

**Mise en place d'un processus permettant
d'inventorier, de suivre et de valoriser les
données de la recherche de l'Institut santé de
la Haute école de santé, HES-SO Valais-Wallis**



**Travail de master réalisé par :
Pauline MELLY**

Sous la direction de :
Pierre-Yves BURGI, Directeur SI adjoint de l'Université de Genève

Sion, le 14 août 2022

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

Remerciements

Je tiens à remercier chaleureusement toutes les personnes qui m'ont apporté leur aide dans la réalisation de ce travail :

Pierre-Yves Burgi, mon directeur de mémoire pour sa disponibilité, ses conseils avisés et son soutien ;

Yohan Van De Looij de la Haute école de santé de Genève, pour avoir accepté d'intervenir en qualité d'expert ;

Alexandre Cotting, responsable du projet Research Data Service Network à la HES-SO Valais-Wallis et Claude-Alexandre Fournier, responsable de l'Institut santé, mes mandants, pour la confiance témoignée, leur disponibilité, leurs conseils ainsi que pour leurs encouragements dans les moments de doute ;

La direction de la HES-SO Valais-Wallis qui encourage le perfectionnement professionnel de ces employé-e-s ;

Un tout grand merci aux chercheuses et chercheurs de l'Institut santé, aux personnes ressources de la HES-SO Valais-Wallis ainsi qu'à toutes les personnes interviewées qui m'ont accordé leur temps pour partager leurs expériences et leurs bonnes pratiques. Sans vous ce travail n'aurait pas été possible ;

Marie-Catherine Beaulieu, responsable des médiathèques santé de la HES-SO Valais-Wallis pour son soutien et sa compréhension ainsi que mes collègues pour leurs encouragements et leur soutien durant ces deux ans ;

Tania Zuber, Céline Dumoulin-Melly et Anne Fournier pour les relectures et les conseils avisés ;

Mes collègues de cours et plus particulièrement Alexia, Myriam, Céline et Eléonore pour les échanges et les encouragements durant cette dernière étape.

Pour terminer je tiens à remercier mon compagnon, ma famille et mes ami-e-s pour leur soutien sans faille durant ces deux ans.

Résumé

Ce projet réalisé pour l'Institut santé (IS) de la Haute école de santé, HES-SO Valais-Wallis, s'inscrit dans le cadre de l'Open Science qui fixe les contours d'une recherche de meilleure qualité, plus transparente, accessible et reproductible. Bien que favorables à ces changements, les universités et hautes écoles sont également soumises à des exigences de protection et de sécurité des données qui complexifient leur partage.

Afin de veiller au respect des exigences légales, réglementaires et contractuelles, et de consacrer les moyens et ressources adéquats pour favoriser la pérennisation, l'accès et la réutilisation des données, un processus d'inventaire des données de recherche a été mis en place. Il permet d'assurer le suivi administratif et scientifique des données. Un canevas d'inventaire a été réalisé en se basant sur : (1) une revue de la littérature, (2) des entretiens d'institutions ayant accompli ce travail, (3) des entretiens avec les personnes ressources de la HES-SO Valais-Wallis qui apportent leur soutien aux équipes de recherche de l'IS en matière de gestion des données. Les chercheuses et chercheurs ont testé l'inventaire pour une sélection de projets de recherche puis ont été interviewé-e-s afin de mieux appréhender leur manière de gérer les données. Une photographie des données et des pratiques au sein de l'IS a été établie et la cohérence du fichier d'inventaire vérifiée. Un processus pour assurer le suivi des données a été proposé ainsi que des recommandations pour renforcer le soutien auprès des équipes.

Le fichier Excel a été choisi pour réaliser et maintenir l'inventaire car il est simple, fonctionnel et répond aux besoins des différentes parties prenantes. Il sera utilisé le temps qu'un nouvel outil de gestion des projets de recherche soit développé. Un cahier des charges basé sur l'analyse des besoins et des pratiques des personnes ressources et des chercheuses et chercheurs ainsi que sur les métadonnées du canevas, pourra orienter le choix du futur outil.

La troisième partie de ce travail traite des solutions de stockage et d'archivage. Une fois les concepts définis, les solutions existantes ont été répertoriées puis comparées à l'aide de grilles d'analyse. Des recommandations ainsi qu'un tableau décisionnel ont été proposés pour le stockage. Pour l'archivage des données, aucune solution ne répond actuellement à l'ensemble des besoins de la HES-SO Valais-Wallis. Si un dépôt ouvert comme OLOS peut être utilisé pour l'archivage des données non confidentielles et personnelles, une autre solution doit être étudiée pour les données plus sensibles. Les plateformes proposées par les Archives de l'Etat du Valais, Docuteam et SWITCH semblent intéressantes mais devront être analysées plus en détail. De plus, les développements prévus pour OLOS devront être suivis avec attention.

Une proposition de communication permettra de valoriser davantage les données de recherche produites au sein de l'IS du début d'un projet jusqu'à son achèvement.

Une politique ainsi qu'une directive sur les questions de responsabilité, de conservation et d'archivage devront être rédigées rapidement afin de définir un cadre clair à la GDR.

Mots clés : inventaire, données de recherche, gestion des données de recherche, archivage, communication, stockage, dépôt de données.

Table des matières

Remerciements	i
Résumé	ii
Liste des tableaux	vi
Liste des figures	vii
Liste des abréviations	viii
1. Introduction	1
1.1 Contexte	1
1.1.1 Haute école de la santé.....	1
1.2 Problématique	3
1.2.1 Question de recherche et objectifs	3
2. Gestion des données de recherche	6
2.1 Données de recherche	6
2.2 Jeu de données	7
2.3 Cycle de vie des données de recherche	7
2.4 Aspects juridiques et éthiques	7
2.4.1 Protection des données	7
2.4.2 Commission d'éthique	10
3. Inventaire des données de recherche	11
3.1 Élaboration du canevas d'inventaire	11
3.1.1 Méthodologie	11
3.2 Analyse des besoins des personnes ressources	14
3.2.1 Synthèse des besoins	14
3.2.2 Métadonnées de l'inventaire.....	15
3.3 Bonnes pratiques liées à l'inventaire des données	16
3.3.1 Planifier l'inventaire	16
3.3.2 Définir les métadonnées.....	17
3.3.3 Compléter l'inventaire.....	17
3.3.4 Mettre à jour l'inventaire	19
3.3.5 Publier l'inventaire	19
3.3.6 S'appuyer sur une politique de données	20
3.4 Mise en place du processus d'inventaire	20
3.4.1 Planifier l'inventaire	20
3.4.2 Définir les métadonnées.....	20
3.4.3 Compléter l'inventaire.....	21
3.5 Gestion des données de recherche au sein de l'IS	22
3.5.1 Photographie des données.....	22
3.5.2 Pratiques et besoins en GDR	24

3.5.3	Inventaire.....	28
3.5.4	Discussions	30
3.6	Recommandations.....	33
3.6.1	Processus d'inventaire	33
3.6.2	Propositions d'amélioration et de soutien.....	36
4.	Outil d'inventaire	37
4.1	Méthodologie	37
4.2	Analyse de solutions existantes	37
4.3	Recommandations.....	38
5.	Stockage et archivage.....	39
5.1	Méthodologie	39
5.2	Définitions des concepts	39
5.3	Stockage.....	40
5.3.1	Contraintes liées au stockage	40
5.3.2	Analyse de solutions existantes	41
5.4	Archivage	43
5.4.1	L'archivage de données de recherche	43
5.4.2	Solutions d'archivage existantes	45
5.4.3	Critères d'analyse.....	48
5.4.4	Analyse de solutions.....	49
5.5	Recommandations.....	52
5.5.1	Solutions pour le stockage	53
5.5.2	Solutions pour l'archivage	54
6.	Communication	55
6.1	Plan de communication	55
6.2	Propositions complémentaires	56
7.	Conclusion	57
7.1	Difficultés rencontrées.....	57
7.2	Perspectives.....	58
	Bibliographie	60
	Annexe 1 : Organigramme de la HES-SO Valais-Wallis	76
	Annexe 2 : Guide d'entretien des personnes ressources	77
	Annexe 3 : Analyse des besoins des personnes ressources	78
	Annexe 4 : Guide d'entretien institutions	86
	Annexe 5 : Description des inventaires	87
	Annexe 6 : Analyse des inventaires de données.....	90
	Annexe 7 : Choix des métadonnées de l'inventaire	102
	Annexe 8 : Guide d'entretien des chercheuses et chercheurs	108

Annexe 9 :	Cahier des charges	110
Annexe 10 :	Descriptif des outils d’inventaire.....	122
Annexe 11 :	Analyse des outils de stockage.....	128
Annexe 12 :	Evaluation et sélection des données à conserver	139
Annexe 13 :	Analyse de solutions d’archivage	142
Annexe 14 :	Glossaire	148

Liste des tableaux

Tableau 1 : Réponses obtenues des institutions contactées	12
Tableau 2 : Liste des sources utilisées	13
Tableau 3 : Liste des outils d'inventaire existants.....	37
Tableau 4 : Solutions de stockage	42
Tableau 5 : Comparatif des solutions de stockage	42
Tableau 6 : Avantages et inconvénients des dépôts.....	46
Tableau 7 : Solutions pour l'archivage à long terme en Suisse	47
Tableau 8 : Volumétrie des projets de l'IS	49
Tableau 9 : Comparatif entre OLOS et SWISSUbase	50
Tableau 10 : Tableau décisionnel pour le stockage des données durant le projet	53
Tableau 11 : Recommandations des institutions sur la durée de conservation	140

Liste des figures

Figure 1 : Présentation d'un projet dans l'inventaire du NCCR.....	19
Figure 2 : Étapes du cycle de vie et interventions des personnes ressources	32
Figure 3 : Schématisation de la mise à jour de l'inventaire.....	35
Figure 4 : Pérennisation des données de recherche	40
Figure 5 : L'archivage des données de recherche	43
Figure 6 : Modèle fonctionnel OAIS	45
Figure 7 : Solutions de stockage et d'archivage pour assurer la pérennisation.....	53

Liste des abréviations

CER-VD	Commission cantonale d'éthique de la recherche sur l'être humain du canton de Vaud
CHUV	Centre hospitalier universitaire vaudois
CINES	Centre informatique national de l'enseignement supérieur
CNRS	Centre national de la recherche scientifique
DAF	Data Asset Framework
DAUNIT	Data Acquisition Unit
DCC	Digital Curation Center
DLCM	Data Life-Cycle Management
DM	Data Manager
DMP	Data Management Plan
DPO	Data Protection Officer
EPFL	École polytechnique fédérale de Lausanne
ETHZ	Eidgenössische Technische Hochschule Zürich
FNS	Fonds national suisse de la recherche scientifique
FORS	Centre de compétence pour les sciences sociales
GDR	Gestion des données de la recherche
HATII	Humanities Advanced Technology and information Institute
HEDS	Haute école de santé
HESAV	Haute école de santé Vaud
HES-SO	Haute école spécialisée de Suisse occidentale
INIST	Institut de l'information scientifique et technique
IS	Institut santé
LESC	Laboratoire d'ethnologie et de sociologie comparative
LIPDA	Loi sur l'information du public, la protection des données et l'archivage
LPD	Loi fédérale sur la protection des données

LRH	Loi fédérale relative à la recherche sur l'être humain
NLPD	Nouvelle loi suisse sur la protection des données
OAIS	Open Archival Information System
OCDE	Organisation de coopération et de développement
OFS	Office fédéral de la statistique
ORD	Open Research Data
OS	Open Science
PF PDT	Préposé fédéral à la protection des données et à la transparence
PI	Principal Investigator
PNR	Programme national de recherche
RDM	Research Data Management
RDSN	Research Data Service Network
RGPD	Règlement général sur la protection des données
SEA	Service d'éthique appliquée
SID	Spécialiste en information documentaire
SINF	Service informatique
UNIBE	Universität Bern
UNIGE	Université de Genève
UNIL	Université de Lausanne

1. Introduction

1.1 Contexte

Le monde de la recherche a connu d'importants changements ces dernières années. Les technologies numériques, l'essor des recherches générant de gros volumes de données ainsi que la montée en puissance du mouvement d'Open Science (OS) ont contribué à transformer en profondeur la recherche afin qu'elle soit de qualité, accessible à tous, transparente et reproductible (Chawinga, Zinn 2021). Elle a été aidée en cela par les organismes de financement comme le Fonds national suisse de la recherche (FNS) ou Horizon 2020 (H2020), qui exigent que les données financées soient mises en libre accès pour réutilisation. Les éditeurs participent également à ce changement en demandant aux auteur-e-s de joindre les données liées aux manuscrits lors du processus de soumission, ou en exigeant, comme PLOS depuis 2014, que « les données sous-jacentes aux résultats présentés dans les articles soient entièrement disponibles au moment de la publication, à moins que certaines restrictions ne s'appliquent » (traduit de Byrne 2017) afin de renforcer la transparence et la reproductibilité (Chawinga, Zinn 2021; 2019; Morgan, Duffield, Walkley Hall 2017). Des initiatives nationales et internationales ont fleuri, incitant les chercheuses et chercheurs à gérer leurs données de recherche en s'appuyant sur un ensemble de principes appelés FAIR : faciles à trouver, accessibles, interopérables et réutilisables (Galonnier et al. 2019). En Suisse, une stratégie nationale en faveur de l'Open Access (OA) en 2016 puis une stratégie Open Research Data (ORD) en 2021 ont fixé le cadre pour une recherche ouverte (Swissuniversities, Fonds national suisse 2016; Délégation Open Science, Groupe de projet ORD, Swissuniversities 2021). En 2022, un pas supplémentaire a encore été fait avec l'adhésion du FNS à la cOAlition S, une « initiative mondiale d'organismes de financement de la recherche en faveur de l'Open Access » (Fonds national suisse 2022a).

De leur côté, les universités édictent aussi des stratégies, comme Open HES-SO, pour répondre aux impératifs nationaux (Haute École spécialisée de Suisse occidentale 2018). Bien qu'elles souhaitent répondre aux objectifs d'ouverture, de partage et de réutilisation des données, elles sont soumises aux exigences légales, réglementaires et contractuelles, notamment avec le Règlement général sur la protection des données (RGPD) et la Loi suisse sur la protection des données (LPD) qui renforcent les mesures liées à la protection des données et complexifient les modalités de partage (Da Sylva et al. 2019). De plus, elles doivent veiller à la sécurité des données dans un contexte où de plus en plus d'universités sont victimes de piratage informatique (Seydtaghia 2022; Steiner 2022).

1.1.1 Haute école de la santé

L'Institut santé (IS) pour qui est réalisé ce travail de Master est rattaché à la Haute école de santé (HEdS). Cette dernière est l'une des cinq hautes écoles de la Haute École spécialisée de Suisse occidentale Valais/Wallis (HES-SO Valais-Wallis) qui comprend également neuf instituts de recherche pour les domaines de l'art, de la gestion, de l'ingénierie, du travail social et de la santé (HES-SO Valais-Wallis 2021a). La HEdS regroupe les filières HES soins infirmiers et physiothérapie qui sont présentes sur trois sites : Sion pour les soins infirmiers francophones, Viège pour les soins infirmiers germanophones et Loèche-les-Bains pour la physiothérapie. La filière ES en soins infirmiers est pour sa part présente sur les sites de Viège et Monthey (HES-SO Valais-Wallis 2021b).

Un certain nombre de services ne sont pas rattachés à une école mais sont proposés de manière centrale pour l'ensemble de la HES-SO Valais-Wallis : le service financier, les ressources humaines, la communication et marketing, l'infrastructure et sécurité, le service académique et le service informatique. Le service d'éthique appliquée, le data protection officer et le service juridique sont rattachés à l'état-major (voir l'annexe 1 pour visualiser l'organigramme complet de la HES-SO Valais-Wallis). Le secrétariat de l'IS et la médiathèque santé sont pour leur part rattachés à la HEdS.

L'Institut santé sous sa forme actuelle a été créé en 2014¹. Vingt-six chercheuses et chercheurs de diverses disciplines – « physiothérapie, sciences infirmières mais aussi sociologie, psychologie et anthropologie » (HES-SO Valais-Wallis 2021a) – travaillent dans l'un des trois laboratoires : le PhysioLab pour la physiothérapie, le LaboSoins pour la filière soins infirmiers et l'InterLabs qui regroupe les activités transverses aux deux filières mais également liées à d'autres instituts ou d'autres disciplines (HES-SO Valais-Wallis 2021c). Depuis février 2022, The Sense, un nouveau centre académique de recherche dédié aux systèmes sensoriels, est rattaché à la HEdS (HES-SO Valais-Wallis 2021b). Il est issu d'une collaboration entre la HES-SO Valais-Wallis, l'Université de Lausanne (UNIL) et le Centre hospitalier universitaire vaudois (CHUV).

1.1.1.1 L'Open Science à la HEdS

Après l'entrée en vigueur de la Stratégie Open HES-SO en 2018 (HES-SO 2018), l'Institut santé a élaboré en collaboration avec la médiathèque santé une « Procédure d'archivage et de valorisation des activités de recherche » qui établit les modalités d'archivage des publications de la recherche pour la HEdS (HES-SO Valais-Wallis. Médiathèque santé 2019). La médiathèque a développé un service de soutien sur les questions d'Open Access via des pages dédiées à la recherche sur son site internet² ainsi que par du coaching et des formations (HES-SO Valais-Wallis. Médiathèque santé 2022). Mais jusqu'à récemment, les questions liées à l'Open data étaient renvoyées soit auprès de la Data Acquisition Unit (DAUnit)³ soit auprès du service Data Life Cycle Management (DLCM)⁴.

En 2020, une analyse des besoins réalisée auprès du personnel de recherche des cinq écoles de la HES-SO Valais-Wallis a fait ressortir des besoins en matière d'information, de formation, de soutien, de matériel et logiciels. Cette analyse a également montré la nécessité, pour toute l'institution, de mettre en place un service de soutien sur les questions d'Open data (Cotting, Calbimonte, Schumacher 2020). En septembre 2021, le Research Data Service Network (RDSN) composé d'un-e data manager de chaque école a été créé, avec pour mandat « d'accompagner les chercheur-euse-s au niveau des besoins métier dans le cadre de la gestion des données Ra&D » (HES-SO Valais-Wallis, RDSN 2021).

En 2021, un travail de Master réalisé par Mondaini (2021) pour l'Institut santé a proposé une méthode et des indicateurs pour analyser la qualité des jeux de données ainsi que des recommandations sur le management des données. Le tableau de bord utilisé par Mondaini sera analysé plus en détail au chapitre 3.3 ainsi que dans l'annexe 6.

¹ Entretien avec C.-A. Fournier, responsable de l'Institut santé de la HEdS Valais, mars 2022.

² <https://www.hevs.ch/fr/mediatheques-hes-so-valais-wallis/rad-et-publications/>

³ <https://www.hevs.ch/fr/pages-minisites/daunit-data-acquisition-unit-18634>

⁴ <https://www.dlcm.ch/>

1.2 Problématique

Actuellement, le responsable de l'Institut santé possède une vue d'ensemble des projets de recherche qui sont menés au sein de l'institution, mais n'a aucune vision des données qui y sont produites ni de la manière dont les données sont gérées. Si les chercheuses et chercheurs de l'IS sont les créatrices et créateurs des données produites dans le cadre de leurs recherches, c'est cependant l'institution qui en a la propriété selon les art. 321b et 332 du Code des obligations (CO ; RS 220) et l'art. 5 du Règlement sur la valorisation des connaissances acquises par la recherche au sein de la HES-SO (Haute école spécialisée de Suisse occidentale 2014a) ou encore selon l'art. 5 du Règlement concernant la valorisation et l'exploitation des résultats acquis par la recherche de la HES-Valais et de la HEVs2 (RS 414.733), ainsi que la responsabilité selon la LPD (art 5, let c LPD ; RS 235.1).

De ce fait, afin de veiller au respect des exigences légales, réglementaires et contractuelles, et de consacrer les moyens et ressources (financières, sécuritaires, de personnel) adéquats pour favoriser la pérennisation, l'accessibilité et la réutilisation des données, l'Institut santé souhaite mettre en place un inventaire ainsi qu'un processus pour assurer un suivi administratif et scientifique des données. Cet inventaire permettra d'une part d'avoir une vue d'ensemble des données de l'institution et de savoir comment ces données sont stockées, gérées et partagées afin de réduire les risques de pertes, de mauvaise gestion ou d'utilisation incorrecte. D'autre part, il mettra en lumière les besoins en matière de soutien et d'infrastructure. Une proposition sera également réalisée sur le choix d'un outil permettant la gestion de cet inventaire.

Différentes solutions de stockage, d'archivage et de partage existent, mais elles peuvent ne pas être connues ou être utilisées de manière inadéquate. De ce fait, le RI souhaite obtenir des recommandations sur ces outils afin de pouvoir mieux orienter les équipes de recherche. Pour terminer, des recommandations en matière de communication sur les données à valoriser sont également attendues.

1.2.1 Question de recherche et objectifs

La question de recherche initiale de ce travail était la suivante : *Quel processus mettre en place pour inventorier, assurer le suivi et la communication des jeux de données possédés par l'Institut santé de la Haute école de santé, HES-SO Valais-Wallis et quelle(s) solution(s) utiliser pour l'inventaire, le stockage, l'archivage et le partage de ces données ?*

En cours de réalisation, il a été nécessaire de revoir la portée des objectifs 3 et 4 car ils ne pouvaient être traités suffisamment en profondeur dans le temps imparti. Pour l'objectif 3, le volet lié au partage des données recherche n'a pas été traité. Seuls le stockage et l'archivage ont été développés. Pour l'objectif 4, une proposition de communication a bien été réalisée, mais elle était basée sur une recherche documentaire sommaire et non sur une revue de la littérature. Les éléments non traités sont présentés entre crochets dans la liste des objectifs ci-dessous.

La question de recherche a été reformulée : *Quel processus mettre en place pour inventorier, assurer le suivi et la communication des jeux de données possédés par l'Institut santé de la Haute école de santé, HES-SO Valais-Wallis et quelle(s) solution(s) utiliser pour l'inventaire, le stockage et l'archivage de ces données ?*

Ce travail est donc composé de quatre objectifs principaux détaillés en sous-objectifs. Pour chacun, la méthode utilisée est mentionnée. Elle sera détaillée plus précisément dans la suite de ce travail.

Objectif 1 : Proposer un processus permettant d'identifier les données liées aux projets de recherche menés par les chercheuses et chercheurs de l'IS afin d'en assurer le suivi administratif et scientifique.

1.1 Répertoire les bonnes pratiques liées à la conception, à la mise en place et à la gestion d'un inventaire de données

- Réaliser une revue de la littérature

1.2 Explorer les inventaires de données proposés par des institutions en Suisse romande et à l'étranger

- Réaliser une revue de la littérature ainsi que des entretiens pour collecter les pratiques existantes

1.3 Définir les champs à conserver pour l'inventaire

- Analyser les différents inventaires pour en extraire les champs à conserver

1.4 Réaliser une analyse des besoins auprès des personnes ressources de la HES-SO Valais-Wallis intervenant dans la gestion des données de recherche et ayant des besoins en termes de suivi des données

- Réaliser des entretiens avec les personnes ressources
- Établir une synthèse de leurs besoins ainsi qu'un processus de leur intervention

1.5 Tester l'utilité et la compréhension des champs proposés et obtenir une première photographie des types de données de l'IS

- Faire compléter l'inventaire à des chercheuses et chercheurs pour une sélection de projets de recherche
- Établir une synthèse des types de données au sein de l'IS

1.6 Récolter les pratiques et les besoins des équipes de recherche en matière de gestion des données ainsi que sur le processus d'inventaire

- Réaliser des entretiens avec une sélection de chercheuses et chercheurs
- Établir une synthèse des pratiques et besoins

1.7 Proposer un canevas d'inventaire ainsi qu'un processus pour la gestion et la mise à jour de l'inventaire

Objectif 2 : Réaliser des propositions pour un outil d'inventaire des données de recherche utilisable pour l'IS

2.1 Identifier des outils d'inventaire existants pouvant être utilisés pour les données de recherche

- Effectuer des recherches dans la littérature ainsi que sur internet et réaliser des entretiens

2.2 Réaliser une analyse des outils d'inventaire de données existants

- Définir les fonctionnalités principales et secondaires attendues

2.3 Proposer un outil pour assurer le maintien de l'inventaire des données de recherche

Objectif 3 : Réaliser des propositions de solutions de stockage, [de partage] et d'archivage des données de recherche qui soient adaptées à l'IS

3.1 Identifier les besoins, les pratiques et les solutions utilisées au sein de l'IS pour le stockage, [le partage] et l'archivage

- Réaliser des entretiens auprès d'une sélection de chercheuses ainsi qu'auprès d'experts

3.2 Identifier les différentes possibilités de stockage, [de partage] et d'archivage qui pourraient être utilisées par les équipes de recherche de l'IS

- Effectuer des recherches dans la littérature ainsi que sur internet

3.3. Réaliser une analyse comparative des solutions de stockage [de partage] et d'archivage des données de recherche

- Élaborer une liste de critères permettant de comparer ces solutions

3.4. Proposer des outils de stockage [de partage] et d'archivage des données de recherche qui soient adaptés à l'IS

Objectif 4 : Réaliser une proposition de communication sur les données à valoriser

[4.1. Identifier les différentes possibilités de communication des données de recherche]

- [Réaliser une revue de la littérature]

4.2. Effectuer des propositions en termes de communication

2. Gestion des données de recherche

Afin de fixer le cadre de notre travail, ce chapitre introduira les notions liées à la gestion des données de recherche qui est « l'ensemble des activités qui visent à organiser, stocker, préserver et donner accès aux données ainsi que l'infrastructure et les outils pour faciliter leur gestion tout au long de leur cycle de vie » (traduit de Chawinga, Zinn 2020, p. 97).

2.1 Données de recherche

La définition des données de la recherche varie en fonction des approches, des institutions et des contextes (Cartier, Moysan, Reymonet 2015). La plus communément citée est celle de l'Organisation de développement et de coopération (OCDE 2007) qui définit les données de recherche comme :

« des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. »
(p. 18)

Mais cette définition reste très générale et, comme le dit Gaillard (2014), elle ne prend pas en considération « la manière dont [les données de recherche] ont été obtenues, la forme sous laquelle elles ont été enregistrées [ni] le traitement qu'elles ont subi » (p. 15). De même, certain-e-s auteur-e-s englobent les métadonnées et la documentation dans la définition, alors que l'OCDE au contraire les en exclut (Australian Research Data Commons 2019; ODCE 2007; University of Melbourne 2013). Les données de recherche peuvent être catégorisées de différentes manières selon que l'on s'intéresse à leur forme, leur nature, leur structure, leur niveau d'élaboration, leur producteur, leur type ou leur provenance, ce qui entraîne autant de définitions que de manières de catégoriser les données (IFLA Big Data Special Interest Group 2018; Bracco et al. 2021; André 2014; Melly 2021).

La définition la plus intéressante trouvée dans la littérature est celle proposée par les parties prenantes de la recherche au Royaume-Uni (High Education Funding Council for England et al. 2016) car elle tend à être la plus complète possible en incluant aussi bien les catégories que des exemples afin de bien expliciter ce terme :

« Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others. Data may be defined as 'relational' or 'functional' components of research, thus signaling that their identification and value lies in whether and how researchers use them as evidence for claims.

They may include, for example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript.

The primary purpose of research data is to provide the information necessary to support or validate a research project's observations, findings or outputs. » (p. 3)

En observant la définition des archives qu'en donne Rousseau (cité dans Couture 1994), on peut également considérer les données de recherche comme des documents d'archives (Da Sylva et al. 2019) :

« [...] l'ensemble des informations, quels que soient leur date, leur nature, ou leur support, organiquement [et automatiquement] réunies par une personne physique ou morale, publique ou privée, pour les besoins mêmes de son existence et l'exercice de ses fonctions, conservées d'abord pour leur valeur primaire, c'est-à-dire administrative, légale, financière ou probatoire, conservées ensuite pour leur valeur secondaire, c'est-à-dire de témoignage ou, plus simplement, d'information générale » (p. 280).

De même, dans le domaine de la gouvernance des données, le terme actif informationnel étant défini comme un « savoir ou donnée représentant de la valeur pour l'organisation » (Baillargeon 2017), il prend également en considération les données de recherche, car elles sont source de valeur (Melly 2021).

Ainsi, des liens peuvent être établis entre « données de recherche, archivistique et plus largement gouvernance de l'information » (Da Sylva et al. 2019, p. 58-59) par les notions d'authenticité, d'intégrité, de confidentialité et de valeur qui sont communes à ces domaines (Da Sylva et al. 2019; Baillargeon 2017).

2.2 Jeu de données

On parle de jeu de données ou *dataset* pour désigner un ensemble cohérent de données « produites dans le cadre d'un même projet, sur un même objet d'étude et/ou recueillies sur un même lieu » (Institut de recherche pour le développement 2021). Il peut y avoir une cohérence soit au niveau technique soit au niveau intellectuel (Cartier et al., 2015, p. 6).

2.3 Cycle de vie des données de recherche

Durant la réalisation d'un projet de recherche, les données passent par différentes étapes « de gestion, de conservation et de diffusion des données de recherche » qui forment le cycle de vie des données (Institut de l'information scientifique et technique [INIST], Centre national de la recherche scientifique [CNRS] 2021) Différents modèles de cycle de vie des données ont été développés afin de décrire l'ensemble des étapes de gestion des données. Celui proposé par le UK Data Service, qui est l'un des cycles les plus couramment présentés, comprend six étapes : la planification, la collecte, l'analyse et le traitement, la publication et le partage, la préservation et pour finir la réutilisation (UK Data Service 2019).

2.4 Aspects juridiques et éthiques

Comme cela a été mentionné, la recherche est à la fois soumise à l'obligation de partager des données et aux exigences juridiques et éthiques qui assurent la protection de la vie privée des participant-e-s aux recherches (Diaz et al. 2022).

2.4.1 Protection des données

Le but de la protection des données est de protéger « la personnalité et les droits fondamentaux » des personnes lors du traitement de leurs données personnelles (art. 1 LPD ; RS 235.1). Sept grands principes encadrent la protection des données : la licéité, la bonne foi, la reconnaissabilité, la proportionnalité, la finalité, l'exactitude et la sécurité (Ehrenström 2021).

Pour la Suisse, trois niveaux de protection ont trait à la protection des données :

- Niveau cantonal : quasiment chaque canton possède une loi sur la protection des données. En Valais, il s'agit de la **Loi cantonale valaisanne sur l'information du public, la protection des données et l'archivage** (LIPDA ; RS 170.2). Elle est actuellement en révision.
- Niveau fédéral : c'est la **Loi fédérale sur la protection des données** (LPD ; RS 235.1) qui règle la protection des données. Une nouvelle version, la nLPD, entrera en vigueur en septembre 2023 (Secrétariat d'État à l'économie [SECO] 2022).
- Niveau européen : c'est le **Règlement général sur la protection des données** (RGPD) qui est appliqué.

À mentionner également la **Loi fédérale relative à la recherche sur l'être humain** (LRH; RS 810.30) ainsi que ses ordonnances d'application, qui ont pour objectif de « protéger la dignité, la personnalité et la santé de l'être humain dans le cadre de la recherche » (art. 1) en fixant les règles de la « recherche médicale sur l'être humain dans le domaine de la santé » (Préposé fédéral à la protection des données et à la transparence [PFPDT] [sans date-a]).

