s afin de motiver leur participation.

Au cours des entretiens, aucun-e chercheu-r-se n'a émis la volonté de participer à des formations autour de la GDR. En effet, l'aspect chronophage et peu pratique des formations a été évoqué comme principales raisons. Deux chercheu-r-se-s ont toutefois noté l'utilité de supports infographiques pour aller chercher l'information en temps voulu. En effet, la volonté de pouvoir obtenir les informations au moment où la difficulté se pose demeure essentielle pour les chercheu-r-se-s. Dans cette optique, deux outils sont proposés dans ce travail pour répondre à ce besoin.

Plusieurs dispositifs sont mis en place au sein des universités et hautes écoles en Suisse romande pour la formation des chercheu-r-se-s en matière de gestion des données : EPFL, UNIL, UNIGE, Journées *Open Science* HES-SO sont quelques exemples des possibilités. Les sites Internet de ces institutions proposent des journées de formation ou formations continues sur inscription pour découvrir ou consolider des compétences en gestion des données de recherche. De plus, certaines institutions proposant des solutions suisses de dépôts de données tels que OLOS ou SWISSUbase encadrent leur solution de formations et de vidéos de découverte en ligne.

5.3.2 Formation de l'équipe Bibliothèque

L'équipe Bibliothèque de l'institution est constituée de trois bibliothécaires et de trois agent-e-s en information documentaire.

Deux bibliothécaires font partie du groupe de travail *Open Science* mais témoignent d'un besoin en formation. Ayant acquis des connaissances de base en gestion de données de recherche, les deux participantes en connaissent la définition, le type, les supports ainsi que le cycle de vie en superficie. Toutefois, elles notent un réel besoin en formation au niveau des formats des données pour la conservation et un besoin en formation pour les compétences techniques nécessaires pour déposer les données sur un dépôt.

De plus, il existe un besoin en formation pour comprendre les avantages et le fonctionnement de la gestion des données de recherche pour réussir à faire adhérer les chercheu-r-se-s à un mouvement *Open Science*, déjà assimilé grâce à l'*Open Data*.

Les principes FAIR sont familiers et connus, toutefois un manque de visibilité des solutions de dépôts et de partage est constaté. La consultation des séances d'informations ou les vidéos de présentations en ligne sur les différentes solutions existantes en Suisse est fortement encouragée pour pallier ce besoin en formation.

Les deux participantes ont été unanimes quant à la nécessité de posséder un cadre clair pour pouvoir intervenir ou auquel cas, rediriger les questions auprès des personnes compétentes.

Enfin, nous pouvons suggérer quelques pistes pour la formation de l'équipe Bibliothèque. En Suisse, les institutions les plus engagées pour la formation continue en *Data Management* sont l'Ecole polytechnique fédérale de Lausanne ainsi que le CERN (Fachinotti et al. 2016 ; HEC 2022). L'EPFL propose en collaboration avec la HEC un CAS en *Data Science & management* d'une durée de seize jours sur neuf mois orienté sur la science des données, l'intelligence artificielle et le machine learning (EPFL 2022e). L'un des objectifs de la formation étant de mettre en œuvre des pratiques efficaces en gestion des données en tant que ressources stratégiques, cette option peut être utile dans le cas d'une formation d'une spécialiste en information documentaire. Un Master en Data Science est également proposé par l'EPFL, mais dure en moyenne trois semestres et propose un plan d'études très pointu donnant un panorama de toutes les possibilités existantes en matière de sciences des données. Enfin, la Haute école de gestion de Genève propose au sein du Master en Sciences de l'information (quatre semestres), une spécialisation de deux semestres sur la gouvernance des données et intègre au sein de ce cursus des éléments orientés sur les données de recherche.

Sur la base de ces diagnostics, nous pouvons procéder aux recommandations correspondant aux besoins de l'institution et de ses chercheu-r-se-s.

6. Recommandations

Les recommandations de ce travail se structurent en trois points : des propositions de gestion des ressources humaines pour la GDR au sein de l'équipe Bibliothèque, des propositions d'aides à l'archivage et au partage des données ainsi qu'une analyse des dépôts sélectionnés selon des critères répondant aux besoins de l'institution.

6.1 Propositions de gestion de ressources humaines

Sont explicitées sous ce point, trois propositions de gestion des ressources humaines pour la gestion des données de recherche. Ces éléments ont été abordés au cours d'entretien dirigés avec deux membres de l'équipe Bibliothèque.

6.1.1 Proposition 1

Formation d'un-e spécialiste en information documentaire en place en collaboration avec le groupe *Open Science*

Sur la base du fonctionnement actuel, la première proposition consiste en la formation d'une bibliothécaire déjà en fonction au sein de l'équipe Bibliothèque. Une orientation de data steward serait avantageuse pour optimiser une gestion du cycle de vie des données dans leur intégralité et permettrait une grande sensibilisation et un accompagnement idéal pour les processus de diffusion, de sécurité et de gestion.

Au sein de l'équipe, cette proposition de ressource a l'avantage de ne pas modifier la structure actuelle de l'équipe. Cette option serait une opportunité pour un-e membre de l'équipe Bibliothèque de se former dans un domaine actuel et moderne, dont les besoins et les retombées sont significatifs. La formation d'une personne déjà en fonction au sein de l'équipe permettrait la diffusion des savoirs en interne. De plus, la personne concernée bénéficierait d'une expérience avantageuse des pratiques internes à l'institution de même que la connaissance des chercheu-r-se-s et de leurs besoins. Néanmoins, notons qu'il est difficile de trouver à la fois une personne intéressée par le domaine de la GDR et prête à s'investir à temps plein ou partiel dans une formation. L'équipe étant essentiellement constituée de pourcentages partiels, il est contraignant et difficile d'augmenter ou diminuer les taux. De plus, un élément supplémentaire constitue un frein dans cette option : le choix de la formation doit être scrupuleusement analysé pour être efficace. De plus, la révision des taux est à envisager et semble difficile selon la structure actuelle.

D'un point de vue financier, l'avantage reviendrait à la personne formée qui verrait sa rémunération augmentée et pourrait peut-être bénéficier d'un changement de classe/statut. Toutefois, une telle formation est onéreuse et cette proposition devrait être prise en charge intégralement ou partiellement par l'institution pour motiver les collaborateurs et collaboratrices à se former.

Les avantages d'une telle proposition sont variés : une seule personne serait référence pour la GDR au sein de l'institution et un-e membre du groupe *Open Science* serait spécialisé-e dans la GDR et confirmerait les pratiques à adopter. Finalement, la formation d'une personne interne à l'institution constituerait l'avantage de l'expérience et de la connaissance du fonctionnement interne pour faciliter les échanges et la mise en place de nouvelles dispositions.

Le tableau ci-dessous résume les éléments positifs ainsi que les éléments à prendre en considération pour l'implémentation d'une telle proposition sous trois angles : la gestion de l'équipe et de ses ressources, l'aspect financier de cette option ainsi que les impacts en découlant.

Tableau 9 : Analyse de la proposition 1

		
Equipe	<p>Aucune modification de la composition actuelle de l'équipe.</p> <p>Motiver un membre de l'équipe dans de nouveaux défis</p> <p>Une unique personne de référence pour le domaine</p> <p>Diffusion des savoirs en interne</p> <p>Connaissances du fonctionnement interne de l'institution</p> <p>Connaissances des chercheur-se-s et de leurs habitudes</p>	<p>Trouver une personne intéressée par le domaine et prête à s'investir dans une formation</p> <p>Choix de la bonne formation (temps, lieu, débouchés, contenu, etc.)</p> <p>Modification et répartition des tâches</p> <p>Redéfinir les pourcentages au sein de l'équipe</p>
Coûts	<p>Changement de rémunération du membre de l'équipe</p> <p>Changement éventuel de statut et/ou de classe salariale</p>	<p>Paiement intégral ou partiel de la formation</p> <p>Révision du salaire du spécialiste en fonction des tâches et du pourcentage</p>
Impacts	<p>Une seule personne de référence au sein de l'institution pour la GDR</p> <p>Un membre du groupe <i>Open Science</i> spécialisé-e dans la GDR</p> <p>Connaissances du fonctionnement interne pour faciliter les échanges et les analyses</p>	<p>Révision des tâches et fonctions d'un-e membre de l'équipe Bibliothèque</p> <p>Investissement personnel non négligeable de la part de la personne formée</p>

6.1.2 Proposition 2

Engagement d'un-e spécialiste en information documentaire spécialisé-e dans la gestion des données de recherche au sein de l'équipe bibliothèque

La deuxième proposition consiste en l'engagement d'une nouvelle personne pour l'intégrer à l'équipe Bibliothèque en vue d'en faire la personne de référence pour la GDR. Une orientation de data steward serait avantageuse pour des raisons identiques exposées dans la proposition 1.

Au sein de l'équipe, cette proposition ne modifierait ni les cahiers des charges ni les pourcentages des personnes en place. Toutefois, l'engagement d'une personne

supplémentaire influe toujours sur la dynamique de l'équipe et cet élément reste à analyser à l'interne. Gardons à l'esprit que la personne engagée serait uniquement impliquée dans les tâches de GDR, ceci impliquant qu'elle ne représenterait pas un soutien pour l'équipe dans les tâches bibliothéconomiques courantes. De plus, le coût d'un engagement entraîne la modification du budget ainsi que la motivation d'une telle demande. Difficile à entreprendre et à mettre en place d'un point de vue administratif, cette demande se révèle pourtant être l'alternative la plus efficace à court terme. En effet, la personne engagée est efficace et opérationnelle dès son arrivée et n'est pas supposée être en besoin de formations. Elle se révèle être l'unique personne de référence pour la GDR au sein de l'institution et aucune révision des tâches et des fonctions n'est nécessaire au sein de l'équipe Bibliothèque.

Le tableau ci-dessous résume les éléments positifs ainsi que les éléments à prendre en considération pour l'implémentation d'une telle proposition sous trois angles : la gestion de l'équipe et de ses ressources, l'aspect financier de cette option ainsi que les impacts en découlant.

Tableau 10 : Analyse de la proposition 2

		
Equipe	Aucune modification du cahier des charges ou des pourcentages actuels	Modification de la structure actuelle de l'équipe Personne dédiée uniquement à la GDR, ne représentant pas un soutien aux tâches bibliothéconomiques courantes
Coûts	Personne spécialisée dans la GDR sans besoin de formation supplémentaire	Modification du budget pour l'engagement d'une personne supplémentaire au sein de l'équipe Bibliothèque Analyse budgétaire pour la définition du pourcentage et du cahier des charges
Impacts	Personne engagée directement opérationnelle Une seule personne de référence au sein de l'institution pour la GDR Un-e membre du groupe <i>Open Science</i> spécialisé-e dans la GDR Aucune révision des tâches et fonctions des membres de l'équipe Bibliothèque	/

6.1.3 Proposition 3

Création d'un service de soutien aux chercheu-r-se-s en collaboration avec d'autres HES vaudoises sur la base du projet RDSN

La dernière proposition de gestion relève d'un projet d'envergure impliquant plusieurs institutions vaudoises dans l'optique de créer un service collaboratif de soutien aux chercheu-r-se-s avec une personne de référence par institution.

Inspirée du projet RDSN en Valais, cette proposition peut être envisagée à long terme *a fortiori* si on l'intègre à plusieurs autres projets en cours. En effet, intégrer un projet de service de soutien aux chercheu-r-se-s dans le cadre du projet *Campus Santé* pourrait être une occasion à saisir.

L'idée de cette proposition est de réunir des hautes écoles vaudoises (HETS, HEIG-VD, ELS, ou autres) et définir un data manager par institution afin d'opérer un travail collaboratif et viser la mise en place d'un service centralisé de soutien aux chercheu-r-se-s.

Partant du principe que le data manager fasse partie de l'équipe Bibliothèque, aucune modification de la composition actuelle de l'équipe n'est à prévoir. Les bénéfices de la connaissance du fonctionnement de l'institution et de ses chercheu-r-se-s n'est pas à prouver. Toutefois, cette solution demeure un véritable investissement personnel et financier. Il se peut aussi qu'une révision des pourcentages ou des cahiers des charges soit à prévoir.