2.4.1.1 Application

Plusieurs critères cumulatifs sont à considérer afin de déterminer la ou les loi(s) qui vont s'appliquer (traduit de Diaz et al. 2022) :

- « **Le lieu d'établissement du responsable de traitement des données** » : c'est la législation (cantonale, fédérale ou européenne) où se trouve l'institution de la chercheuse ou du chercheur qui s'applique. Pour l'IS c'est la législation cantonale valaisanne qui s'applique.
- « **Le statut juridique du responsable du traitement** » : les organes publics fédéraux comme les écoles polytechniques fédérales, les chercheuses et chercheurs privés ainsi que l'organe public fédéral sont soumis à la LPD (FORS 2018). Les organes publics cantonaux comme les hautes écoles, les universités, les hôpitaux sont soumis aux lois cantonales (Lorenzetti 2022). En cas de recherche impliquant une collaboration entre institutions cantonales et fédérales, les deux lois s'appliqueront (FORS 2018).
- « **La localisation géographique de la collecte des données** » : le lieu où se trouvent les personnes au moment où leurs données sont collectées ainsi que « l'intention ou non de cibler certaines populations spécifiques » sont les éléments importants de ce critère (traduit de Diaz et al. 2022). Il faut également prendre en considération la notion de territorialité et d'extraterritorialité du RGPD : institution qui effectue le traitement, établie en UE, traitement effectué dans l'UE, traitement sur des résident-e-s de l'UE (Lorenzetti 2022).
- « **Le secteur d'activité** » : pour les scientifiques faisant de la recherche sur « les maladies humaines » et sur « la structure et le fonctionnement du corps humain » (art. 2), la LRH (RS 810.30) s'applique. On peut encore mentionner la Loi sur la statistique fédérale (LSF; RS 431.01) pour les recherches qui travaillent avec des données statistiques fournies par les institutions fédérales et/ou cantonales ainsi que ses ordonnances d'application. Au niveau valaisan, c'est la Loi cantonale valaisanne sur les statistiques (LStat ; Réf.-2022-005) qui fait foi. Elle était soumise au référendum facultatif jusqu'au 7 juillet 2022.

2.4.1.2 Données personnelles et sensibles

Au regard de la loi suisse, les **données personnelles** selon l'art. 3 let. a de la LPD (RS 235.1) concernent « toutes les informations qui se rapportent à une personne identifiée ou identifiable ». On y trouve des identifiants directs – comme le nom, le numéro de téléphone, le

numéro AVS, les empreintes digitales, etc. – ou indirects qui, une fois croisés, peuvent révéler l'identité d'une personne.

Les **données sensibles** sont des données personnelles qui portent sur « les opinions ou activités religieuses, philosophiques, politiques ou syndicales, la santé, la sphère intime ou l'appartenance à une race, des mesures d'aide sociale, des poursuites ou sanctions pénales et administratives » selon l'art. 3 let. c de la LPD. La nouvelle Loi fédérale sur la protection des données (nLPD) qui devrait entrer en vigueur durant le deuxième semestre de 2023 ajoute à cette liste les données génétiques et les données biométriques (art. 5 let. c nLPD).

Les données de santé peuvent être définies comme « toutes les données médicales et/ou relatives aux déterminants généraux de santé, et à la santé d'une personne, d'un groupe de personnes (couple, famille, quartier, ville, région, ethnie, pays, etc.) ou de populations » (Dillaerts 2016). On peut y trouver des données de santé personnelles comme non personnelles. En droit suisse, l'art. 3 let. f de la LRH (RS 810.30) parle de **données personnelles liées à la santé** et les définit comme « les informations concernant une personne déterminée ou déterminable qui ont un lien avec son état de santé ou sa maladie, données génétiques comprises ».

On peut encore définir le **traitement** qui comprend toute opération effectuée sur les données personnelles qui va de la collecte jusqu'à l'archivage ou la destruction des données, « quels que soient les moyens et procédés utilisés » (art. 3 let e LPD ; RS 235.1).

2.4.1.3 Données anonymisées ou codées

Afin d'utiliser et de partager des données collectées sur des personnes tout en limitant le risque de divulgation, il est nécessaire de dépersonnaliser les données en les anonymisant ou en les codant (Groupe de travail sur la COVID-19 du Réseau Portage et al. 2020, p. 3).

- Les données sont considérées comme **anonymisées** lorsqu'il n'est plus possible de les rattacher à une personne déterminée « sans engager des efforts démesurés » (art. 3 let. i LRH). Une attention particulière doit être portée aux « nom, adresse, date de naissance et numéros d'identification caractéristiques » afin qu'ils soient détruits ou rendus méconnaissables (art. 24 ORH). On parle de données **anonymes** lorsqu'elles sont collectées directement sans « identificateur direct » (Groupe d'experts sur les données sensibles 2020, p. 4).
- Les **données personnelles codées** sont des données identifiantes directes ou indirectes qui ne peuvent être rattachées à « une personne déterminée qu'au moyen d'une clé » (art. 3 let. h LRH) appelée aussi code dans l'ORH (art. 26). Cette clé ne doit être accessible qu'à un nombre limité de personnes et selon le degré de sensibilité elle pourra même être confiée à une personne externe au projet (Préposé fédéral à la protection des données et à la transparence [PFPDT] [sans date-b]).

Les données personnelles, sensibles ou personnelles liées à la santé qui ont été anonymisées ne sont plus soumises aux réglementations de la LIPDA, LPD, RDGP, LRH, etc. (Université de Genève, ResearchData 2018). En revanche, si elles sont codées, elles restent des données personnelles. Les données « devront [alors] être anonymisées le plus rapidement possible » et les résultats de recherche devront être publiés « sous forme anonyme » (PFPDT [sans date-b]).

2.4.1.4 Durée de conservation

Les données ne doivent en principe pas être conservées « au-delà de la réalisation du but fixé dans le projet de recherche ». Une « limite temporelle à la conservation » doit être fixée et, passé ce délai, les données devront être soit détruites soit anonymisées (PFPDT [sans date-b]). Cependant, certaines données ont l'obligation d'être conservées pendant 10 ans pour répondre aux exigences de la LRH. Il s'agit des données d'essais cliniques (art 45 de l'OClin; RS 810.305) ainsi que de dispositifs médicaux (art. 89 de l'ODim; RS 812.213). Selon l'art. 60 chapitre II du CO (RS 220), le délai de prescription pour les dommages corporels étant de 20 ans, certaines données devraient être conservées durant ce laps de temps. Cependant, la CER-VD suit généralement les exigences de la LRH mais peut, en fonction des risques liés à l'étude, exiger une durée de conservation allant dans le sens du CO⁵.

2.4.1.5 Consentement

Avant toute collecte de données personnelles, il est nécessaire d'obtenir le consentement des personnes concernées après les avoir informées de manière complète sur l'étude ainsi que sur la finalité du traitement de leurs données (art. 4 al. 5 LPD). Le consentement doit être libre, spécifique, éclairé et univoque dans le cas de données personnelles (Lorenzetti 2022) et explicite pour une collecte de données sensibles (art. 4 al. 5 LPD). Dans le cadre de la LRH, l'art. 7 (RS 810.30) stipule que pour toute recherche sur l'être humain, il est nécessaire d'obtenir le consentement éclairé de la personne ou que cette dernière n'ait pas fait valoir son droit d'opposition après avoir été suffisamment informée. Elle peut en tout temps refuser ou revenir sur sa décision sans avoir besoin de se justifier. Le consentement doit être donné par écrit (art. 16 LRH ; RS 810.30).

La personne doit être informée par oral et par écrit de manière compréhensible sur :

« la nature, le but, la durée et le déroulement du projet, sur les risques et contraintes prévisibles, les bénéfices escomptés du projet de recherche, notamment pour elle-même ou d'autres personnes, sur les mesures pour assurer la protection de ses données personnelles et ses droits » (art. 16 LRH ; RS 810.30).

2.4.2 Commission d'éthique

Pour les projets soumis à la LRH, une autorisation doit également être demandée à la Commission cantonale d'éthique de la recherche sur l'être humain du canton de Vaud (CER-VD) qui délivre des autorisations pour les projets de recherche sur l'être humain et les essais cliniques qui se déroulent dans les cantons de Vaud, Fribourg, Neuchâtel et du Valais. Elle suit les standards de l'organisation faitière SwissEthics (Association suisse des Commissions d'éthique de la recherche 2022), veille à la protection des sujets de recherche et évalue la conformité des projets de recherche sur l'être humain aux exigences éthiques, juridiques et scientifiques, ainsi qu'aux bonnes pratiques de la recherche » (CER-VD 2022).

⁵ Entretien J.G. Piguet, responsable éthique à la HES-SO Valais-Wallis, 18.07.2022.

3. Inventaire des données de recherche

Un inventaire de données est :

« une liste de jeux de données [ou d'ensembles de données], annotée avec des informations importantes (connues sous le nom de métadonnées) qui peuvent aider les utilisateurs à comprendre pourquoi les données ont été collectées, ce qu'elles contiennent, comment elles sont gérées et comment elles seront mises à la disposition des autres »
(traduit de Beale et al. 2021, p. 3).

Il permet de « garder une trace » des données possédées par l'institution et facilite l'archivage et le partage des données grâce aux informations déjà répertoriées (UK Data Service 2012).

3.1 Élaboration du canevas d'inventaire

3.1.1 Méthodologie

L'objectif de cette première partie consiste, pour rappel, à proposer un processus permettant d'identifier les données liées aux projets de recherche menés par les chercheuses et chercheurs de l'IS afin d'en assurer le suivi administratif et scientifique. Il s'agit de la partie principale de ce travail.

Afin de répertorier les bonnes pratiques liées à la conception, à la mise en place et à la gestion d'un inventaire de données, une revue de la littérature a été réalisée dans les bases de données LISTA, LISA, Sage, PubMed ainsi que Google Scholar. Des recherches complémentaires ont ensuite été menées sur internet.

Dans un deuxième temps, des entretiens ont été menés auprès des personnes ressources de la HES-SO Valais-Wallis afin de comprendre à quel(s) moment(s) et de quelle manière elles apportent leur aide. Il s'agissait également de connaître leurs difficultés et leurs besoins dans cette tâche de soutien, leur intérêt par rapport au projet de création d'un inventaire et leurs besoins en matière de suivi des données. Les interviews ont été réalisées soit à distance via l'outil de visioconférence Teams soit en présentiel à la HEdS afin d'établir une synthèse et un processus de leur intervention. Le guide d'entretien utilisé pour mener les interviews est disponible dans l'annexe 2.

En parallèle, plusieurs universités et hautes écoles de Suisse romande ont été contactées afin de savoir si un inventaire des données de recherche était réalisé au sein de leur institution. Des courriels ont été envoyés à tous les responsables Ra&D de la HES-SO, ainsi qu'aux services de data management des universités de Genève, Lausanne, Neuchâtel, Fribourg ainsi que de l'École polytechnique fédérale de Lausanne (EPFL). Le Centre de coopération internationale en recherche agronomique pour le développement (CIRAD) en France a également été contacté. Sur les 17 réponses obtenues, trois institutions ont réalisé un inventaire de leurs données : la Haute école de santé Vaud (HESAV), le National Centre of Competence in Research (NCCR) one the move et le Cirad.

Tableau 1 : Réponses obtenues des institutions contactées

Institution	Entretien	Inventaire
Bibliothèque de l'Université de Genève (UNIGE) – Service Research Data	Non	Non
Centre de coopération internationale en recherche agronomique pour le développement (CIRAD) Entretien avec S. Fortuno, chargée mission science ouverte et données de la recherche, 6 avril 2022.	Oui	Oui
Ecole hôtelière de Lausanne (EHL)	Non	Non
École polytechnique fédérale de Lausanne (EPFL), Bibliothèque Entretien avec A. Borel et F. Varrato, data managers, 11 avril 2022.	Oui	Non
Haute école Arc conservation-restauration	Non	Non
Haute école Arc santé	Non	Non
Haute école de la Source	Non	Non
Haute école de musique et conservatoire de Lausanne (HEMU)	Non	Non
Haute école de santé Vaud (HESAV) Entretien avec L. Amiotte-Suchet répondant Open Science, 14 avril 2022.	Oui	Oui
Haute école de travail social et de la santé Lausanne (HETSL) Entretien avec N. Pulzer, soutien à la recherche, 29 avril 2022.	Oui	Non
Haute école de travail social Fribourg (HETS-FR)	Non	Non
Haute école de viticulture et œnologie de Changins	Non	Non
Haute école d'ingénierie et de gestion du canton de Vaud (HEIG-VD) Entretien avec M. de Lapparent, responsable de l'institut interdisciplinaire du développement de l'entreprise, 14 avril 2022.	Oui	Non
Haute école du paysage, d'ingénierie et d'architecture de Genève (HEPIA)	Non	Non
La Manufacture – Haute école des arts de la scène	Non	Non
Université de Neuchâtel – Recherche et Innovation	Non	Non
National Centre of Competence in Research (NCCR) one the move, Université de Neuchâtel Entretien avec A. Perret, data manager, 26 avril 2022.	Oui	Oui

Des entretiens ont ensuite été menés à distance via l'outil de visioconférence Teams afin d'obtenir des bonnes pratiques aussi bien sur la mise en place du processus d'inventaire qu'en matière d'archivage et de partage de données. Un guide d'entretien a été rédigé et a servi à mener les interviews (annexe 4).

Après trois entretiens d'institutions ne réalisant pas d'inventaire (EPFL, HETSL, HEIG-VD), la portée des interviews a été revue afin de se concentrer uniquement sur les institutions réalisant un inventaire au risque de perdre l'objectif principal de vue.

L'expérience de huit institutions a permis de décrire les bonnes pratiques en matière d'inventaire. Six d'entre elles ont fourni un inventaire détaillé qu'il a été possible d'analyser plus en profondeur.

Tableau 2 : Liste des sources utilisées

Pays	Institution	Type de documentation	Inv.	B.P.
Royaume-Uni	Humanities Advanced Technology and information Institute (HATII), University of Glasgow, Digital Curation Center	Rapport : Data audit framework methodology (Jones, Ruusalepp, Ross 2009) Rapport : Data audit framework : implementation guide (Humanities Advanced Technology and information Institute, University of Glasgow, Digital Curation Center 2009) Article : The Data audit framework : a first step in the data management challenge (Jones, Ball, Ekmekcioglu 2008) Acte de conférence : The Data audit framework : a toolkit to identify research assets and improve data management in research led institutions (Jones, Ross, Ruusalepp 2008).	X	X
États-Unis	CABI, Bill & Melinda Gates Fondation	Guide : How to create a data inventory de Beale et al. (2021)	X	X
États-Unis	Johns Hopkins University	Site internet : Data inventory guide (Johns Hopkins University. Center for Government Excellence 2019)		X
France	Cirad	Présentation PowerPoint : L'inventaire de jeux de données scientifiques : une démarche d'aide à la décision vers l'ouverture de données (Fortuno 2015a) Document de travail : Guide pour l'inventaire et la saisie des données numériques scientifiques du Cirad (Fortuno 2015b) Guide : Dataverse : guide utilisateur (Fortuno 2018) Compte-rendu d'entretien ⁶	X	X
France	Laboratoire d'ethnologie et de sociologie comparée (LESC)	Podcast : Politique des données basée sur un inventaire des jeux de données d'un laboratoire (Heintz, Gautier 2022b)	X	X

⁶ Entretien S. Fortuno, chargée mission science ouverte et données de la recherche au Cirad, Teams, 6 avril 2022.

		Présentation PowerPoint : Retour d'expérience sur la mise en place d'un inventaire et d'une politique de données (Heintz, Gautier 2022a) Site internet : Lesc : présentation (Laboratoire d'ethnologie et de sociologie comparative [sans date])		
Suisse	Alessandro Mondaini	Travail de master : Vers un data warehouse pour des datasets en santé (Mondaini 2021)	X	X
Suisse	HESAV	Compte-rendu d'entretien ⁷	X	X
Suisse	National Centre of Competence in Research (NCCR) one the move – Université de Neuchâtel	Compte-rendu d'entretien		X

L'annexe 5 présente une courte description de chaque institution et de sa démarche d'inventaire et l'annexe 6 la liste des métadonnées analysées.

3.2 Analyse des besoins des personnes ressources

Au sein de la HES-SO Valais-Wallis, plusieurs personnes peuvent être sollicitées pour apporter conseil et soutien lors de la réalisation d'un projet de recherche : le responsable de l'Institut (RI), la responsable administrative de l'IS, le responsable architecture et sécurité du service informatique (SINF), le responsable éthique, le délégué à la protection des données (DPO), la juriste, la spécialiste en information documentaire (SID) de la médiathèque et, depuis le début de l'année, la data manager (DM). L'annexe 3 présente, pour chaque personne, son rôle dans l'institution, son intervention auprès des équipes de recherche, ses besoins et problématiques, les outils utilisés et les métadonnées qu'il serait intéressant de collecter dans l'inventaire.

3.2.1 Synthèse des besoins

Si, pour l'Open Access, un cadre a été clairement défini au sein de l'IS avec la « Procédure d'archivage et de valorisation des activités de recherche »⁸, il est manquant pour l'Open data. Aujourd'hui, les chercheuses et chercheurs n'ont pas ou ont peu connaissance du soutien qu'elles et ils pourraient recevoir de la part des personnes ressources. N'ayant pas l'obligation de les contacter, les équipes travaillent de manière autonome, ne les sollicitant qu'en cas de difficultés (Data management plan [DMP] pas validé, problème de sécurité, différents contractuels, budget insuffisant...). De ce fait, les personnes ressources ne peuvent s'assurer que toutes les démarches nécessaires à la bonne gestion des données et du projet ont bien été entreprises (sécurité, solutions de stockage adaptées et volumétrie suffisante, respect des exigences éthiques, légales et contractuelles, budget adéquat), et agissent après coup, une fois le problème survenu, alors qu'il aurait pu être évité si elles avaient été sollicitées en amont du projet.

⁷ Entretien, L. Amiotte-Suchet, répondant Open Science à HESAV, Teams, 14 avril 2022.

⁸ <https://www.hevs.ch/media/document/20/procedure-arodes-version-definitive.pdf>

Le schéma de l'annexe 3 décrit pour chaque étape du cycle de vie des données de recherche les interactions avec les personnes ressources. Il sera utilisé dans la suite de ce travail lors des entretiens avec les chercheuses et chercheurs de l'IS. Comme on peut le constater, l'étape clé du processus est la planification, car c'est durant cette phase que le projet est conceptualisé, que les bonnes questions doivent se poser pour préparer le terrain et limiter au maximum les risques, les complications et les pertes de temps. Lorsqu'un DMP est rédigé, il fournit aux personnes ressources toutes les indications nécessaires pour comprendre le projet et s'assurer que les moyens et ressources mis en œuvre sont suffisants. De ce fait, sa rédaction devrait être encouragée pour chaque projet.

Dans l'idéal, les personnes ressources souhaiteraient être informées dès qu'un nouveau projet touchant à leur domaine d'expertise est en préparation. Actuellement, il existe une bonne collaboration entre le DPO, le responsable éthique, la juriste et le service informatique qui permet un échange d'informations et un premier suivi des projets. Avec le nouveau rôle de data manager, les équipes de recherche pourront passer au préalable par cette dernière qui transmettra les bonnes pratiques et assurera le relais et le suivi auprès des différentes personnes ressources en cas de besoin. Elle devra identifier les besoins des équipes puis, en fonction de la problématique, intégrer dans la boucle la ou les personnes ressources pour la ou les tenir informée-s de tout nouveau projet touchant son ou leur domaine d'expertise. Comme cela a été relevé, il serait intéressant que le DPO, le responsable éthique et la juriste soient, toujours au besoin, contacté-e-s en premier lieu pour répondre aux questions liées aux démarches à entreprendre. Dans un deuxième temps, le SINF peut être contacté pour le choix de solutions techniques appropriées.

3.2.2 Métadonnées de l'inventaire

Toutes les personnes ressources interrogées ont fait part de leur intérêt pour la réalisation de cet inventaire ainsi que pour y accéder une fois le processus mis en place. Les métadonnées à collecter, selon elles, sont :

- **Informations de base sur le projet** : n° du projet, titre du projet de recherche, acronyme, filière, chef de projet, financeurs, date de début et de fin du projet, état d'avancement du projet.
- **Informations sur le jeu de données** : nom du ou des jeux de données, type de données, volumétrie, personne responsable de la gestion des données, date de collecte des données.
- **Protection des données** : présence de données personnelles ou sensibles, présence de données anonymisées / pseudonymisées, soumission à un comité d'éthique + date de validation de la part du comité.
- **Stockage et conservation** : lieu et méthode de stockage, durée de conservation des données, date de destruction, alerte lorsque l'on arrive à l'échéance.
- **Partage** : lien sur la plateforme de partage, lien sur la publication archivée dans ArODES.
- **Administratif** : présence d'un DMP ? si oui, en obtenir une copie ; présence d'un contrat ? si oui, copie à la juriste.

Ces métadonnées seront utiles dans la suite de ce travail lors de l'analyse et de la définition des métadonnées à conserver pour notre inventaire.

3.3 Bonnes pratiques liées à l'inventaire des données

Les entretiens ainsi que la revue de littérature ont permis d'établir une méthodologie et des bonnes pratiques pour la réalisation et la mise à jour de l'inventaire dont voici la synthèse.

3.3.1 Planifier l'inventaire

La première étape liée à la création de l'inventaire est la planification qui va permettre de définir avec le responsable de la recherche l'étendue du travail, l'objectif, les attentes ainsi que le niveau de description de l'inventaire.

Définir l'objectif de l'inventaire. Souhaite-t-on améliorer la qualité, la disponibilité ou la gestion des données ; identifier et atténuer les risques ; identifier les opportunités de réutilisation ou encore s'assurer de la conformité des pratiques (Beale et al. 2021) ? Autant d'objectifs qu'il est nécessaire de définir préalablement car ils auront un impact sur les décisions qui devront être prises dans la suite du travail.

Définir les données et le niveau de description. Avant de remplir l'inventaire, il est nécessaire de préciser ce que l'on entend par données « afin d'avoir une compréhension commune de la portée de l'inventaire » (traduit de Beale et al. 2021, p. 7). On réfléchira ensuite à la granularité ou la profondeur de description. En effet, il est possible de décrire les données au niveau de « des fichiers individuels, des jeux de données ou des projets » (traduit de HATII, University of Glasgow, DCC 2009, p. 8). On peut également les décrire en fonction du « type de données, du périmètre, de l'état de finalisation ou du lieu de stockage » (Heintz, Gautier 2022a, p. 6). Certain-e-s auteur-e-s recueillent des informations détaillées pour accroître la découverte des données (Beale et al. 2021)⁹ tandis que d'autres préfèrent une description restant au niveau du projet ou du terrain (Heintz, Gautier 2022b)¹⁰. L'Université Johns Hopkins (2019) conseille d'inventorier toutes les données, même celles « sensibles, privées ou peu susceptibles d'être publiées ».

Établir le périmètre de l'inventaire en fonction de son organisation. Chaque institution va proposer un inventaire adapté à sa réalité, à la connaissance et à l'adhésion des équipes de recherche sur les questions d'ORD, aux ressources à disposition pour accompagner et réaliser l'inventaire, aux législations en vigueur et aux souhaits et possibilités de partage (Johns Hopkins University. Center for Government Excellence 2019). Certaines institutions font le choix de commencer l'inventaire dans un département ou laboratoire « test » comme au Cirad¹¹, tandis que d'autres proposent de l'intégrer directement au sein de toute l'institution. Certaines ont choisi une fenêtre temporelle (5 ans à l'HESAV¹², 10 ans au Cirad¹³), alors que d'autres ont laissé la liberté aux équipes de recherche ; dans ces cas, il s'agissait majoritairement de projets archivés (Heintz, Gautier 2022b). Jones, Ruusalepp, Ross (2009) ont pour leur part proposé une classification sur trois niveaux afin de déterminer quels jeux de données feront l'objet d'une description détaillée : « vital » pour les données issues de projets « en cours de réalisation » ; « important » pour les données de projets terminés mais « dont l'organisation est responsable de la conservation », ou celles qui pourraient être utilisées à

⁹ Entretien S. Fortuno, 6 avril 2022.

¹⁰ Entretien L. Amiotte-Suchet, 14 avril 2022.

¹¹ Entretien S. Fortuno, 6 avril 2022.

¹² Entretien L. Amiotte-Suchet, 14 avril 2022.

¹³ Entretien S. Fortuno, 6 avril 2022.

l'avenir ; « mineur » pour des données « dont elle n'a soit pas explicitement besoin » soit pas la responsabilité de conservation à long terme (notre traduction, p. 27).

Impliquer les parties prenantes. Une fois le cadre du projet d'inventaire établi conjointement avec la direction de la recherche, une première communication doit être réalisée afin de présenter le projet aux équipes de recherche ainsi qu'aux personnes qui utiliseront l'inventaire (Jones, Ruusalepp, Ross 2009; Beale et al. 2021). Au Lesc, la démarche a été présentée aux membres du laboratoire en leur expliquant que remplir l'inventaire ne signifiait pas ouvrir les données, mais plutôt signaler les données sensibles afin de les protéger (Heintz, Gautier 2022b). Comme cela a été souligné dans le projet GovEx, l'inventaire des données sera l'occasion d'impliquer les parties prenantes dans ce projet en étant à l'écoute de leurs besoins et préoccupations, en échangeant sur les bonnes pratiques de GDR, ce qui peut renforcer leur adhésion (Johns Hopkins University. Center for Government Excellence 2019).

Réaliser des recherches préliminaires. Des recherches préliminaires sont réalisées afin de se familiariser avec l'organisation et le contexte du service, les différentes personnes qui y travaillent, les pratiques actuelles, les problèmes rencontrés ainsi que la documentation qui fournira des informations sur les jeux de données (Jones, Ruusalepp, Ross 2009).

Être efficace et organisé-e lors de la réalisation de l'audit. La personne responsable de l'audit devra faire preuve d'efficacité et d'organisation et posséder « de bonnes compétences relationnelles » pour mettre les gens à l'aise, bien communiquer et encourager la participation (traduit de Jones, Ruusalepp, Ross 2009, p. 17). Si cette personne est interne à l'organisation, elle doit « être familière avec la gestion des données de l'organisation et avoir un accès complet aux documents et systèmes internes » (p. 18).

3.3.2 Définir les métadonnées

La deuxième étape consiste à décider quelles métadonnées seront intégrées dans l'inventaire. Chaque institution a défini un certain nombre de champs plus ou moins importants en fonction de l'objectif de son inventaire. Le tableau comparatif de l'annexe 6 présente en détail les champs des différents inventaires, mis à part celui du projet et du NCCR qui ne proposaient pas de liste complète de leurs métadonnées. Une dernière colonne comprend les champs mentionnés dans l'analyse des besoins des personnes ressources de la HES-SO Valais-Wallis relevés au chapitre 3.2.

3.3.3 Compléter l'inventaire

Durant cette étape, on souhaite avoir une photographie la plus complète possible des types de données existantes au sein de l'institution.

Choisir la ou les méthodes de collecte. Différentes méthodes de collecte peuvent être utilisées selon « le volume de données, la facilité d'accès aux sources d'information et l'objectif du processus » (traduit de Beale et al. 2021, p. 12) : la recherche documentaire, le questionnaire, l'entretien individuel ou le focus group ainsi que l'ajout d'informations par la ou le responsable des données. Cette personne peut être soit la ou le PI (principal investigator), c'est-à-dire la chercheuse ou le chercheur principal-e, soit une personne désignée responsable des données. Une combinaison de ces méthodes est la meilleure des approches selon les conclusions tirées des études de cas du DAF (HATII, University of Glasgow, DCC 2009).

Le questionnaire a l'avantage de toucher un large éventail de personnes, mais le taux de réponse est souvent bas. Il est donc important d'être clair et concis pour obtenir les informations nécessaires (Beale et al. 2021; Jones, Ruusalepp, Ross 2009).

Les recherches documentaires fournissent des informations sur la manière dont ont été créées et traitées les données par le biais des articles scientifiques et des métadonnées des projets, ainsi qu'à l'aide des informations administratives sur les projets (Jones, Ruusalepp, Ross 2009).

Les entretiens sont plus chronophages mais permettent d'obtenir des informations qualitatives plus détaillées, de connaître la manière dont sont gérées les données, d'en comprendre le contexte, d'évaluer la compréhension des problèmes. Ils permettent également de récolter les besoins et de sensibiliser les chercheuses et chercheurs aux bonnes pratiques en les amenant à réfléchir sur leur propre manière de gérer leurs données (HATII, University of Glasgow, DCC 2009; Beale et al. 2021). Les auteurs recommandent d'utiliser le modèle du cycle de vie des données pour faciliter les entretiens. Des focus groups ou des ateliers peuvent également être proposés pour permettre un échange entre chercheuses et chercheurs (Heintz, Gautier 2022b)¹⁴.

L'ajout d'informations dans l'inventaire par la ou le PI a été utilisé par plusieurs institutions : à la HESAV¹⁵, au Cirad¹⁶ au Lesc (Heintz, Gautier 2022b) ainsi que dans le guide GovEx (Johns Hopkins University. Center for Government Excellence 2019). À la HESAV, un fichier Excel contenant un exemple de réponses attendues a été transmis aux répondant-e-s, tandis qu'au Cirad le fichier était accompagné d'un guide détaillant la procédure et donnant une définition et un exemple pour chaque champ de l'inventaire.

Prévoir suffisamment de temps pour la réalisation des entretiens. La collecte d'informations étant une tâche chronophage, il est nécessaire de fixer au plus tôt les entretiens afin d'éviter les pertes de temps inutiles. Lors des interviews, il s'agit d'informer clairement les répondant-e-s sur ce qui sera réalisé ainsi que sur les avantages à retirer pour les scientifiques et pour l'institution (HATII, University of Glasgow, DCC 2009).

Une fois les données collectées, une évaluation doit être réalisée afin de déterminer si les jeux de données répertoriés sont conservés et gérés de manière efficace, si les ressources sont suffisantes et si certaines données pourraient être réutilisées (Jones, Ross, Ruusalepp 2008)¹⁷. Il sera ensuite possible de dresser un premier bilan des jeux de données, de faire ressortir les types de données prédominants et de proposer des recommandations pour améliorer leur gestion.

Ce travail d'inventaire a permis aux équipes de recherche de prendre conscience qu'elles possèdent plus de données que supposé, que leurs gestion, conservation et partage ne sont pas optimaux et que des services liés à la gestion des données de recherche doivent être mis en place¹⁸. Il leur a également permis de prendre conscience de la nécessité d'anticiper dès le début du projet les questions liées à la gestion et au partage (Heintz, Gautier 2022b).

¹⁴ Entretien S. Fortuno, Teams, 6 avril 2022.

¹⁵ Entretien L. Amiotte-Suchet, Teams, 14 avril 2022.

¹⁶ Entretien S. Fortuno, Teams, 6 avril 2022.

¹⁷ Entretien S. Fortuno, Teams, 6 avril 2022.

¹⁸ Entretien S. Fortuno, Teams, 6 avril 2022.

3.3.4 Mettre à jour l'inventaire

L'inventaire doit être tenu à jour et être intégré dans le processus de gouvernance et de gestion des données de l'institution. De ce fait, la responsabilité de l'actualisation de l'inventaire doit être clairement définie, tout comme le lieu d'hébergement et les conditions d'accès (Beale et al. 2021). Ce rôle pourrait être joué par le responsable Open Science de l'institution¹⁹. De plus, la mise à jour de l'inventaire doit être planifiée lors de chaque projet, dans l'idéal au moment de la rédaction du DMP, afin que cette tâche ne soit pas oubliée, mais dans tous les cas en début de projet ainsi qu'à son achèvement (Beale et al. 2021; Heintz, Gautier 2022b). Il serait possible de prévoir une actualisation annuelle mais, comme l'a constaté l'Université Johns Hopkins (2019), une mise à jour régulière est préférable.

Beale et al. (2021) préconisent d'automatiser la mise à jour ou la collecte des métadonnées de l'inventaire. Ils conseillent « d'introduire des protocoles pour que, chaque fois qu'un document, un ensemble de données ou une base de données est ajouté à un système de gestion de contenu, [la-le PI] soit invité-e à compléter les métadonnées de base qui sont automatiquement ajoutées à l'inventaire » (notre traduction, p. 15).

3.3.5 Publier l'inventaire

La publication de l'inventaire, en interne de l'institution ou en externe, permet de rendre les données de l'institution plus facilement découvrables et utilisables. Ce n'est pas parce que l'inventaire des données est publié que les jeux de données seront forcément accessibles. Seules les métadonnées de l'inventaire le seront. Mais cela permet d'amener de la visibilité et de la transparence sur les données détenues par l'institution (Beale et al. 2021). Il serait possible de ne publier qu'un résumé de l'inventaire qui contiendrait des informations sur le niveau d'ouverture des jeux de données (Johns Hopkins University. Center for Government Excellence 2019). On peut mentionner l'exemple du National Center of Competence in Research (2022) qui publie sur son site un inventaire des projets en format PDF comprenant pour chacun des liens sur : la page du site internet du projet, le projet créé dans SWISSUbase, le codebook si existant, les données déposées dans l'un des dépôts proposés et les visualisations interactives :

Figure 1 : Présentation d'un projet dans l'inventaire du NCCR

The screenshot displays a project entry for 'The Impact of the COVID-19 Pandemic on Bordering Discourses Regarding Migration and Mobility in Europe' by Sandra Lavenex (University Geneva). It includes analysis details, a list of related projects with their respective codes (E, C, Z, T) and titles, and a table of dataset availability options. The table lists various repositories and their corresponding codes: P (Proprietary), D (Dataverse), Z (ZENODO), O (OSF), E (EUI CADMUS), F (FORS), Standardized Codebook, and Interactive Visualization.

Project name (links to its publications list on Zotero.org)	Project Title (links to project homepage) Project leader(s)
	Project description (links to entry on a FAIR repository)
	P (Dataset is proprietary)
	D (Dataset at Dataverse)
	Z (Dataset at ZENODO)
	O (Dataset at OSF)
	E (Dataset at EUI CADMUS)
	F (Dataset at FORS)
	Standardized Codebook
	Interactive Visualization

(NCCR 2022, p. 9, 16)

¹⁹ Entretien L. Amiotte-Suchet, Teams, 14 avril 2022.

3.3.6 S'appuyer sur une politique de données

Il n'est pas nécessaire d'avoir une politique de données pour réaliser un inventaire. Ni le Cirad²⁰, ni l'HESAV²¹, ni le Lesc (Heintz, Gautier 2022b) n'avait de politique au moment de la réalisation de l'inventaire. Au Lesc, l'inventaire a même été la première étape de la rédaction de la politique de données ; il a permis de connaître les types de données et leur localisation pour ensuite pouvoir rédiger une politique qui soit adaptée. Dans le cadre du projet GovEx labs, les auteur-e-s ont interrogé les municipalités à ce sujet et il ressort « qu'avoir une politique de données ouvertes qui appelle à un inventaire des données est utile pour compléter l'inventaire en temps opportun et démontrer son importance dans l'ensemble de l'organisation, mais pas nécessaire » (traduit de Johns Hopkins University. Center for Government Excellence 2019).

3.4 Mise en place du processus d'inventaire

Il s'agit maintenant d'appliquer les étapes et les bonnes pratiques relevées dans la revue de littérature à ce travail.

3.4.1 Planifier l'inventaire

Les objectifs et le périmètre du projet ont été définis lors d'entretiens avec le responsable de l'Institut santé et le mandant du TM.

Nous avons choisi de suivre la classification proposée par le DAF pour sélectionner les projets à analyser (Jones, Ruusalepp, Ross 2009). Nous nous sommes limités aux projets vitaux, c'est-à-dire aux projets en préparation ou en cours de réalisation ainsi qu'à des projets importants, terminés depuis 2020. Ce choix permettait de répertorier les projets les plus importants pour la suite du projet, d'obtenir des informations et des discussions sur les pratiques en matière de gestion, d'archivage et de partage, mais également de tester tous les champs de l'inventaire. Le niveau de description des jeux de données a été laissé au libre choix des participant-e-s.

En accord avec le responsable de l'IS, 51 projets de recherche de 16 PI ont été sélectionnés dans la liste Excel tenue par la responsable administrative : un projet en préparation, 23 projets en cours, sept projets terminés scientifiquement et/ou administrativement et 20 projets archivés. 20 projets proviennent de six PI du domaine physiothérapie et 31 de 10 PI du domaine soins infirmiers.

Afin d'impliquer les parties prenantes comme mentionné dans la revue de littérature (chapitre 3.3.1), le projet de travail de Master a été présenté aux équipes de recherche lors d'une séance de l'IS. L'objectif était de les informer du projet et de solliciter leur collaboration pour remplir le fichier d'inventaire ainsi que pour réaliser des entretiens avec certain-e-s d'entre eux. De plus, comme nous l'avons vu au chapitre 3.2, des entretiens ont également été menés avec les personnes ressources afin notamment de récolter leurs besoins par rapport à l'inventaire.

3.4.2 Définir les métadonnées

L'inventaire a été réalisé dans un fichier Excel en s'appuyant sur l'analyse des inventaires de l'annexe 6. Il contient quatre onglets :

²⁰ Entretien S. Fortuno, Teams, 06 avril 2022.

²¹ Entretien L. Amiotte-Suchet, Teams, 14 avril 2022.

- **Inventaire** : répertorie pour chaque projet de recherche le ou les jeux de données associés.
- **Démarche** : décrit le but et les étapes du projet, établit le périmètre de l'inventaire et donne la définition des termes données de recherche, jeu de données et niveau de description afin d'avoir une compréhension commune.
- **Descriptif des champs** : détaille pour chaque champ jugé pertinent pour le projet : une définition, un type de champ (texte libre, liste déroulante ou liste à choix multiple), une liste d'options si le type choisi était déroulant ou à choix multiple ainsi qu'une colonne remarque pour ajouter des informations complémentaires en cas de besoin. De plus, les champs ont été regroupés en sept catégories afin de les organiser de manière cohérente : Description du projet, Description du jeu de données, Provenance des données, Modalités de partage, Modalités de stockage et d'archivage, Gestion et Qualité.
- **Listes déroulantes** : cet onglet contient les différentes listes utilisées pour simplifier la saisie des informations dans l'inventaire.

Les onglets « Démarches » et « Descriptif des champs » sont détaillés dans l'annexe 7.

3.4.3 Compléter l'inventaire

Afin de répondre à l'objectif de l'inventaire, deux modes de collecte ont été utilisés : la collecte directe par l'envoi du fichier Excel d'inventaire au responsable du projet et la réalisation d'entretiens.

Le fichier Excel d'inventaire contenant les projets de chaque PI a été envoyé par courriel individuellement afin de respecter la confidentialité des participant-e-s et des données collectées. Un formulaire d'information et de consentement accompagnait le message et devait être signé et retourné. Une relance a été effectuée après deux semaines. À trois occasions, la collecte des données de l'inventaire et les entretiens ont été réalisés conjointement et dans trois autres cas les personnes ont souhaité obtenir de l'aide pour remplir le document.

Huit entretiens ont ensuite été réalisés afin de vérifier l'intérêt des champs proposés dans l'inventaire, de connaître les pratiques et les besoins des chercheuses et chercheurs en matière de GDR – plus particulièrement sur les aspects de stockage, d'archivage et de partage –, et d'obtenir des pistes pour l'intégration de l'inventaire dans le processus plus large de l'accompagnement à la GDR. Le choix de l'entretien a été conseillé dans le DAF (HATII, University of Glasgow, DCC 2009) mais également par Duprat (2019) qui souligne qu'étant donné la méconnaissance du sujet par les chercheuses et chercheurs, une enquête qualitative donne des résultats plus intéressants et plus précis qu'un questionnaire où un fort taux d'abandon a été relevé.

Les participant-e-s ont été choisi-e-s avec l'aide du responsable de l'IS en veillant à la bonne représentation des laboratoires. Les choix se sont portés sur des personnes ayant différents niveaux d'expérience dans la recherche, des profils variés (sciences humaines et sociales, médicales et paramédicales, etc.), réalisant des recherches de différents types (rétrospectives, prospectives, expérimentales...), ayant des projets avec des financements et des exigences variés afin d'obtenir un panel représentatif des projets de l'IS.

Des entretiens semi-structurés ont été réalisés soit en présentiel soit par Teams en suivant le guide d'entretien élaboré selon la démarche proposée par Blanchet (2015). Les thématiques

abordées sont basées sur les questionnaires du DAF (HATII, University of Glasgow, DCC 2009), de Bonneville et al. (2021) et de Dillaerts et al. (2020). Le guide d'entretien est disponible en annexe 8. Il était accompagné, comme suggéré dans le DAF (HATII, University of Glasgow, DCC 2009), du schéma de l'annexe 3 représentant le cycle de vie des données, les étapes d'un projet de recherche et les interactions possibles avec les personnes ressources. Un formulaire d'information et de consentement a été transmis aux participant-e-s avant le début de la rencontre. Les séances ont été enregistrées afin de faciliter les retranscriptions. Une fois la synthèse des entretiens réalisée, les enregistrements ont été supprimés. Une grille d'analyse a été utilisée pour extraire les informations de chaque entretien selon une liste de thématiques retraçant les étapes du cycle de vie des données :

- **Cycle de vie des données** : modélisation des interactions, description des étapes.
- **Planification** : planification préalable, DMP, contrats liés aux données, personnes ressources, soutien.
- **Collecte** : types de données, logiciels, organisation et partage durant le projet, documentation, données personnelles et sensibles, stockage et sauvegarde, soutien.
- **Traitement et analyse** : types de traitement et d'analyse, soutien.
- **Planification et partage** : partage, soutien.
- **Préservation** : tri et évaluation, archivage, soutien.
- **Réutilisation** : réutilisation, soutien.
- **Mise en place de l'inventaire** : métadonnées, processus.

3.5 Gestion des données de recherche au sein de l'IS

Treize chercheuses et chercheurs ont renvoyé leur fichier complété, ce qui a permis de récolter des informations pour 33 projets. Une personne a décrit ses données au niveau du jeu de données pour deux projets, les autres sont restées au niveau du projet. Afin d'avoir une cohérence, l'analyse sera faite sur les projets. Ce fichier complété permettra d'analyser la pertinence des champs de l'inventaire et d'établir une photographie des types de données de l'IS.

3.5.1 Photographie des données

Les projets menés au sein de l'IS sont financés par la HES-SO (13), les associations et fondations (7), les organismes publics (7), la HEdS (3), le FNS (2), le PNR (1) et autres (2).

La majorité des données de l'IS sont des données d'observation (35) que l'on peut subdiviser en données d'entretiens (16), de questionnaires et sondages (15) et d'observation de terrain (4). Viennent ensuite les données de référence (10) puis expérimentales (4) et dérivées (3). Aucune donnée de simulation n'a été enregistrée.

La majorité des données sont textuelles (17), de tableurs (14) ou de bases de données (15). On en trouve également sous forme audio (9), vidéo (1) et de codes sources (1). Au vu de leur forme, il n'est pas étonnant que la plupart des projets ne dépassent pas les 500 Mo (27). Seul un projet fait entre 500 Mo et 1 Go et deux entre 1 Go et 200 Go. Deux projets n'ayant pas encore récolté de données n'ont pu être renseignés.

Les projets ayant des données primaires produites dans le cadre de la recherche sont majoritaires (30). Mais il est intéressant de se pencher sur les projets utilisant des données

secondaires (7). Les données de ces projets sont de deux types : soit des données de référence (6) soit des données dérivées (3). Les données de référence sont toutes des données issues de revues de la littérature. Soit le projet en lui-même était une revue systématique (3) soit une partie du projet a consisté en une revue de la littérature. Il est probable que ce type de données aurait pu être ajouté dans plus de projets. À l'avenir, il sera important de sensibiliser les équipes sur ce point, d'autant plus que ces données peuvent facilement être exportées et partagées grâce aux logiciels de gestion de références bibliographiques. La deuxième catégorie de données réutilisées concerne les données dérivées. Il s'agit de données qui résultent « du traitement, de la sélection, de la compilation ou de l'agrégation » (André 2014) de données préexistantes afin « de créer une nouvelle interprétation des données » (National Library of Medicine [sans date]).

Il n'est pas surprenant, étant donné le domaine de recherche de l'IS, de constater que 25 projets contiennent des données personnelles (3), sensibles (8) ou soumises à la LRH (6) ; sept ont des données confidentielles et seuls huit projets n'ont aucune sensibilité. Cette composante a un impact sur le stockage, l'archivage, le partage et la réutilisation.

Concernant l'emplacement de stockage, plusieurs solutions peuvent être utilisées selon les besoins et contraintes du projet. Les solutions les plus régulièrement utilisées sont SWITCHdrive (8), le serveur externe d'un partenaire de projet (REDCap ou autres) (8), ordinateur portable personnel (6), OneDrive (9), disque dur externe/clé USB (4), serveur interne de l'école (3), service externe privée (Dropbox...) (2), autre (1). On constate que les solutions utilisées sont en adéquation avec les données soumises à la LRH, ces dernières étant toujours stockées sur des serveurs sécurisés (serveur interne de l'école ou serveur externe d'un partenaire de projet y compris REDCap). Pour les données personnelles, sensibles ou confidentielles, les solutions sont plus variées et pas toujours propices : SWITCHdrive (6), OneDrive (5), disques durs externes/clés USB/ordinateur portable personnel (5), serveurs sécurisés interne/externe (3). Le fichier Excel ne fournissant pas d'informations concernant le chiffrement, l'anonymisation et le codage des données, il n'est pas possible de vérifier si le niveau de sécurité est suffisant. On constate que pour les solutions de stockage, des directives devraient être proposées.

La durée de conservation des données est une question sur laquelle les chercheuses et chercheurs ne sont pas au clair, et cela se ressent dans l'analyse des inventaires où 15 projets n'ont pas de durée définie. Les durées avancées sont 10 ans (9), 20 ans (4), 30 ans (3) et 5 ans (2). Les durées de 10 et 20 ans peuvent s'expliquer comme nous l'avons vu au chapitre 2.4.1.4 car elles sont liées aux données soumises à la LRH. La durée de 5 ans est appliquée dans certaines universités, notamment à l'UNIL . À noter qu'il est possible que des clauses contractuelles fixent des délais de conservation spéciaux.

Pour 13 projets de recherche, des contrats ont été signés. La majorité concerne des accords de transfert et d'utilisation (6), des contrats de confidentialité (2), des contrats de collaboration entre institutions (2) et autres (4) qui concernent des contrats de prestations. Si l'on se réfère aux propos de la juriste, il serait nécessaire à l'avenir que chaque projet fasse l'objet d'un contrat écrit afin d'éviter toute mauvaise surprise. De ce fait, au regard de ces chiffres, un travail de sensibilisation doit être effectué.

Actuellement, aucun jeu de données n'a encore été déposé sur un dépôt de données ouvert pour les projets listés dans cet inventaire, bien qu'un projet soit signalé comme tel. Cependant,

cela devrait évoluer si l'on se réfère aux DMP rédigés (4) ou en cours de préparation (2). Au regard du champ « Niveau d'accès », on constate que les chercheuses et chercheurs sont ouverts au partage pour 25 projets, avec accès soit sur demande (16), soit public (9). Seuls six projets sont privés et deux n'ont encore aucune indication. Les chercheuses et chercheurs ont également conscience que leurs données ont un fort potentiel de réutilisation car seuls quatre projets n'ont aucune possibilité de réutilisation. La réutilisation pour de futures recherches est la plus importante (19), suivie de l'enseignement (14), pour des partenariats institutionnels (11) et pour finir auprès du grand public (5).

En matière de documentation, le protocole est le type le plus utilisé (12). Viennent ensuite les carnets de bord/terrain (2), et les manuels d'utilisation (2). Pour les métadonnées, c'est le codebook qui est privilégié (9), puis les métadonnées intégrées directement dans le fichier de données (5) et enfin le Readme (1). Sept projets n'ont pas (encore) de documentation.

3.5.2 Pratiques et besoins en GDR

Le chapitre suivant présente une synthèse des pratiques et besoins en GDR au sein de l'IS. Le schéma de l'annexe 3 a été repris afin d'y intégrer les éléments fournis par les chercheuses et chercheurs ainsi que la nouvelle fonction de data management qui sera mise en place progressivement. Il permet de bien comprendre les étapes par lesquelles il est nécessaire de passer et les contraintes qui y sont liées en mettant un accent particulier sur la planification, qui est l'étape la plus importante de la gestion des données et la plus fastidieuse en matière de démarches à réaliser. Il est représenté dans la figure 2 de la page 32 dont voici en détail l'explication.

3.5.2.1 Planification

Généralement, le projet débute, une fois la problématique trouvée, par l'esquisse du projet qui comprend la question de recherche, la population concernée et les méthodes de collecte des données. Des contacts sont pris auprès de collègues à l'interne ou à l'externe afin de créer des synergies. Le responsable de l'Institut et/ou le responsable du laboratoire sont généralement informés des intentions de la ou du PI. Ils peuvent apporter des conseils pour le choix de partenaires ou au niveau scientifique par exemple. Les partenaires de terrain peuvent être contactés afin d'obtenir des lettres de soutien pour appuyer le dossier de candidature.

Une fois la première esquisse de projet établie, la ou le PI se tourne en cas de besoin vers les différentes personnes ressources pour vérifier ou compléter le dossier. La collaboratrice administrative (CA) est toujours la dernière personne contactée pour la vérification du budget avant le dépôt du dossier de soumission. En fonction du bailleur de fonds, un DMP devra être joint au dossier. Actuellement, il est rédigé seulement en cas d'exigence et le modèle utilisé est celui du FNS. À chaque nouveau projet, un DMP existant est repris et réadapté en fonction des nouvelles contraintes (contrats, anonymisation, type de données) afin de gagner du temps. Le niveau de description reste simple car sinon l'investissement serait trop important pour un projet qui risque d'être refusé. Il n'est pas mis à jour régulièrement par manque de temps, mais complété à la fin du projet afin de respecter les exigences. Avoir un canevas avec des propositions de réponses simplifierait sa rédaction. Il est parfois difficile de prévoir tous les types de données produites ou collectées ainsi que les volumétries nécessaires car ils peuvent évoluer au fil du projet, surtout dans le domaine SHS.

Une fois le projet accepté et selon le type données collectées, une validation de la CER-VD est nécessaire. Un dossier complet avec une description de la gestion des données est préparé puis soumis pour approbation. Le responsable éthique est parfois contacté pour apporter son aide. Des contrats peuvent être signés avec les partenaires et/ou le bailleur de fonds.

Pour les études cliniques « financées par des fonds privés et publics » (traduit de U.S. National Library of Medicine [NLM], U.S. National Institutes of Health [NIH] 2021), dès la validation de la CER-VD, le protocole est enregistré sur une base de données internationale d'études cliniques comme clinicaltrials.gov²² qui signale les études en cours. Les informations doivent être mises à jour tout au long du projet. Pour les revues systématiques (RS), on peut citer notamment Prospero²³. Si pour les RS cet enregistrement est vivement conseillé, pour les études cliniques il s'agit d'une obligation, car sans cet enregistrement il ne sera pas possible de publier.

Au niveau administratif, une fois la validation du bailleur reçue, la CA crée un projet dans SageX afin d'y insérer toutes les informations nécessaires (budget, heures...) et donner les accès aux membres de l'équipe de recherche. Elle s'occupe également de l'archivage une fois le projet terminé scientifiquement et administrativement.

La plupart des répondant-e-s n'avaient pas connaissance de l'existence du DPO, de la juriste et des canevas qu'elle propose ainsi que du nouveau rôle de data manager. Certain-e-s ont déjà fait appel au responsable éthique, au SINF ou à la DAUnit selon le type de projet et de données collectées, durant la phase de planification principalement ou dans d'autres phases en cas de problème. Seuls, la collaboratrice administrative, le responsable de l'IS et la spécialiste ID sont systématiquement contactés lors du projet. Une seule personne a connaissance de toutes ces ressources et les contacte régulièrement lors de la planification pour des relectures (DMP, protocole, dossier pour la CER-VD). Elle a le sentiment que cela donne une garantie lors du passage devant la commission d'éthique. Quelques personnes ont trouvé de l'aide auprès des institutions partenaires, soit car la ou le PI n'était pas de l'école soit car les ressources étaient plus importantes ou plus adéquates par rapport à leur domaine de recherche. La majorité des participant-e-s, une fois le rôle de la data manager présenté, trouvent utile d'avoir une interlocutrice de premier niveau vers qui se tourner, qui se renseigne à leur place, synthétise l'information et transmet les bonnes pratiques. Certain-e-s préfèrent contacter directement les personnes expertes pour éviter la mauvaise transmission d'informations. De ce fait, pour ces dernières, il serait suffisant d'avoir une liste des personnes ressources.

3.5.2.2 Collecte

Une majorité de répondant-e-s effectuent des recherches quantitatives, voire mixtes. Pour le quantitatif, elles ou ils ont utilisé des questionnaires (données statistiques), des tests physiques ou cognitifs (données expérimentales)... Pour le qualitatif, les méthodes sont plus variées : observation participante, notes visuelles (photos, films, captures d'écran), notes de terrain (données observationnelles), entretiens, focus groups (données narratives), recension et analyse d'écrits (données de référence). Seules quelques personnes utilisent des données secondaires, principalement pour réaliser des statistiques. Ces données proviennent

²² <https://clinicaltrials.gov/>

²³ <https://www.crd.york.ac.uk/prospero/>

d'institutions étatiques ou cantonales (OFS, registres d'hôpitaux...). Ce type d'acquisition est soumis, comme nous l'avons vu au chapitre 2.4.1.1, à la LSF ainsi qu'à ses ordonnances d'application. L'OFS peut fournir des données appariées (D'Angelo 2020). Selon l'art. 13h de l'Ordonnance sur les relevés statistiques (RS 431.012.1), l'appariement des données est « le fait de relier des données provenant de sources différentes, telles que données d'enquêtes, données de registres, données administratives ou données de mesures ». Leur utilisation est intéressante car elle permet d'obtenir rapidement un grand nombre de données, avec de grandes couvertures temporelles tout en mobilisant un minimum de ressources (Taksler et al. 2021). Mais ce type de données entraîne également des contraintes importantes au niveau de la sécurité des données, dont un concept de sécurité très important et une obligation contractuelle de détruire les données à la fin du projet.

Plusieurs personnes ont utilisé REDCap pour la collecte de données, principalement pour les projets contenant des données sensibles. L'outil est complexe à utiliser et demande du soutien, mais ses possibilités de paramétrage sont très développées. L'annexe 11 présente en détail cette solution logicielle pour l'acquisition de données personnelles et/ou sensibles.

Les règles de nommage des fichiers ne sont presque jamais établies. En revanche, l'organisation des fichiers et des dossiers est souvent plus ou moins structurée. La responsabilité de la bonne gestion des données est souvent implicitement attribuée au PI ou alors déléguée à un-e assistant-e du projet. Les droits d'accès sont définis assez précisément pour les projets utilisant l'outil REDCap ou pour des projets ayant des données hautement sensibles provenant d'hôpitaux. Mais pour les autres plateformes de stockage, les membres de l'équipe ont généralement tous les mêmes accès.

La documentation la plus souvent utilisée est le protocole. Certain-e-s ont également mentionné utiliser ou vouloir utiliser le fichier Readme ou le DMP. Au niveau des métadonnées, le codebook est majoritairement cité. Ce dernier peut être généré facilement avec REDCap. Les métadonnées sont parfois également ajoutées dans les fichiers de données. Cette notion de documentation n'est pas toujours claire et son intérêt au regard du temps nécessaire pour sa rédaction est souvent remis en question.

Les données personnelles et sensibles sont toujours codées avant l'analyse et anonymisées avant le partage. La plupart du temps, seule une voire deux personnes ont accès aux données identifiantes. Selon les cas, une fois l'étude terminée, les coordonnées des participant-e-s sont conservées afin qu'il soit possible de les recontacter. L'anonymisation et le codage des données qualitatives avant l'analyse peuvent entraîner une perte d'éléments permettant une meilleure interprétation et être contre-productifs pour la recherche. Certain-e-s se demandent s'il ne serait pas possible d'avoir un contrat de confiance au sein de l'équipe et d'anonymiser/coder les données seulement avant de les rendre publiques. Les formulaires de consentement intègrent la plupart du temps les notions de partage et de réutilisation des données. Ils sont au format papier car ils doivent être signés selon la LRH, puis sont scannés et conservés dans les archives (voir le chapitre 2.4.1.5).

S'il n'y a pas de gros volumes de données, les solutions de stockage les plus utilisées sont OneDrive et SWITCHdrive. Cette dernière est utilisée pour collaborer avec des institutions hors de la HES-SO, mais plusieurs personnes ont rencontré des difficultés. Les solutions qui évitent de devoir exporter et réimporter les données après utilisation sont préférées. En raison de données hautement sensibles ou de gros volumes de données, des solutions particulières

ont dû être fournies par le SINF. Elles ont l'avantage d'être très sécurisées et les sauvegardes y sont automatiques. Presque personne n'utilise le serveur interne de l'école. Les arguments avancés sont le fait de devoir demander au SINF l'ouverture d'un espace, la facilité des autres solutions, l'accès hors ligne qui n'est pas possible, la nécessité d'être une collaboratrice ou un collaborateur de l'école. Certaines personnes ne partagent pas d'espace de stockage commun avec leurs collègues, et leur envoient les fichiers par courriel. Quelques répondant-e-s seulement sécurisent leurs fichiers et dossiers à l'aide d'un mot de passe, mais personne ne chiffre ses données. Cette notion de chiffrement a par ailleurs particulièrement intéressé les participant-e-s.

Certaines personnes réalisent des sauvegardes ponctuelles sur des disques durs externes ou sur une autre solution que celle utilisée habituellement, mais plusieurs ont également signalé ne pas le faire car elles pensent que des sauvegardes automatiques sont réalisées sur OneDrive et SWITCHdrive. Peu de personnes réalisent des changements de versions pour garder un historique. Plusieurs chercheuses et chercheurs travaillent sur Mac et non sur PC, ce qui peut entraîner des complications car le SINF équipe ses collaboratrices et collaborateurs avec des PC à l'inverse d'autres institutions. Ceci amène des difficultés supplémentaires au niveau des versions de logiciels.

3.5.2.3 Traitement et analyse

Le nettoyage des données est une tâche qui prend énormément de temps car elle est fastidieuse et peu d'aide est disponible à l'interne. Certaines personnes ont pu obtenir une aide extérieure pour certains projets (FORS, doctorants, assistants de recherche...), mais cela représente un coût non négligeable, parfois difficile à supporter. Le besoin le plus fortement exprimé par la plupart des participant-e-s concerne le soutien d'un statisticien pour les aider dans l'analyse, une personne qui comprenne la recherche, ce qui est fait et qui soit à disposition pour les projets.

3.5.2.4 Partage et publication

Seules quelques personnes connaissent des noms de dépôts ouverts. Les dépôts OSF, Zenodo, FORSbase, SwissUbase ont été cités. Une seule personne a déjà déposé ses données sur le dépôt OSF, mais plusieurs ont prévu de le faire prochainement. Comme une grande partie des données contiennent des données personnelles, sensibles ou de santé, un important travail devra être effectué pour anonymiser ces données tout en ayant conscience du risque de perte d'information. Les exigences contradictoires au niveau du partage entre les bailleurs de fonds, les éditeurs de revues (BMJ, JAMA), la CER-VD et les hôpitaux complexifient le travail des équipes. Pour l'instant, si des données hautement sensibles sont collectées, il n'est pas nécessaire de les publier ; mais pour combien de temps encore ? Dans certains contrats, il peut même y avoir une clause qui oblige les équipes à supprimer les données une fois le projet terminé.

La majorité des participant-e-s sont favorables au partage des données, en accès ouvert ou restreint, même si elles et ils ne savent pas encore comment procéder. Les freins au partage sont le temps, la difficulté pour anonymiser les données surtout si elles sont qualitatives, le coût, la disponibilité pour le faire. Un partage, même à l'interne de l'école, pourrait être intéressant. Mais actuellement, on ne sait pas qui travaille sur quoi, quelles données sont détenues par qui. Une meilleure valorisation, une meilleure visibilité serait importante, même si les données ne seront pas réutilisées.

3.5.2.5 Préservation et archivage

De manière générale, les participant-e-s conservent toutes leurs données soit par peur de jeter quelque chose d'important, par souci de conserver le processus de création, soit par manque de temps pour réaliser du tri. Certain-e-s enregistrent les anciennes versions dans un dossier « Archives » mais rien n'est jeté. Quelques personnes ont également des archives papier. Plusieurs personnes ne sont pas au clair par rapport à la durée de conservation des données : faut-il les garder 5, 10 ans ?

Les formats pérennes et leur importance pour le partage et l'archivage sont peu connus. La majorité des personnes n'ont pas prévu d'autres lieux de stockage que ceux utilisés en cours de recherche. Elles se demandent si l'institution propose des solutions (disque dur externe conservé sous clé à l'institution). D'autres questions se posent : en cas de départ à la retraite ou de fin d'activité, quelles seront les conditions d'accès, les obligations légales/éthiques, les démarches à entreprendre ?

3.5.2.6 Réutilisation

La majorité des interviewé-e-s n'ont jamais réutilisé de données existantes pour leurs projets. En revanche, certaines de leurs données ont été réutilisées dans le cadre des cours qu'elles ou ils donnent, moyennant dans certains cas des accords de confidentialité signés par les étudiant-e-s. Quelques un-e-s savent qu'elles et ils peuvent découvrir des données en passant par les articles, mais personne n'a effectué de recherches sur des dépôts ou via des répertoires de dépôts.

3.5.3 Inventaire

L'inventaire a pu être complété sans trop de difficulté. Cependant, un aspect principal est à revoir. Il s'agit de la notion de granularité. Le choix avait été laissé aux participant-e-s et ces derniers ont toujours décrit les données au niveau du projet, sauf une personne. Toutefois, ce niveau de détail n'est pas suffisant ; une description au niveau du jeu de données est bien plus intéressante, même si elle est plus chronophage.

Il aurait été intéressant d'ajouter une ligne d'exemple dans le fichier d'inventaire pour les aider à mieux comprendre les éléments attendus.

3.5.3.1 Métadonnées de l'inventaire

L'analyse du fichier d'inventaire ainsi que celle des entretiens ont permis de faire ressortir un certain nombre de problèmes qu'il est nécessaire de corriger.

Titre du jeu de données. Plusieurs participant-e-s ont rencontré de la difficulté pour donner un titre au jeu de données. S'agit-il du nom du fichier, d'un titre qui reflète le contenu du jeu, du titre du projet ?

Type de données. Si, dans un premier temps, il avait été décidé d'utiliser la catégorisation des données proposée par André (2014), il a semblé pertinent d'apporter davantage de précisions sur la nature des données d'observation, ce type de données étant très majoritairement représenté au sein de l'IS. Il a donc été décidé, en s'inspirant des types proposés par Bonneville et al. (2021) et Duprat (2019), de séparer ces données selon trois types : les données d'entretiens et les données d'observation de terrain qui permettent

d'obtenir des données qualitatives et les données de questionnaires et sondages qui concernent des données quantitatives.

Forme des données. La distinction entre tableurs et bases de données n'a pas toujours été bien comprise. Ce constat avait déjà été relevé par Baud-Vittoz et Varrato (2022a; 2022b) lors de leur intervention aux Journées Open Science de la HES-SO en mai 2022. Quelques exemples permettraient d'améliorer la compréhension. Les colonnes « Formes des données » et « Format(s) de fichier » sont plus ou moins redondantes et un seul champ pourrait être conservé.