Sur la base du projet RDSN, un projet d'une telle envergure s'établit sur une année ou plus. Les coûts sont à prendre en considération sur plusieurs niveaux : choix des data managers pour chaque institution et modification de leur cahiers des charges en vue d'un tel projet, formation éventuelle (DMP, logiciels, etc.), définition des objectifs, communication, élaboration de documentation, établissement de formations destinées aux chercheu-r-se-s, identification des besoins pour chaque institution et construction d'un accompagnement en fonction des besoins. De ce fait, cette proposition se révèle certainement être la plus onéreuse et donc la moins applicable à l'heure actuelle.

Les impacts d'une telle proposition sont les suivants : une unique personne de référence au sein de chaque institution pour la GDR, un service GDR centralisé pour plusieurs institutions vaudoises et les avantages dont découlent le travail collaboratif. Toutefois, notons qu'un projet d'une telle envergure nécessite une préparation importante, dont les étapes et objectifs s'agencent sur une année ou plus.

Le tableau suivant résume les éléments positifs ainsi que les éléments à prendre en considération pour l'implémentation d'une telle proposition sous trois angles : la gestion de l'équipe et de ses ressources, l'aspect financier de cette option ainsi que les impacts en découlant.

Tableau 11 : Analyse de la proposition 3

		
Equipe	Aucune modification de la structure de l'équipe actuelle Diffusion des savoirs en interne Connaissances du fonctionnement interne de l'institution Connaissances des chercheu-r-se-s et de leurs habitudes	Investissement important pour le data manager choisi Projet considérable demandant potentiellement une augmentation du pourcentage ou une révision partielle du cahier des charges
Coûts	/	Coûts se répercutant sur plusieurs niveaux Projet onéreux
Impacts	Une seule personne de référence au sein de chaque institution pour la GDR Un membre du groupe <i>Open Science</i> spécialisé-e dans la GDR Un service GDR centralisé pour plusieurs institutions vaudoises Travail collaboratif	Préparation du projet en amont Durée du projet d'une année ou plus

Recommandation finale

Au terme de cette analyse, nous pouvons proposer à l'institution la seconde proposition qui s'avère être la plus proche des besoins de l'institution à l'heure de ce mandat. Elle se révèle être la plus pratique au niveau de la structure actuelle de l'équipe, la plus efficace à court terme car aucune formation n'est supposée être nécessaire.

6.2 Aides à l'archivage et au partage des données

6.2.1 L'archivage des données

L'archivage est une étape qui reflète de réels besoins de soutien et d'encadrement. Ainsi, les outils proposés (Bezzi 2020 ; Rebord 2015 ; CNIL 2020a ; EPFL 2022d ; EPFL 2022h) peuvent aider dans les processus d'archivage. Nous formulons dans ce travail quelques recommandations pour l'archivage des données en prêtant attention à la sélection des données à préserver, ainsi qu'à leurs formats.

Le choix de la conservation

Comme mentionné par l'Université de Genève, toutes les données ne peuvent être conservées et ce, notamment pour des raisons de coûts. Les critères de sélection de conservation sont à la charge des chercheu-r-se-s, car ce sont les acteurs principaux capables de définir la valeur des données récoltées (Gaillard 2014 ; EPFL 2022h). Pour faire ce choix, les chercheu-r-se-s doivent s'interroger sur les obligations de conservation, sur le cadre légal

existant, sur la possibilité ou non de réplication, sur la valeur de la donnée ainsi que sur la réutilisation (Université de Genève [sans date]b ; EPFL 2022h ; Bezzi 2020). Pour ce faire, ils peuvent s'appuyer sur les interrogations suivantes :

- Existe-t-il une obligation à préserver les données en fonction de l'institution ou du bailleur de fonds ?
- Les données sont-elles en adéquation avec la nLPD ? Si tel n'est pas le cas, veiller à l'anonymisation et à la restriction des accès
- La réplication des données est-elle possible ? Et cette réplication est-elle moins onéreuse que la préservation ?
- Quelle est la valeur des données ?
- La réutilisation est-elle possible ?

Les définitions de préservation ne font pas apparaître explicitement la possibilité de réutilisation, toutefois notons que la tendance s'oriente vers l'inutilité de la conservation si elle n'est pas compatible avec une réutilisation (Bezzi 2020 ; Université de Genève 2018)

Le format

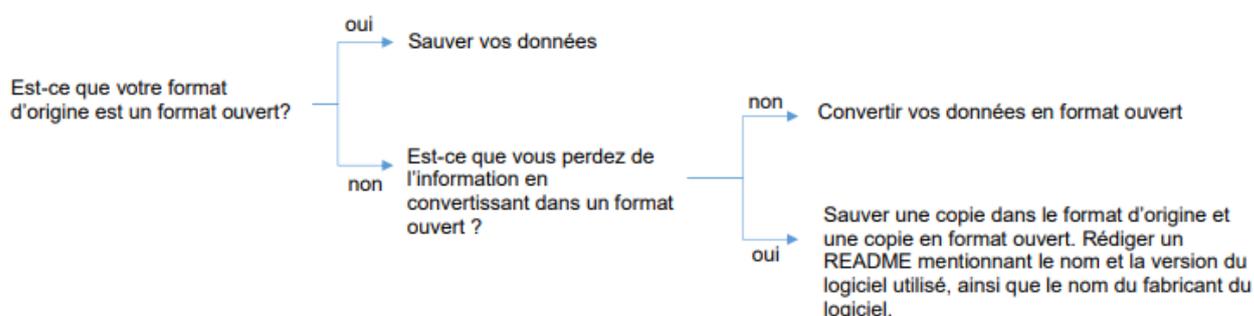
Le formatage adéquat des données a pour but la conservation à long terme des données ainsi que leur réutilisation.

La bonne pratique à cette étape du cycle de vie est de privilégier un format qui soit (Bezzi 2020) :

- Non-propriétaire
- Non-crypté
- Non- compressé
- Couramment utilisé dans le domaine de recherche
- Format ouvert dans le cas d'une conservation de plus de 10 ans

L'arbre décisionnel issu de la réflexion de Bezzi (2020) permet de visualiser le processus de formatage des données :

Figure 20 : Arbre décisionnel pour le formatage des données



(Source : Bezzi 2020, p. 43)

Pour aider les chercheu-r-se-s dans le formatage de leurs données, L'EPFL a fourni dans ses fiches techniques un tableau réunissant les formats adaptés ou non à la conservation des données. Il se révèle être un outil crucial dans le processus de formatage. Nous conseillons vivement aux chercheu-r-se-s de la HESAV de s'appuyer sur ce tableau, sachant que le formatage atteint un niveau de maturité basique ou faible chez la plupart des chercheu-r-se-s :

Tableau 12 : Formats adaptés à l'archivage

TYPE OF DATA	APPROPRIATE	ACCEPTABLE	DEPRECATED
Tabular (extensive metadata)	CSV – HDF5	TXT – HTML – TEX – FASTQ ^[3] – POR	
Tabular (minimal metadata)	CSV – TAB – ODS – SQL – TSV	XML (if appropriate DTD) – XLSX	XLS – XLSB
Textual / Presentation	TXT – PDF – ODT – ODM – TEX – MD – HTM – XML – EXTXYZ ^[4] – ODF	PPTX – RTF – DOCX – PDF (with embedded forms) – EPS – IPF	DOC – PPT – DVI – PS
Code / Computation	M – R – PY – IYPNB – RSTUDIO – RMD – NETCDF – AIDL	SDD	MAT – RDATA
Image & Spectroscopy	TIF – PNG – SVG – JPEG – FITS	JCAMP – JPG – JP2 – TIF – TIFF – PDF – GIF – BMP – DM3 – OIR – LSM ^[5]	INDD – AIT – PSD – SPC
Audio	FLAC – WAV – OGG – MXL – MIDI – MEI – HUMDRUM	MP3 – AIF	
Video	MP4 – MJ2 – AVI – MKV	OGM – MP4 – WEBM	WMV – MOV – QT
Geospatial	NETCDF – tabular GIS attribute data – SHP – SHX – DBF – PRJ – SBX – SBN – POSTGIS – TIF – TFW – GEOJSON	MDB – MIF	
3D structures & images	X3D – X3DV – X3DB – PDF3D – POV – PDBML	DWG – DXF – PDB	PXP
Generic	XML – JSON – RDF		

(Source : EPFL 2022d)

6.2.2 Le partage des données

Le partage des données est une étape qui s'est révélée source de nombreuses interrogations de la part des chercheu-r-se-s de la HESAV. Ainsi, nous exposons ici quelques recommandations pour le partage des données en détaillant notamment la documentation contextuelle lors du partage ainsi que la capture des métadonnées.

Le contexte

La contextualisation des données permet de rendre les données compréhensibles et réutilisables. Elle permet l'identification et la compréhension des données pour les chercheu-r-se-s de l'équipe de recherche, mais aussi pour les chercheu-r-se-s réutilisant les données.

Documenter la collecte des données, le système d'organisation des fichiers, leur nomenclature, les traitements opérés sur les données ainsi que le système de conservation est une pratique à mettre en place pour contextualiser les données.

La rédaction d'un fichier README ou d'un *data paper* permet la compréhension des données mais aussi la façon dont elles sont générées, ceci dans un but de réplification et de réutilisation. Ainsi, ci-dessous un modèle de README inspiré de plusieurs dispositifs existants (Bezzi 2020 ; UNIL, sans date ; Cornwell University, sans date) :

Tableau 13 : Rédaction d'un fichier README

Informations générales	<ul style="list-style-type: none"> • Titre du jeu • Description des données • Noms des chercheu-r-se-s dans l'équipe de recherche • Nommage des fichiers • Formats des fichiers • Arborescence • Contact • Institution de financement
Méthodologie	<ul style="list-style-type: none"> • Méthodes de collecte • Méthodes d'analyse • Date de la collecte • Explication des codes et des symboles si utilisés • Liste et description des logiciels utilisés • Description des standards utilisés
Résultats et données	<ul style="list-style-type: none"> • Définitions des intitulés • Date de création du fichier • Mention des unités de mesure • Définition des codes et symboles pour les données manquantes • Mention des formats • Date de modification du fichier et mention des explications
Partage et accès	<ul style="list-style-type: none"> • Mention des conditions d'accès durant le projet et après partage • Lien pérenne à la publication des résultats liés au jeu de données • Citation du jeu

Les métadonnées

La création de métadonnées permet l'identification et la réutilisation des données. Pour ce faire, il est important d'utiliser le standard de métadonnées spécifique à la discipline¹ et de déposer les données sur un dépôt figurant dans la liste des propositions.

Il existe plusieurs familles de métadonnées (EPFL 2022f) :

- Métadonnées d'utilisation (nombre de téléchargements)
- Métadonnées descriptives (titre, auteur, mots-clés)
- Métadonnées administratives (date de publication, licence utilisée)
- Métadonnées techniques (version de production)

¹ <https://www.dcc.ac.uk/guidance/standards/metadata>

- Métadonnées de préservation (documentation de conservation, date du dernier contrôle)

Il existe des métadonnées obligatoires selon les dépôts. En général, le titre, la description, la date de publication, les contributeurs ainsi que les niveaux d'accès sont des champs nécessaires.

Afin de construire ses propres métadonnées, il est possible de s'appuyer sur différents outils et normes :

Figure 21 : Construction de métadonnées



(Source : EPFL 2022f)

La construction de métadonnées s'est révélée méconnue de la part de la majorité des chercheurs interrogés. Ainsi pour la sélection des dépôts qui intervient ci-après, nous tentons de proposer des dépôts fournissant des services liés aux métadonnées, afin de répondre au mieux aux besoins des chercheurs de la HESAV.

6.3 Sélection de dépôts pour la HESAV

Sur la base des besoins des équipes de recherche et de leurs connaissances/pratiques en gestion des données de recherche, une liste de dépôts a été établie et analysée selon quatre critères de sélection. Pour définir ces derniers, plusieurs éléments ont été pris en compte.