Description. Il serait pertinent de préciser dans la définition que la description du jeu de données doit également contenir des informations sur le type d'échantillon ainsi que sur le type de recherche.

Sensibilité. Les différentes notions de sensibilité n'étaient pas suffisamment claires pour les répondant-e-s, mais particulièrement la notion de données confidentielles. L'expliciter dans l'onglet « Description des champs ».

Mesures liées à la protection des données. L'ajout de ce champ permettrait de connaître les mesures prises durant la phase de réalisation du projet pour assurer la protection des données. Liste à choix multiple comprenant : chiffrement, anonymisation, codage, autre.

Langues. Seules trois langues étaient proposées dans le fichier Excel alors que certains projets contenaient d'autres langues (portugais, suisse allemand) que celles proposées.

Emplacement(s) actuel(s). Il aurait été intéressant de ressortir REDCap comme solution de stockage à part entière plutôt que de l'associer au serveur externe d'un partenaire de projet car cet intitulé peut amener de la confusion. Il serait préférable de créer une catégorie « Modalités de partage » contenant les dépôts de données ouverts et les dépôts d'éditeur scientifique.

Licences. Il aurait été nécessaire de préciser que le champ licences était lié aux données et non à la publication et préférable de le mettre dans la catégorie « Modalités de partage ».

Contrats. Ce champ manquait de clarté car il n'était pas précisé si l'on prenait en compte tous les types de contrats ou seulement ceux en lien avec les données de recherche.

Erreurs de formatage. Certains champs n'avaient pas été correctement formatés. Les champs « Niveau de sensibilité » et « Langues » n'étaient pas à choix multiple. Dans les listes à choix multiple, une variable sélectionnée par erreur une deuxième fois est à nouveau rajoutée dans le champ.

Organisation des champs. La catégorie « Modalité de stockage et d'archivage » devrait venir avec les informations liées aux « Modalités de partage » car les champs concernant le partage n'ont pas toujours été bien compris. S'agissait-il de partage durant le projet ou à la fin du projet ?

Onglet « Démarche d'inventaire ». Une définition plus précise doit être ajoutée pour les termes « données dérivées », « niveau de description » et pour « données de la recherche ».

3.5.3.2 Mise à jour de l'inventaire

Une partie des personnes interrogées pensent que le tableau pourrait être complété une fois par année en même temps que le tableau récapitulatif des activités de recherche tenu par la SID (voir annexe 3). Soit le tableau pourrait être rempli directement par les chercheuses et chercheurs soit il pourrait être complété au préalable par la DM puis vérifié par elles et eux.

Une autre possibilité serait une intervention en deux temps, lorsque le projet a été accepté puis lorsque ce dernier est définitivement terminé. La DM pourrait être informée au même moment que le secrétariat lorsqu'un projet a été accepté. Ceci pourrait être fait soit par les chercheuses et chercheurs soit par la collaboratrice administrative. Cette dernière a une vision sur les projets qui débutent et ceux qui se terminent et la DM doit également avoir cette connaissance. À ce moment-là, une entrée dans l'inventaire pourrait être créée en s'appuyant sur le protocole ainsi que sur les informations de SageX. Concernant sa mise à jour, plusieurs personnes préfèrent que la DM les contacte afin de l'effectuer ensemble.

3.5.4 Discussions

On constate que les pratiques en matière de gestion des données de recherche ne sont pas inexistantes, mais sont lacunaires par manque de cadre et de connaissance du soutien qui pourrait être apporté. De ce fait, la première tâche importante à réaliser pour l'IS est de rédiger une politique, une directive ainsi que des bonnes pratiques pour encadrer la GDR. Les personnes ressources représentent un soutien important et ces personnes doivent être présentées et connues de toutes et tous. Un workflow doit être proposé afin de les intégrer dans le processus de recherche, principalement durant l'étape de planification. Le schéma de la page 31 pourra servir de point de départ.

La planification est une étape incontournable pour la GDR. Or pour l'instant, seules les personnes qui doivent rédiger des DMP ou des dossiers pour la CER-VD mènent une réflexion sur la gestion de leurs données. De ce fait, une checklist détaillant précisément les tâches à réaliser en matière de GDR devrait être réalisée rapidement afin que chacun et chacune sache ce qu'elle ou il doit entreprendre.

Comme nous l'avons vu, les données de recherche de l'IS sont très majoritairement des données primaires observationnelles, avec une répartition quasi équivalente entre les données issues de questionnaires (quantitatif) et celles issues de données d'entretiens (qualitatif). Ces données étant principalement textuelles et numériques, elles ne nécessitent pas de grandes capacités de stockage. La majorité des projets contient des données soit personnelles, soit sensibles, soit personnelles liées à la santé, ce qui engendre des contraintes supplémentaires en matière de sécurité des données ainsi que pour le partage des données.

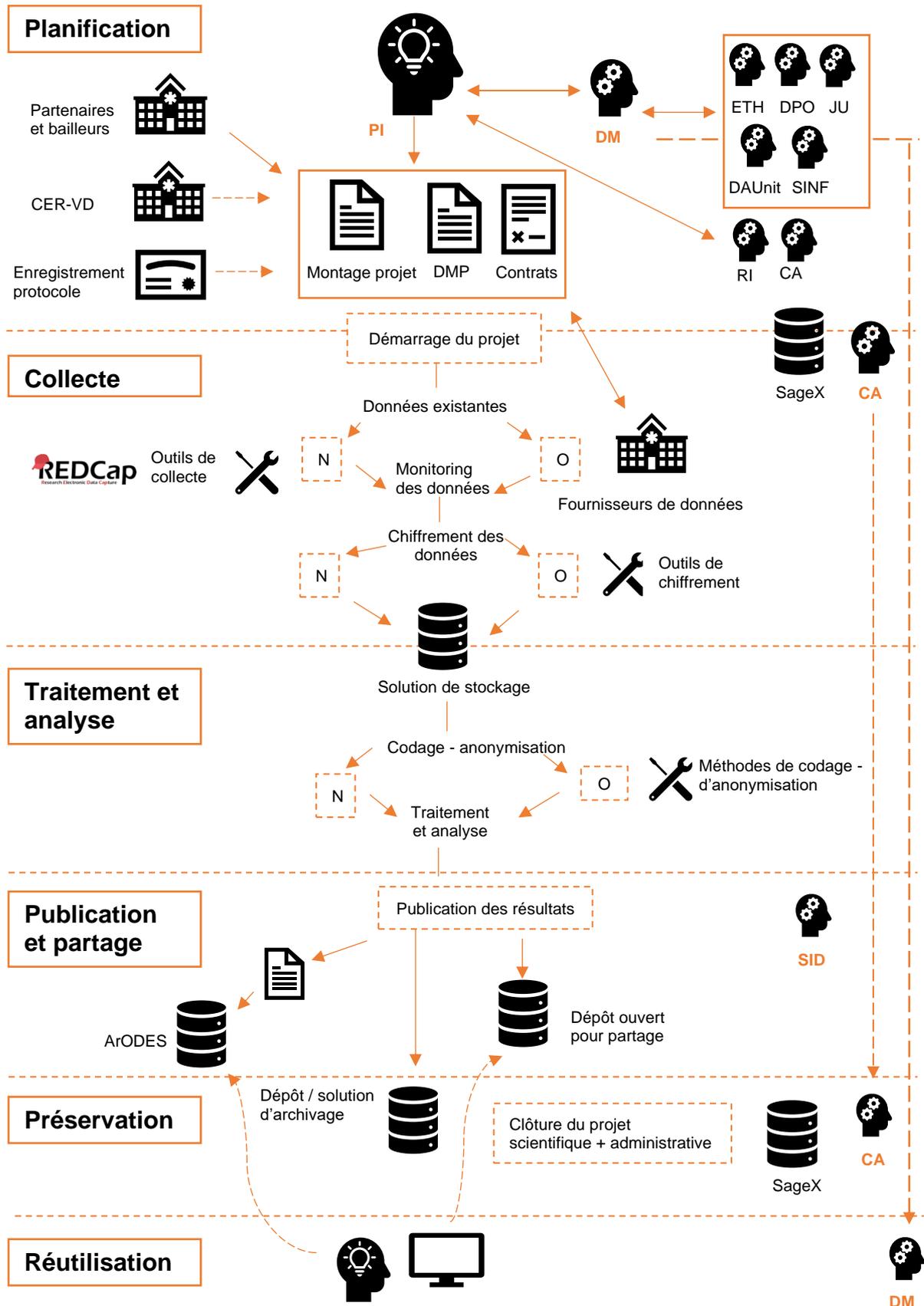
Les solutions de stockage choisies ne sont pas toujours adaptées aux contraintes des projets car il y a une méconnaissance par rapport aux solutions existantes et à leurs avantages et inconvénients. De ce fait, un tableau décisionnel devra être proposé pour aider les équipes à choisir la solution la plus adaptée à leur projet. Pour cela, voir le chapitre 5.3. La gestion des données sensibles nécessite une protection particulière, par du chiffrement, pour apporter une couche de protection supplémentaire. Or, pour l'instant, personne ne sécurise les données de cette manière.

L'étape de préservation et d'archivage avec les notions de tri, d'archivage, de formats ouverts, de durées de conservation est un des grands points noirs de la GDR au sein de l'IS. Aucune

information ni directive n'étant disponible à ce sujet, tout est conservé, ce qui est problématique. De plus, aucune solution d'archivage à long terme n'est actuellement mise à disposition. Des règles et des outils vont devoir être rapidement proposés car il n'est pas envisageable de continuer de cette manière.

Les équipes n'ayant encore quasiment jamais déposé de données dans un dépôt, un soutien particulier devra leur être offert, notamment en matière de procédure de dépôt, d'anonymisation et de documentation. Ce soutien sera d'autant plus important au vu du nombre de projets ayant des données personnelles, sensibles et de santé ainsi que de la proportion de recherches qualitatives réalisées à l'IS. Un bon accompagnement pourra influencer positivement leur vision du partage des données et les encourager à renouveler l'expérience.

Figure 2 : Étapes du cycle de vie et interventions des personnes ressources



DPO : délégué à la protection des données, CA : collaboratrice administrative, SID : spécialiste en information documentaire, DM : data manager, ETH : éthique, SINF : service informatique, RI : responsable d'institut, DAUnit : data acquisition unit, JU : juriste, PI : chercheuse-chercheur. Les flèches en trait-tiré sont des éléments facultatifs, les ressources en orange sont obligatoirement contactées durant la phase du projet

(Inspiré de Delafontaine, 2020)

3.6 Recommandations

3.6.1 Processus d'inventaire

Étant donné le rôle de la DM dans l'accompagnement des équipes de recherche par rapport aux données, lui confier également la gestion de l'inventaire a tout son sens. Mais afin de pouvoir mener adéquatement cette tâche, un processus doit être mis en place. Ce processus est représenté dans la figure 3 de la page 35.

Au vu des retours des chercheuses et chercheurs et de la collaboratrice administrative, ainsi que du processus proposé pour la collecte des activités de recherche et du retour d'expérience faite par la spécialiste ID à ce sujet (voir annexe 3), il est possible de proposer quatre temps de collecte d'informations pour compléter l'inventaire : deux temps obligatoires (T. 2 et T. 4 sur le schéma) et deux temps facultatifs (T. 1 et T. 3).

T. 1. Il est possible, lors de la rédaction de l'esquisse du projet, que la DM soit contactée par la chercheuse ou le chercheur afin de fournir un soutien pour le montage du projet. Elle pourra alors récolter les premières informations sur le futur projet, en sachant qu'à ce stade, le projet peut être refusé. Selon les exigences du projet, certaines personnes ressources pourront être informées qu'un projet nécessitant leur attention est en cours de montage. Il ne s'agit pas de les informer de chaque nouveau projet, mais seulement de ceux pour lesquels leur aide et leurs conseils pourraient être bienvenus. La collaboratrice administrative devra obligatoirement être vue afin de valider le budget avant l'envoi du dossier.

T. 2. Une fois le projet accepté par le bailleur de fonds, la collaboratrice administrative sera informée par la ou le PI et ouvrira un projet dans SageX. Elle pourra transmettre l'information ainsi que le protocole à la DM afin que cette dernière prenne connaissance du projet et crée une entrée dans le fichier d'inventaire. Cette transmission d'informations sera obligatoire. La DM pourra en cas de besoin contacter la ou le PI afin d'obtenir des informations complémentaires et/ou pour rappeler ses services. Elle informera les personnes ressources concernées. Selon les possibilités de l'outil d'inventaire, une alerte automatique pourra être envoyée aux personnes ressources, dans l'idéal en ciblant uniquement la ou les personnes concernées. Un accès en lecture seule au fichier d'inventaire pourra leur être attribué. Le SINF sera obligatoirement contacté soit par la DM soit par la ou le PI pour l'ouverture d'un espace sur le serveur NAS (voir chapitre 5.5.1).

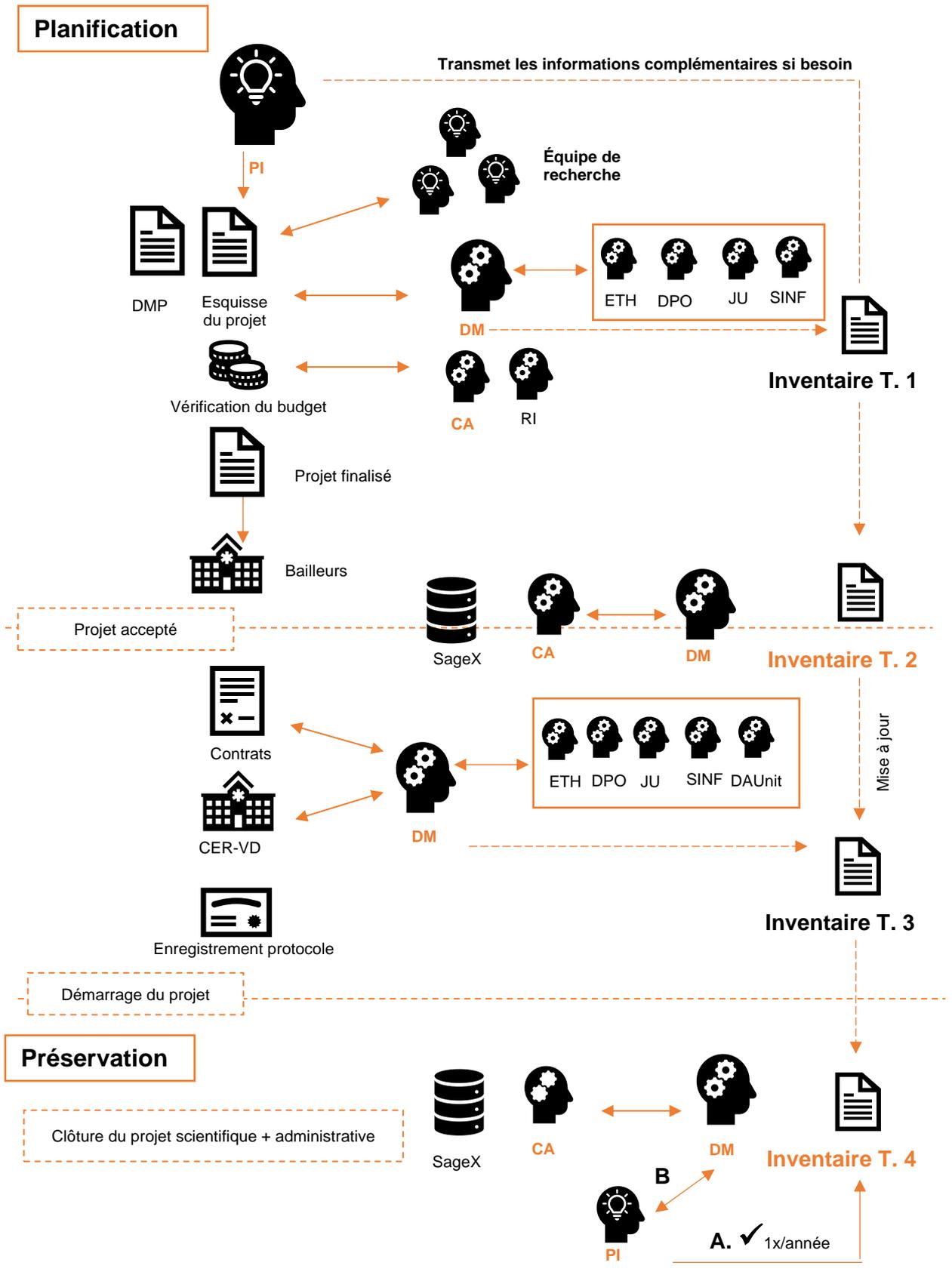
T. 3. Certains projets nécessitent une validation de la CER-VD. Pour ce faire, un dossier doit être préparé et des informations concernant la gestion des données doivent y figurer. À nouveau, les équipes pourront en cas de besoin contacter la DM et/ou les personnes ressources pour obtenir de l'aide. L'inventaire pourra être mis à jour si nécessaire. Cette phase est facultative.

T. 4. Au moment de l'archivage du projet dans SageX, la CA informera la DM qui pourra mettre à jour l'inventaire. À ce stade, il y a deux possibilités : soit le fichier Excel sera envoyé une fois par année à l'ensemble des collaboratrices et collaborateurs pour leur demander de vérifier et compléter les informations de l'inventaire (lettre A sur le schéma), soit la DM contactera chaque PI pour vérifier et compléter les informations conjointement (lettre B) lors d'un court entretien. Cette dernière solution est à privilégier car elle permet un vrai échange avec les scientifiques. De plus, comme il a été constaté lors de ce travail, les informations fournies dans le fichier Excel par la ou le PI ou conjointement lors de l'entretien peuvent sensiblement

changer selon le niveau de compréhension des informations attendues. Ces échanges fourniront l'occasion de dialoguer autour de la gestion des données, de récolter les besoins des PI et de leur rappeler le soutien qui peut leur être proposé.

Ce T. 4 sera également l'occasion de s'assurer que les données du projet ont bien été déposées sur le serveur interne de l'école selon les recommandations établies par l'IS (voir chapitre 5.5). Une évaluation de la qualité des données et des métadonnées sera réalisée avec la ou le scientifique en utilisant les critères (principes FAIR, data quality dimensions et critères infonomics) proposés par Mondaini (2021) dans son travail afin de déterminer la valeur des données. Une évaluation sera effectuée afin de décider quelles données seront conservées et pour quelle durée. Les points d'attention de l'annexe 12 pourront aider à la sélection. La DM pourrait être désignée pour veiller à la suppression des données une fois le délai de conservation échu.

Figure 3 : Schématisation de la mise à jour de l'inventaire



DPO : délégué à la protection des données, CA : collaboratrice administrative, SID : spécialiste en information documentaire, DM : data manager, ETH : éthique, SINF : service informatique, RI : responsable d'institut, DAUnit : data acquisition unit, JU : Juriste, PI : chercheuse-chercheur. Les ressources en orange sont obligatoirement contactées durant la phase du projet

3.6.2 Propositions d'amélioration et de soutien

Les propositions suivantes sont tirées des entretiens ainsi que de l'analyse de la GDR au sein de l'IS. Elles concernent la GDR tout au long du projet, mais également des aspects liés à l'organisation de l'IS. Bien que ces derniers sortent quelque peu du cadre de ce travail, il était malgré tout important de les signaler. Il s'agit de proposer pour chaque étape du cycle de vie :

- **Des guidelines** (l'anonymisation/le codage, les formats pérennes, les modalités de stockage, sécuriser les données, documentation...).
- **Une liste d'outils** (convertisseurs de fichiers, chiffrement, questionnaires...).
- **Des canevas** (DMP, Readme...).
- **De courtes formations** sur des thématiques liées à la gestion des données, qui pourraient être données par des chercheuses et chercheurs (statistique, nettoyage...), par les personnes ressources (DM, éthique, DPO...) en fonction de l'expertise de chacun-e, notamment sur le partage des données, la réutilisation (rechercher des données, plateformes existantes, conditions).
- **Une checklist** des tâches à accomplir lors des étapes d'un projet qui mentionne à quel(s) moment(s) la DM et les autres personnes ressources peuvent ou doivent être contactées.
- **Du soutien** notamment pour partager et documenter les données dans un dépôt ouvert si ce travail ne peut être transféré à la DM, ainsi que du soutien de la part d'un-e statisticien-ne pour réaliser les analyses statistiques.

Proposer une politique ainsi qu'une directive sur la GDR contenant des informations par rapport à la durée de conservation, au choix des données à conserver ou à supprimer, aux plateformes de stockage et d'archivage, aux conditions d'accès, aux exigences en matière de protection des données et aux responsabilités et obligations des PI dans la GDR.

Organiser de courtes séances régulières au sein de l'IS, par Teams ou en présentiel, afin de se rencontrer, d'échanger, de présenter l'avancement de son travail et ainsi favoriser la cohésion d'équipe. Elles pourraient également être l'occasion de faire intervenir les personnes ressources pour présenter rapidement des informations importantes pour les équipes.

Favoriser l'échange de savoir-faire entre seniors et juniors en proposant des solutions comme une équipe Teams qui pourrait être utilisée pour poser des questions et recevoir des conseils de la part des collègues.

Délivrer une attestation ou un label aux chercheuses et chercheurs ayant suivi un certain nombre de formations sur la GDR afin de les encourager à développer leurs compétences. Un tel label pourrait être mis en avant lors des soumissions auprès des bailleurs de fonds et reporté dans les CVs.

Améliorer la visibilité des recherches et des équipes sur le site de l'école. Promouvoir la plateforme des compétences de la HES-SO [people@hes-so²⁴](mailto:people@hes-so.ch) car elle est peu connue.

Développer un réseau d'interlocutrices et interlocuteurs externes que l'institution encouragerait à contacter afin de bénéficier d'un appui en cas de besoin et de favoriser une émulation scientifique.

²⁴ <https://people.hes-so.ch/fr/>

4. Outil d'inventaire

4.1 Méthodologie

Le deuxième objectif de ce projet consiste à formuler des propositions d'outils pour la gestion de l'inventaire. Pour ce faire, les informations tirées de l'analyse des besoins des personnes ressources (chapitre 3.2), les pratiques et besoins des chercheuses et chercheurs (chapitre 3.5) ainsi que les métadonnées validées de l'inventaire ont permis de définir une version simplifiée d'un cahier des charges en se basant sur le processus proposé par Constantinidis (2020; 2015). Il est disponible dans l'annexe 9. Des recherches dans la littérature, des entretiens ainsi que la consultation de différents sites internet ont ensuite été réalisés afin de répertorier des solutions existantes puis d'en faire une analyse. Enfin, en mettant en parallèle le cahier des charges et l'analyse des outils, une proposition adaptée à l'IS a été effectuée.

4.2 Analyse de solutions existantes

Selon Beale et al. (2021), le format de l'inventaire dépendra de l'objectif de l'inventaire ainsi que du nombre de métadonnées collectées. En s'appuyant sur ces propos, une liste de solutions envisageables pour inventorier les données de recherche a été dressée :

Tableau 3 : Liste des outils d'inventaire existants

Outil	Institutions utilisant la solution
Tableur Excel	Il a été utilisé notamment par le Cirad, le Lesc, la HESAV, le NAA ²⁵ ainsi que par plusieurs autres institutions ayant réalisé des registres d'actifs informationnels. On peut citer notamment : NHS East Staffordshire Clinical Commissioning Group ²⁶ , State of Victoria, Department of Health and Human Services ²⁷ , Organisation des Nations Unies ²⁸ , UK Government ²⁹ (Bischoff et al. 2021)
Base de données de type Microsoft Access	NCCR, UK Data Service (propose un canevas d'inventaire pour les données de recherche)
SageX	Utilisé au sein de la HES-SO

Chaque outil a été décrit de manière détaillée dans l'annexe 10, en fournissant les informations suivantes : but, coût, données techniques, fonctionnalités, maintenance, propriétaire, institutions utilisant cette solution, avantages, contraintes, connaissances nécessaires, autres versions, lien, sources. Pour le champ « fonctionnalités », la liste des fonctionnalités principales et secondaires de l'annexe 9 a été reprise pour l'analyse.

Le dépôt de données ouvert peut être considéré comme un inventaire de données. Mais, si l'on s'appuie sur les propos de Beale et al. (2018), cette solution n'inclut pas toutes les métadonnées choisies dans notre inventaire. De ce fait, elle ne sera pas intégrée à l'analyse.

²⁵ <https://www.naa.gov.au/information-management/information-governance/conducting-information-review/designing-and-maintaining-information-asset-register#information-asset-register-template>

²⁶ <https://eaststaffscg.nhs.uk/publications/information-assets-register/>

²⁷ <https://www.dhhs.vic.gov.au/department-health-and-human-services-information-asset-register-public>

²⁸ <https://archives.un.org/fr/content/field-information-management>

²⁹ <https://www.gov.uk/government/publications/information-asset-register>

Le DMP est également une solution qui semblait intéressante. Il contient une bonne partie des informations nécessaires à la tenue de l'inventaire, mais uniquement pour un projet. Il ne permet pas d'avoir une vue d'ensemble des jeux de données ni d'en assurer le suivi (alertes automatiques, envoi de courriels) de manière adéquate. Dans tous les cas, un outil aurait été nécessaire pour cela.

Actuellement, l'outil SageX n'est pas une solution envisageable car il n'est plus développé et va être changé prochainement. Cependant, il a semblé pertinent de connaître le nombre de fonctionnalités déjà existantes ainsi que celles restant à développer afin d'avoir une idée de l'importance des développements.

Le tableur et la base de données répondent tous deux aux besoins de l'inventaire. Le tableur a l'avantage d'être simple d'utilisation et tout de suite exploitable et de répondre à quasiment toutes les fonctionnalités. La base de données de type Access est une solution plus évoluée qui demande un travail de formatage. Le canevas proposé par le UK Data Service pourrait servir de base.

4.3 Recommandations

Le fichier Excel utilisé dans le cadre de ce travail est tout à fait exploitable pour la tenue de l'inventaire. Il est simple d'utilisation, répond aux besoins et peut même être développé pour recevoir des alertes automatiques. Il offre une vision d'ensemble des projets de recherche et des jeux de données et permet de rechercher et d'afficher uniquement certains projets ou certains jeux en fonction de critères sélectionnés.

Cependant, il s'agit d'un nouvel outil à ajouter à la liste déjà existante. Il serait encore plus pertinent de profiter du changement annoncé concernant l'outil de gestion des projets de recherche pour transmettre au service concerné un cahier des charges contenant les champs et fonctionnalités à intégrer afin de pouvoir l'utiliser également comme outil d'inventaire. Car, comme nous l'avons vu, la collaboratrice administrative ainsi que la spécialiste ID utilisent toutes deux un fichier Excel, l'une pour la gestion des projets de recherche de l'IS et l'autre pour la gestion des activités de recherche. Le nouvel outil de gestion des projets de recherche ne pourrait-il pas intégrer également leurs besoins afin que l'ensemble des activités de recherche soit géré par le même outil ? Les DMP, les contrats, les publications et autres activités de recherche pourraient être archivées dans la solution, ainsi toutes les personnes assurant le suivi pourraient accéder aux documents nécessaires à la bonne compréhension du projet. Le cahier des charges pourrait être complété en intégrant les besoins de la coordinatrice administrative et de la SID. Cette solution sort quelque peu du cadre de ce travail mais il était important de prendre un peu de hauteur et ne pas rester focalisé uniquement sur la problématique de l'inventaire. Bien que l'outil SageX ne soit pas uniquement utilisé par la HES-SO Valais-Wallis, les développements proposés pourraient très certainement intéresser d'autres écoles. Avec la mise en œuvre de la stratégie Open HES-SO, toutes les HES auront besoin tôt ou tard d'une solution pour assurer le suivi de leurs données. Le développement de cet outil pour l'ensemble des HES permettrait un gain de temps et unifierait les pratiques.

Si cette solution ne peut se réaliser, la base de données de type Access pourrait être développée pour rassembler en un seul outil les différents fichiers Excel en parallèle de SageX.

5. Stockage et archivage

Un certain nombre de solutions existent en matière de stockage, d'archivage et de partage, mais toutes ne sont pas adaptées au contexte de la HEdS. Elles dépendent en outre du projet de recherche et des types de données collectées.

5.1 Méthodologie

Ce chapitre traite de l'objectif 3 de ce travail qui consiste à formuler des propositions de solutions de stockage et d'archivage pour les données de recherche de l'IS.

Dans un premier temps, il était prévu de réaliser un benchmark des outils de stockage, d'archivage et de partage puis d'effectuer des propositions pour chaque partie en se basant sur une liste de critères permettant de les comparer. Au vu du temps imparti et de l'envergure de cette étape, le volet lié au partage des données n'a pas été traité.

Des recherches dans la littérature et sur le web ainsi que des entretiens auprès d'experts du domaine ont été réalisés afin de répertorier des solutions existantes ainsi que des critères permettant d'analyser les solutions de stockage et d'archivage. Les pratiques et solutions existantes mises en lumière lors des entretiens avec les scientifiques ainsi qu'avec les personnes ressources sont également venues compléter cette partie. Des grilles d'analyse ont ensuite été élaborées afin de réaliser un comparatif des solutions. Des recommandations adaptées au contexte de l'IS ont été émises.

5.2 Définitions des concepts

Chaque institution présente les notions de stockage, d'archivage et de partage de manière différente.

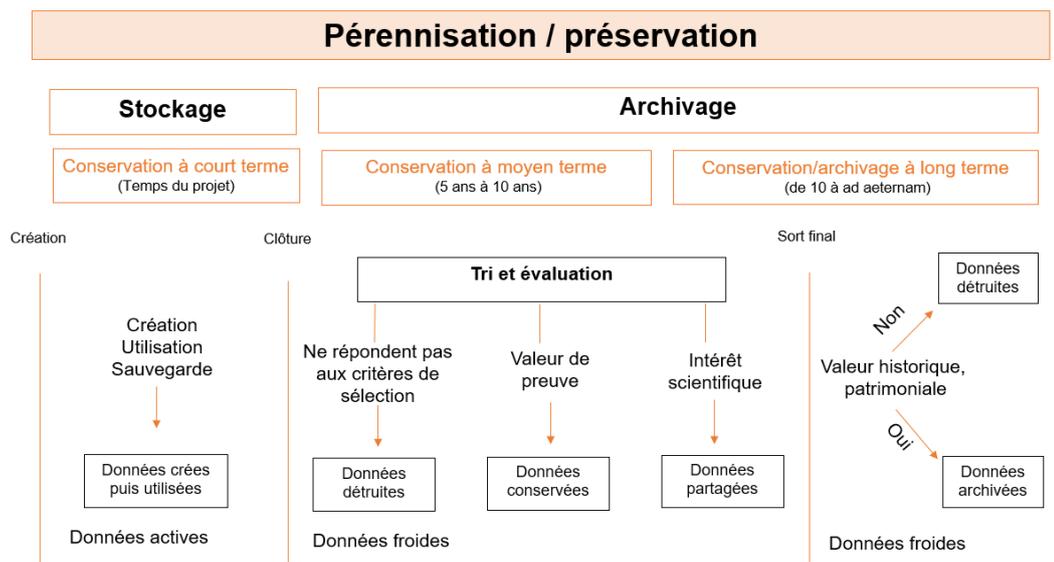
Le FNS (2017) dans son formulaire de DMP comprend une partie sur le stockage et la préservation des données. La notion de stockage intervient dans la phase de réalisation du projet et englobe la question de la sauvegarde. Vient ensuite la préservation qui intervient à la fin de la réalisation du projet. Les chercheuses et chercheurs sont amené-e-s, lors de la rédaction du DMP, à « définir un plan de conservation à long terme » qui détaille quelles données seront « gardées, partagées et archivées » à la fin du projet (p. 3).

La bibliothèque de l'EPFL (2022) et l'Université de Genève (UNIGE 2020) suivent la même dénomination que celle du FNS, au contraire de l'Université de Lausanne (UNIL 2020a; 2021b) qui propose trois phases différentes sur son site internet : le stockage courant qui a lieu dans la phase de réalisation du projet et peut se poursuivre jusqu'à 5 ans après le projet ; le stockage à long terme qui se fait une fois le projet terminé et concerne des données « froides » qui ne sont plus modifiées et dont la durée de conservation n'excède généralement pas les 10 ans ; et l'archivage à long terme qui vise une conservation pérenne (ad aeternam) dans des dépôts d'archivage pérenne et FAIR. Dans la directive 4.5 sur le traitement et la gestion des données de recherche, la direction de l'Université de Lausanne (2021) parle pour sa part de stockage pour « l'utilisation et la sauvegarde » durant la réalisation du projet et d'archivage pour « la conservation des données » une fois le projet terminé.

Avant d'entrer plus en avant dans ce chapitre, il semble donc pertinent de reclarifier chacune de ces notions afin d'éviter toute confusion. La définition de chaque concept est présente dans l'annexe 14. En s'appuyant sur ces définitions ainsi que sur les différents exemples présentés

ci-dessus, il est possible de schématiser la pérennisation et l'utilisation des données dans le temps.

Figure 4 : Pérennisation des données de recherche



(inspiré de Lorenzetti sans date)

La pérennisation ou préservation englobe tout le processus sur le court, moyen et long terme. Le stockage comprend l'utilisation et la sauvegarde des données durant la phase de réalisation du projet (à court terme). Une fois le projet terminé, on entre dans la phase d'archivage. Une première étape comprend la conservation sur le moyen terme qui permet de garantir l'accès aux données « à des fins de réutilisation, de preuve ou de reproductibilité » (Sbeih, Dedet, et al. 2020, p. 6). Les données pourront dans certains cas être partagées, d'autres seront uniquement conservées ou gardées selon la dénomination du FNS (2017) et d'autres encore seront détruites car elles ne correspondent pas aux critères de sélection (voir l'annexe 12). Généralement, les données sont partagées et réutilisées pour 5 à 10 ans (Lai 2019). La dernière étape consiste en une conservation à long terme de plus de 10 ans jusqu'à une conservation ad aeternam. On y trouve des données conservées entre 10 et 20 ans à titre de preuve, comme les données de santé, ou liées à des brevets ou des contrats, ainsi que des données ayant une valeur patrimoniale ou historique et qui seront conservées ad aeternam ou selon une durée définie, de plus de 30 ans (Lai 2019) jusqu'à plus de 50 ans (Sbeih, Dedet, et al. 2020).

5.3 Stockage

5.3.1 Contraintes liées au stockage

Plusieurs aspects exercent une influence sur la stratégie de stockage.

Espace de stockage. Le volume de données collectées a un impact sur la solution de stockage choisie. Si des données vidéo ou audio sont collectées, elles nécessiteront un espace de stockage plus important que des données textuelles (Summers, Corti 2020).

Sécurité des données personnelles et sensibles. Une des contraintes concerne la sensibilité des données collectées durant le projet (Summers, Corti 2020; Recker, CESSDA Training Team 2017). En effet, comme nous l'avons vu au chapitre 2.4, les données personnelles, sensibles et personnelles liées à la santé nécessitent un niveau de sécurité plus

élevé ; de ce fait, seules certaines solutions qui assurent un niveau de protection suffisant pourront être utilisées. Cela peut être une exigence du fournisseur de données. L'OFS (2021), par exemple, impose des mesures de haute sécurité pour les données appariées par ses soins : pas d'environnement cloud accessible au public, stockage en Suisse, chiffrement des données lors du transport, du stockage et de la sauvegarde, accès physique limité, authentification à doubles facteurs et droits d'accès limité et contrôlé, etc.

Gouvernance et hébergement. Un autre aspect important pour la sécurité des données personnelles et sensibles concerne l'hébergement des serveurs et la transmission des données à l'étranger. Selon l'art. 6 de la LPD, les données personnelles ne peuvent être communiquées à l'étranger si la législation en vigueur dans le pays ne garantit pas « un niveau de protection adéquat » (RS 235.1). Afin de s'assurer « de la licéité de la communication transfrontalière de données » (PFPDT 2021a, p. 1), le PFPDT propose un guide ainsi qu'une liste « des Etats ayant une législation assurant un niveau de protection adéquat » (PFPDT 2021b, p. 1). Les Etats-Unis font partie des pays qui ne respectent pas ces exigences (PFPDT 2021b, p. 11). De ce fait, GoogleDrive ou Dropbox qui sont soumis aux lois américaines ne peuvent être utilisés en Suisse.

Partage et accès. Selon que les chercheuses et chercheurs collaborant sur le projet proviennent de la HES-SO Valais-Wallis uniquement, d'autres institutions membres de la HES-SO ou sont affilié-e-s à d'autres universités en Suisse ou à l'étranger, toutes les solutions ne pourront convenir (Recker, CESSDA Training Team 2017; Summers, Corti 2020; Diaz et al. 2022). Il s'agira également de s'assurer que les solutions permettent le contrôle des accès (Summers, Corti 2020).

Sauvegarde. Il est indispensable de connaître les stratégies de backup qui sont mises en place pour chaque outil (Summers, Corti 2020). Une sauvegarde, comme nous l'avons vu dans la définition du glossaire (annexe 14), consiste à réaliser une copie sur un autre support. Ainsi, en cas d'attaque, le risque que les deux versions soient touchées est plus limité. En revanche, certains outils proposent un historique des versions et une restauration accessible durant une certaine période. Cette fonctionnalité est pratique en cas de suppression par mégarde d'un fichier mais en cas d'attaque, bien souvent, ces versions sont également touchées³⁰.

5.3.2 Analyse de solutions existantes

En analysant des questionnaires sur la gestion des données de recherche (Dennie, Guindon 2017; Bonneville et al. 2021; Dillaerts et al. 2020; Duprat 2019; Donati 2019; Tenopir et al. 2015) ainsi qu'en interviewant les scientifiques de l'IS (voir chapitre 3.5), et le responsable architecture et sécurité du SINF, il a été possible de dresser un panorama de solutions utilisées pour le stockage. Le tableau de la page suivante présente ces solutions catégorisées selon la proposition de Diaz et al. (2022).

Chaque solution a été analysée plus précisément dans l'annexe 11 en utilisant les critères mentionnés précédemment : but, gouvernance, hébergement, accès et sécurité, partage, sauvegarde, espace de stockage, avantages, contraintes, remarques et sources.

³⁰ Entretien P. Schneider, responsable de l'équipe architecture et sécurité au sein du Service informatique de la HES-SO Valais-Wallis, 14 juillet 2022.

Tableau 4 : Solutions de stockage

Catégorie	Type de solutions
Local	Ordinateur portable professionnel/personnel Disque dur externe/Clé USB
Cloud	Serveur cloud privé (Dropbox, Google Drive) ³¹ One Drive Enterprise Teams Sharepoint SWITCHdrive personnel et collaboratif REDCap
Lecteur réseau	Network Attached Storage Machine virtuelle

Le tableau suivant présente les solutions qui peuvent être utilisées en fonction des critères de volume, de sensibilité des données, de sauvegarde ainsi que des personnes participant au projet.

Tableau 5 : Comparatif des solutions de stockage

	Ordinateur portable	Disque dur externe	OneDrive	Teams Sharepoint	SWITCHDrive perso.	SWITCHDrive collaboratif	REDCap	NAS	VM
Individuel	X	X	X	X	X	X	X	X	X
Groupe			X	X	X	X	X	X	X
Interne VS	X	X	X	X	X	X	X	X	X
Externe			X	X	X	X	X		X
< 1 TB	X	X	X	X	X	X	X	X	X
> 1 TB				X		C	X	X	X
Pas sensibilité	X	X	X	X	X	X	X	X	
Données personnelles			C	C	X	X	X	X	
Données sensibles					C	C	X	C	
Données très sensibles									X
Sauvegardes automatiques				C			X	X	X

X : utilisable, C : sous conditions

Diaz et al. (2022) recommandent pour les données personnelles sensibles « d'utiliser une combinaison de chiffrement et de mot de passe fort pour les solutions locales, de chiffrer toutes

³¹ Les serveurs de cloud privé n'étant plus autorisés au sein de la HES-SO Valais-Wallis, ils ne seront pas analysés dans la suite de ce travail.

- **Préserver l'intelligibilité** : des métadonnées permettant de contextualiser, comprendre et identifier l'information à travers le temps sont nécessaires (CINES 2016).

En cela, le facteur temps est un élément primordial ; un archivage entre 5 et 10 ans ne devrait pas poser de problèmes d'accessibilité ou de lisibilité, l'évolution technologique n'étant pas suffisamment importante. En revanche, un archivage sur une plus longue période engendre des risques d'obsolescence matérielle, logicielle et de format de fichier ainsi qu'une perte de signification du contenu (CINES 2020a).

Le coût pour l'archivage à long terme doit également être pris en considération, à plus forte raison si des données doivent être conservées 10 à 20 ans (Beagrie, Jones, Digital Preservation Coalition 2021).

5.4.1.1 Situation en Suisse

Gastl-Hartmann (2022) s'est intéressée à l'archivage des données de recherche dans le contexte suisse. Si les universités confient encore souvent aux archives de l'État l'archivage de leurs documents (universités de Berne, Fribourg, Bâle et Saint-Gall), depuis quelques années un certain nombre d'universités (Genève, Lausanne, Lucerne, Zurich) et l'Eidgenössische Technische Hochschule Zürich (ETHZ) disposent d'un service d'archives universitaires géré par le service IT, en collaboration avec les bibliothèques pour Genève et l'ETHZ, par le service des ressources informationnelles et archives UNIRIS pour l'UNIL ou par le secrétariat général des universités.

Les universités de Bâle, Berne, Fribourg... Lausanne, Saint-Gall et Zurich n'archivent pas leurs données de recherche (Gastl-Hartmann, 2022). Lucerne a, pour sa part, clairement explicité dans son concept d'archivage ne pas conserver de données de recherche sauf cas particulier. À l'UNIBE, les données sont conservées pour 5 ans, puis seule une sélection correspondant aux critères de collecte des archives de l'État et non au domaine d'activité de l'université est archivée. Si à Fribourg et Lausanne des calendriers de conservation sont utilisés, dans la première, aucun critère de conservation ne concerne les données de recherche, au contraire de l'UNIL qui, dans sa directive 4.5, mentionne la possibilité d'un archivage à long terme pour certaines données (Gastl-Hartmann 2022). Pour les durées de conservation, voir la deuxième partie de l'annexe 12.

Actuellement, au sein de l'IS ainsi qu'à la HES-SO Valais-Wallis il n'y a pas de pratique d'archivage, mais du stockage à long terme. Des environnements de production normalement prévus pour de l'hébergement sont utilisés à des fins de conservation. De plus, aucun processus n'a encore été mis en place pour assurer la découverte, la réutilisation, l'intégrité, l'authenticité, l'accessibilité, la qualité, la confidentialité et la pérennité des données à long terme³².

³² Entretien P. Schneider, responsable de l'équipe architecture et sécurité au sein du Service informatique de la HES-SO Valais-Wallis, 14 juillet 2022.

5.4.2 Solutions d'archivage existantes

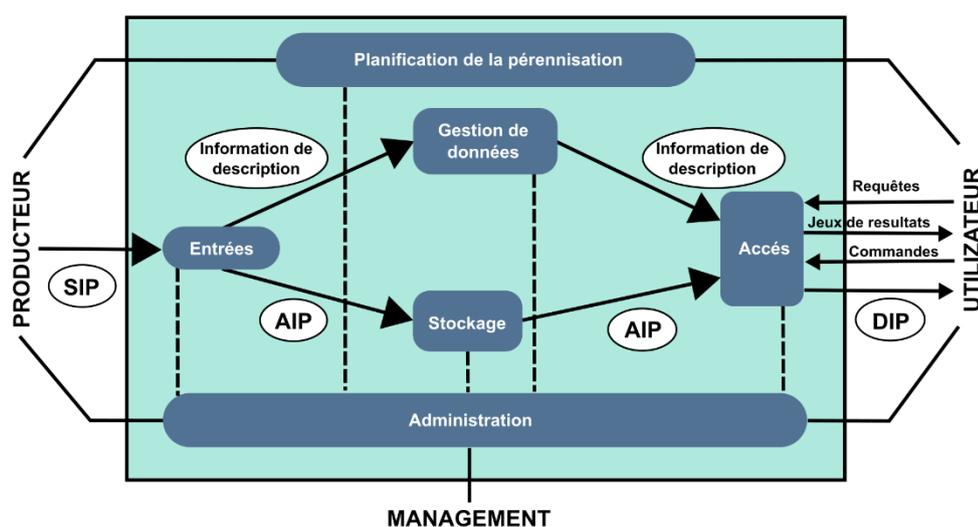
Différentes solutions sont utilisées pour l'archivage à long terme des données de recherche : les dépôts de données et les plateformes d'archivage numérique. Elles seront exposées dans les chapitres suivants. Mais avant tout, il est nécessaire de présenter le modèle OAIS sur lequel se basent ces infrastructures.

5.4.2.1 Le modèle OAIS

L'architecture des dépôts ou des plateformes d'archivage destinés à la préservation à long terme est généralement basée sur le modèle défini dans la norme ISO 14721:2012 pour « un système ouvert d'archivage électronique (OAIS) » (Comité consultatif pour les systèmes de données spatiales [CCSDS] 2017). OAIS est un modèle conceptuel qui décrit les principaux acteurs, les fonctions ainsi que les flux d'informations nécessaires à un système d'archivage (Maday 2016; CINES 2019) :

- Les **acteurs de l'archivage** : les producteurs qui fournissent les objets à archiver, les utilisateurs qui accèdent aux copies des objets archivés, le management qui a une fonction de décideur politique, et l'archive qui est « l'acteur interne, c'est-à-dire l'opérateur du système d'archivage » (CINES 2019).
- Les **domaines fonctionnels**, appelés entités : Entrées, Stockage, Gestion des données, Administration, Planification de la préservation et Accès (Équipe interministérielle du programme Vitam [sans date]; CINES 2019).
- Le **modèle d'information** : « chaque objet à archiver est contenu dans un paquet d'informations qui regroupe plusieurs types d'informations à pérenniser afin de garantir une conservation sur le long terme » : des informations de représentation (structure, sémantique) et de pérennisation (provenance, contexte, identification, intégrité) (CINES 2019).
- À l'entrée du processus, le **paquet d'information** se nomme un SIP (submission information package), « une fois archivés il devient un AIP (archival information package) » puis lors de la mise à disposition pour l'utilisateur, il devient un DIP (dissemination information package) (CINES 2019).

Figure 6 : Modèle fonctionnel OAIS



(Mathieualexhache 2021)

5.4.2.2 Dépôt de données

Un dépôt ou entrepôt de données, *research data repository* ou *data repository* en anglais, est généralement utilisé pour partager les données. Il s'agit d'une « base de données destinée à accueillir, conserver, rendre visibles et accessibles des données de recherche. Son rôle est de permettre le dépôt ou la collecte de données, leur description, leur accès, et leur partage en vue de leur réutilisation » en s'appuyant sur les principes FAIR (Dedieu, Barale 2020). Certains dépôts possèdent une infrastructure suffisante pour assurer également l'archivage à long terme (University of New Hampshire 2020). Dans la suite de ce travail, seuls ces dépôts seront analysés.

Dans la littérature, on trouve généralement trois différents types de dépôts de données :

- **Institutionnel**, qui s'adresse uniquement aux chercheur-euse-s de l'institution.
- **Disciplinaire**, portant sur une seule discipline de recherche ou sur un domaine de recherche qui regroupe plusieurs disciplines comme les sciences médicales.
- **Généraliste ou multidisciplinaire**, qui propose des données de différents domaines.

Chaque solution a des avantages et des inconvénients qui sont relevés ici :

Tableau 6 : Avantages et inconvénients des dépôts

	Avantages	Inconvénients
Institutionnel	<p>Conforme aux exigences de l'institution</p> <p>Peut rendre les données visibles via la diffusion et la promotion</p> <p>Adaptation plus facile aux besoins et pratiques des chercheurs et chercheuses</p> <p>Ajout de données quelle que soit leur valeur</p> <p>Transfert des données facilité (interopérabilité systèmes)</p> <p>Pas de frais pour le dépôt</p> <p>Dépôt de confiance</p>	<p>Ressources financières, humaines pour développement et maintenance</p> <p>Pas de colocation avec des données similaires à l'extérieur de l'institution</p> <p>Difficulté à offrir un accès durable à long terme aux données</p> <p>Manque d'expertise par rapport aux normes de (méta)données pour une préservation et un accès à long terme</p>
Disciplinaire	<p>Spécialisé dans un domaine, expertise en GDR (documentation)</p> <p>Connaissance approfondie du domaine, expertise spécialisée</p> <p>La préservation, la curation sont faites selon des normes communes (migration fichiers)</p> <p>Capacité à contrôler l'accès aux données personnelles (sensibles)</p> <p>Visibilité des données via la diffusion et la promotion</p> <p>Accepte des ensembles de données complets et non des sous-ensembles</p>	<p>Travail important de curation (métadonnées, documentation, préparation des données)</p> <p>Peut être sélectif sur le type de données accepté</p> <p>Coût possible</p>

Généraliste	Convient à un grand nombre de types de données Fonctionnalités avancées (recherche, navigation, visualisation) Grande visibilité car large public Convient aux données pluridisciplinaires	Pas nécessairement conforme aux exigences de l'institution ou des bailleurs de fonds Pas de contrôle sur la qualité, sur les métadonnées Métadonnées simples, pas suffisant pour une réutilisation
-------------	---	--

(Bezzi 2020; Guirlet 2020; Diaz et al. 2022; Bezjak, Vipavc Brvar, CESSDA Training Team 2017)

5.4.2.3 Plateforme d'archivage numérique à long terme

Une plateforme d'archivage numérique est également une solution intéressante pour les données de recherche. Ces plateformes sont généralement proposées par « des sociétés d'archivage et par des institutions publiques » (Sbeih, Dedet, et al. 2020, p. 8). En France, le programme interministériel VITAM (Valeurs immatérielles transmises aux archives pour mémoire) propose une « solution logicielle libre d'archivage numérique pour la prise en charge, la conservation et la consultation sécurisée de très gros volumes d'archives numériques » (Équipe interministérielle du programme Vitam [sans date]). Le CINES (2020b) de son côté s'occupe de l'archivage des données issues de « l'Enseignement Supérieur et de la Recherche française (ESR) ». Il utilise notamment la solution PAC (Plateforme d'archivage au CINES)³³ et est en cours d'intégration du logiciel Vitam. Plusieurs grandes institutions collaborent avec le CINES, dont l'Institut national de la santé et de la recherche médicale (Inserm) et l'Institut national de la recherche agronomique (Inrae) (Sbeih, Dedet, et al. 2020; Sbeih, Loaec, et al. 2020). Au niveau européen, différentes solutions sont également disponibles pour l'archivage à long terme : B2SAFE³⁴ de l'European Data for e-science (EUDAT), Arkivum du consortium du même nom et Labdrive du consortium Libnova qui devrait voir le jour prochainement (Sbeih, Dedet, et al. 2020; Archiving and Preservation for Research Environments 2019).

5.4.2.4 Situation en Suisse

En Suisse, plusieurs institutions, associations ou entreprises proposent une solution pour l'archivage des données de recherche :

Tableau 7 : Solutions pour l'archivage à long terme en Suisse

Institution	Nom de la plateforme	Informations
EPFL	ACOUA	Academic Output Archive, une archive institutionnelle pour la conservation à long terme des données (École polytechnique fédérale de Lausanne 2021)
ETHZ	ETH Research Collection ETH Data Archive	Dépôt pour l'archivage et le partage des publications et des données de recherche (Eidgenössische Technische Hochschule Zürich. Bibliothek 2022) ETH Research Collection pour la publication et ETH Data Archive pour l'archivage

³³ <https://www.cines.fr/archivage/nos-solutions-darchivage/pac/>

³⁴ <https://sp.eudat.eu/catalog/resources/5d81cb5b-3640-4430-b46e-fc652e06a4db>

UNIGE	Yareta	Dépôt de données ouvert pour l'archivage et les partages (2019a)
Association OLOS, DLCLM	OLOS	Dépôt de données ouvert pour l'archivage et partage (OLOS 2022a). L'UNIFR, la HEG-GE et la HES-SO sont membres fondateurs du projet. EPFL, ETHZ, HES-SO, SWITCH, UNIL, UNIBAS UNIZH, UNIGE et le FNS sont partenaires du projet Membres : Bern FH (Gold) et HEP FR (bronze) (Krug, Noirjean 2022)
FORS, UNIL, UNIZH, SWITCH, DaSCH	SWISSUbase	Dépôt de données ouvert pour l'archivage et partage (FORS 2022).
UNIBAS, DaSCH, FNS	DaSCH	Swiss National Data and Service Center for the Humanities pour les données en sciences humaines
Archives de l'Etat du Valais (AEV)	Plateforme	« Plate-forme de pérennisation et d'archivage des documents et données électroniques » (AEV [sans date])
Docuteam	Docuteam Cosmos	Propose des solutions pour l'archivage numérique à long terme (Docuteam 2022)

L'université de Genève propose une solution de stockage froid sur bandes pour les données « qui n'ont pas de potentiel de réutilisation, et qui ne sont pas nécessaires ni pour documenter et valider vos résultats ni pour la reproductibilité » mais qui doivent être conservées « pour des raisons juridiques ou de conformité » (UNIGE 2020). Une solution de ce type est également proposée par l'entreprise SWITCH, SWITCHcoldstor pour le stockage à long terme des données de recherche (SWITCH 2022a).

La HES-SO Valais-Wallis ne dispose d'aucune plateforme d'archivage à long terme ni de dépôt institutionnel pour les données de recherche. Au niveau de la HES-SO, il existe la plateforme d'archivage institutionnelle ARODES pour les publications, mais cette dernière n'a pas une infrastructure suffisante pour archiver des données de recherche³⁵. La HES-SO est partenaire du projet de dépôt de données ouvert OLOS. Cependant, pour l'instant, le choix est laissé à chaque école d'utiliser ou non ce dépôt (HES-SO 2021).

5.4.3 Critères d'analyse

Il existe un certain nombre de principes, de méthodes d'audit, de certifications et d'auto-évaluations qui permettent d'évaluer la pertinence et la fiabilité des infrastructures de préservation numérique. On peut citer (Beagrie, Jones, Digital Preservation Coalition 2021) :

- Les principes TRUST : *Transparency, Responsibility, User focus, Sustainability, Technology* (Lin et al. 2020).

³⁵ R. Grolimund, Journées Open Science HES-SO, 10.05.2022.

- Le *Trustworthy Repositories Audit & Certification: Criteria and Checklist* (TRAC) une certification qui liste les paramètres qui devraient être assurés par l'archive numérique pour être en conformité avec la norme OAIS.
- Le *Digital Repositories Audit Method Based on Risks Assessment* (DRAMBORA) qui propose une méthode d'auto-évaluation des risques (Banat-Berger, Hub 2011).
- Plus récemment le *CoreTrustSeal* (CTS) (Research Data Alliance 2018) une auto-évaluation approfondie ainsi qu'un audit externe qui se basent sur la norme allemande DIN 31644 (Information et documentation – critères pour les archives numériques fiables) ainsi que sur l'ISO 16363 (Audit et certification des référentiels numériques de confiance). Ce dernier a remplacé le *Data Seal of Approval*.

On peut encore mentionner *Le cadre COAR de bonnes pratiques pour les entrepôts* (2020) ainsi que *Les niveaux de préservation numérique* de la National Digital Stewardship Alliance (NDSA 2020) deux ressources qui permettent une première évaluation. Guirlet (2020) a réalisé une analyse des principales certifications afin de proposer une grille pour évaluer les dépôts sous l'angle de l'accès.

Il aurait été intéressant de réaliser une grille axée « préservation » qui intègre les fonctionnalités pertinentes proposées dans le cadre COAR, le CoreTrustSeal et la grille de Guirlet. Mais étant donné l'envergure de ce travail et le temps à disposition, seule la grille du cadre COAR a été utilisée.

5.4.3.1 Estimation de la volumétrie

Afin d'avoir une estimation de la volumétrie concernée par l'archivage nous pouvons reprendre les chiffres obtenus lors de l'inventaire des projets en cours et terminés depuis 2020. Les données sans date de conservation n'ayant pas de données soumises à la LRH, on peut partir du principe qu'elles n'excéderont pas les 10 ans de conservation.

Tableau 8 : Volumétrie des projets de l'IS

Critères	Nbre de projets	Volumétrie
Conservation de 5 ans	2	100 Mo
Conservation de 10 ans	24	403.3 Go
Conservation de 20 ans et plus	7	202.3 Go
Ensemble des durées de conservation	33	605.7 Go

Comme on peut le constater le volume des données à archiver et/ou à partager est peu important. Il est même surestimé car le fichier propose des fourchettes et non des valeurs exactes. De plus, les chiffres concernent l'ensemble des données et non les données conservées après l'étape de tri.

5.4.4 Analyse de solutions

Pour rappel, dans le cadre de l'IS, sur 33 projets seuls huit n'ont aucune sensibilité (voir le chapitre 3.5.1). Toutes les données ne pouvant pas être anonymisées, un grand nombre ne seront pas partagées. Il sera donc nécessaire de prévoir des solutions qui permettent d'assurer l'archivage avec et sans partage.

5.4.4.1 Dépôts de données permettant la préservation

En se basant sur les dépôts analysés par Guirlet (2020), seules les solutions OLOS, SWISSUbase et DaSCH possèdent une infrastructure conçue pour la préservation, hébergent leurs serveurs en Suisse et peuvent être utilisés pour les jeux de données de l'IS.

Des dépôts comme Zenodo ou OSF assurent le partage mais non l'archivage à long terme basé sur le modèle OAIS (Guirlet 2020). C'est pourquoi ils ne seront pas pris en compte dans l'analyse de l'annexe 13. De même, le dépôt de données DaSCH ne sera pas analysé car il devrait rejoindre prochainement la plateforme SWISSUbase (SWISSUbase 2022a).

De l'analyse des deux dépôts, les avantages et inconvénients suivants peuvent être relevés :

Tableau 9 : Comparatif entre OLOS et SWISSUbase

	OLOS	SWISSUbase
Avantages	<p>Accepte toutes les disciplines et tous les types de données</p> <p>Ajout de métadonnées supplémentaires selon son propre schéma</p> <p>Métadonnées disponibles dans des formats lisibles par les hommes/machines</p> <p>Indexation par les moteurs de recherche</p> <p>Choix d'une durée de conservation</p> <p>Versement de gros volumes de données en une seule fois</p> <p>Tous les formats de fichiers acceptés</p> <p>Vérification automatique du niveau de conformité + modération optionnelle + annotations de la qualité et de l'utilité des données par les utilisatrices et utilisateurs</p> <p>Choix du niveau de sensibilité à l'aide de DataTags</p> <p>Possible de déposer des données chiffrées</p> <p>Plateforme de test</p> <p>Simple d'utilisation</p>	<p>Gratuit pour l'archivage et le soutien à la GDR</p> <p>Métadonnées riches</p> <p>Grande expérience dans le partage de données</p> <p>Vérification et validation des données, des métadonnées et de la documentation par des experts du domaine.</p> <p>Limitation de la réutilisation possible pour la recherche/enseignement</p> <p>Protection des données : contrats de dépôt et d'utilisation, conditions d'accès limité et anonymisation des données</p> <p>Chiffrement du stockage</p> <p>Métadonnées toujours disponibles même si les données ne le sont pas ou plus</p>
Inconvénients	<p>Payant pour l'archivage et le soutien à la GDR si non inclus dans le forfait</p> <p>Stockage non chiffré</p> <p>Jeunesse du dépôt : peu de données archivées, peu d'institutions qui l'utilisent</p>	<p>Actuellement orienté sciences sociales et linguistique mais d'autres domaines seront progressivement intégrés</p> <p>Métadonnées ciblées sur les sciences sociales. N'existe pas encore de schéma pour les autres disciplines.</p> <p>Liste de formats limitée</p> <p>Accepte des données anonymisées uniquement (sauf exception)</p> <p>Utilisation plus complexe</p>

Le dépôt SWISSUbase est davantage axé sur le partage des données avec des données qui doivent être anonymisées pour être déposées, sauf exception. Ses principales qualités sont :

- La gratuité de l'archivage et du soutien à la GDR pour les chercheuses et chercheurs ;
- Le recours à des experts du domaine des sciences sociales et de la linguistique pour s'assurer de la qualité des données, des métadonnées et de la documentation. Pour l'instant les autres domaines n'ont pas encore de schéma de métadonnées ni d'expert-e-s définis mais à terme cela devrait être proposé ;
- Un concept de protection des données qui comprend notamment des contrats de dépôts et d'utilisation ;
- La possibilité de signaler un projet dans le dépôt sans obligatoirement y déposer les données associées;
- Une grande expérience dans le partage des données.

Ses points faibles sont principalement liés à son orientation « sciences sociales » mais les développements annoncés devraient progressivement gommer ses faiblesses. Les données de l'IS étant majoritairement des données d'observation cette contrainte est peu significative pour l'institution. Si la richesse des métadonnées est une force du dépôt, elle complexifie cependant la démarche de versement et peut être un frein à l'archivage sur cette plateforme.

Le dépôt OLOS a l'avantage d'être multidisciplinaire. Il accepte tous les types et formats de données et permet le versement de gros volumes de données. Il propose un grand nombre de fonctionnalités intéressantes et répond davantage aux critères du COAR.

Ses principales forces sont :

- La gestion fine du niveau de sensibilité des données à l'aide du système DataTags. Le dépôt gère l'archivage des données non confidentielles et personnelles uniquement ; la prise en charge des données sensibles n'étant pour l'instant que partiellement assurée. Cependant des fonctionnalités complémentaires pour les trois derniers niveaux seront développés prochainement avec l'aide d'un expert en cybersécurité ;
- La possibilité de choisir différentes durées de conservation ;
- La possibilité de déposer des données chiffrées. Cependant les clés d'encodage doivent être gérées par la-le responsable des données et non par le fournisseur de la plateforme ;
- La simplicité d'utilisation et la convivialité de l'outil.

Ses points faibles concernent la jeunesse de sa solution, son coût pour le soutien à la GDR ainsi que l'absence d'expert-e-s pour la vérification de la qualité des données. Cependant pour ce dernier point, il compense cette faiblesse par trois fonctionnalités intéressantes : la vérification automatique du niveau de conformité des formats, un rôle de modérateur qui peut être attribué à un-e membre de l'unité organisationnelle et l'annotation de la qualité et de l'utilité des métadonnées par les utilisatrices et utilisateurs. Dans notre cas, la modération pourrait être réalisée par la DM.

Les possibilités offertes par OLOS ainsi que sa simplicité d'utilisation font de ce dépôt une option vraiment intéressante.

5.4.4.2 Plateformes d'archivage à long terme

Pour les données qui doivent être conservées mais ne peuvent être partagées, trois solutions peuvent être envisagées :

- Les Archives de l'Etat du Valais (AEV) disposent d'une plateforme d'archivage numérique à long terme pour leur administration ainsi que pour les communes et quelques privés. La HES-SO Valais-Wallis pourrait rentrer dans cette dernière catégorie. Les deux institutions étant actuellement en discussion sur l'archivage de documents analogiques, la partie numérique pourrait tout à fait y être intégrée³⁶.
- L'entreprise SWITCH propose SWITCHcoldstor pour le stockage à long terme des données de recherche qui doivent être conservés pour « soutenir la reproductibilité » ou pour respecter « les exigences d'audit ou de réglementation » (SWITCH 2022a). Le responsable architecture et sécurité du SINP est actuellement en discussion avec l'entreprise pour utiliser cette solution pour l'ensemble de la HES-SO Valais-Wallis³⁷.
- L'entreprise Docuteam fournit « des services d'archivage et de conseil en gestion de l'information » (Docuteam [sans date]) et propose notamment « une solution cloud d'archivage numérique à long terme » Docuteam Cosmos (Docuteam 2022)³⁸.