Premièrement, la correspondance aux principes FAIR demeure le premier critère de sélection. Sur la base des recommandations des hautes écoles et universités suisses, nous avons effectué un premier tri des dispositifs existants. Pour nous aider dans le survol des solutions recommandées en Suisse, nous nous sommes appuyés sur un tableau comparatif proposé par Guirlet :

Tableau 14 : Recommandations de dépôts par neuf institutions académiques suisses

UNIGE		UNIL		HES-SO	
Existants Archive ouverte (p) Yareta OLOS	Recommandations disc. avec Re3data Yareta Zenodo Dryad FORSbase OSF	Existants SERVAL (p) FORSbase la PlaTec	Recommandations Re3data avec checklist FNS Zenodo FORSbase la PlaTec	Existants ARODES (p) OLOS	Recommandations top 10 par projets FNS Re3data, OpenAIRE, liste Nature FORSbase, DaSCH, GenBank OLOS, Zenodo, Dryad Figshare, autres FNS Yareta
En projet SWISSUBase					
EPFL		ETHZ		UNIZ	
Existant Infoscience (p)	Recommandations disc. ou com. avec Re3data Zenodo Dryad Figshare	Existants Research Collection (p+d) ETHZ Data Archive (préservation)	Recommandations selon directives Research Collection ETHZ Data Archive conseils de PLOS Re3data selon formats des données	Existants Zora (p)	Recommandations liste FNS disc. disc. et FAIR avec Re3data liste dépôts disc. utilisés par les chercheurs et chercheuses de l'UNIZ
En projet ACOUA				En projet SWISSUBase	
UNIBAS		UNIBE		UNIFR	
Existants edoc (p) DaSCH (Digital Humanities Lab)	Recommandations Zenodo Dryad disc. avec Re3data, OADirectory conseils de PLOS DaSCH FORSbase	Existant BORIS (p+d) BerDA (préservation)	Recommandations disc. avec Re3data, OADirectory conseils de PLOS Zenodo, Dryad, B2SHARE, Harvard Dataverse BORIS	Existant edoc (p)	Recommandations FORSbase, DaSCH, GenBank Re3data, OpenAIRE, liste Nature Figshare, Zenodo, autres FNS
		En projet BORIS Research Data			

(Source : Guirlet 2020a)

Deuxièmement, la grande sensibilité des données collectées au sein de la HESAV a orienté le choix du second critère de sélection. Dans l'optique d'optimiser la sécurité des données, nous souhaitons proposer des dépôts dont les serveurs se trouvent sur sol helvétique ou en seconde possibilité sur sol européen.

Puis, les pratiques des chercheur-se-s ont orienté un critère de sélection sur le niveau de curation proposé par la plateforme. En effet, le nettoyage des données n'étant pas une opération systématique à l'heure actuelle, le niveau de curation doit être pris en compte pour l'établissement de cette proposition. De plus, dans une perspective de conservation à long terme, le niveau de curation se doit d'être important. Plus le dépôt propose un service de pérennisation long, plus le niveau de curation doit correspondre aux besoins de curation y relatifs (Guirlet 2020b).

Enfin, les chercheur-se-s de la HESAV ont besoin d'un cadre pour le choix d'un dépôt pour le partage des données, mais aussi et surtout un dispositif d'archivage ou de pérennisation des dites données. De ce fait, le dernier critère de sélection s'oriente vers une proposition d'archivage ou de pérennisation correspondant au modèle OAIS.

En résumé, les quatre critères de sélection sont les suivants :

- Correspondance aux principes FAIR
- Localisation des serveurs
- Niveau de curation
- Correspondance au modèle de référence OAIS

Correspondances aux principes FAIR

Sur la base du tableau exposé plus haut (dispositifs existants avant 2020), nous avons listé les recommandations de plusieurs institutions universitaires et hautes écoles suisses et retenu

uniquement les deux les plus recommandés (7 institutions sur 9) correspondant aux principes FAIR. Nous avons ajouté à ces deux solutions, deux solutions suisses supplémentaires opérationnelles depuis 2020 et 2022 :

- Zenodo : dépôt généraliste et dont la maintenance est assurée par le CERN, cette solution est facile et instinctive, recommandée par bon nombre d'institutions suisses. (Zenodo, sans date ; Guirlet, 2020a)
- Dryad : offre un lieu d'accueil polyvalent pour une grande variété de données. Né d'une initiative d'un groupe de revues et de sociétés scientifiques visant à adopter une politique commune d'archivage des données (JDAP) pour leurs publications. De là a découlé une plateforme de données ouverte, facile à utiliser et sans but lucratif. (Dryad, 2022)
- SWISSUbase (données migrées depuis FORSbase) : solution nationale destinée aux universités suisses et aux autres organismes de recherche ayant besoin de déposer leurs données et s'appuie sur des normes et des processus d'archivage internationaux pour garantir la conservation et l'accessibilité des données à long terme. (Swissubase, 2022)
- OLOS : FAIR, adaptable et facilement paramétrable, OLOS offre aux chercheurs de différentes disciplines un dépôt pérenne déployé dans un cloud national. (Swissuniversities, 2022)

Localisation des serveurs

Sur les quatre dispositifs proposés plus haut, la localisation des serveurs a été analysée pour proposer des dépôts dont les serveurs sont localisés en Suisse.

Le tableau ci-dessous regroupe les localisations des serveurs de la liste préliminaire.

Tableau 15 : Localisation des serveurs des dépôts sélectionnés

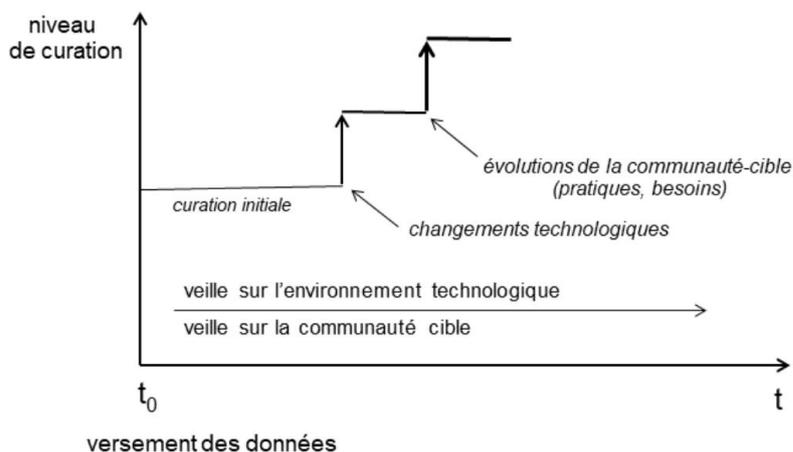
Dépôt	Localisation des serveurs	Adéquation avec le critère de sélection
Dryad	Etats-Unis et Europe	×
OLOS	Suisse	✓
SWISSUbase	Suisse (UNIL)	✓
Zenodo	Suisse et Hongrie	×

Niveau de curation

Puis, ces trois dépôts ont été analysés sous l'angle de la curation. Pour ce faire, le niveau de curation a été analysé afin de déterminer quel dépôt pourrait correspondre au mieux aux besoins de la HESAV.

Il est nécessaire, dans une volonté de pérennisation, de prêter attention à la curation des données afin d'optimiser leur accessibilité ainsi que leur réutilisation. Plus le temps de conservation des données est long, plus les services, les contrôles et les ressources associées doivent être nombreuses (L'Hours, Kleemola et de Leeuw, 2019). Le schéma suivant explicite la dépendance entre niveau de curation et temps de conservation des données.

Figure 22 : Niveau de curation nécessaire pour le maintien de l'accessibilité et la réutilisation des données



(Source : L'Hours, Kleemola et de Leeuw, 2019, p. 3)

De plus, pour continuer l'analyse, nous nous sommes appuyés sur la certification CoreTrustSeal (cité dans Guirlet 2020a, p. 29), qui nous indique les différentes caractéristiques correspondantes aux niveaux de curation :

Tableau 16 : Niveaux de curation – Core Trust Seal

Niveau de curation	Caractéristiques
Inexistant	Contenu accessible tel que déposé
Basique	Contrôles initiaux, ajout de métadonnées, ajout de documentation de collecte/gestion
Approfondi	Contrôles des formats, documentation de collecte/gestion/conservation/analyse et.
Appliqué aux données	Curation approfondie et édition des données déposées

Sur la base de cette catégorisation, nous pouvons analyser les niveaux de curation des dépôts.

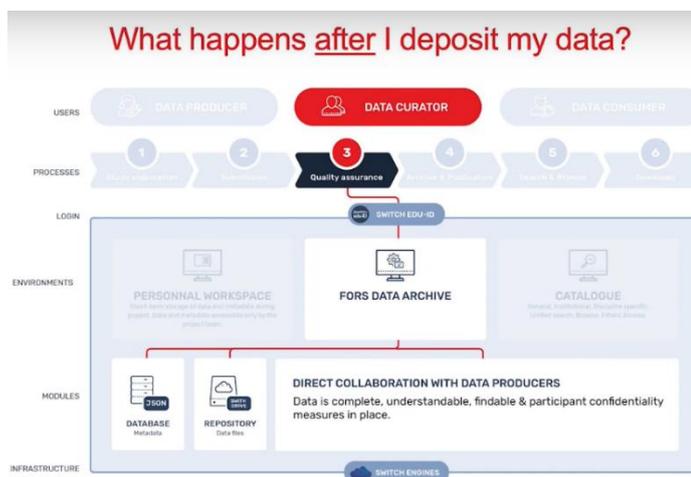
OLOS propose un service technique élevé. La maintenance et le niveau de sécurité des données sont des éléments essentiels et mis en avant par le dépôt. Possédant différents niveaux de sensibilité et de confidentialité, cette solution répond aux besoins spécifiques de chaque chercheur-ses (Swissuniversities 2022).

Du point de vue des services ajoutés sur les données, le format des données déposées est analysé automatiquement et propose au dépositaire un système de classification visuel grâce à des étoiles pour connaître l'adéquation des formats déposés avec le temps de conservation choisi. En d'autres termes, si lors du dépôt le chercheur observe 3 étoiles, le formatage des données déposées est conforme et conseillé pour le temps de conservation sélectionné. À l'inverse, si aucune étoile n'apparaît en marge du fichier, le format est déconseillé et non propice à la conservation. Toutefois, le formatage définitif est à la charge du dépositaire, les services du dépôt ne proposant aucun formatage automatique, et ce pour pouvoir accueillir le plus de données de recherche possible.

Quant à la documentation supplémentaire fournie par le service, OLOS intègre Formly, une bibliothèque dynamique (alimentée par JSON) utilisée pour concevoir des formulaires. Elle permet de personnaliser une documentation des métadonnées descriptives (DataCite) et facilite la compréhension et l'identification de l'ensemble du set de données (OLOS 2022).

La plateforme SWISSUbase permet de déposer les données et au fil du processus de compléter le fichier grâce à des métadonnées ainsi que des documentations contextuelles (SWISSUbase 2022b). À la suite du dépôt, des *data curators* prennent en charge l'analyse des métadonnées et de la documentation fournie par le dépositaire. Ils contrôlent que les données soient propres et que l'anonymisation soit suffisante. Une fois l'analyse terminée, ceux-ci publient le set au sein du catalogue (SWISSUbase 2022b). Lors de l'analyse, les *data curators* contrôlent la lisibilité des formats, mais les dépositaires ne savent pas, au moment du dépôt, si le format accordé à leurs données est compatible à la conservation à long terme ou non. Ainsi nous pouvons qualifier le niveau de curation de la plateforme comme basique. La figure suivante propose une schématisation des interventions des *data curators* après le dépôt des données.

Figure 23 : Curation de la plateforme SWISSUbase



(Démonstration SWISSUbase 2022)

Tableau 17 : Niveau de curation des dépôts sélectionnés

Dépôt	Niveau de curation	Adéquation avec le critère de sélection
OLOS	Approfondi	✓
SWISSUbase	Basique	✓

Correspondance au modèle OAIS

Dans une optique de conservation des données à long terme, le dernier critère de sélection repose sur la correspondance au modèle OAIS.

L'analyse de la correspondance au modèle OAIS permet d'obtenir une recommandation définitive du-des dépôt-s correspondant aux besoins de l'institution.

Sur le site de la plateforme, la conformité au modèle est explicite et mise en avant : OLOS est conforme à la norme ISO 14721 de l'ingest à l'accès. Les formats sont ouverts afin d'assurer l'interopérabilité et éviter les problèmes de lecture (OLOS, 2022).