La solution choisie ne sera pas uniquement utilisée par l'IS mais par tous les instituts de la HES-SO Valais-Wallis. De ce fait, il sera nécessaire d'estimer en premier lieu la volumétrie des projets concernés pour chaque école. Si pour l'IS le volume est peu important comme nous l'avons vu précédemment (chapitre 5.4.3.1), les instituts en ingénierie, en informatique, en gestion pourraient avoir des projets contenant des données confidentielles (de brevets, liées à des contrats) ou sensibles avec des volumes bien plus conséquents qu'à l'IS.

5.5 Recommandations

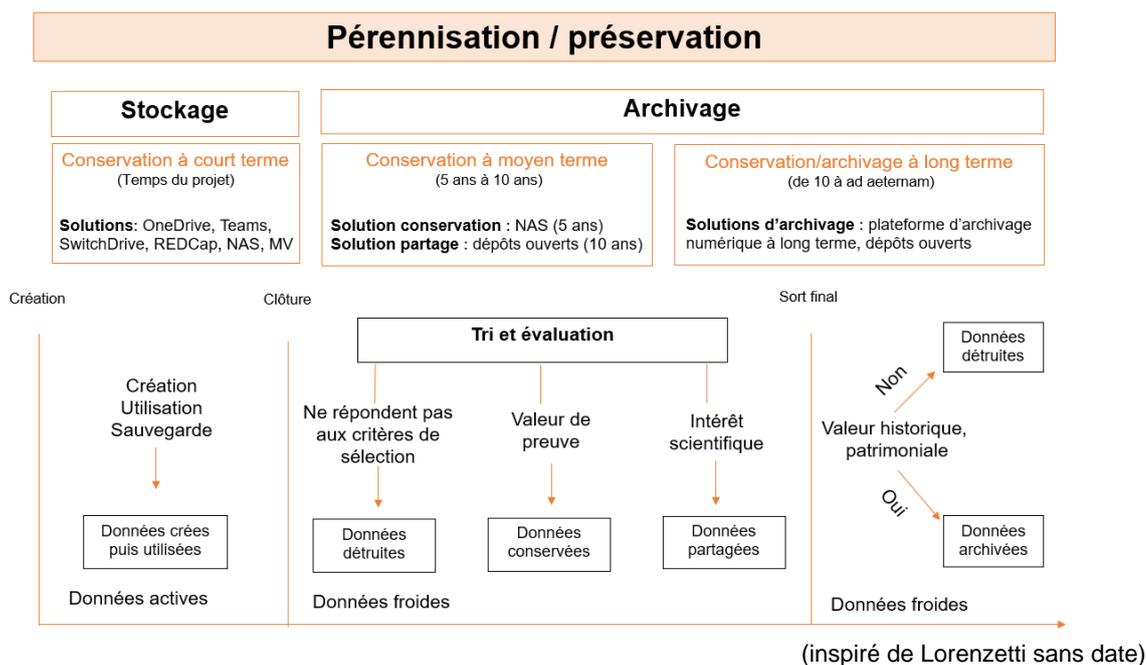
Dans cette partie seront présentés les solutions proposées en termes de stockage et d'archivage. En reprenant le schéma du début de ce chapitre il est maintenant possible de rajouter une strate sur les solutions à utiliser en fonction de la phase de conservation (voir la figure 7 en page suivante)

³⁶ Entretien F. Vionnet, Archiviste-RM aux Archives de l'Etat du Valais, 12.07.2022

³⁷ Entretien, P. Schneider, SINP HES-SO Valais-Wallis, 14 juillet 2022

³⁸ Courriel F. Lorenzetti, DPO à la HES-SO Valais-Wallis, 20.06.2022

Figure 7 : Solutions de stockage et d'archivage pour assurer la pérennisation



5.5.1 Solutions pour le stockage

En reprenant le tableau de la partie 5.3.2 et en conservant uniquement les critères conseillés pour chaque solution, on obtient un tableau décisionnel qui orientera les équipes de recherche dans le choix de la meilleure solution en fonction des contraintes de chaque projet :

Tableau 10 : Tableau décisionnel pour le stockage des données durant le projet

Solution	NAS	Teams Sharepoint	SWITCHDrive collaboratif	REDCap	VM
Individuel	X			X	X
Groupe	X	X	X	X	X
Interne VS	X			X	X
Externe	C	X	X	X	X
< 1 TB	X	X	X	X	X
> 1 TB	X	X	C	X	X
Pas sensibilité	X	X	X		
Données personnelles	X	C	X		
Données sensibles	C		C	X	
Données très sensibles					X
Sauvegardes automatiques	X	C		X	X

X : conseillé, C : sous conditions.

Pour chaque nouveau projet, l'ouverture d'un espace sur le NAS devra être demandé au SINF afin de l'utiliser comme espace principal pour les projets internes à la HES-SO Valais-Wallis ou comme espace de sauvegarde durant le projet s'il touche des personnes externes à la

HES-SO Valais-Wallis. Cette solution devra être privilégiée et rendue obligatoire car elle est la plus sécurisée. L'attribution de droits d'accès ainsi que le chiffrement devront être utilisés en fonction des contraintes de chaque projet.

Pour les projets en groupe avec des membres externes à la HES-SO Valais-Wallis, Teams Sharepoint, SWITCHDrive collaboratif sont à privilégier, tout en réalisant des sauvegardes régulières sur le NAS. Le chiffrement des données ainsi que l'attribution de droits d'accès devront être utilisés en complément afin de garantir la sécurité des données. Pour Teams Sharepoint, le SINF peut instaurer des sauvegardes automatiques sur une équipe Teams sur demande. Pour SWITCHDrive collaboratif, la création d'un ou de plusieurs dossiers doit faire l'objet d'une demande au SINF.

La VM est une solution sur mesure qui peut être adaptée en fonction des exigences du projet. Elle est généralement utilisée pour des projets contenant des données très sensibles avec un niveau d'exigence en matière de sécurité très avancé. Une demande doit être faite au SINF.

Pour finir REDCap peut être utilisé pour l'acquisition de données sensibles quelles que soient les autres contraintes du projet. Une fois le projet terminé, les données devront être sauvegardées sur le NAS.

5.5.2 Solutions pour l'archivage

En fin de projet, les données devront être triées afin de ne conserver que celles jugées pertinentes. Les données pourront être stockées sur le serveur NAS pour une durée maximum de 5 ans comme conseillé par le SINF³⁹ puis elles devront être archivées.

L'archivage à long terme est une problématique qui ne concerne pas uniquement l'Institut santé mais toute la HES-SO Valais-Wallis. De ce fait, la solution choisie pour l'archivage des données devra prendre en considération les volumes ainsi que la sensibilité des données de chaque école.

Actuellement, il n'existe pas une solution unique qui réponde à l'ensemble des besoins en matière d'archivage des données. Plusieurs solutions doivent être utilisées en fonction de la sensibilité des données.

Le dépôt ouvert SWISSUbase peut être utilisé pour l'archivage des données anonymisées et OLOS pour les données non confidentielles ou personnelles (les trois premiers niveaux du système DataTags). Cependant, étant donné les développements annoncés par OLOS pour renforcer la sécurité des trois derniers niveaux de son système DataTags, ce dépôt semble prometteur et pourra peut-être convenir pour toutes les données à l'avenir. Mais pour l'instant, les données confidentielles ou sensibles ne peuvent y être archivées c'est pourquoi une autre solution doit être étudiée. Comme nous l'avons vu dans l'analyse, les Archives de l'Etat du Valais, SWITCH et Docuteam proposent des solutions qui pourraient répondre à ce besoin. Mais une analyse plus approfondie devra être menée car elle n'a pu être réalisée dans le cadre de ce travail.

³⁹ Entretien, P. Schneider, responsable de l'équipe architecture et sécurité au sein du Service informatique de la HES-SO Valais-Wallis, 14 juillet 2022

6. Communication

Le dernier objectif de ce travail consiste à émettre des propositions en matière de communication sur les données à valoriser. Pour cela, une revue de la littérature devait être réalisée pour formuler des propositions. Etant donné le temps à disposition et aux vues de l'importance du sujet seules une recherche documentaire plus modeste dans la littérature ainsi que des éléments ressortis durant les différents entretiens ont été utilisés pour réaliser la proposition ci-après.

6.1 Plan de communication

Durant la phase de réalisation du projet. Comme cela a été soulevé durant les entretiens avec les chercheuses et chercheurs, il serait intéressant de présenter aux collègues de l'IS la recherche en cours, afin de mettre en avant les données collectées, de susciter l'intérêt des collègues et de mener peut-être à une demande de réutilisation de leur part.

Avant la soumission de l'article. Si les équipes envisagent de partager leurs données, un projet devrait être créé dans le dépôt destiné à les accueillir avant la soumission de l'article auprès d'un éditeur. Ainsi le DOI du projet pourra être ajouté dans l'article afin d'augmenter la visibilité des données. Selon Colavizza et al. (2020), les publications qui incluent un lien vers des données déposées dans un dépôt ont « un taux de citation supérieur de 25 % » par rapport à celles qui n'en ont pas. La création du dépôt pourra être faite de concert avec la DM.

Lors de la publication de l'article. Cette phase est très importante car elle va permettre de communiquer via différents canaux pour faire la promotion de l'article ainsi que des données déposées en Open data : communiqué de presse, rédaction de billets de blog, interview réalisée par le service communication de l'institution et relayée sur le site internet de l'école ainsi que sur les réseaux sociaux de l'école et de la chercheuse ou du chercheur (Twitter, ResearchGate, LinkedIn) (Groupe de Travail Réutilisabilité et al. 2022). Le lien sur la publication ainsi que sur le dépôt de données devra être ajouté sur la page people HES-SO⁴⁰, le répertoire de compétences de la HES-SO, sur le compte ORCID⁴¹, ainsi que sur la page internet dédiée au projet sur le site internet de la HES-SO Valais-Wallis⁴².

Lors du dépôt des données. Quelques bonnes pratiques sont à mettre en place pour favoriser la promotion de ses données : déposer ses données en accès ouvert, dans un format ouvert et pérenne, choisir une licence creative commons la plus ouverte possible, citer ses données, utiliser un standard de métadonnées, documenter de manière complète et précise son jeu de données, rédiger un fichier Readme.txt qui soit lisible aussi bien par les humains que par les machines (Bezjak, Vipavc Brvar, CESSDA Training Team 2017; Fressengeas 2020).

Une fois les données déposées. Afin de valoriser davantage les données, il est possible de rédiger un *data paper* et de le publier dans un *data journal*. Il s'agit d'« une publication scientifique, soumise à la validation des pairs, dont le but principal est de décrire un ou plusieurs jeux de données, afin de permettre leur réutilisation » (Bibliothèque universitaire

⁴⁰ <https://people.hes-so.ch/fr/>

⁴¹ <https://orcid.org/>

⁴² <https://www.hevs.ch/fr/recherche-appliquee-et-developpement-a-la-hes-so-valais-wallis/institut-sante-valais/collaborateurs/>

Claude Bernard Lyon 1, Bibliothèque universitaire Grenoble Alpes [sans date]). Le data paper fournit la méthodologie, le processus de collecte et de traitement, l'originalité, la portée et la réutilisation possible. Publier un data paper favorise la visibilité, la réutilisation, le repérage et la citation et permet de valoriser les données et son travail (Hadrossek et al. 2021).

À l'archivage de la publication. Lors de l'archivage sur ArODES, le DOI des données pourra être ajouté à la notice de la publication afin d'en augmenter la visibilité⁴³.

Dans sa pratique professionnelle. Il est également possible de continuer à promouvoir ses données en les citant, en les utilisant lors de conférences ou lors des enseignements qui sont donnés (Open Knowledge Foundation [sans date]).

6.2 Propositions complémentaires

Les propositions faites ci-après sortent du cadre du plan de communication, mais elles ont le mérite de proposer une manière moins conventionnelle de communiquer sur les données de recherche.

Un billet paru en mai de cette année sur le blog de l'Institut Pasteur (2022) a présenté une manière originale de mettre en valeur aussi bien ses données que sa publication par la création d'articles exécutables proposés notamment par la revue eLife⁴⁴. Il s'agit de formater la publication « sous la forme d'un logiciel dynamique » afin qu'elle devienne interactive et qu'il soit possible « [d']interagir avec les données et rejouer (ou même modifier) les codes utilisés ». Il est ainsi possible pour les lectrices et lecteurs « d'étudier le code, le modifier et le ré-exécuter directement dans le navigateur » pour en tester la reproductibilité puis de télécharger tous les éléments (la publication, les codes et les données) afin de pouvoir les réutiliser.

Un postdoctorant de l'EPFL a créé un site internet pour présenter son projet pNEUMA⁴⁵ qui vise à expliquer les mécanismes de congestion du trafic urbain à l'aide de données massives récoltées par drones (Barmounakis, Geroliminis 2020). Les données y sont déposées en Open data. Un blog intégré au site lui permet de communiquer régulièrement sur le développement de son projet ainsi que sur des projets qui réutilisent ses données.

On peut également remonter à l'exemple du National Center of Competence in Research (NCCR) on the move de l'Université de Neuchâtel qui développe pour chaque projet de recherche une page spécifique sur son site internet. Chaque publication (articles, posters) se rapportant au projet y est signalée, de même qu'un lien sur la page de chaque chercheuse et chercheur collaborant sur le projet ainsi que des liens sur des graphiques et des visualisations interactives permettent d'explorer les données⁴⁶. Les liens sur le codebook ainsi que sur les dépôts de données sont proposés dans l'inventaire des projets uniquement (National Center of Competence in Research 2022).

⁴³ Confirmation reçue de V. Keller, spécialiste ID à la HEG Genève sur le forum ArODES, Cyberlearn, 18.07.2022.

⁴⁴ Lien sur la collection d'articles exécutables de la revue : <https://elifesciences.org/collections/d72819a9/executable-research-articles>

⁴⁵ <https://open-traffic.epfl.ch/>

⁴⁶ Voici un exemple de visualisation : <https://public.tableau.com/app/profile/nccr.on.the.move/viz/Migration-MobilitySurvey-Exploration/Survey>

7. Conclusion

Ce travail de Master réalisé pour l'Institut santé de la HEdS Valais-Wallis a permis une réflexion approfondie sur la réalisation d'un inventaire des données recherche ainsi que sur la gestion des données de recherche, tout en l'insérant dans le contexte et la problématique d'une institution en particulier.

Un inventaire simple et fonctionnel a été créé ainsi qu'un processus permettant d'assurer le suivi administratif et scientifique des données de recherche de l'IS. Cet inventaire a pu être testé et complété pour une sélection de projets de recherche. Il a ainsi été possible de dresser une photographie des données possédées par l'IS ainsi que de la manière dont elles sont gérées. Les entretiens menés aussi bien avec les chercheuses et chercheurs qu'avec les personnes ressources ont également permis de dresser un état des lieux des pratiques. Des propositions pour renforcer le soutien sur la GDR et l'implication des personnes ressources ont été formulées. Un fichier Excel sera utilisé pour tenir à jour l'inventaire, le temps qu'un nouvel outil de gestion des projets de recherche soit choisi. Le cahier des charges proposé dans ce travail sera transmis aux personnes chargées de ce choix afin qu'elles puissent prendre en compte les fonctionnalités et besoins de l'inventaire pour le nouvel outil.

Des solutions de stockage ont ensuite été analysées afin de proposer aux équipes de recherche des solutions adaptées en fonction des spécificités de leurs projets. Un tableau décisionnel permet d'accompagner les équipes de recherche dans ce choix.

Bien que les dépôts de données OLOS et SWISSUbase puissent être utilisés pour l'archivage des données anonymes, non confidentielles et personnelles, OLOS semble plus prometteur étant donné les développements annoncés. Mais pour l'instant, les dépôts de données ouverts ne répondent que partiellement à l'archivage des données. De ce fait, une solution complémentaire doit être trouvée. Les plateformes proposées par les AEV, SWITCH et Docuteam semblent intéressantes mais elles devront être analysées plus en détail afin de s'assurer de leur pertinence, de connaître leur coût en fonction du volume de données pour l'ensemble de la HES-SO Valais-Wallis.

Pour terminer, une proposition de communication a été émise. Elle permettra de valoriser davantage les données de recherche produites au sein de l'Institut santé du début d'un projet jusqu'à son achèvement.

7.1 Difficultés rencontrées

Ce travail était particulièrement intéressant car il a permis, mis à part la conception d'un outil et d'un processus qui pourra être utilisé dans mon institution, de développer mes connaissances et compétences en matière de GDR en les axant plus particulièrement sur le domaine de la santé, ce qui ce sera vraiment utile pour la suite de mon activité professionnelle. Les entretiens avec les collègues de la recherche m'ont permis de mieux comprendre le processus de recherche, leurs pratiques et leurs besoins. Cependant, j'ai eu de la difficulté à concilier mon rôle d'étudiante avec mon nouveau rôle de data manager de l'école. J'ai parfois manqué de neutralité dans mes questions ainsi que dans ma posture. Néanmoins, le fait d'être membre de l'institution, de connaître et de collaborer avec la plupart d'entre elles et eux pour la partie Open Access a permis de travailler dans un climat de confiance et de bienveillance envers ma démarche.

La réalisation de l'inventaire a nécessité plus de temps que prévu. De ce fait, il n'a pas été préalablement testé alors que cela aurait permis de résoudre les bugs de paramétrage. De plus, la compréhension des champs n'était pas toujours correcte malgré le fait d'avoir donné une définition pour chaque champ. Proposer un exemple aurait pu les aider. En cela, j'ai surestimé leurs connaissances. Il était finalement beaucoup plus pertinent et intéressant de remplir ce fichier ensemble et ainsi pouvoir échanger directement. J'ai pu ainsi avoir une meilleure compréhension de leur projet, de l'intérêt ou non des champs, ainsi que du contenu à redéfinir.

La principale autre difficulté rencontrée pour l'inventaire concerne la délimitation du périmètre du travail. J'ai sous-estimé son envergure et il a été nécessaire, en cours de réalisation, de revoir certains objectifs par manque de temps. La problématique étant cruciale pour ma pratique professionnelle, tout me semblait important, et de ce fait j'ai eu du mal à fixer des priorités et à m'y tenir.

La partie sur l'archivage des données de recherche a été complexe à mener. Il est délicat de proposer des recommandations pour les dépôts de données OLOS et SWISSUbase, étant donné la jeunesse de ces solutions et les développements attendus. Il s'agit plutôt d'un préavis. De plus, certaines questions sont restées en suspens, les personnes contactées étant en vacances au moment de la prise de contact. Par manque de temps il n'a pas été possible d'analyser en détail les solutions proposées par les AEV, SWITCH et Docuteam ce qui est vraiment regrettable car cette problématique de l'archivage est une question vraiment importante pour l'institution. Le périmètre de cet objectif aurait dû être revu au départ du mandat pour le cibler sur cette notion.

7.2 Perspectives

Peu d'institutions à la HES-SO réalisent actuellement des inventaires, mais les retours reçus montrent un intérêt pour cette démarche. Si le canevas suggéré dans ce travail ainsi que le processus proposé pouvaient être repris et réadaptés en fonction des besoins de chaque institution, le futur outil gestion de projets de recherche pourrait être une solution encore plus intéressante et qui aurait le mérite d'apporter à l'ensemble des institutions de la HES-SO une solution logicielle à la mise en place d'un inventaire. Comme cela a également été mentionné, l'outil pourrait inclure un développement intégrant la gestion des activités de recherche et notamment des publications archivées sur ArODES.

La photographie des données de l'IS, le déroulement proposé, les synthèses des entretiens avec les chercheuses et chercheurs ainsi qu'avec les personnes ressources pourront être réutilisés afin de développer des processus plus globaux de GDR puis de cibler le soutien sur les éléments les plus importants.

En matière d'archivage des données de recherche, une solution d'archivage à long terme pour l'ensemble de la HES-SO Valais-Wallis doit encore être choisie pour les données qui ne peuvent être archivées sur un dépôt ouvert. Une analyse complémentaire sera nécessaire afin de déterminer la volumétrie concernée pour chaque école. Une fois cette information connue il sera possible de reprendre contact avec les AEV, SWITCH et Docuteam pour déterminer la solution la plus adaptée et obtenir une estimation des coûts.

La question de la communication sur les jeux de données à valoriser n'a malheureusement été que survolée tout comme le partage des données. Étant donné l'importance de ces sujets

et l'intérêt de ces questions pour les institutions, il serait utile de développer ces thématiques dans un futur travail.

La rédaction d'une politique et surtout d'une directive qui aborde les questions de responsabilité, de conservation et d'archivage (durée, critères) est vraiment l'étape la plus pressante car elle permettra de fixer un cadre clair.

Bibliographie

AGNES TUTOS ET FORMATIONS, 2021. Créer une alerte sur Excel avec message de rappel, annotations et mise en forme conditionnelle [enregistrement vidéo]. *YouTube*. [en ligne]. 16 avril 2021. [Consulté le 25 juillet 2022]. Disponible à l'adresse: <https://www.youtube.com/watch?v=ifozUeGU4T8>

ANDRÉ, Francis, 2014. Déluge des données de la recherche ? Petit manuel d'immersion. Curation, infrastructures et partage. In: CALDERAN, Lisette, LAURENT, Pascale, LOWINGER, Hélène et MILLET, Jacques (éd.), *Big data : nouvelles partitions de l'information. Actes du séminaire IST Inria, octobre 2014*. Louvain-la-Neuve: De Boeck Supérieur. pp. 77-95. Information et stratégie. ISBN 978-2-8041-8915-0.

ARCHIVES DE L'ETAT DU VALAIS, [sans date]. Chapitre 7 : Gérer les documents et les archives dans le monde électronique. *Canton du Valais*. [en ligne]. [Consulté le 29 juillet 2022]. Disponible à l'adresse: <https://www.vs.ch/web/culture/guide-administration-chapitre-7>

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS, 2019. ARCHIVER & EOSC, the European Open science cloud. *Archiver-project*. [en ligne]. 6 décembre 2019. [Consulté le 23 juillet 2022]. Disponible à l'adresse: <https://archiver-project.eu/archiver-eosc>

ASSOCIATION SUISSE DES COMMISSIONS D'ÉTHIQUE DE LA RECHERCHE, 2022. *Swissethics*. *swissethics*. [en ligne]. 2022. [Consulté le 20 juillet 2022]. Disponible à l'adresse: <https://swissethics.ch/fr/>

AUSTRALIAN RESEARCH DATA COMMONS, 2019. What is research data? *ARDC*. [en ligne]. 2019. [Consulté le 3 août 2022]. Disponible à l'adresse: <https://ardc.edu.au/wp-content/uploads/2020/01/What-is-research-data.pdf>

BAILLARGEON, Diane, 2017. La catégorisation des actifs informationnels: l'expérience de l'université de Montréal. *Congrès de l'Association des archivistes du Québec*. [en ligne]. 2 juin 2017. [Consulté le 23 novembre 2021]. Disponible à l'adresse: http://congres.archivistes.qc.ca/wp-content/uploads/2017/08/V10_Diane_Baillargeon.pdf

BANAT-BERGER, Françoise et HUB, Claude, 2011. Section 5 - Modèle de référence OAI et normes associées : version 2. *PIAF*. [en ligne]. 22 novembre 2011. [Consulté le 23 juillet 2022]. Disponible à l'adresse: https://www.piaf-archives.org/sites/default/files/bulk_media/m07s05/co/section05_web.html

BARMPOUNAKIS, Emmanouil Manos et GEROLIMINIS, Nikolas, 2020. pNEUMA : open-traffic – Une initiative de données ouvertes. *pNEUMA*. [en ligne]. 2020. [Consulté le 16 juillet 2022]. Disponible à l'adresse: <https://open-traffic.epfl.ch/>

BAUD-VITTOZ, Guilaine et VARRATO, Francesco, 2022a. Gestion des données : ça commence avant d'en avoir. *HES-SO*. [en ligne]. 9 mai 2022. [Consulté le 19 juin 2022]. Disponible à l'adresse: https://www.hes-so.ch/fileadmin/documents/HES-SO/Documents_HES-SO/pdf/open-science/2-EPFL_Library_RDM_presentation_20220509.pdf

BAUD-VITTOZ, Guilaine et VARRATO, Francesco, 2022b. Gestion des données : ça commence avant d'en avoir [enregistrement vidéo]. *YouTube*. [en ligne]. 9 mai 2022. [Consulté le 19 juin 2022]. Disponible à l'adresse: <https://www.youtube.com/watch?v=g7qvUN3Wjsk>

BEAGRIE, Neil, 2019. *What to Keep: A Jisc research data study*. [en ligne]. Bristol: JISC. [Consulté le 22 mai 2022]. Disponible à l'adresse: https://repository.jisc.ac.uk/7262/1/JR0100_WHAT_RESEARCH_DATA_TO_KEEP_FEB2019_v5_WEB.pdf

BEAGRIE, Neil, JONES, Maggie, et DIGITAL PRESERVATION COALITION, 2021. *Manuel de préservation numérique*. [en ligne]. 2e éd. Glasgow: DPC. [Consulté le 22 juillet 2022]. Disponible à l'adresse: <https://www.dpconline.org/docs/digital-preservation-handbook2/2519-handbook-2021-fr/file>

BEALE, Tim, SMITH, Fiona, DODDS, Leigh, L'HENAFF, Pauline et YATES, Deborah, 2021. How to create a data inventory. *Gates Open Research*. [en ligne]. 11 mars 2021. [Consulté le 3 août 2022]. Disponible à l'adresse: <https://doi.org/10.21955/gatesopenres.1116749.1>

BEZJAK, Sonja, VIPAVC BRVAR, Irena, et CESSDA TRAINING TEAM, 2017. 6. Archive & publish. *CESSDA Data Management Expert Guide*. [en ligne]. 2019 2017. [Consulté le 10 juin 2022]. Disponible à l'adresse: <https://dmeq.cessda.eu/Data-Management-Expert-Guide/6.-Archive-Publish>

BEZZI, Manuela, 2020. *Préservation des données de recherche: proposer des services de soutien 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 3 août 2022]. Disponible à l'adresse: <https://sonar.ch/hesso/documents/315066>

BIBLIOTHÈQUE UNIVERSITAIRE CLAUDE BERNARD LYON 1 et BIBLIOTHÈQUE UNIVERSITAIRE GRENOBLE ALPES, [sans date]. Publier un data paper : où et comment ? *DATAACC*. [en ligne]. [Consulté le 16 juillet 2022]. Disponible à l'adresse: <https://www.dataacc.org/vos-besoins/valoriser-ses-donnees/data-papers/>

BISCHOFF, Sébastien, MELLY, Pauline, RACINE, Céline et TROMBERT, Alexia, 2021. *Conception d'un registre d'actifs informationnels pour le réseau hospitalier neuchâtelois (RHNe)* [document PDF]. Travail réalisé à la HEG-GE, dans le cadre du cours M7c Gouvernance des données, donné au semestre d'automne par Mme B. Makhoul-Shabou

BLANCHET, Alain, 2015. *L'entretien*. 2e éd. Paris: A. Colin. Tout le savoir en 128 pages. ISBN 978-2-200-60191-1.

BONNEVILLE, Anne, TUCCI, Ingrid, VION, Antoine et GIGLIO, Laurent, 2021. Données de la recherche : pratiques et besoins dans un laboratoire pluridisciplinaire SHS : rapport final. *HAL*. [en ligne]. mai 2021. [Consulté le 6 décembre 2021]. Disponible à l'adresse: <https://hal.archives-ouvertes.fr/hal-03265603>

BRACCO, Laetitia, CHEVIRON, Stéphanie, DONATI, Caroline-Sophie, DU PASQUIER, Delphine et MARTIN, France, 2021. Comment sensibiliser les chercheur.e.s en SHS à la gestion des données ? *Zenodo*. [en ligne]. 22 mars 2021. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.5281/zenodo.4639590>

BYRNE, Meg, 2017. Making progress toward open data: reflections on data sharing at Plos One. *EveryONE*. [en ligne]. 8 mai 2017. [Consulté le 3 février 2022]. Disponible à l'adresse: <https://everyone.plos.org/2017/05/08/making-progress-toward-open-data/>

CARTIER, A, MOYSAN, M et REYMONET, N, 2015. Réaliser un plan de gestion de données. *BU Paris Descartes*. [en ligne]. 9 janvier 2015. [Consulté le 3 août 2022]. Disponible à l'adresse: http://www.bu.parisdescartes.fr/doc/documents/Realiser_un_DMP_V1.pdf

CENTRE DE COOPÉRATION INTERNATIONALE EN RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT, 2021. Unités de recherche. *CIRAD*. [en ligne]. 16 juin 2021. [Consulté le 21 avril 2022]. Disponible à l'adresse: <https://www.cirad.fr/nous-connaître/unites-de-recherche>

CENTRE DE COOPÉRATION INTERNATIONALE EN RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT, 2022. Le Cirad en bref. *CIRAD*. [en ligne]. 24 mars 2022. [Consulté le 21 avril 2022]. Disponible à l'adresse: <https://www.cirad.fr/nous-connaître/le-cirad-en-bref>

CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR, 2016. L'archivage numérique, qu'est-ce que c'est ? *CINES*. [en ligne]. 19 septembre 2016. [Consulté le 12 juillet 2022]. Disponible à l'adresse: <https://www.cines.fr/archivage/un-concept-des-problematiques/>

CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR, 2019. Le modèle de référence : l'OAIS. *CINES*. [en ligne]. 3 avril 2019. [Consulté le 23 juillet 2022]. Disponible à l'adresse: <https://www.cines.fr/archivage/un-concept-des-problematiques/le-modele-de-reference-loais/>

CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR, 2020a. Le concept d'archivage numérique pérenne. *CINES*. [en ligne]. 6 janvier 2020. [Consulté le 23 juillet 2022]. Disponible à l'adresse: <https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/>

CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR, 2020b. Archivage. *CINES*. [en ligne]. 7 mai 2020. [Consulté le 23 juillet 2022]. Disponible à l'adresse: <https://www.cines.fr/archivage/>

CHAWINGA, Winner Dominic et ZINN, Sandy, 2020. Research data management at an African medical university: Implications for academic librarianship. *The Journal of Academic Librarianship*. 1 juillet 2020. Vol. 46, no. 4, pp. 102161. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.1016/j.acalib.2020.102161>. [accès par abonnement]

COLAVIZZA, Giovanni, HRYNASZKIEWICZ, Iain, STADEN, Isla, WHITAKER, Kirstie et MCGILLIVRAY, Barbara, 2020. The citation advantage of linking publications to research data. *PLOS ONE*. 22 avril 2020. Vol. 15, no. 4, pp. e0230416. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.48550/arXiv.1907.02565>

COMITÉ CONSULTATIF POUR LES SYSTÈMES DE DONNÉES SPATIALES, 2017. *Modèle de référence pour un Système ouvert d'archivage d'information (OAIS) : pratique recommandée CCSDS 650.0-M-2 (F)* [en ligne]. Washington D. C.: CCSDS. [Consulté le 11 juin 2022]. Disponible à l'adresse: <https://public.ccsds.org/Pubs/650x0m2%28F%29.pdf>

COMMISSION CANTONALE D'ÉTHIQUE DE LA RECHERCHE SUR L'ÊTRE HUMAIN, 2022. Commission cantonale d'éthique de la recherche sur l'être humain. *CER-VD*. [en ligne]. 14 juillet 2022. [Consulté le 20 juillet 2022]. Disponible à l'adresse: <https://www.cer-vd.ch>

CONFEDERATION OF OPEN ACCESS REPOSITORIES, 2020. Cadre commun COAR de bonnes pratiques en matière d'entrepôts. *Zenodo*. [en ligne]. 6 novembre 2020. [Consulté le 7 juillet 2022]. Disponible à l'adresse: <https://doi.org/10.5281/zenodo.4118380>

CONSTANTINIDIS, Yves, 2015. *Expression des besoins pour le SI: guide d'élaboration du cahier des charges*. 3e éd. Paris: Eyrolles. Solutions d'entreprise. ISBN 978-2-212-14331-7.

- CONSTANTINIDIS, Yves, 2020. *Cahier des charges informatique*. [en ligne]. 4e éd. Paris: Eyrolles. [Consulté le 7 mai 2022]. Mémento. ISBN 978-2-212-67988-5. Disponible à l'adresse: <https://hesge.scholarvox.com/book/88902881>. [accès par abonnement]
- COTTING, Alexandre, CALBIMONTE, Jean-Paul et SCHUMACHER, Michael, 2020. *Projet de création d'un Service de gestion des données de recherche HES-SO Valais-Wallis*. [fichier Microsoft Word]. Document interne à la HES-SO Valais-Wallis
- COUTURE, Carol, 1994. *Les fondements de la discipline archivistique*. [en ligne]. Québec: Presses de l'Université du Québec. [Consulté le 21 novembre 2021]. ISBN 978-2-7605-0781-4. Disponible à l'adresse: <https://hesge.scholarvox.com/catalog/book/docid/88801141>. [accès par abonnement]
- COUTURE, Carol, 1996. L'évaluation des archives: état de la question. *Archives*. 1997 1996. Vol. 28, no. 1, pp. 3-31. [Consulté le 9 août 2022]. Disponible à l'adresse: https://www.archivistes.qc.ca/revuearchives/vol28_1/28-1-couture.pdf
- DA SYLVA, Lyne, MAUREL, Dominique, BRUYÈRE, Marc, SAINT-GERMAIN, Marielle et GAREAU, Geneviève, 2019. Écosystème de la gestion de données de recherche et professionnels de l'information : présentation des enjeux, de la méthodologie et des solutions préconisées d'une enquête canadienne. *Études de communication. langages, information, médiations*. 1 juin 2019. No. 52, pp. 51-70. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.4000/edc.8615>
- D'ANGELO, Marc, 2020. Directives sur l'appariement (version 1.2; 05.03.2020). *OFS*. [en ligne]. 18 juin 2020. [Consulté le 18 mai 2022]. Disponible à l'adresse: <https://www.bfs.admin.ch/asset/fr/2760553>
- DEDIEU, Laurence et BARALE, Martine, 2020. Déposer des données de recherche dans un entrepôt. *Cirad*. [en ligne]. 2020. [Consulté le 9 avril 2022]. Disponible à l'adresse: <https://doi.org/10.18167/coopist/0070>
- Définition, 2021. *SageX Wiki*. [en ligne]. [Consulté le 27 mai 2022]. Disponible à l'adresse: <https://agp-wiki.hes-so.ch/xwiki/bin/view/Aide%20m%C3%A9tier/D%C3%A9finition/>
- DELAFONTAINE; Serge. 2020. *Schémas pour proposition de structure*. [Document Microsoft Visio]. Document interne à la HES-SO Valais-Wallis.
- DÉLÉGATION OPEN SCIENCE, GROUPE DE PROJET ORD, et SWISSUNIVERSITIES, 2021. Stratégie nationale suisse Open research data. *Swissuniversities*. [en ligne]. 23 avril 2021. [Consulté le 9 octobre 2021]. Disponible à l'adresse: https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/Swiss_National_ORD_Strategy_fr.pdf
- DENNIE, Danielle et GUINDON, Alex, 2017. Résultats d'une enquête sur les pratiques et attitudes des chercheurs de l'Université Concordia en matière de gestion des données de recherche. *Documentation et bibliothèques*. 5 décembre 2017. Vol. 63, no. 4, pp. 59-72. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.7202/1042311ar>
- DIAZ, Pablo, HEERS, Marieke, STAM, Alexandra, et FORS DATA ARCHIVE, 2022. *Top 10 data management questions* [présentation PowerPoint]. Formation organisée par FORS les 20-21 juin 2022
- DILLAERTS, Hans, 2016. Qu'est-ce qu'une donnée de santé ? *DLIS*. [en ligne]. 15 septembre 2016. [Consulté le 3 mars 2022]. Disponible à l'adresse: <https://dlis.hypotheses.org/704>

DILLAERTS, Hans, PAGANELLI, Céline, VERLAET, Lise et CATHERINE, Hugo, 2020. Usages et pratiques en lien avec les données de recherche. Une enquête menée auprès des chercheurs de l'université Paul-Valéry Montpellier 3. *HAL*. [en ligne]. juillet 2020. [Consulté le 8 mai 2022]. Disponible à l'adresse: <https://halshs.archives-ouvertes.fr/halshs-02902710>

DOCUTEAM, [sans date]. Notre entreprise et notre réseau professionnel. *Docuteam*. [en ligne]. [Consulté le 28 juillet 2022]. Disponible à l'adresse: <https://www.docuteam.ch/fr/entreprise/>

DOCUTEAM, 2022. Docuteam cosmos : solution cloud pour l'archivage numérique à long terme : version 3.0. *Docuteam*. [en ligne]. 2022. [Consulté le 28 juillet 2022]. Disponible à l'adresse: https://www.docuteam.ch/wp-content/uploads/2022/06/Produktbeschreibung_cosmos_fr.pdf

DONATI, Caroline Sophie, 2019. Données de la recherche : quelles pratiques ? Quels besoins ? Enquête à Aix-Marseille Université. *HAL*. [en ligne]. 2019. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://hal-amu.archives-ouvertes.fr/hal-02493679>

DUPRAT, Julie, 2019. Les données de la recherche à l'Université Bordeaux Montaigne. *HAL*. [en ligne]. février 2019. [Consulté le 6 décembre 2021]. Disponible à l'adresse: <https://hal.archives-ouvertes.fr/hal-02020141>

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, 2021. ACOUA : une nouvelle archive pour préserver les données de recherche. *EPFL*. [en ligne]. 8 février 2021. [Consulté le 8 juillet 2022]. Disponible à l'adresse: <https://actu.epfl.ch/news/acoua-une-nouvelle-archive-pour-preserver-les-donn/>

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE. BIBLIOTHÈQUE, 2022. Store vs back up vs archive research data [enregistrement vidéo]. *YouTube*. [en ligne]. 14 février 2022. [Consulté le 13 mai 2022]. Disponible à l'adresse: <https://www.youtube.com/watch?v=ifLJPs4cFFM>

EHRENSTRÖM, Philippe, 2021. Les principes en droit suisse de la protection des données. *Le droit du travail et de la protection des données en Suisse*. [en ligne]. 1 août 2021. [Consulté le 11 juillet 2022]. Disponible à l'adresse: <https://droitdutravailensuisse.com/2021/08/01/les-principes-en-droit-suisse-de-la-protection-des-donnees/>

EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE ZÜRICH, 2022. Guidelines for research data management at ETH Zurich (RDM guidelines). *Rechtssammlung der ETH Zürich*. [en ligne]. 1 juillet 2022. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://rechtssammlung.sp.ethz.ch/Dokumente/414.2en.pdf>

EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE ZÜRICH. BIBLIOTHEK, 2022. About Research Collection. *Research Collection*. [en ligne]. 2022. [Consulté le 23 juillet 2022]. Disponible à l'adresse: <https://www.research-collection.ethz.ch/about>

ÉQUIPE INTERMINISTÉRIELLE DU PROGRAMME VITAM, [sans date]. Focus archivistes. *Programme Vitam*. [en ligne]. [Consulté le 23 juillet 2022 a]. Disponible à l'adresse: https://www.programmevitam.fr/pages/presentation/pres_archivistes/

ÉQUIPE INTERMINISTÉRIELLE DU PROGRAMME VITAM, [sans date]. Le programme. *Programme Vitam*. [en ligne]. [Consulté le 23 juillet 2022 b]. Disponible à l'adresse: <https://www.programmevitam.fr/pages/presentation/>

EXTENDOFFICE, 2022. Comment envoyer automatiquement un e-mail en fonction de la valeur de la cellule dans Excel? *ExtendOffice*. [en ligne]. 2022. [Consulté le 25 juillet 2022]. Disponible à l'adresse: <https://fr.extendoffice.com/documents/excel/4656-excel-send-email-based-on-cell-value.html>

FONDS NATIONAL SUISSE, 2015. Règlement d'exécution général relatif au règlement des subsides. *FNS*. [en ligne]. 9 décembre 2015. [Consulté le 22 juillet 2022]. Disponible à l'adresse: <https://www.snf.ch/fr/Zi65otScMDqzyALN/page/encouragement/documents-telechargements/undefined/fr/Zi65otScMDqzyALN/page/>

FONDS NATIONAL SUISSE, 2017. Data management plan : contenu du formulaire mySNF. *FNS*. [en ligne]. 26 juin 2017. [Consulté le 13 mai 2022]. Disponible à l'adresse: https://www.snf.ch/media/fr/4i9AE5YEIf7tqhGz/DMP_content_mySNF-form_fr.pdf

FONDS NATIONAL SUISSE, 2022a. Le FNS rejoint la cOAlition S pour le libre accès immédiat aux articles scientifiques. *FNS*. [en ligne]. 1 juin 2022. [Consulté le 19 juillet 2022]. Disponible à l'adresse: <https://www.snf.ch/fr/sOggLwEnxtgZsCD8/news/undefined/fr/sOggLwEnxtgZsCD8/news/>

FONDS NATIONAL SUISSE, 2022b. Open research data : FAQ. *FNS*. [en ligne]. 2022. [Consulté le 7 juillet 2022]. Disponible à l'adresse: <https://www.snf.ch/fr/dMILj9t4LNk8NwyR/dossier/undefined/fr/dMILj9t4LNk8NwyR/dossier/>

FORS, [sans date]. SWISSUBase. *Swissuniversities*. [en ligne]. [Consulté le 21 juillet 2022]. Disponible à l'adresse: https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Organisation/SUK-P/SUK_P-2/DOC_SWISSUBase_final.pdf

FORS, 2018. Data management. *FORS Center*. [en ligne]. 19 juin 2018. [Consulté le 10 mars 2022]. Disponible à l'adresse: <https://forscenter.ch/data-services/data-management/?lang=fr>

FORS, 2021. Déposer des données. *FORS Center*. [en ligne]. 14 décembre 2021. [Consulté le 3 août 2022]. Disponible à l'adresse: <https://forscenter.ch/deposer-des-donnees/?lang=fr>

FORS, 2022. SWISSUbase. *SWISSUbase*. [en ligne]. 2022. [Consulté le 21 juillet 2022]. Disponible à l'adresse: <https://www.swissubase.ch/fr/>

FORS DATA SERVICE, 2022a. Acquisition policy : version 5.1. *FORS Center*. [en ligne]. juillet 2022. Disponible à l'adresse: https://forscenter.ch/wp-content/uploads/2022/07/acquisition-policy_2022.pdf

FORS DATA SERVICE, 2022b. Preservation policy : version 4.1. *FORS Center*. [en ligne]. juillet 2022. Disponible à l'adresse: https://forscenter.ch/wp-content/uploads/2022/07/preservation-policy_2022.pdf

FORTUNO, Sophie, 2015a. L'inventaire de jeux de données scientifiques: une démarche d'aide à la décision vers l'ouverture de données [présentation PowerPoint]. *FOSTER*. [en ligne]. 2015a. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://www.fosteropenscience.eu/sites/default/files/pdf/2021.pdf>

FORTUNO, Sophie, 2015b. Guide pour l'inventaire et la saisie des données numériques scientifiques du Cirad. *Cirad*. [en ligne]. 2015b. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://agritrop.cirad.fr/577459/>. [accès restreint]

FORTUNO, Sophie, 2018. Dataverse : guide utilisateur V1. *Cirad*. [en ligne]. 2018. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://agritrop.cirad.fr/587424/1/2018-02-Dataverse-Guide-utilisateur-V1.1.pdf>

FRESSENGEAS, Nicolas, 2020. Les réseaux et la visibilité de la recherche: de la science ouverte aux réseaux sociaux par la pratique. *HAL*. [en ligne]. 2020. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://hal.univ-lorraine.fr/hal-02501151>

GAILLARD, Rémi, 2014. *De l'Open data à l'Open research data: quelle (s) politique (s) pour les données de recherche?* [en ligne]. Villeurbanne: Université de Lyon. Enssib. Mémoire d'études. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf>

GALONNIER, Juliette, LE COURANT, Stefan, PECQUEUX, Anthony et NOÛS, Camille, 2019. Ouvrir les données de la recherche ? *Tracés*. 31 décembre 2019. No. 19, pp. 17-33. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.4000/traces.10588>

GASTL-HARTMANN, Angela, 2022. Forschen für die Ewigkeit: Bewertung und Archivierung von Forschungsdaten als neues Arbeitsgebiet für Hochschularchive. *Informationswissenschaft: Theorie, Methode und Praxis*. 22 juin 2022. Vol. 7, no. 1, pp. 396-429. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.18755/iw.2022.22>

GRANT, Rebecca, 2017. Recordkeeping and research data management: a review of perspectives. *Records Management Journal*. 17 juillet 2017. Vol. 27. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.1108/RMJ-10-2016-0036> [accès par abonnement]

GROUPE DE TRAVAIL RÉUTILISABILITÉ, CAHIER CONSORTIUM, AUCAGNE, Julie, BORDRY, Marguerite, DESILES, Camille, FILOCHE, Francine, GARCIA-FERNANDEZ, Anne, ELISABETH, Greslou, KOSKAS, Camille, PATAT, Gwenaëlle, WALTER, Richard et WILLAIME, Pierre, 2022. Vademecum pour la réutilisabilité des données. *HAL*. [en ligne]. 2022. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://hal.archives-ouvertes.fr/hal-03698675/document>

GROUPE DE TRAVAIL SUR LA COVID-19 DU RÉSEAU PORTAGE, THOMPSON, Kristi, CLARY, Erin, COSTANZO, Lucia, KNAZOOK, Beth, ROCHLIN, Nick, TAYLER, Felicity, FRY, Jane, RIPP, Chantal, SZIGETI, Kathy, ZHANG, Qian, REKA, Roger, WANG, Minglu, DICKSON, Rebecca, LEGGOTT, Mark et PARLETTE-STEWART, Melanie, 2020. Directives sur la dépersonnalisation des données. *Zenodo*. [en ligne]. 23 septembre 2020. [Consulté le 22 mai 2022]. Disponible à l'adresse: <https://doi.org/10.5281/zenodo.4042022>

GROUPE D'EXPERTS SUR LES DONNÉES SENSIBLES, 2020. Boîte à outils pour les données sensibles destiné aux chercheurs : partie 1: glossaire terminologique sur l'utilisation des données sensibles à des fins de recherche. *Zenodo*. [en ligne]. 14 octobre 2020. [Consulté le 22 mai 2022]. Disponible à l'adresse: <https://zenodo.org/record/4088986>

GUIRLET, Marielle, 2020. *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. Travail de Master. [Consulté le 4 mars 2022]. Disponible à l'adresse: <https://doc.rero.ch/record/329696/files/Guirlet-Moire-Vdef.pdf>

HADROSSEK, Christine, JANIK, Joanna, LIBES, Maurice, LOUVET, Violaine, QUIDOZ, Marie-Claude, RIVET, Alain et ROMIER, Geneviève, 2021. Guide de bonnes pratiques sur la gestion des données de la recherche. *HAL*. [en ligne]. février 2021.

[Consulté le 24 mai 2022]. Disponible à l'adresse: <https://hal.archives-ouvertes.fr/hal-03152732>

HAUTE ÉCOLE SPÉCIALISÉE DE SUISSE OCCIDENTALE, 2014a. Règlement sur la valorisation des connaissances acquises par la recherche au sein de la HES-SO. *HES-SO*. [en ligne]. 11 décembre 2014. [Consulté le 21 novembre 2021]. Disponible à l'adresse: <https://www.hes-so.ch/data/documents/Reglement-valorisation-connaissances-acquises-par-la-recherche-4516.pdf>

HAUTE ÉCOLE SPÉCIALISÉE DE SUISSE OCCIDENTALE, 2014b. *Directives de traitement pour AGP HES-SO* [document PDF]. Document interne à la HES-SO

HAUTE ÉCOLE SPÉCIALISÉE DE SUISSE OCCIDENTALE, 2018. Stratégie Open HES-SO. *Open HES-SO*. [en ligne]. 2018. [Consulté le 8 mai 2021]. Disponible à l'adresse: <https://openscience.hes-so.ch/data/documents/2018-Strategie-Open-HES-SO-12321.pdf>

HAUTE ÉCOLE SPÉCIALISÉE DE SUISSE OCCIDENTALE, 2021. Archiver ses données de recherche. *HES-SO*. [en ligne]. 2021. [Consulté le 8 mai 2022]. Disponible à l'adresse: <https://www.hes-so.ch/recherche-innovation/open-science/open-data/archiver-ses-donnees-de-recherche>

HEINTZ, Monica et GAUTIER, Helene, 2022b. Politique des données basée sur un inventaire des jeux de données d'un laboratoire [enregistrement vidéo]. *Canal U*. [en ligne]. 2022b. [Consulté le 14 avril 2022]. Disponible à l'adresse: <https://www.canal-u.tv/chaines/renatis/cfe-renatis-politique-des-donnees-basee-sur-un-inventaire-des-jeux-de-donnees-d-un>

HEINTZ, Monica et GAUTIER, Helene, 2022a. Retour d'expérience sur la mise en place d'un inventaire et d'une politique de données. *Canal U*. [en ligne]. 2022a. Disponible à l'adresse: https://www.canal-u.tv/sites/default/files/medias/fichiers/2022/04/20220407_Cafe_Renatis.pdf

HES-SO FRIBOURG, SERVICE INFORMATIQUE, 2020. Microsoft OneDrive. *Sinfo*. [en ligne]. 2020. [Consulté le 27 mai 2022]. Disponible à l'adresse: <https://confluence.hefr.ch/display/public/ORGSIFR/Microsoft+OneDrive>

HES-SO VALAIS-WALLIS, [sans date]. Mini guide pour OneDrive Enterprise. *Cyberlearn HES-SO*. [en ligne]. [Consulté le 27 mai 2022]. Disponible à l'adresse: https://cyberlearn.hes-so.ch/pluginfile.php/3676677/mod_resource/content/1/Mini-guide%20OneDrive%20V1.pdf

HES-SO VALAIS-WALLIS, 2021a. Collaborateurs·trices. *hevs.ch*. [en ligne]. 2021. [Consulté le 18 avril 2022]. Disponible à l'adresse: <https://www.hevs.ch/fr/recherche-appliquee-et-developpement-a-la-hes-so-valais-wallis/institut-sante-valais/collaborateurs/>

HES-SO VALAIS-WALLIS, 2021b. The Sense. *hevs.ch*. [en ligne]. 2021. [Consulté le 13 juillet 2022]. Disponible à l'adresse: <https://www.hevs.ch/fr/mini-sites/projets-produits/the-sense-201672>

HES-SO VALAIS-WALLIS, 2022. Organigramme. *hevs.ch*. [en ligne]. 2022. [Consulté le 18 avril 2022]. Disponible à l'adresse: <https://intranet.hevs.ch/src/SQ/getDoc.aspx?ID=3648>

HES-SO VALAIS-WALLIS. MÉDIATHÈQUE SANTÉ, 2019. Procédure d'archivage et de valorisation des activités de recherche. *hevs.ch*. [en ligne]. 2019. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://www.hevs.ch/media/document/20/procedure-arodes-version-definitive.pdf>

HES-SO VALAIS-WALLIS. MÉDIATHÈQUE SANTÉ, 2022. Ressources et services. *hevs.ch*. [en ligne]. 2022. [Consulté le 10 février 2022]. Disponible à l'adresse: <https://www.hevs.ch/fr/mediatheques-hes-so-valais-wallis/rad-et-publications/>

HES-SO VALAIS-WALLIS, RESEARCH DATA SERVICE NETWORK, 2021. *Projet stratégique DG «Research Data Service Network - RDSN» et «data managers»* [présentation PowerPoint]. Documentation interne à la HES-SO Valais-Wallis

HIGH EDUCATION FUNDING COUNCIL FOR ENGLAND, RESEARCH COUNCILS UK, UNIVERSITIES UK, et WELLCOME, 2016. Concordat on Open Research Data. *UKRI*. [en ligne]. 28 août 2016. [Consulté le 26 mai 2022]. Disponible à l'adresse: <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>

HUMANITIES ADVANCED TECHNOLOGY AND INFORMATION INSTITUTE, UNIVERSITY OF GLASGOW et DIGITAL CURATION CENTER, 2009. Data asset framework : implementation guide. *data-audit.eu*. [en ligne]. 2009. [Consulté le 14 novembre 2021]. Disponible à l'adresse: https://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf

IFLA BIG DATA SPECIAL INTEREST GROUP, 2018. A concept data science framework for libraries. *IFLA*. [en ligne]. 2018. [Consulté le 10 janvier 2022]. Disponible à l'adresse: https://www.ifla.org/wp-content/uploads/2019/05/assets/big-data/publications/a_concept_data_science_framework_for_libraries.pdf

INSTITUT DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE et CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, 2021. Le cycle de vie des données de recherche. *DoRANum*. [en ligne]. 4 février 2021. [Consulté le 11 juillet 2022]. Disponible à l'adresse: https://doranum.fr/enjeux-benefices/le-cycle-de-vie-des-donnees-de-recherche_10_13143_gzi2-i593/

INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT, 2021. Définitions. *IRD data*. [en ligne]. 2021. [Consulté le 21 novembre 2021]. Disponible à l'adresse: <https://data.ird.fr/definitions/>

INSTITUT PASTEUR, 2022. Les articles exécutables. *Open science : évolutions, enjeux et pratiques*. [en ligne]. 13 mai 2022. [Consulté le 14 juin 2022]. Disponible à l'adresse: <https://openscience.pasteur.fr/2022/05/13/les-articles-executables/>

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH et UNIVERSITY OF MICHIGAN, 2022. Details on appraisal criteria. *ICPSR*. [en ligne]. 2022. [Consulté le 25 mars 2022]. Disponible à l'adresse: <https://www.icpsr.umich.edu/web/pages/datamanagement/lifecycle/details.html>

IONOS, 2022. Alternatives à Microsoft Access : existe-il des programmes gratuits qui proposent les mêmes fonctionnalités ? *IONOS Digitalguide*. [en ligne]. 9 mai 2022. [Consulté le 27 mai 2022]. Disponible à l'adresse: <https://www.ionos.fr/digitalguide/serveur/outils/aperçu-des-alternatives-a-microsoft-access/>

JOHNS HOPKINS UNIVERSITY. CENTER FOR GOVERNMENT EXCELLENCE, 2019. Data inventory guide. *GovEx Labs*. [en ligne]. 8 mars 2019. [Consulté le 25 mars 2022]. Disponible à l'adresse: <https://labs.centerforgov.org/data-governance/data-inventory/#%3A%7E%3Atext%3DA%20data%20inventory%20is%20a%2Csource%2C%20and%20other%20relevant%20details>

JONES, Sarah, BALL, Alexander et EKMEKCIOGLU, Çuna, 2008. The data audit framework: a first step in the data management challenge. *International Journal of Digital*

Curation. 2 décembre 2008. Vol. 3, no. 2, pp. 112-120. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.2218/ijdc.v3i2.62>

JONES, Sarah, ROSS, Seamus et RUUSALEPP, Raivo, 2008. The data audit framework: a toolkit to identify research assets and improve data management in research led institutions. In: *5th International IPRES Conference, 29-30.09.2008*. [en ligne]. London: British Library. 2008. pp. 213-219. [Consulté le 8 août 2022]. Disponible à l'adresse: <http://www.bl.uk/ipres2008/ipres2008-proceedings.pdf>

JONES, Sarah, RUUSALEPP, Raivo et ROSS, Seamus, 2009. Data audit framework: methodology : draft for discussion. *data-audit.eu*. [en ligne]. 26 mai 2009. [Consulté le 24 octobre 2021]. Disponible à l'adresse: https://www.data-audit.eu/DAF_Methodology.pdf

KRUG, Silas et NOIRJEAN, Aimée, 2022. Atelier sur l'utilisation du dépôt de données OLOS [enregistrement vidéo]. *YouTube*. [en ligne]. 14 juin 2022. [Consulté le 12 juillet 2022]. Disponible à l'adresse: https://www.youtube.com/watch?v=JqEkz_SVcLM

KUNG, Janice Yu Chen et CAMPBELL, Sandy, 2016. What not to keep: not all data has future research value. *Journal of the Canadian Health Libraries Association*. [en ligne]. 7 août 2016. Vol. 37, no. 2. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.5596/c16-013>

LABORATOIRE D'ETHNOLOGIE ET DE SOCIOLOGIE COMPARATIVE, [sans date]. Lesc : présentation. *lesc-cnrs.fr*. [en ligne]. [Consulté le 24 avril 2022]. Disponible à l'adresse: <https://lesc-cnrs.fr/fr/laboratoire/lesc>

LAI, Paolo, 2019. Cycle de vie des données [enregistrement vidéo]. *POD UGA*. [en ligne]. 20 novembre 2019. [Consulté le 11 juin 2022]. Disponible à l'adresse: <https://videos.univ-grenoble-alpes.fr/video/19538-cycle-de-vie-des-donnees/>

LAPLANTE, Catherine, 2019. Les données de recherche : comment évaluer pour mieux conserver? *Papyrus*. [en ligne]. 2019. [Consulté le 6 août 2022]. Disponible à l'adresse: https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/22340/laplane_catherine_2019_donnees_recherche_comment_evaluer.pdf?sequence=1&isAllowed=y

LIN, Dawei, CRABTREE, Jonathan, DILLO, Ingrid, DOWNS, Robert R., EDMUNDS, Rorie, GIARETTA, David, DE GIUSTI, Marisa, L'HOURS, Hervé, HUGO, Wim, JENKYNS, Reyna, KHODIYAR, Varsha, MARTONE, Maryann E., MOKRANE, Mustapha, NAVALE, Vivek, PETERS, Jonathan, SIEMAN, Barbara, SOKOLOVA, Dina V., STOCKHAUSE, Martina et WESTBROOK, John, 2020. The TRUST Principles for digital repositories. *Scientific Data*. 14 mai 2020. Vol. 7, no. 1, pp. 144. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.1038/s41597-020-0486-7>

Loi cantonale valaisanne sur la statistique publique cantonale du 11 mars 2022 : référendum (LStat ; Réf.-2022-005). *lex.vs.ch*. [en ligne]. 11 mars 2022. [Consulté le 8 août 2022]. Disponible à l'adresse: <https://lex.vs.ch/frontend/materials/658?locale=fr>

Loi fédérale du 9 octobre 1992 sur la statistique (= LSF; RS 431.01). *Fedlex*. [en ligne]. 09 octobre 1992. Mise à jour le 01 janvier 2016. [Consulté le 18 mai 2022]. Disponible à l'adresse: https://www.fedlex.admin.ch/eli/cc/1993/2080_2080_2080/fr

Loi fédérale du 19 juin 1992 sur la protection des données (LPD ; RS 235.1). *Fedlex*. [en ligne]. 19 juin 1992. Mise à jour le 01 mars 2019. [Consulté le 3 mars 2022]. Disponible à l'adresse: https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/fr#a3

Loi fédérale du 30 mars 1911 complétant le code civil suisse (Livre cinquième: Droit des obligations) (CO ; RS 220). *Fedlex*. [en ligne]. 30 mars 1911. Mise à jour le 01 janvier 2022. [Consulté le 17 juillet 2022]. Disponible à l'adresse: https://www.fedlex.admin.ch/eli/cc/27/317_321_377/fr

Loi fédérale du 30 septembre 2011 relative à la recherche sur l'être humain (Loi relative à la recherche sur l'être humain), (LRH ; RS 810.30). *Fedlex*. [en ligne]. 30 septembre 2011. Mise à jour le 26 mai 2021. [Consulté le 3 mars 2022]. Disponible à l'adresse: <https://www.fedlex.admin.ch/eli/cc/2013/617/fr>

LORENZETTI, Franco, 2022. Aspects juridiques et éthiques de l'Open data [enregistrement vidéo]. *YouTube*. [en ligne]. 14 juin 2022. [Consulté le 10 juillet 2022]. Disponible à l'adresse: <https://www.youtube.com/watch?v=l4Qi8jOQC8Q>

LORENZETTI, Franco, sans date. Open data : notions de base. *hevs.ch* [en ligne]. Sans date. [Consulté le 25 août 2022]. Disponible à l'adresse: <https://www.hevs.ch/fr/mediatheques-hes-so-valais-wallis/rad-et-publications/open-data-202382>

MADAY, Charlotte, 2015. L'apport de la gestion des documents d'activité (records management) à l'ouverture des données. Réflexions basées sur les pratiques en France. *Les cahiers du numérique*. 30 juin 2015. Vol. 11, no. 2, pp. 149-166. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.3166/lcn.11.2.149-166>

MADAY, Charlotte, 2016. OAIS - ISO 14721 : forces, faiblesses, opportunités, menaces. *Spark Archives*. [en ligne]. 2016. [Consulté le 10 mai 2022]. Disponible à l'adresse: <https://www.spark-archives.com/fr/OAIS-ISO14721-Analyse-SWOT>

MADAY, Charlotte et MOYSAN, Magalie, 2014. Records management for scientific data. *Archives and Manuscripts*. 4 mai 2014. Vol. 42, no. 2, pp. 190-192. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.1080/01576895.2014.911686>

MATHIEUJALEXHACHE, 2021. Modèle fonctionnel de OAIS. *Wikimedia Commons*. [en ligne]. 15 janvier 2021. [Consulté le 23 juillet 2022]. Disponible à l'adresse: [https://commons.wikimedia.org/wiki/File:OAIS_Functional_Model_\(fr\).svg](https://commons.wikimedia.org/wiki/File:OAIS_Functional_Model_(fr).svg)

MELLY, Pauline, 2021. *Identification et traçabilité des données de recherche : nature et dispositifs* [document PDF]. Travail réalisé à la HEG-GE, dans le cadre du cours M7c Gouvernance des données, donné au semestre d'automne par Mme B. Makhoul-Shabou

MICROSOFT, 2022c. Partager des fichiers et des dossiers OneDrive. *Support Microsoft*. [en ligne]. 2022c. [Consulté le 28 mai 2022]. Disponible à l'adresse: <https://support.microsoft.com/fr-fr/office/partager-des-fichiers-et-des-dossiers-onedrive-9fcc2f7d-de0c-4cec-93b0-a82024800c07>

MICROSOFT, 2022a. Microsoft 365 et Office : Ressources et informations produits. *Support Microsoft*. [en ligne]. 2022. [Consulté le 7 mai 2022]. Disponible à l'adresse: <https://www.microsoft.com/fr-fr/microsoft-365/microsoft-365-and-office-resources>

MICROSOFT, 2022b. Microsoft Access : logiciels et applications de base de données |. *Support Microsoft*. [en ligne]. 2022. [Consulté le 27 mai 2022]. Disponible à l'adresse: <https://www.microsoft.com/fr-ch/microsoft-365/access>

Microsoft Access, 2022. *Wikipédia*. [en ligne]. [Consulté le 27 mai 2022]. Disponible à l'adresse: https://fr.wikipedia.org/w/index.php?title=Microsoft_Access&oldid=191944571Page

Microsoft Excel, 2022. *Wikipédia*. [en ligne]. [Consulté le 7 mai 2022]. Disponible à l'adresse: https://fr.wikipedia.org/w/index.php?title=Microsoft_Excel&oldid=193216270Page

MONDAINI, Alessandro, 2021. *Vers un data warehouse pour des datasets en santé*. Lausanne: HES-SO Master. Travail de Master.