Enfin, la plateforme Swissubase avance elle aussi sa conformité au modèle OAIS. (Swissuniversities 2022 ; SWISSUbase demo 2022)

Tableau 18 : Correspondance au modèle OAIS pour les dépôts sélectionnés

Dépôt	Préservation des données selon le modèle OAIS	Adéquation avec le critère de sélection
OLOS	Conforme au modèle OAIS (OLOS 2022)	✓
SWISSUbase	Conforme au modèle OAIS (Swissuniversities 2022)	✓

En définitive, nous proposons les deux solutions suivantes pour le partage et la pérennisation des données de la HESAV :

- OLOS
- SWISSUbase

Ces deux solutions suisses, compatibles aux principes FAIR, possèdent des serveurs sur sol suisse. De plus, le niveau de curation proposé est suffisant pour permettre aux chercheu-r-se-s de la HESAV de compléter leurs sets de données avec des documentations ainsi qu'un formulaire d'aide pour les métadonnées. Des indications pour le formatage sont disponibles sur la plateforme d'OLOS, bien qu'aucun formatage automatique ne soit proposé. Cet aspect permet aux chercheu-r-se-s d'adapter, si besoin, le format de leurs données si celui-ci est déconseillé par l'analyse système. Enfin, correspondant au modèle OAIS, ces deux dépôts proposent un modèle d'archivage des données numériques sur le long terme et sur des bases normatives internationales solides.

Afin de proposer une solution unique, notre choix s'oriente sur la solution suisse OLOS, qui en plus d'être en adéquation avec l'intégralité de nos critères de sélection, bénéficie aussi d'un système de contrôles des formats afin de donner un aperçu de leur conservation sur le long terme. Bien qu'entourés d'une équipe de *data curators*, la plateforme SWISSUbase effectue le contrôle après le dépôt, alors qu'OLOS permet au dépositaire de contrôler au moment du dépôt la compatibilité de ses formats avec la conservation à long terme.

6.4 Propositions d'aides à la GDR

6.4.1 Schéma du cycle de vie des données à la HESAV

Le schéma du cycle de vie des données défini au point 3.1.4 permet de comprendre l'importance de la connaissance d'un tel outil dans une institution. L'utilisation et la connaissance d'un tel outil est la clé du partage des données au sein de la communauté scientifique (DLCM, 2018).

Nous proposons à l'Annexe 4 un schéma adapté à la HESAV qui présente aux chercheu-r-se-s l'intégralité du cycle en six grandes phases, durant lesquelles des procédés-clés à appliquer sont mentionnés.

Au centre du schéma s'intègrent les trois phases de vie des documents (active, semi-active et définitive) afin de connaître quelle méthode et quel lieu de stockage et/ou d'archivage les chercheu-r-se-s doivent utiliser :

- Phase active : Stockage sécurisé sur disque réseau U
- Phase intermédiaire : Stockage sur le disque réseau S aux formats adaptés au stockage à long terme
- Phase définitive : Archivage et partage sur un dépôt correspondant aux principes FAIR

Ce schéma permet à l'ensemble des équipes de recherche d'obtenir rapidement des informations utiles pour chaque phase et de gérer les données au mieux tout au long de leur cycle de vie afin d'assurer leur validité et leur qualité. Afin d'optimiser la gestion du cycle de vie des données, nous proposons au point suivant une checklist afin de contrôler la gestion de ses données durant l'intégralité d'un projet.

6.4.2 Checklist pour gérer ses données tout au long d'un projet de recherche

La checklist proposée, disponible à l'Annexe 5, est une version augmentée du schéma de cycle de vie des données du point précédent. Elle est construite sur la base du diagnostic établi ainsi que sur les outils proposés par le projet DLCM, notamment la *Data Management Checklist* (2018). Cette checklist ne contient toutefois pas d'éléments détaillés sur la

réutilisation des données. Nous avons préféré expliciter les éléments relatifs à la préservation et au partage des données.

En suivant le cycle de vie des données, cet outil regroupe les éléments à ne pas oublier, les questions à se poser ainsi que les processus à appliquer de façon détaillée.

Ainsi dans l'étape de planification, les chercheu-r-se-s veilleront à :

- Lister tous les intervenants et définir les accès
- Définir un responsable de la gestion des données
- Définir le niveau de confidentialité des données
- Définir une politique de nommage
- Choix du dépôt pour le partage et/ou l'archivage

Lors de la collecte, les chercheu-r-se-s garderont à l'esprit de :

- Définir le type de données collectées
- Réutilisation de données et expliciter leur extraction
- Définir le volume des données
- Définir spécifiquement des contrôles de qualité
- Encadrer le contexte de collecte de façon stricte

Lors du traitement des données, il est important de :

- Documenter le traitement des données
- Lister l'intégralité des logiciels et outils utilisés
- Capturer les métadonnées
- Documenter les standards de métadonnées utilisés
- Utiliser des formats adaptés en vue de l'analyse des données
- Identifier quelle(s) version(s) conserver pour l'analyse

Pour l'analyse des données, les chercheu-r-se-s doivent rester attentifs-ves au fait de :

- Documenter l'analyse des données
- Lister l'intégralité des logiciels et outils utilisés
- Identifier quelle(s) version(s) conserver pour la préservation
- Prévoir la publication des résultats

L'étape de préservation demeure essentielle. Pour optimiser la préservation des données, il est conseillé aux chercheu-r-se-s de :

- Différencier les phases de vie des documents
- Formater les données de façon adéquate pour leur préservation
- Stocker les données de façon sécurisée tout au long des phases
- Choix des données pour la conservation
- Éliminer
- Archiver les données en fin de projet

Enfin, lorsque le partage est prévu, il est bon de veiller à :

- Vérifier les questions éthiques en vue du partage
- Vérifier les conditions de confidentialités et les consentements
- Vérifier la profondeur de l'anonymisation des données
- Vérifier les éléments de sécurité du dépôt sélectionné

Enfin, le dernier outil proposé permet de consolider la gestion des données au sein de l'institution en établissant les éléments principaux à faire figurer dans une future politique de gestion des données de recherche.

6.4.3 Vers une politique de gestion des données de recherche

En premier lieu, il est bon de déterminer les principales caractéristiques d'une politique de gestion des données de recherche.

Les politiques sont des documents de planification incluant des processus systématiques adoptés dans l'ensemble d'une institution par la direction ou un ensemble directeur. On peut définir la politique comme suit :

« A high-level overall plan, containing a set of principles that embrace the general goals of the organization and are used as a basis for decisions. A policy can include some specifics of processes allowed and not allowed. » (Smallwood 2014)

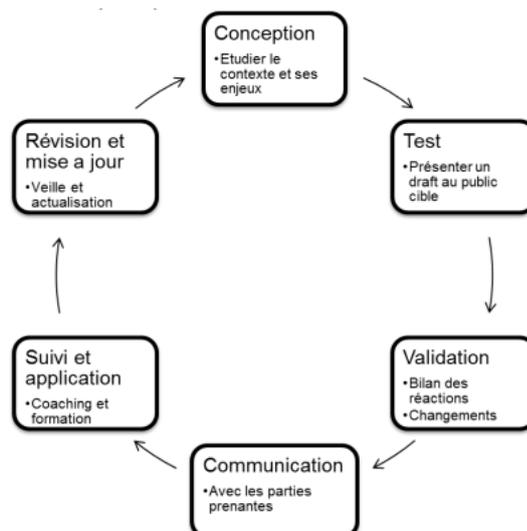
Quant aux politiques de gouvernance informationnelle, elles s'appliquent à la suite d'un engagement pris à un haut niveau de l'organisation et prend forme au travers d'un document officiel. Elles donnent cadre et légitimité à des exigences et des règles applicables uniformément au sein de toute l'institution (Perrein 2013). De plus, ces politiques cadrent le cycle de vie des informations mais encadrent aussi des éléments plus spécifiques tels que la confidentialité des informations, la protection des données personnelles, les questions autour de la sécurité ou encore celles relatives aux métadonnées (Jules et al., 2013).

Les objectifs de ces documents sont les suivants (Makhlouf Shabou 2021e) :

- Consigner les règles et les conventions organisationnelles
- Communiquer les règles et les conventions organisationnelles
- Engager les différents acteurs, aussi bien les collaborateurs que la direction
- Uniformiser les pratiques informationnelles afin d'optimiser les processus

Ils ne s'auto-construisent pas et doivent s'intégrer dans le contexte organisationnel et au niveau stratégique, s'organisent autour d'un processus de création propre dont les parties prenantes sont des acteurs techniques, des acteurs stratégiques ainsi que le-s responsable-s de la gouvernance.

Figure 24 : Processus de création d'une politique de gouvernance



(Source : Maklhouf Shabou 2021e)

Sur la base des outils proposés par le projet DLCM (Data Management Policy Template 2018), celle fournie par les modèles existants (Université de Genève [sans date]c) ainsi que sur les éléments susmentionnés, nous pouvons proposer une structure la structure suivante :

- Préambule
- Objectifs opérationnels
- Champs d'application et public cible
- Rôles et responsabilités
- Cadre légal
- Principes
- Processus et documents opérationnels de mise en œuvre
- Glossaire

Faisons à présent figurer les éléments principaux de la construction d'une politique future :

Préambule

Nous conseillons de faire figurer dans ce chapitre les principes généraux de la gestion des données de recherche et de mettre en exergue l'importance de maîtriser le cycle de ces dernières.

Deuxièmement, il est important de mentionner ce que couvre la politique, à savoir quel type de données, leur provenance, acquises ou créées dans le cadre de projets de recherches conduits par la Haute Ecole de Santé Vaud et financés par des fonds internes ou externes.

Objectifs opérationnels

Une politique de gouvernance des données de recherche au sein de l'institution devrait viser les objectifs suivants :

1. Mettre en place un cadre pratique précis (directives, méthodes) permettant de traiter efficacement et de manière efficiente les données de recherche tout au long de leur cycle de vie ;
2. Promouvoir une gestion des données de recherche mutualisée entre les chercheu-r-se-s et les services intervenants ;
3. Doter l'institution d'un cycle de vie des données, d'un système d'archivage systématique des données et d'un guide de gestion des données tout au long de leur cycle de vie ;
4. Faciliter l'accès et rendre plus visibles les informations relatives à la gestion des données ;
5. Promouvoir et renforcer l'autonomie des chercheu-r-se-s en matière de gestion des données de recherche et les doter d'une culture *Open Data*.

Champs d'application et public cible

Ce document s'adresse au personnel de l'Unité de recherche en santé de la Haute Ecole de Santé Vaud, au personnel de l'Unité juridique en santé, au Service informatique ainsi qu'à la Direction de l'institution.

Rôles et responsabilités

Afin d'assurer la bonne mise en œuvre d'une politique, les rôles et responsabilités de chaque acteur doivent être clairement définis.

- La direction de l'institution
- La direction de l'unité
- Le-la responsable de la gestion des données de recherche
- Les chercheu-r-se-s

Cadre légal

Nous exposons ici les principaux règlements et lois à prendre en compte pour l'élaboration d'une politique :

- Loi sur le droit d'auteur
- Loi sur la protection des données
- Loi fédérale relative à la recherche sur l'être humain
- Règlement général sur la protection des données

Principes

Il est important de faire figurer ici les principes à appliquer pour la bonne réussite de cette politique. L'ordre dans lequel ils sont présentés ne reflète pas leur importance informationnelle, mais leur suite logique.

- Gouverner les données tout au long de leur cycle de vie
- Reconnaître les données de recherche comme un actif institutionnel gage de qualité scientifique

- Construire collectivement la gestion des données de recherche
- Garantir un système de gestion des données de recherche
- Promouvoir une culture de l'*Open Science* passant par l'*Open Data*

Processus et documents opérationnels de mise en œuvre

Pour assurer la bonne mise en œuvre d'une gouvernance informationnelle, il convient de tenir compte des trois aspects suivants :

- L'administration des documents ;
- L'accessibilité des informations ;
- La sécurité et protection des données.

Glossaire

- Données de recherche
- Cycle de vie des données de recherche
- Dépôt de données
- Principes FAIR
- Etc.

Ainsi, les éléments-clés à faire figurer dans une future politique de gouvernance ont été présentés. Nous recommandons d'encadrer la gestion des données de recherche d'un tel document au sein de l'institution, afin de consolider les processus, de créer une dynamique autour des *Open Data* et d'encourager les bonnes pratiques de gestion.

Finalement, toutes les propositions figurant au point 6 visent à aider l'institution dans la mise en place de pratiques systématisées en matière de GDR. Ces propositions amènent aussi des pistes de réflexions pour la gestion future, ainsi que des outils utilisables directement, car construits pour répondre aux besoins actuels des chercheu-r-se-s.

7. Conclusion

Rappel des résultats

Au terme de ce travail, nous avons abouti aux résultats suivants :

Nous avons identifié les types de données et les types de formats majoritaires collectés et traités au sein de l'institution ;

Nous avons collecté, identifié et analysé les comportements et les pratiques des chercheur-se-s de l'unité de recherche pour en effectuer un diagnostic sur la base d'un modèle de maturité ;

Nous avons établi un diagnostic du cycle de vie des données de recherche produites par les chercheur-se-s et identifié quelles étapes du cycle méritait une attention ;

Nous avons défini les outils et recommandations pouvant répondre au besoin d'optimisation de la gestion des étapes problématiques.

Sur la base des éléments susmentionnés, nous avons pu fournir plusieurs types de propositions :

- Propositions de gestion des ressources humaines
- Propositions d'aides au partage et à l'archivage
- Analyse des solutions implémentables et proposition de dépôts compatibles aux besoins des chercheur-se-s
- Propositions d'outils : schéma du cycle de vie des données de la HESAV, une checklist pour optimiser la gestion des données tout au long de leur cycle de vie ainsi qu'une liste des éléments-clés à faire figurer dans une politique de gestion

Arrivés au terme des résultats de ce travail, relevons toutefois quelques éléments qui ont limité la profondeur des éléments apportés dans ce mandat.

Limites

Le cadre légal, notamment la responsabilité et l'appartenance des données aux différentes étapes des projets de recherche aurait apporté des éclairages importants à faire figurer dans une politique de gouvernance. Toutefois, par manque de temps et d'espace, ces éléments n'ont pas pu être intégrés dans notre analyse.

Deuxièmement, l'échantillon de quinze chercheur-se-s espéré s'est vu diminué de moitié. Un échantillon plus important aurait permis de dévoiler d'autres pratiques de collecte, de gestion ou de sécurité des données. Nous réitérons toutefois nos remerciements aux sept participant-e-s ayant accordé leur temps pour répondre aux questions d'entretien.

De plus, la quatrième partie du guide d'entretien n'a pas apporté autant d'éléments concrets qu'espéré. En effet, avec une volonté d'accéder en compagnie des chercheur-se-s à l'organisation des fichiers, leur nommage, et même la volonté de consulter un DMP et d'en comprendre sa construction, les réponses ont été essentiellement orales. Un-e-chercheur-se sur 7 a montré l'arborescence de ses fichiers et le système de nommage. Les autres ont préféré expliciter oralement leur gestion pour le projet spécifiquement analysé.

Éléments conclusifs

La connaissance du cycle de vie des données de recherche ainsi que l'application des bonnes pratiques à chaque étape du cycle demeurent indispensables pour optimiser la gestion des données de recherche au sein de la HESAV. Les éléments d'aides à l'archivage et au partage des données fournissent le socle d'une pratique institutionnelle systématisée. Les propositions de dépôts et les outils fournis en vue du partage et de l'archivage des données rendent ces pratiques plus aisées et le transfert plus fluide sur les plateformes.

Gardons finalement à l'esprit que l'environnement des données de recherche est encore évolutif ; les solutions se construisent, les dispositifs existants s'affinent et s'adaptent aux besoins de la communauté scientifique. De ce fait, il reste important d'adapter les outils et les pratiques proposés dans ce travail au fil de l'évolution du mouvement *Open Data*.

Bibliographie

ALIPRANDI, Simone, 2012. Open licensing and databases. *Journal of Open Law, Technology & Society* [en ligne]. 9 mai 2012. Vol. 4, n°1, pp. 5-18. [Consulté le 15.07.2022]. Disponible à l'adresse : <https://www.jolts.world/index.php/jolts/article/view/62>

AMIOTTE-SUCHET, Laurent, 2022. Les données qualitatives, que peut-on en FAIR ? Les recherches qualitatives au défi de l'Open Data. *Traverse*. 2022. 1/2022, pp. 1-6.

BALONG, Yang, HONG Wu et HAODONG Zhang, 2018. *Research and application of data management based on Data Management Maturity Model (DMM)*. [en ligne]. ICMLC 2018: Proceedings of the 2018 10th International Conference on Machine Learning and Computing, pp. 157-160. [Consulté le 27.05.2022]. Disponible à l'adresse : <https://doi.org/10.1145/3195106.3195177>

BAYLE, Aurélie, 2020. RGPD et protection des données : panorama d'une réglementation et ses impacts [Document PPT]. Support de cours : « *Open Science* ». Haute école de gestion de Genève. Master en Science de l'information, année académique 2020-2021. 16 octobre 2020.

BEZZI, Manuela, 2020. *Préservation des données de recherche : proposer des services de soutiens aux chercheurs du site Uni Arve de l'Université de Genève*. [en ligne]. Genève : Haute école de gestion de Genève. Travail de Master. [Consulté le 10.06.2022]. Disponible à l'adresse : <https://sonar.ch/hesso/documents/315066>

BROOKES, Naomi and CLARK, Robin, 2009. Using maturity models to improve project management practice. In: Production and Operation Management Society (POMS). 20th Annual Conference, Orlando, Floride, 1er au 4 mai 2009 [en ligne]. [Consulté le 25 mai 2022]. Disponible à l'adresse : <https://pomsmeetings.org/ConfProceedings/011/FullPapers/011-0288.pdf>

COMMISSION EUROPENNE, sans date. *Open Innovation, Open Science, Open to the World*. [en ligne]. 22 juin 2015. [Consulté le 10.05.2022]. Disponible à l'adresse : https://ec.europa.eu/commission/presscorner/detail/fr/SPEECH_15_5243

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, [sans date]a *Guide pratique : les durées de conservations*. [Document PDF]. [en ligne]. [Consulté le 18.06.2022]. Disponible à l'adresse : https://www.cnil.fr/sites/default/files/atoms/files/guide_durees_de_conservation.pdf

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, [sans date]b. Les durées de conservation des données. [en ligne]. 28 juillet 2020 [Consulté le 20.06.2022]. Disponible à l'adresse : <https://www.cnil.fr/fr/les-durees-de-conservation-des-donnees>

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, [sans date]c. Glossaire. Donnée personnelle. [en ligne]. [Consulté le 03.04.2022]. Disponible à l'adresse : <https://www.cnil.fr/fr/definition/donnee-personnelle>

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, [sans date]d. Glossaire. Donnée sensible. [en ligne]. [Consulté le 03.04.2022]. Disponible à l'adresse : <https://www.cnil.fr/fr/definition/donnee-sensible>

CORETRUSTSEAL STANDARDS AND CERTIFICATION BOARD, 2020. CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020-2022, version 2.0 [en ligne]. 2020. S.l. : s.n. [Consulté le 15.07.2022]. Disponible à l'adresse : <https://www.coretrustseal.org/why-certification/requirements/>, <https://zenodo.org/record/3632533>

CORNWELL UNIVERSITY, [sans date]. Research Data Management Service Group. Guide to writing « readme » style metadata. [en ligne]. [Consulté le 15.07.2022]. Disponible à l'adresse : <https://data.research.cornell.edu/content/readme>

CORTI, Louise, 2011. The European Landscape of Qualitative Social Research Archives: Methodological and Practical Issues. *Forum Qualitative Sozialforschung*. [en ligne] Septembre 2011. [Consulté le 15.07.2022]. Disponible à l'adresse : https://www.researchgate.net/publication/277851250_The_European_Landscape_of_Qualitative_Social_Research_Archives_Methodological_and_Practical_Issues

DATA LIFE-CYCLE MANAGEMENT, 2020a. Research Data Management Policy Template. [en ligne]. version 1.3, 09.2016. [Consulté le 10.04.2022]. Disponible à l'adresse : <https://www.dlcm.ch/resources/dlcm-policy>

DATA LIFE-CYCLE MANAGEMENT, 2020b. Data Management Plan. Data Management Checklist. [en ligne] [Consulté le 10.06.2022] Disponible à l'adresse : <https://www.dlcm.ch/resources/dlcm-dmp>

DATA LIFE-CYCLE MANAGEMENT, 2020c. The DLCM project. [en ligne] [Consulté le 15.09.2022] Disponible à l'adresse : <https://www.dlcm.ch/about-us/dlcm-project>

DATA SHARING TOOLKIT, 2021. How to create a data inventory. [Document PDF]. Module 7. [en ligne]. Mars 2021. [Consulté le 15.03.2022]. Disponible à l'adresse : <https://www.datasharingtoolkit.org/wp-content/uploads/2021/03/CABI-Mod7-3-01-Checklist.pdf>.

DILLO, Ingrid, DE LEEUW, Lisa, 2018. CoreTrustSeal. [en ligne] Mitteilungen der VÖB 71 (2018) Nr. 1: Repositorien in Österreich, pp. 162-170. [Consulté le 25.05.2022] Disponible à l'adresse: <https://doi.org/10.31263/voebm.v71i1.1981>

DOCUTEAM, [sans date]a. Docuteam packer. [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : <https://docs.docuteam.ch/packer/6.0/en/index>

DOCUTEAM, [sans date]b. Prestations. Archivage électronique à long terme. [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : <https://www.docuteam.ch/fr/archivage-electronique-a-long-terme/>

DRYAD, 2022a. FAIR Data. Best practices for creating reusable data publications. datadryad.org [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : https://datadryad.org/stash/best_practices

DRYAD, 2022b. Frequently Asked Questions. datadryad.org [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : <https://datadryad.org/stash/faq>

DRYAD, 2022c. Our mission. datadryad.org [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : https://datadryad.org/stash/our_mission

DRYAD, 2022d. Our Platform. datadryad.org [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : https://datadryad.org/stash/our_platform

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, 2022a. RDM guides & templates. EPFL Library. [en ligne]. [Consulté le 05.03.2022]. Disponible à l'adresse : <https://www.epfl.ch/campus/library/services-researchers/rdm-guides-templates/>

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, 2022b. Data Masking. [Document PDF] Fast guides. [en ligne]. [Consulté le 10.05.2022]. Disponible à l'adresse : https://infoscience.epfl.ch/record/265349/files/09_Data_Masking_EPFL_Library_RDM_FastGuide.pdf?ln=en

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, 2022c. FAIR Data Principles. [Document PDF] Fast guides. [en ligne]. [Consulté le 10.05.2022]. Disponible à l'adresse : https://infoscience.epfl.ch/record/265349/files/02_FAIR_EPFL_Library_RDM_FastGuide.pdf?ln=en

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, 2022d. File formats. [Document PDF]. Fast guides. [en ligne]. [Consulté le 10.05.2022]. Disponible à l'adresse : https://infoscience.epfl.ch/record/265349/files/04_File_Formats_EPFL_Library_RDM_FastGuide.pdf?ln=en

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, 2022e. Master en Data Science [en ligne]. [Consulté le 10.07.2022]. Disponible à l'adresse : <https://www.epfl.ch/schools/ic/fr/education-fr/master-fr/data-science/>

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, 2022f. Metadata. [Document PDF]. Fast guides. [en ligne]. [Consulté le 10.05.2022]. Disponible à l'adresse : https://infoscience.epfl.ch/record/265349/files/05_Metadata_EPFL_Library_RDM_FastGuide.pdf?ln=en

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, EPFL, 2022g. Research Data : the basics. [Document PDF]. Fast guides. [en ligne]. [Consulté le 10.05.2022]. Disponible à l'adresse : https://infoscience.epfl.ch/record/265349/files/01_The_Basics_EPFL_Library_RDM_FastGuide.pdf?ln=en

ECOLE POLYTECHNIQUE FEDERALE LAUSANNE, EPFL, 2022h. Storage, publication and preservation. [Document PDF]. Fast guides. [en ligne]. [Consulté le 10.05.2022]. Disponible à l'adresse : https://infoscience.epfl.ch/record/265349/files/10_Storage_Publication_Preservation_EPFL_Library_RDM_FastGuide.pdf?ln=en

ERARD, Frédéric, 2021. Les données codées dans le contexte de la recherche : personnelles ou anonymes ? *Pratique juridique actuelle*. 2021. 5/2021, pp. 606-616.

EXECUTIVE EDUCATION HEC LAUSANNE, 2022. Formations continues. Data science & management. [en ligne]. [Consulté le 10.07.2022]. Disponible à l'adresse : <https://execed.unil.ch/formation-continue/certificat-data/>

FACHINOTTI, Elena, GOZZELINO, Eva, LONATI, Sara, 2016. *Les bibliothèques scientifiques et les données de recherche : défis et enjeux*. [en ligne]. Genève : Haute école de gestion de

Genève. Mémoire de recherche. [Consulté le 10.05.2022]. Disponible à l'adresse : https://doc.rero.ch/record/258991/files/M8_MemoireReroDoc_FachinottiGozzelinoLonati.pdf

FONDS NATIONAL SUISSE, 2022. Open Research Data. Les principes des données FAIR : [en ligne]. [Consulté le 20.09.2022]. Disponible à l'adresse : <https://www.snf.ch/fr/dMILj9t4LNk8NwyR/dossier/open-research-data>

FORTIN, Marie-Fabienne et GAGNON, Johanne, 2016. Fondements et étapes du processus de recherche : méthodes quantitatives et qualitatives. 3ème. Montréal, Québec : Chenelière éducation. ISBN 978-2-7650-5006-3.

GAILLARD, Rémi, 2014. De l'Open Data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? [en ligne]. Thèse Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques. [Consulté le 20.07.2022]. Disponible à l'adresse : <http://eprints.rclis.org/22746/>

GENEVA GRADUATE INSTITUT, [mis à jour le 13 juin 2022]. Données de recherche. Droit d'auteur [en ligne]. [Consulté le 04.07.2022]. Disponible à l'adresse : <https://libguides.graduateinstitute.ch/droit-dauteur/donnees-recherche>

GOUVERNEMENT DU QUÉBEC, 2009. Politique-cadre sur la gouvernance et la gestion des ressources informationnelles des organismes publics. [Document PDF]. Québec. Secrétariat du Conseil du trésor. [en ligne] 2021. [Consulté le 25.05.2022]. Disponible à l'adresse : https://www.tresor.gouv.qc.ca/fileadmin/PDF/ressources_informationnelles/Politique_lois/politique_cadre.pdf.

GUIRLET, Marielle, 2020a. *Guide décisionnel et vade-mecum pour la mise à disposition d'un dépôt de données de recherche ouvertes en Suisse*. [en ligne]. Genève : Haute école de gestion de Genève. Travail de Master. [Consulté le 05.02.2022]. Disponible à l'adresse : https://doc.rero.ch/record/329696/files/Guirlet-M_moire-Vdef.pdf

GUIRLET, Marielle, 2020b. 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*. [en ligne]. 18 décembre 2020. Numéro 21, décembre 2020. [Consulté le 15.04.2022]. Disponible à l'adresse : <http://www.ressi.ch/num21/article182>

HAUTE ECOLE DE SANTE VAUD, 2022a. HESAV. Direction et organisation. [en ligne]. [Consulté le 05.03.2022]. Disponible à l'adresse : <https://hesav.ch/organisation/>

HAUTE ECOLE DE SANTE VAUD, 2022b. HESAV. Unité Juridique Santé. [en ligne]. [Consulté le 05.03.2022]. Disponible à l'adresse : <https://hesav.ch/recherche/unite-juridique-sante/>

HAUTE ECOLE DE SANTE VAUD, 2022c. RA & D. Nos axes de recherche. [en ligne]. [Consulté le 05.03.2022]. Disponible à l'adresse : <https://hesav.ch/recherche/axes-de-recherche/>

HAUTE ECOLE DE SANTE VAUD, 2022d. RA & D. Notre Unité de recherche. [en ligne]. [Consulté le 05.03.2022]. Disponible à l'adresse : <https://hesav.ch/recherche/>

HAUTE ECOLE DE SANTE VAUD, 2022e. Un campus pour de nouvelles synergies. [en ligne]. [Consulté le 10.07.2022]. Disponible à l'adresse : <https://hesav.ch/campus-sante/>

HAUTE ECOLE DE SUISSE OCCIDENTALE VALAIS-WALLIS, 2021. Services Ra&D. services RDSN. [en ligne]. [Consulté le 15.06.2022]. Disponible à l'adresse : <https://www.hevs.ch/fr/mediatheques-hes-so-valais-wallis/rad-et-publications/services-rad-202163>

HAUTE ECOLE DE SUISSE OCCIDENTALE, 2021a. Qu'est-ce que l'Open Data ? Recherche et Innovation. [en ligne]. [Consulté le 10.06.2022]. Disponible à l'adresse : <https://www.hes-so.ch/recherche-innovation/open-science/open-data/quest-ce-que-lopen-data>

HAUTE ECOLE DE SUISSE OCCIDENTALE, 2021b. Unité de recherche en santé. [en ligne]. [Consulté le 10.03.2022]. Disponible à l'adresse : <https://www.hes-so.ch/recherche-innovation/instituts-de-recherche/detail-institut/unite-de-recherche-en-sante-dhesav>

HODSON, Simon, JONES, Sarah et al., 2018. Turning FAIR data into reality [en ligne]. S.I. [Consulté le 23.06.2022]. Disponible à l'adresse : <http://hdl.handle.net/11299/188654>

HUYGHE, Marie, CAILLY, Laurent, OPPENCHAIM, Nicolas, 2018. *Partager des données d'enquêtes en recherche qualitative : approche juridique, éthique, déontologique et épistémologique*. [en ligne] 2018. [Consulté le 05-07.2022]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/halshs-01790983>

JACQUEMOT-PERBAL, Marie-Christine et CIOLEK-FIGIEL, Anne, 2016. Gestion et diffusion des données de la recherche [document drive Google]. Formation URFIST Lyon. 22 mars 2016. [Consulté le 10.06.2022]. Disponible à l'adresse : <https://drive.google.com/file/d/0BxKZLWq08xX-TW5VOEUtd2FSRE0/view>

JOHNSTON, Lisa R, CARLSON, Jake, HUDSON-VITALE, Cynthia, IMKER, Heidi, KOZLOWSKI, Wendy, OLENDORF, Robert et STEWART, Claire, 2016. Data Curation Terms and Activities [en ligne]. 23 octobre 2016. [Consulté le 22.06.2022]. Disponible à l'adresse : <https://hdl.handle.net/11299/188638>.

JOHNSTON, Lisa R., 2017. Data Curation Handbook Steps. In: JOHNSTON, Lisa R., 2017. Curating Research Data Volume Two: A Handbook of Current Practice. Ed: American Library Association, 2017. [en ligne]. [Consulté le 23.06.2022]. Disponible à l'adresse : https://conservancy.umn.edu/bitstream/handle/11299/183502/Data%20Curation%20Handbook%20Steps_v2.pdf?sequence=1

JULES, Arnaud, MOURAIN, Jean, BLANGER, Jean-Pierre, MASSE, Claudine, CHABIN, Marie-Anne, 2013. Processus, outils, méthodes : les modes opératoires de la gouvernance. Documentaliste-sciences de l'information [en ligne]. 2013. Vol. 50, n° 1, pp. 50-59. [Consulté le 02.07.2022]. Disponible à l'adresse : <https://doi.org/10.3917/docsi.501.0050>

L'HOURS, Hervé, KLEEMOLA, Mari et de LEEUW, Lisa, 2019. CoreTrustSeal: From academic collaboration to sustainable services. In : IASSIST Quarterly. 10.05.2019. Vol. 43, n 1, p. 1-17. DOI 10.29173/iq936

LANGSTON, Craig, GHANBARIPOUR, Amir, Naser, 2016. A Management Maturity Model (MMM) for project-based organisational performance assessment. Construction Economics and Building [en ligne]. Vol. 16, n° 4, pp. 68-85. [Consulté le 25 mai 2022]. Disponible à l'adresse : <http://dx.doi.org/10.5130/AJCEB.v16i4.5028>

MAKHLOUF SHABOU, Basma, 2020. Méthodologies de recherche. Fiche descriptive. Support de cours « Méthodologie ». [document PDF]. Haute Ecole de Gestion de Genève, filière Information documentaire, année académique 2020-2021.

MAKHLOUF SHABOU, Basma, 2021a. Data processing. Fiche descriptive. Support de cours « Gouvernance des données ». [document PDF]. Haute Ecole de Gestion de Genève, filière Information documentaire, année académique 2020-2022.

MAKHLOUF SHABOU, Basma, 2021b. Introduction à l'OAIS. Fiche descriptive. Support de cours « Gouvernance des données ». [document PDF]. Haute Ecole de Gestion de Genève, filière Information documentaire, année académique 2020-2022.

MAKHLOUF SHABOU, Basma, 2021c. Modèles de maturité. Fiche descriptive. Support de cours « Gouvernance des données » [document PDF]. Haute Ecole de Gestion de Genève, filière Information documentaire, année académique 2021-2022.

MAKHLOUF SHABOU, Basma, 2021d. Open Data. Fiche descriptive. Support de cours « Gouvernance des données ». [document PDF]. Haute Ecole de Gestion de Genève, filière Information documentaire, année académique 2020-2022.

MAKHLOUF SHABOU, Basma, 2021e. Politique de gouvernance. Fiche descriptive. Support de cours « Gouvernance des données ». [document PDF]. Haute Ecole de Gestion de Genève, filière Information documentaire, année académique 2020-2022.

MAKHLOUF SHABOU, Basma, NOBS, Gregory, NICOLET, Aurèle, 2019. Analyse de politiques de gouvernance informationnelle : rapport de recherche [en ligne]. 26 juin 2019. [Consulté le 15.07.2022]. Disponible à l'adresse : <https://arodes.hes-so.ch/record/3773?ln=fr>

MEIER, Philippe, 2011. Protection des données : fondements, principes généraux et droit privé. Berne : Stämpfli Editions SA. Précis de droit Stämpfli. ISBN 978-3-7272-2372-3.

MÉTILLE, Sylvain, 2014. Protection des données : tout savoir sur le consentement. Sylvain Métille [en ligne]. 20 novembre 2014. [Consulté le 19.07.2022]. Disponible à l'adresse : <https://smetille.ch/2014/11/20/protection-des-donnees-tout-savoir-sur-le-consentement/>

OLOS, 2022a. Documentation. [en ligne]. [Consulté le 10.06.2022]. Disponible à l'adresse : <https://olos.swiss/documentation>

OLOS, 2022b. Researchers. [en ligne]. [Consulté le 10.06.2022]. Disponible à l'adresse : <https://olos.swiss/researchers>

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE), 2007. Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics [en ligne]. Paris. [Consulté le 17.03.2022]. Disponible à l'adresse : <http://www.oecd.org/fr/sti/sci-tech/38500823.pdf>

PERREIN, Jean-Pascal, 2013. Définition de la gouvernance de l'information par des mots : Extrait du livre GouvInfo "Océan bleu". 3org – Points de vue sur le flux Information [en ligne]. 2 avril 2013. [Consulté le 02.07.2022]. Disponible à l'adresse : http://www.3org.com/news/gouvernance_de_linformation/definition-de-la-gouvernance-de-linformation-par-des-mots-extrait-du-livre-gouvinfo-ocean-bleu/

PFUND, Anouck, KAECH, Christelle, CHAUTEMS, Caroline, PERRENOUD, Patricia, 2022. *Préparer & déposer les données qualitatives sur un repository : un retour sur expérience* [en ligne] Journées Open Science HES-SO. Mai 2022. [Présentation PPT] [Consulté le 05.07.2022] Disponible à l'adresse : https://www.hes-so.ch/fileadmin/documents/HES-SO/Documents_HES-SO/pdf/open-science/5-b-Perrenoud-URS_Open_data_mai_2022_HES-SO.pptx

R3DATA.ORG, sans date. Registry of research data repositories. [en ligne]. [Consulté le 03.04.2022]. Disponible à l'adresse : <https://www.re3data.org/>

REBORD, Véronique, 2015. *Guide de gestion des documents et archives pour les personnes privées*. [en ligne]. Genève : Haute école de gestion de Genève. Travail de Bachelor. [Consulté le 25.06.2022]. Disponible à l'adresse : <https://sonar.ch/hesso/documents/314490>

ROSENTHAL, David, STUDER, Samira, LOMBARD, Alexandre, 2020. La nouvelle loi sur la protection des données. *Jusletter*. [en ligne] 16 novembre 2020, pp. 1-78. ISSN : 1424-7410 [Consulté le 15.05.2022]. Disponible à l'adresse : <https://www.rosenthal.ch/downloads/Rosenthal-Studer-Lombard-nouvelleLPD.pdf>.

SAGE X WIKI, [sans date]. Site d'aide en ligne wiki pour AGP-SageX. Extractions. [en ligne]. [Consulté le 10.09.2022]. Disponible à l'adresse : <https://agp-wiki.hes-so.ch/xwiki/bin/view/Aide%20structurelle/4.%20Extractions>

SANTOS, Anouk, 2020. *Données de la recherche : cadre juridique et licences*. [en ligne]. Genève : Haute école de gestion de Genève. Travail de Master. [Consulté le 17.04.2022]. Disponible à l'adresse : https://doc.rero.ch/record/329704/files/Santos-M_moire-Vdef.pdf

SAUTHIER, Géraldine, PIRINOLI, Christine, 2018. Stratégie Open HES-SO. [en ligne] Décembre 2018. [Consulté le 03.04.2022] Disponible à l'adresse : https://www.hes-so.ch/fileadmin/documents/HES-SO/Documents_HES-SO/pdf/open-science/2018-Strategie-Open-HES-SO-12321.pdf.

SMALLWOOD, Robert F., 2014. Information governance: concepts, strategies and best practices. Hoboken : Wiley. Wiley CIO series. ISBN 9781118218303.

STANFORD UNIVERSITY, 1997. Retention of and Access to Research Data. Research Policy Handbook [en ligne]. 30 octobre 1997. [Consulté le 15.03.2022]. Disponible à l'adresse : <https://doresearch.stanford.edu/policies/research-policy-handbook/conduct-research/retention-and-access-research-data>

SWISS NATIONAL SCIENCE FOUNDATION, 2017. Examples of data repositories [en ligne]. 27.04.2017. [Consulté le 07.03.2022]. Disponible à l'adresse : <https://www.snf.ch/fr/dMILj9t4LNk8NwyR/dossier/open-research-data>

SWISSUBASE, 2022a. Le guide « Metadata Guide ». [Document PDF]. Chercheur. Guide d'utilisation. [en ligne]. Mai 2022. [Consulté le 10.07.2022]. Disponible à l'adresse : https://resources.swissubase.ch/wp-content/uploads/2022/05/SWISSUbase-Metadata-Guide_May-2022.pdf

SWISSUBASE, 2022b. Le guide d'utilisation de swissubase. [Document PDF]. Chercheur. Guide d'utilisation. [en ligne]. Mai 2022. [Consulté le 10.07.2022]. Disponible à l'adresse :

https://resources.swissubase.ch/wp-content/uploads/2022/05/SUB_User-Guide-update-May-2022.pdf

SWISSUNIVERSITIES, 2022a. Stratégie nationale et plan d'action. [en ligne]. [Consulté le 25.05.2022]. Disponible à l'adresse : <https://www.swissuniversities.ch/fr/themes/digitalisation/open-access-campagne/strategie-nationale-et-plan-daction>

SWISSUNIVERSITIES, 2022b. SWISSUbase. Résumé du projet. [en ligne]. [Consulté le 25.05.2022]. Disponible à l'adresse : <https://www.swissuniversities.ch/fr/themes/digitalisation/p-5-information-scientifique/projets/swissubase-1>

THE UNIVERSITY OF EDINBURGH, 2020. Our definitions. Information Services, Research Data Service. ed.ac.uk [en ligne]. 10.07.2020 [Consulté le 12.04.2022]. Disponible à l'adresse: <https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/definitions>

UK DATA SERVICE, 2012. Research data mangement. [en ligne]. 2012. [Consulté le 04.05.2022]. Disponible à l'adresse : <https://ukdataservice.ac.uk/learning-hub/research-data-management/>

UNIVERSITE DE GENEVE, [sans date]a. Données de recherche. Glossaire [en ligne]. [Consulté le 24.05.2022] Disponible à l'adresse : <https://www.unige.ch/researchdata/fr/propos/glossaire>

UNIVERSITE DE GENEVE, [sans date]b. Données de recherche. Pourquoi gérer ses données de recherche ? [en ligne]. [Consulté le 20.06.2022] Disponible à l'adresse : <https://www.unige.ch/researchdata/fr/accueil/>

UNIVERSITÉ DE GENÈVE, [sans date]c. Politique institutionnelle sur la gestion des données de recherche [en ligne]. 25 juin 2018. [Consulté le 10.06.2022]. Disponible à l'adresse : <https://www.unige.ch/researchdata/fr/footer/politique/>

UNIVERSITÉ DE LAUSANNE, [sans date]a. Archivage & partage. L'Open Science à l'UNIL [en ligne]. [Consulté le 26.06.2022]. Disponible à l'adresse : <https://www.unil.ch/.../archivage--partage.html>

UNIVERSITÉ DE LAUSANNE, [sans date]b. Cycle de vie et types de données. L'Open Science à l'UNIL [en ligne]. [Consulté le 20.06.2022]. Disponible à l'adresse : <https://www.unil.ch/openscience/fr/home/menuinst/open-research-data/les-donnees-de-recherche/cycle-de-vie-et-types-de-donnees.html>

UNIVERSITÉ DE LAUSANNE, [sans date]c. Données ouvertes & FAIR. L'Open Science à l'UNIL [en ligne]. [Consulté le 26.06.2022]. Disponible à l'adresse : <https://www.unil.ch/.../conformite--exigences/donnees-ouvertes--fair.html>

UNIVERSITÉ DE LAUSANNE, [sans date]d. Politique de Records Management et d'archivage pour une gouvernance informationnelle. [Document PDF]. UNIRIS – Service des ressources informationnelles et archives. Documents de références. [en ligne]. 19 juin 2014 [Consulté le 25.05.2022]. Disponible à l'adresse : <https://www.unil.ch/uniris/fr/home/menuguid/documents-a-telecharger/documents-de-reference.html>.

VAN GULICK, Ana 2016. Three stakeholders in one: Steering the evolution of research data management services from many perspectives. CODATA et ICSU WORLD DATA SYSTEM. SciDataCon 2016, Denver, 11-13 septembre 2016 [Consulté le 20.05.2022]. Disponible à l'adresse :

https://kithub.cmu.edu/articles/presentation/Three_stakeholders_in_one_Steering_the_evolution_of_research_data_management_services_from_many_perspectives/5558050

VERNAZ, Nicolas, 2021. Protection des données – RGPD, LPD et cybersécurité. [Document PDF]. Support de cours : « Gouvernance des données ». Haute école de gestion de Genève. Master en Science de l'information, année académique 2020-2021. 11 novembre 2021.

ZENODO, [sans date]a. Frequently Asked Questions. zenodo.org [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : <https://help.zenodo.org/>

ZENODO, [sans date]b. General policies. zenodo.org [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : <https://about.zenodo.org/policies/>

ZENODO, [sans date]c. Infrastructure. zenodo.org [en ligne]. [Consulté le 01.07.2022]. Disponible à l'adresse : <https://about.zenodo.org/infrastructure/>

Documentation interne à l'institution

AMIOTTE-SUCHET, Laurent. HESAV, 2019. *Rapport d'activités 2019. Open Science et recherches en sciences sociales : entre injonction institutionnelle et spécificités disciplinaires. Éléments de cadrage et de réflexion.* 11 mars 2020. Document interne à l'institution HESAV.

AMIOTTE-SUCHET, Laurent. HESAV, 2021. *Rapport d'activités Open Science – Rapports sur les activités liées à l'Open Science à la HESAV.* 26 janvier 2022. Document interne à l'institution HESAV.

AMIOTTE-SUCHET, Laurent. HESAV, 2022. *Open Science : état de la situation et stratégie HESAV.* 15 mars 2022. Document interne à l'institution. [Présentation PPT]

GROUPE DE TRAVAIL OPEN SCIENCE, HESAV, 2020 *Aide-mémoire juridique à l'attention des chercheuses et chercheurs de HESAV : les projets de recherche.* Février 2020. Document interne à l'institution HESAV.

GROUPE DE TRAVAIL OPEN SCIENCE, HESAV, 2020. *Rédaction du Data Management Plan (DMP) pour les projets de recherche FNS : partage de réflexions et mutualisation des informations à destination des chercheuses et chercheurs de HESAV.* Février 2020. Document interne à l'institution HESAV.

GROUPE DE TRAVAIL OPEN SCIENCE, HESAV, 2022. *La sécurité des données de recherche à la HESAV.* Janvier 2022. Document interne à l'institution HESAV.

Textes juridiques

LDA Loi fédérale du 9 octobre 1992 sur le droit d'auteur et les droits voisins (LDA ; RS 231.1). Confédération suisse [en ligne]. 9 octobre 1992. Mise à jour le 1er avril 2020. [Consulté le 02.07.2022]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/19920251/>

- LPD Loi fédérale du 19 juin 1992 sur la protection des données (LPD ; RS 235.1). Confédération suisse [en ligne]. 19 juin 1992. Mise à jour le 1er mars 2019. [Consulté le 01.07.2022]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/19920153/index.html>
- LRH Loi fédérale du 30 septembre 2011 relative à la recherche sur l'être humain (LRH ; RS 810.30). Confédération suisse [en ligne]. 30 septembre 2011. Mise à jour le 1er janvier 2020. [Consulté le 07.04.2022]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/20061313/index.html>
- RGPD Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE). EUR-Lex [en ligne]. 27 avril 2016. Mise à jour le 4 mai 2016. [Consulté le 02.07.2022]. Accessible à l'adresse : <http://data.europa.eu/eli/reg/2016/679/oj/fra>

Annexe 1 : Guide d'entretien vierge : Chercheu-r-se-s

Questions liminaires sur le-la chercheu-r-se

Nom et prénom : _____

Fonction au sein d'HESAV : _____

Cursus : _____

Thèmes de recherches : _____

Compréhension et identification des connaissances et intérêts pour la gestion des données de recherches

Portez-vous un intérêt à la gestion des données de recherches ?

À quoi pensez-vous lorsque l'on vous parle de « données de recherches » ?

Avez-vous suivi des formations sur cette thématique ?

La gestion des données de recherches demeure pour vous une pratique utile ?

Connaissez-vous des dispositifs ou outils encadrant la gestion des données de recherches ? (Tels que le DMP, les dépôts de données ou des outils de nettoyage de données, etc.) Si oui, leur utilisation vous est-elle familière ?

Quels seraient pour vous les points forts et les points négatifs d'une gestion systématique des données de recherches ?

La notion de principes FAIR vous est-elle familière ?

Connaissez-vous des dépôts ou *data repositories* correspondant aux principes FAIR ?

Compréhension et identification des pratiques de gestion du cycle de vie des données de recherches

Nous allons retracer l'ensemble du cycle de vie des données pour identifier les mécanismes ou habitudes de gestion.

Planification de la recherche

Avez-vous certaines habitudes pour gérer vos données lorsque vous élaborez un projet de recherche ?

Au commencement d'un projet de recherche, planifiez-vous la gestion des données ?

Avez-vous déjà rédigé un *Data management plan* ?

Prévoyez-vous un système de nommage ou un système d'organisation des fichiers ?

Prévoyez-vous, si la recherche le demande, le consentement pour le partage des données et la réutilisation de celles-ci ?

Comment planifiez-vous la collecte de vos données ?

Prévoyez-vous des contrôles de vos données aux différentes étapes de la recherche ?

Création et collecte

Dans le déroulement d'un projet de recherche, comment collectez-vous ou créez-vous vos données ?

Portez-vous de l'intérêt aux métadonnées ? Quels standards connaissez-vous ou utilisez-vous pour les métadonnées ?

Sécurisez-vous d'une quelconque façon vos données dans le cadre de vos recherches ?

Avez-vous déjà acquis des données de tiers existantes sur des dépôts ?

Traitement et analyse des données

Lors du traitement et de l'analyse des données produites/collectées, faites-vous un traitement spécifique tel que saisie dans un tableau Excel, transcription, traduction etc. ?

Quels outils utilisez-vous pour traiter vos données ? Comment en avez-vous eu connaissance ?

Nettoyez-vous vos données ? (Données non pertinentes supprimées, doublons, erreurs, valeurs manquantes supprimées ou utilisées, traitement des valeurs aberrantes etc.)

Anonymisez-vous les données ?

Faites-vous une description ou une documentation de tout ce que vous avez produit/collecté ?

Dans le cas d'une acquisition d'un set de données existant, citez-vous les sources des données acquises ?

Partage et publication

Avez-vous pour habitude de partager vos données ?

Dans le cadre du partage de données, quelles entités/personnes d'HESAV interviennent pour l'aide au partage des données ?

Dans le cadre des publications, quelles entités/personnes d'HESAV interviennent pour l'aide à la publication ?

Comment définissez-vous l'accès approprié à vos données lors du partage ?

Créez-vous ou avez-vous déjà créé une documentation à destination des utilisateurs pour la compréhension de set de données (fichier readme) ?

Créez-vous ou avez-vous déjà créé des métadonnées et un système de classement ?

Archivage et pérennisation

La notion d'archivage vous est-elle familière ?

Quelles habitudes avez-vous en termes d'archivage de données de recherches ?

Dans le cadre de l'archivage de vos données, quelles entités/personnes d'HESAV interviennent pour l'aide à l'archivage ?

Miguez-vous ou avez-vous déjà migré vos données vers un format conseillé ou que vous considérez comme adapté à l'archivage ? Si oui, dans quel format ?

Réutilisation

Avez-vous déjà utilisé des sets de données existants pour vos projets ?

Avez-vous déjà analysé des résultats d'autres recherches ?

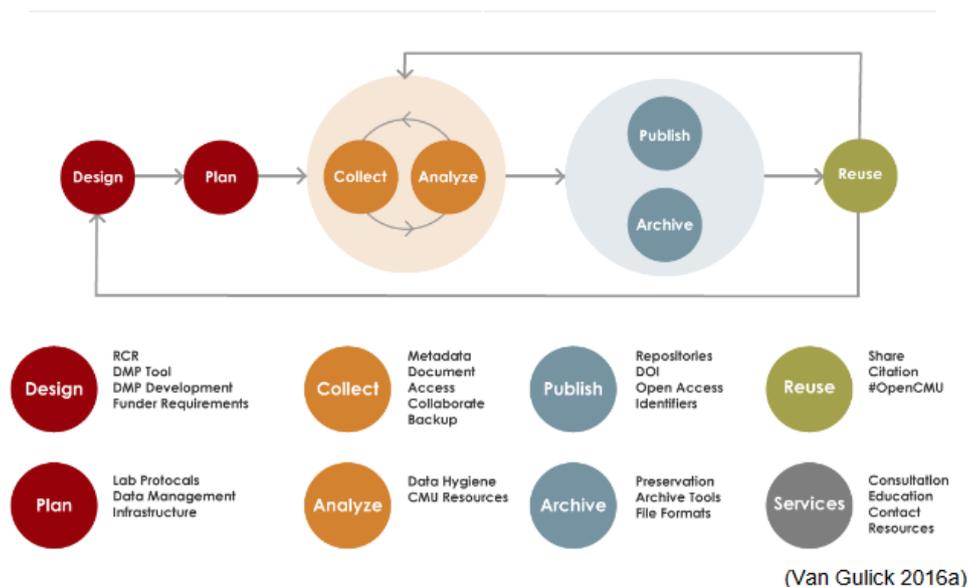
Analyse des pratiques sur une recherche spécifique

Titre de la recherche : _____

Date de la recherche : _____

Composition de l'équipe : _____

Sur une recherche spécifique, l'étudiante et le-la chercheur-euse analyseront à l'aide du schéma ci-dessous les pratiques de gestion de données de recherches.



Design

Avez-vous élaboré un Data Management Plan pour cette recherche ?

Si oui, comment ?

Collect

Quel type de données avez-vous collecté ?

Quel volume de données avez-vous collecté pour cette recherche ?

Comment avez-vous collecté vos données ?

Comment les avez-vous stockées et sécurisées ?

Avez-vous désigné un chercheur responsable de la gestion des données ?

Comment vous êtes-vous assuré-e de conserver les données intègres et intégrales tout au long du projet ?

Analyze

Pour l'analyse de vos données, les avez-vous nettoyées ? Si oui, comment ?

Ont-elles subi un traitement quel qu'il soit ?

Publish

Les avez-vous publiées sur un dépôt ? Si oui, lequel et pourquoi ?

Avez-vous attribué un doi (= *Digital object identifier*) ?

Les avez-vous publiées en Open Access ?

Archiv

Avez-vous archivé, en plus des résultats de votre recherche, les données collectées tout au long du projet ?

Qui est intervenu dans l'archivage ?

Avez-vous été aidé-e, appuyé-e, pour l'archivage de vos données ?

Avez-vous formaté vos données pour l'archivage ?

Connaissez-vous les obligations légales relatives au temps d'archivage des données et les avez-vous respectées ? (Études observationnelles, essais cliniques, etc.) Avez-vous été entouré-e pour le respect de ces obligations ?

Conclusion et ouverture

Au terme de cet entretien, éprouvez-vous un besoin de formation dans une ou plusieurs étapes du cycle de vie des données de recherches ?

Quel sentiment possédez-vous face à la gestion des données de recherche après toutes ces interrogations ?

Avez-vous des attentes spécifiques quant aux résultats de ce mandat ?

Annexe 2 : Guide d'entretien vierge : Bibliothèque

Questions liminaires

Nom et prénom : _____

Fonction au sein d'HESAV : _____

Compréhension et identification des connaissances pour la gestion des données de recherche

À quoi pensez-vous lorsque l'on vous parle de « données de recherches » ?

Portez-vous un intérêt à la gestion des données de recherches ?

Avez-vous suivi des formations sur cette thématique ?

La gestion des données vous semble-t-elle importante ? Et, pourquoi ?

Connaissez-vous des dispositifs ou outils encadrant la gestion des données de recherches ?
Si oui, leur utilisation vous est-elle familière ?

Quels seraient pour vous les points forts et les points négatifs d'une gestion des données de recherche durant l'intégralité de leur cycle de vie ?

La notion de principes FAIR vous est-elle familière ?

Connaissez-vous des dépôts ou *data repositories* correspondant aux principes FAIR ?

Analyse des besoins en formation

Nous allons retracer l'ensemble du cycle de vie des données pour identifier les besoins en formation.

Planification de la recherche

Êtes-vous familière avec le *Data Management Plan*, document qui initie la gestion des données de recherche ?

Sauriez-vous conseiller ou orienter un chercheur pour cet outil ?

Création et collecte

Quelles sont vos connaissances en métadonnées appliqués au domaine de la recherche ?

Avez-vous des compétences en matière de sécurité des données de recherches ?

Savez-vous qu'il est possible d'acquérir des sets de données d'autres chercheurs sur des dépôts ?

Sauriez-vous conseiller un chercheur sur toutes ces questions ?

Traitement et analyse des données

Sauriez-vous conseiller un chercheur dans le traitement, l'anonymisation, le nettoyage de ses données ?

Partage

Dans le cadre du partage des données des chercheurs-euses, quelles entités/personnes d'HESAV interviennent pour l'aide au partage des données ?

Sauriez-vous conseiller un chercheur pour un accès adéquat aux données ?

Sauriez-vous conseiller un chercheur dans l'élaboration d'un fichier readme ?

Publication

Verriez-vous un avantage à partager un set de données lié à un résultat de recherche en même temps que la publication des résultats eux-mêmes ?

Archivage et pérennisation

La notion d'archivage de données vous est-elle familière ?

Dans le cadre de l'archivage des données des chercheurs-euses, quelles entités/personnes d'HESAV interviennent pour l'aide à l'archivage ?

Sauriez-vous conseiller un chercheur pour le formatage adéquat des données en vue de leur pérennisation ?

Proposition de modèles de gestion liés à l'équipe bibliothèque

- Formation d'une bibliothécaire en place en vue de lui attribuer un pourcentage et des tâches dans son cahier des charges liés au soutien aux chercheurs-euses pour la gestion des données de recherche tout au long de leur cycle de vie en collaboration avec le groupe de travail *Open Science*.
 - Pensez-vous cette option implémentable ? (Budget, équipe en place, intérêt pour la question, flux de travail actuel etc.)
 - Quels seraient les avantages et les inconvénients de cette option en l'état actuel ?
- Engagement d'un spécialiste en information documentaire spécialisé dans la gestion des données de recherche ou *Data Manager* au sein de l'équipe bibliothèque dont le pourcentage et les tâches dans son cahier des charges sont exclusivement consacrés au soutien aux chercheurs-euses pour la gestion des données de recherche tout au long de leur cycle de vie.
 - Pensez-vous cette option implémentable ? (Budget, équipe en place, intérêt pour la question, flux de travail actuel etc.)
 - Quels seraient les avantages et les inconvénients de cette option en l'état actuel ?
- Sur les mêmes principes que le projet RDSN en Valais, élaboration d'un projet d'envergure et création d'un service de soutien aux chercheurs-euses en collaboration avec d'autres HES vaudoises, pour définir à un *Data Manager* répondant pour chaque institution ?

- Pensez-vous cette option implémentable ? (Budget, équipe en place, intérêt pour la question, collaboration avec la HETS, la HEIG-VD, ELS, etc.)
- Quels seraient les avantages et les inconvénients de cette option en l'état actuel ?

Conclusion et ouverture

Quel ressenti possédez-vous sur la gestion des données de recherche suite à cet entretien ?

Avez-vous des attentes spécifiques quant aux résultats de ce mandat ?

Annexe 3 : Modèle de maturité

Table 1. Data Management Maturity Model of Project S

Level	Collection capability	Management capability	Security capability	Service capability
The initial level	Data collection is mainly completed in collection team spontaneously, without a unified data management process about pictures, audio, video and other unstructured data.	Provide simple data storage management, MS Excel indexing.	Mobile hard disk to collect data, weak security capability	Passive service, to solve the problem case by case.
The managed level	Develop data collecting flow charts, independent project management office, data management team and data research team	Preliminary rules of data collecting, cleaning and using	Centralized unified data management, data security has been protected by developing a data backup and monitoring strategy,	Provide offline data borrowing. Clarity the searching scope and authority for manager, researcher and the public.
The steady level	A series of standardized management processes have been developed around data collecting, management and using	Provide different levels of data management tools at different levels. data quality is under control	Upgrade hardware facilities regularly, set up VPN tunnels. The data initiative safety has been promoted.	Meeting the needs for data research with online device.
The	Provide data relative offline software to	Establish a unified metadata model and	Develop Classification	Develop a general service interface for web portal.

159

quantitative management level	improve the collected data quality	standard of data acquisition, cleaning and using	protection system, provide quantitative measure of policy for data storage, data accessing and data services.	
The optimization level	Provide online feedback mechanism through the platform, and adjust inappropriate data acquisition rules in time.	Establish full life cycle data quality management.	Establish the Disaster Recovery Data Center to enhance the data security furtherly.	Provide full-text search capability for massive unstructured data, and provide online real-time service for date research.

(Source : Balong et al. 2018)

Annexe 4 : Schéma du cycle vie des données pour la HESAV



Annexe 5 : Checklist aux différentes étapes du cycle de vie

 Checklist aux différentes étapes du cycle de vie des données	
PLANIFIER <ul style="list-style-type: none">✓ Lister tous les intervenants et définir les accès✓ Définir un responsable de la gestion des données✓ Définir le niveau de confidentialité des données✓ Définir une politique de nommage✓ Choix du dépôt pour le partage et/ou l'archivage	COLLECTER <ul style="list-style-type: none">✓ Définir le type de données collectées✓ Réutilisation de données et expliciter leur extraction✓ Définir le volume des données✓ Définir spécifiquement des contrôles de qualité✓ Encadrer le contexte de collecte de façon stricte
TRAITER <ul style="list-style-type: none">✓ Documenter le traitement des données✓ Lister l'intégralité des logiciels et outils utilisés✓ Capturer les métadonnées✓ Documenter les standards de métadonnées utilisés✓ Utiliser des formats adaptés en vue de l'analyse des données✓ Identifier quelle(s) version(s) conserver pour l'analyse	ANALYSER <ul style="list-style-type: none">✓ Documenter l'analyse des données✓ Lister l'intégralité des logiciels et outils utilisés✓ Identifier quelle(s) version(s) conserver pour la préservation✓ Prévoir la publication des résultats
PRESERVER <ul style="list-style-type: none">✓ Différencier les phases de vie des documents✓ Formater les données de façon adéquate pour leur préservation✓ Stocker les données de façon sécurisée tout au long des phases✓ Choix des données pour la conservation✓ Élimination✓ Archiver les données en fin de projet	PARTAGER <ul style="list-style-type: none">✓ Vérifier les questions éthiques en vue du partage✓ Vérifier les conditions de confidentialités et les consentements✓ Vérifier la profondeur de l'anonymisation en fonction des données✓ Vérifier les éléments de sécurité du dépôt sélectionné

Hes·so