NATIONAL CENTER OF COMPETENCE IN RESEACH, 2022. Inventory of projects. *NCCR - on the move*. [en ligne]. 5 juillet 2022. [Consulté le 19 juillet 2022]. Disponible à l'adresse: <https://nccr-onthemove.ch/DataManagement/Inventory.pdf>

NATIONAL DIGITAL STEWARDSHIP ALLIANCE., 2020. Utilisation des niveaux de préservation numérique : aperçu de la V2.0. *HAL*. [en ligne]. 23 avril 2020. [Consulté le 22 juillet 2022]. Disponible à l'adresse: <https://hal-bnf.archives-ouvertes.fr/hal-02551807>

OFFICE FÉDÉRAL DE LA STATISTIQUE, 2021. Formulaire pour les demandes d'appariement de données. *OFS*. [en ligne]. 3 mai 2021. [Consulté le 28 mai 2022]. Disponible à l'adresse: <https://www.bfs.admin.ch/asset/fr/17084399>

OLOS, 2022a. FAQ. *OLOS*. [en ligne]. 2022. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://olos.swiss/documentation/faq>

OLOS, 2022b. DLCM Integration Guide : DLCM Solution v2.1.7. *admin.olos.swiss*. [en ligne]. 22 juillet 2022. [Consulté le 28 juillet 2022]. Disponible à l'adresse: <https://admin.olos.swiss/administration/docs/DLCM-IntegrationGuide.html#dlcm-architecture>

OLOS, 2022. Pricing. *OLOS*. [en ligne]. 2022. [Consulté le 7 août 2022]. Disponible à l'adresse: <https://olos.swiss/pricing>

OPEN KNOWLEDGE FOUNDATION, [sans date]. So I've opened up some data, now what? *Open data handbook*. [en ligne]. [Consulté le 22 juillet 2022]. Disponible à l'adresse: <https://opendatahandbook.org/guide/en/following-up/>

Ordonnance du 1er juillet 2020 sur les dispositifs médicaux (ODim; RS 812.213). *Fedex* [en ligne]. 01 juillet 2020. Mise à jour le 26 mai 2022. [Consulté le 10 juin 2022]. Disponible à l'adresse: <https://www.fedlex.admin.ch/eli/cc/2020/552/fr>

Ordonnance du 20 septembre 2013 relative à la recherche sur l'être humain à l'exception des essais cliniques (Ordonnance relative à la recherche sur l'être humain (ORH; RS 810.301). *Fedex* [en ligne]. 20 septembre 2013. Mise à jour le 26 mai 2022. [Consulté le 10 juin 2022]. Disponible à l'adresse: <https://www.fedlex.admin.ch/eli/cc/2013/642/fr>

Ordonnance du 20 septembre 2013 sur les essais cliniques hors essais cliniques de dispositifs médicaux (Ordonnance sur les essais cliniques (OClin; RS 810.305). *Fedex* [en ligne]. 20 septembre 2013. Mise à jour le 26 mai 2022. [Consulté le 10 juin 2022]. Disponible à l'adresse: <https://www.fedlex.admin.ch/eli/cc/2013/643/fr>

Ordonnance du 30 juin 1993 concernant l'exécution des relevés statistiques fédéraux (Ordonnance sur les relevés statistiques) (RS 431.012.1). *Fedlex*. [en ligne]. 30 juin 1993. Mise à jour le 01 janvier 2022. [Consulté le 18 mai 2022]. Disponible à l'adresse: https://www.fedlex.admin.ch/eli/cc/1993/2100_2100_2100/fr

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES, 2007. *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*. [en ligne]. Paris: Ed. OCDE. [Consulté le 6 novembre 2021]. ISBN 978-92-

64-03402-0. [Consulté le 9 août 2022]. Disponible à l'adresse:
<https://doi.org/10.1787/9789264034020-en-fr>

ORGANISATION INTERNATIONALE DE NORMALISATION, 2001. *ISO 15489-1 : Information et documentation : «Records management» : partie 1: principes directeurs*. [en ligne]. Genève. ISO. [Consulté le 27 novembre 2021]. 15489-1:2001(fr). Disponible à l'adresse: <https://www.iso.org/obp/ui/#iso:std:iso:15489:-1:ed-1:v1:fr>

OTT, Florence, 2015. *Apport de la gestion documentaire à la gouvernance de l'information*. Paris]; Cachan: Hermès ; Lavoisier. ISBN 978-2-7462-4713-0.

PRÉPOSÉ FÉDÉRAL À LA PROTECTION DES DONNÉES ET À LA TRANSPARENCE, [sans date-a]. Recherche médicale. *PPPDT*. [en ligne]. [sans date-a]. [Consulté le 3 mars 2022]. Disponible à l'adresse:
<https://www.edoeb.admin.ch/edoeb/fr/home/datenschutz/statistik--register-und-forschung/forschung/medizinische-forschung.html>

PRÉPOSÉ FÉDÉRAL À LA PROTECTION DES DONNÉES ET À LA TRANSPARENCE, [sans date-b]. Protection des données et recherche en général. *PPPDT*. [en ligne]. [sans date-b]. [Consulté le 14 juillet 2022]. Disponible à l'adresse:
<https://www.edoeb.admin.ch/edoeb/fr/home/datenschutz/statistik--register-und-forschung/forschung/datenschutz-und-forschung-im-allgemeinen.html>

PRÉPOSÉ FÉDÉRAL À LA PROTECTION DES DONNÉES ET À LA TRANSPARENCE, 2021a. Guide pour l'examen de la licéité de la communication transfrontière de données (art. 6, al. 2, let. a, LPD). *PPPDT*. [en ligne]. juin 2021. [Consulté le 6 août 2022]. Disponible à l'adresse:
<https://www.edoeb.admin.ch/dam/edoeb/fr/dokumente/2021/Anleitung%20f%C3%BCr%20die%20Pr%C3%BCfung%20von%20Daten%20C3%BCbermittlungen%20mit%20Auslandbezug%20FR%20V1.1.pdf.download.pdf/Anleitung%20f%C3%BCr%20die%20Pr%C3%BCf>

PRÉPOSÉ FÉDÉRAL À LA PROTECTION DES DONNÉES ET À LA TRANSPARENCE, 2021b. État de la protection des données dans le monde. *PPPDT*. [en ligne]. 15 novembre 2021. [Consulté le 8 août 2022]. Disponible à l'adresse:
https://www.edoeb.admin.ch/dam/edoeb/fr/dokumente/2021/20211115_Staatenliste_f.pdf.download.pdf/20211115_Staatenliste_f.pdf

RE3DATA.ORG, 2021. OLOS. *Re3data.org*. [en ligne]. 2021. [Consulté le 24 juillet 2022]. Disponible à l'adresse: <https://www.re3data.org/repository/r3d100013500>

RE3DATA.ORG, 2022. SWISSUbase. *Re3data.org*. [en ligne]. 2022. [Consulté le 24 juillet 2022]. Disponible à l'adresse:
<https://www.re3data.org/repository/r3d100013746>

RECKER, Jonas et CESSDA TRAINING TEAM, 2017. 4. Store. *CESSDA Data Management Expert Guide*. [en ligne]. 2019 2017. [Consulté le 10 juin 2022]. Disponible à l'adresse:
<https://dmeq.cessda.eu/Data-Management-Expert-Guide/4.-Store>

Règlement du 16 janvier 2008 concernant la valorisation et l'exploitation des résultats acquis par la recherche de la HES-Valais et de la HEVs2 (RS 414.733). *lex.vs.ch* [en ligne]. 16 janvier 2008. Mise à jour le 01 janvier 2010, [Consulté le 17 juillet 2022]. Disponible à l'adresse: https://lex.vs.ch/app/fr/texts_of_law/414.733

RESEARCH DATA ALLIANCE, 2018. Entrepôts de données de confiance : critères de conformité. *RDA Alliance*. [en ligne]. 2018. [Consulté le 21 juillet 2022]. Disponible à l'adresse: <https://www.rd-alliance.org/system/files/documents/CoretrustsealFR.pdf>

SBEIH, Lina, DEDET, Fanny, MOREAU, Patrick et DZALE, Esther, 2020. *L'archivage des données de la recherche à l'Inra : éléments de réflexion, démarche et perspectives*. [en ligne]. Paris: INRA. [Consulté le 12 mai 2022]. Cahier des techniques de l'INRA. Disponible à l'adresse: <https://hal.inrae.fr/hal-02861909>

SBEIH, Lina, LOAEC, Mikael, CAT, Sébastien, MOREAU, Patrick, CAHUZAC, Eric, DEDET, Fanny, FOURY, Jacques, DEHNE-GARCIA, Alexandre, SZABO, Dimitri, DZALÉ-YEUMO, Esther et LOAEC, Mikael, 2020. Fiche Archivage. *Datapartage Inrae*. [en ligne]. juillet 2020. [Consulté le 23 juillet 2022]. Disponible à l'adresse: <https://datapartage.inrae.fr/content/download/3769/39973/version/1/file/Fiche-Archivage.pdf>

SCHNEIDER, Patrice, 2022. *SWITCHdrive : concept d'exploitation : V 0.3* [document PDF]. Document interne au Service informatique de la HES-SO Valais-Wallis

SECRETARIAT D'ETAT À L'ÉCONOMIE, 2022. Nouvelle loi sur la protection des données (nLPD). *Portail PME*. [en ligne]. 31 mars 2022. [Consulté le 10 juillet 2022]. Disponible à l'adresse: <https://www.kmu.admin.ch/kmu/fr/home/fakten-und-trends/digitalisierung/datenschutz/neues-datenschutzgesetz-revdsq.html>

SEYDTAGHIA, Anouch, 2022. EXCLUSIF – Le piratage de l'Université de Neuchâtel est tentaculaire. *Le Temps*. [en ligne]. 9 mars 2022. [Consulté le 19 juillet 2022]. Disponible à l'adresse: <https://www.letemps.ch/economie/exclusif-piratage-luniversite-neuchatel-tentaculaire>

SPADA, Fabrice, 2021. Manuel d'utilisation de AGP SageX. *SageX Wiki*. [en ligne]. novembre 2021. [Consulté le 27 mai 2022]. Disponible à l'adresse: <https://agp-wiki.hes-so.ch/xwiki/bin/download/Main/WebHome/2021%20Manuel%20utilisation%20SageX%20v1.4.pdf?rev=1.1>

STEINER, Alexandre, 2022. La Haute Ecole Arc victime d'une cyberattaque. *Le Temps*. [en ligne]. 4 juillet 2022. [Consulté le 19 juillet 2022]. Disponible à l'adresse: <https://www.letemps.ch/suisse/haute-ecole-arc-victime-dune-cyberattaque>

SUMMERS, Scott et CORTI, Louise, 2020. Storing and moving data. In: CORTI, Louise, HAAKER, Maureen, SUMMERS, Scott, VAN DEN EYNDEN, Veerle, BISHOP, Libby et WOOLLARD, Matthew, *Managing and sharing research data: a guide to good practice*. 2e ed. Los Angeles: SAGE. pp. 135-158. ISBN 978-1-5264-6025-7.

SWISSUBASE, 2022a. Discover SWISSUbase presentation and demo [enregistrement vidéo]. *YouTube*. [en ligne]. 9 mai 2022. [Consulté le 24 juillet 2022]. Disponible à l'adresse: <https://www.youtube.com/watch?v=BN5UqpErvP8>

SWISSUBASE, 2022b. Metadata guide : version 2.1.0. *SWISSUbase*. [en ligne]. mai 2022. [Consulté le 26 juillet 2022]. Disponible à l'adresse: https://resources.swissubase.ch/wp-content/uploads/2021/12/Metadata-Guide_V1.pdf

SWISSUBASE, 2022c. User guide: submit a study and deposit data : version 2.1.0. *SWISSUbase*. [en ligne]. mai 2022. [Consulté le 26 juillet 2022]. Disponible à l'adresse: <https://resources.swissubase.ch/wp-content/uploads/2021/12/SWISSUbase-User-Guide-V1.pdf>

SWISSUNIVERSITIES et FONDS NATIONAL SUISSE, 2016. Stratégie nationale suisse sur l'Open Access. *Swissuniversities*. [en ligne]. 2016. [Consulté le 8 mai 2021]. Disponible à l'adresse: https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/Open_Access/Open_Access_strategy_final_f.pdf

SWITCH, 2022a. SWITCHColdStor : long term storage for research and education. *SWITCH*. [en ligne]. 2022. [Consulté le 28 juillet 2022]. Disponible à l'adresse: <https://www.switch.ch/fr/coldstor/>

SWITCH, 2022b. FAQ : bases. *SWITCH*. [en ligne]. 2022. [Consulté le 14 juillet 2022]. Disponible à l'adresse: <https://help.switch.ch/drive/faq/Basics/#collapse-4e69618a-6685-11e4-b6ef-5254009dc73c-16>

T GOLDEN EYE, 2014. How to create automatic email reminder from Ms Access [enregistrement vidéo]. *YouTube*. [en ligne]. 7 novembre 2014. [Consulté le 18 juillet 2022]. Disponible à l'adresse: https://www.youtube.com/watch?v=KlZrWoH_Z-0

TAKSLER, Glen B., DALTON, Jarrod E., PERZYNSKI, Adam T., ROTHBERG, Michael B., MILINOVICH, Alex, KRIEGER, Nikolas I., DAWSON, Neal V., ROACH, Mary J., LEWIS, Michael D. et EINSTADTER, Douglas, 2021. Opportunities, pitfalls, and alternatives in adapting electronic health records for health services research. *Medical Decision Making*. 1 février 2021. Vol. 41, no. 2, pp. 133-142. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.1177%2F0272989X20954403>. [accès par abonnement]

TENOPIR, Carol, DALTON, Elizabeth D., ALLARD, Suzie, FRAME, Mike, PJESIVAC, Ivanka, BIRCH, Ben, POLLOCK, Danielle et DORSETT, Kristina, 2015. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*. 26 août 2015. Vol. 10, no. 8, pp. e0134826. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.1371/journal.pone.0134826>

UK DATA SERVICE, 2012. Data inventory. *UK Data Service*. [en ligne]. 2022 2012. [Consulté le 25 mars 2022]. Disponible à l'adresse: <https://ukdataservice.ac.uk/learning-hub/research-data-management/collaborative-research/data-inventory/>

UK DATA SERVICE, 2019. Research data lifecycle [enregistrement vidéo]. *YouTube*. [en ligne]. 13 août 2019. [Consulté le 19 avril 2022]. Disponible à l'adresse: <https://www.youtube.com/watch?v=-wjFMMQD3UA>

UNIVERSITÉ DE GENÈVE, 2019a. Le dépôt cantonal des données de recherche Yareta. *UNIGE E-Research*. [en ligne]. 30 juillet 2019. [Consulté le 21 juillet 2022]. Disponible à l'adresse: <https://www.unige.ch/eresearch/fr/projets/yareta/>

UNIVERSITÉ DE GENÈVE, 2019b. Intégrité dans la recherche scientifique. *UNIGE Memento*. [en ligne]. 2 octobre 2019. [Consulté le 12 juillet 2022]. Disponible à l'adresse: <https://memento.unige.ch/doc/0003>

UNIVERSITÉ DE GENÈVE, RESEARCHDATA, 2018. Données personnelles ou sensibles ? *UNIGE ResearchData*. [en ligne]. 11 décembre 2018. [Consulté le 10 mars 2022]. Disponible à l'adresse: <https://www.unige.ch/researchdata/fr/stocker/personnelles-sensibles/>

UNIVERSITÉ DE GENÈVE. RESEARCHDATA, 2020. Préservation à l'UNIGE. *UNIGE ResearchData*. [en ligne]. 22 décembre 2020. [Consulté le 13 mai 2022]. Disponible à l'adresse: <https://www.unige.ch/researchdata/fr/perserver/preservation-lunige/>

UNIVERSITÉ DE LAUSANNE, 2020a. Archivage & partage. *L'Open Science à l'Unil*. [en ligne]. 19 août 2020. [Consulté le 13 mai 2022]. Disponible à l'adresse: <https://www.unil.ch/openscience/fr/home/menuinst/open-research-data/gerer-les-donnees-de-recherche-research-data-management/archivage--partage.html>

UNIVERSITÉ DE LAUSANNE, 2021b. Stockage & sécurité. *L'Open Science à l'Unil*. [en ligne]. 11 mai 2021. [Consulté le 13 mai 2022]. Disponible à l'adresse:

<https://www.unil.ch/openscience/fr/home/menuinst/open-research-data/gerer-les-donnees-de-recherche-research-data-management/stockage--securite.html>

UNIVERSITÉ DE LAUSANNE, 2021. Directive de la Direction 4.5 sur le traitement et la gestion des données de recherche. *UNIL*. [en ligne]. 15 juin 2021. [Consulté le 4 juin 2022]. Disponible à l'adresse: <https://www.unil.ch/files/live/sites/central/files/textes-leg/4-rech/dir4-5-donnees-rech3.pdf>

UNIVERSITÉ DE LAUSANNE. FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES, [sans date]. Stocker, partager, synchroniser des fichiers avec SWITCHdrive : guide d'utilisation. *UNIL*. [en ligne]. [Consulté le 24 mai 2022]. Disponible à l'adresse: https://www.unil.ch/ssp/files/live/sites/ssp/files/shared/procedures_internes/SwitchDrive.pdf

UNIVERSITÉ DE LAUSANNE. UNIRIS, 2014. Politique de records management : glossaire. *UNIL*. [en ligne]. 11 juin 2014. [Consulté le 22 novembre 2021]. Disponible à l'adresse: https://www.unil.ch/files/live/sites/uniris/files/documents/references/UNIL_POL_RM_Glossaire_VF.pdf

UNIVERSITÉ DE LAUSANNE. UNIRIS, 2019. Guide pour l'établissement du fichier Readme : description des données de recherche. *L'Open Science à l'Unil*. [en ligne]. 30 octobre 2019. [Consulté le 8 août 2022]. Disponible à l'adresse: https://www.unil.ch/openscience/files/live/sites/openscience/files/Donnees_de_recherche/Files/Guidelines_Long%20Term_Storage_Data_Description.docx

UNIVERSITY OF MELBOURNE, 2013. Management of Research Data and Records Policy (MPF1242). *Melbourne policy library*. [en ligne]. 30 novembre 2013. [Consulté le 8 mai 2022]. Disponible à l'adresse: <https://policy.unimelb.edu.au/MPF1242#section-5>

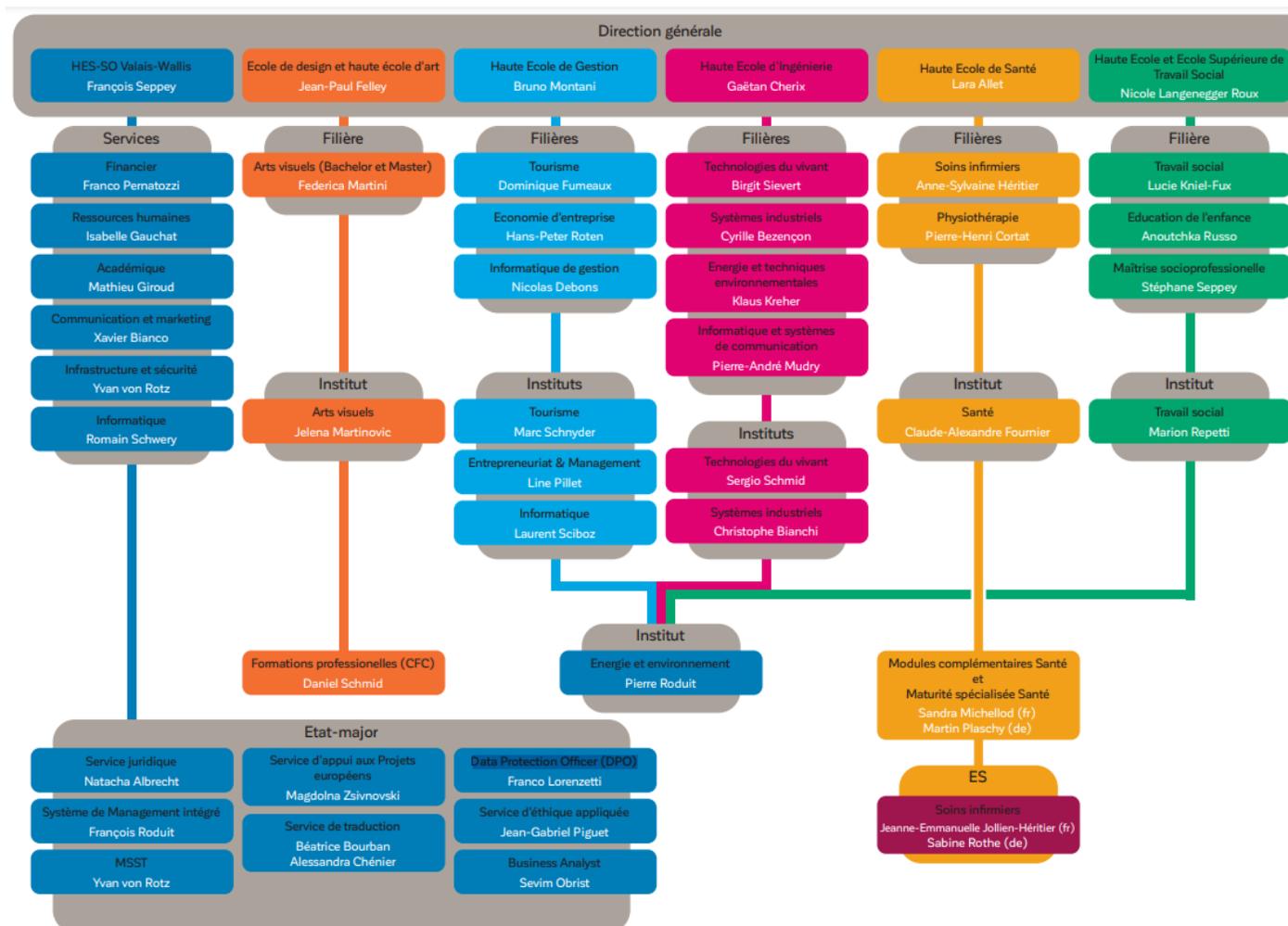
UNIVERSITY OF NEW HAMPSHIRE, 2020. Long-term preservation. *Data Services*. [en ligne]. 17 février 2020. [Consulté le 14 mai 2022]. Disponible à l'adresse: <https://www.usnh.edu/data-services/research-data/long-term-preservation>

U.S. NATIONAL LIBRARY OF MEDICINE et U.S. NATIONAL INSTITUTES OF HEALTH, 2021. ClinicalTrials.gov Background. *ClinicalTrials.gov*. [en ligne]. mai 2021. [Consulté le 30 juin 2022]. Disponible à l'adresse: <https://clinicaltrials.gov/ct2/about-site/background>

WHYTE, Angus et WILSON, Andrew, 2010. How to appraise and select research data for curation. *DCC*. [en ligne]. 25 octobre 2010. [Consulté le 24 octobre 2021]. Disponible à l'adresse: <https://www.dcc.ac.uk/guidance/how-guides/appraise-select-data>

WIKIHOW, [sans date]. Comment utiliser Microsoft Access. *wikiHow*. [en ligne]. [Consulté le 27 mai 2022]. Disponible à l'adresse: <https://fr.wikihow.com/utiliser-Microsoft-Access>

Annexe 1 : Organigramme de la HES-SO Valais-Wallis



(HES-SO Valais-Wallis 2022)

Annexe 2 : Guide d'entretien des personnes ressources

Fonction :

Date de la séance :

SOUTIEN APPORTÉ AUX CHERCHEUSES ET CHERCHEURS

En quoi consiste votre poste ?

A quels moments d'un projet de recherche intervenez-vous ?

Quel support aux chercheuses et chercheurs mettez-vous à disposition ?

Est-ce que les chercheurs ont une obligation de vous contacter ?

Combien de demandes avez-vous par mois/ par année ?

Devez-vous assurer un suivi de ces demandes ?

Si oui, utilisez-vous un outil spécifique pour assurer ce suivi ?

MISE EN PLACE D'UN INVENTAIRE

Si nous mettons en place un outil d'inventaire, auriez-vous un intérêt à y accéder/utiliser ?

Quelles informations vous seraient utiles ?

Selon vous quel(s) moment(s) les informations devraient-elles être collectées ?

Avez-vous connaissance d'autres institutions qui réalisent un inventaire de leurs données ?

AUTRES

Autres questions spécifiques au domaine d'activité de la personne interviewée

Annexe 3 : Analyse des besoins des personnes ressources

Responsable de l'Institut santé (RI)

Il a une vision globale des projets en cours de réalisation ou étant terminés. En tant que RI il est informé des projets soumis aux bailleurs de fonds. Il fournit des conseils aux chercheuses et chercheurs qui le sollicite et valide les engagements financiers.

Intervention	Planification . Fournit des conseils et valide les engagements financiers
Besoins	Savoir où, comment sont gérées les données Avoir une vue d'ensemble des données Connaitre les types, le volume des données
Problématique	Responsabilité liée aux données Manque de connaissance des données produites Améliorer la valorisation des données Manque de réutilisation des données
Outils	Outil SageX utilisé pour la gestion des projets de recherche, dès le moment où il a été accepté. La-le PI (principal investigator) est responsable de l'enregistrement des informations liées au projet de recherche dans l'outil.
Métadonnées	Titre du projet de recherche Nom du ou des jeux de données Qui est responsable des données Où sont stockées les données Présence de données personnelles ou sensibles Type de données Famille de données

Le responsable de l'équipe architecture et sécurité au sein du service informatique (SINF)

Le responsable architecture et sécurité de la HES-SO Valais-Wallis a pour tâches au niveau sécurité de mettre en place et/ou de valider les concepts de sécurité afin de garantir la protection des informations au niveau de l'intégrité, de la disponibilité et de la confidentialité. Pour cela il réalise un diagnostic avec une analyse de risques qui vérifie la cohérence avec les besoins du projet. Pour la partie architecture, il est responsable du montage de solutions spécifiques, il veille au maintien de la cohérence avec le reste du système d'information et valide les solutions mises en place.

Intervention	Phase de planification : contacté soit par les équipes soit par les personnes ressources sur question de stockage et concept sécurité pour la relecture DMP, clauses contrats.
---------------------	--

	Autres phases : le plus souvent, lorsqu'une équipe rencontre des difficultés (concept sécurité manquant/incomplet, solution technique pas adaptée au projet, solutions techniques non adéquates par rapport aux clauses contrats)
Besoins	<p>Être contacté dès la planification afin de clarifier les besoins, d'analyser les risques si besoin, de proposer des solutions adaptées à leur projet ainsi que des conseils sur les bonnes pratiques.</p> <p>Être informé si un nouveau projet nécessite une infrastructure particulière</p> <p>Pouvoir accéder au DMP est idéal car il a ainsi toutes les informations nécessaires à la bonne compréhension du projet</p> <p>Avoir connaissance du type de données collectées et de la volumétrie de stockage à prévoir pour chaque nouveau projet</p> <p>Connaitre la durée de conservation des données pour chaque projet</p> <p>Gestion des documents déposés sur les serveurs</p>
Problématique	<p>Comme il n'y a pas d'obligation de le contacter, il ne peut garantir que tous les projets aient le bon niveau de protection.</p> <p>Anticiper les coûts, les budgétiser pour éviter des temps d'attente</p> <p>Dossiers liés à des projets de recherche qui sont stockés indéfiniment sur les serveurs de l'école</p> <p>Toutes les équipes ne savent pas qu'elles peuvent le contacter</p>
Outils	Pas de procédure, checklist sur les questions d'architecture et de sécurité.
Métadonnées	<p>Durée du projet</p> <p>Durée de conservation</p> <p>Volumétrie des données</p> <p>Type de données</p>

Le responsable d'éthique appliquée (ETH)

Il apporte son soutien pour tous les projets de recherche qui doivent passer devant le comité d'éthique (CER-VD), dans l'écriture des requêtes et sur les questions légales liées à l'éthique.

Intervention	<p>Phase de planification : relecture du DMP et apporter des conseils, CER-VD (oui/non) et aide à la préparation du dossier.</p> <p>Autres phases : aide ponctuelle si difficultés, conseils sur anonymisation.</p> <p>Il n'a pas pour mission de contrôler mais de conseiller</p> <p>Propose des formations sur les questions liées à l'éthique, va à la rencontre des équipes de recherche pour les sensibiliser aux questions d'éthique.</p>
Besoins	Chaque nouveau membre de l'IS devrait suivre une formation sur la gestion des données et sur les aspects liés à l'éthique.

	<p>Avoir un processus de travail qui renseigne pour chaque phase du projet sur les personnes à contacter afin de recevoir les conseils au moment adéquat et éviter tous retards et imprévus</p> <p>Avoir un suivi des données qui arrivent en fin de protection et qui doivent faire l'objet d'une destruction</p> <p>Définition d'alertes automatiques</p> <p>Etre informé lorsqu'un projet a été validé par le comité d'éthique ou dans le cas de projets contenant des données anonymisées.</p>
Problématique	<p>Pas d'obligation de le contacter, les scientifiques peuvent passer directement par le comité d'éthique.</p> <p>N'a pas la vision d'ensemble des projets soumis à la LRH</p> <p>N'a pas connaissance si les données ont bien été correctement anonymisées – pseudonymisées</p> <p>Les données qui doivent être détruites selon accords/LRH sont-elles vraiment détruites ? Suivi des données</p> <p>Conservation suffisamment sécurisée pour les données sensibles /personnelles</p>
Outils	<p>Pas de solution particulière mentionnée</p>
Métadonnées	<p>Présence de données anonymisées / pseudonymisées</p> <p>Données soumises à un comité d'éthique</p> <p>Date de validation du comité d'éthique</p> <p>Date de fin de conservation des données + alerte lorsque l'on arrive à l'échéance</p> <p>Lieu de stockage des données</p> <p>Un DMP a-t-il été rédigé, et si oui en obtenir une copie</p>

Le délégué à la protection des données (DPO)

Il s'occupe de tous les aspects de protection de données liés à la loi sur la protection des données. Il a pour tâche d'inventorier les traitements de données personnelles qui sont faits au sein de l'école dans un registre de traitements pour la HES-SO Valais-Wallis.

Intervention	<p>Intervient uniquement s'il y a des données personnelles dans le projet.</p> <p>Planification : rédaction et relecture du DMP</p> <p>Peut également être amené à intervenir durant d'autres phases si besoin. Vérifie les formulaires de consentement, Il fournit également des conseils théoriques sur les aspects sécurité de l'information, chiffrement, anonymisation et intervient pour des formations lors des séances des instituts. La dernière formation donnée pour l'IS remonte à deux ans.</p> <p>Le suivi des demandes est réalisé via un registre qui permet de tenir à jour une documentation sur tous les traitements. Il décrit qui sont les personnes concernées, quand a lieu la collecte, comment a été réalisé le traitement, à quelle date le projet se termine et durant quelle période les données doivent être conservées.</p>
---------------------	---

Besoins	Selon lui il est vraiment nécessaire de le contacter le plus tôt possible dès la planification. Il souhaiterait qu'une checklist soit donnée aux scientifiques qui les informerait sur les personnes à contacter selon le type de données contenues dans le projet. Dans l'idéal cette checklist devrait être validée et signée par la-le responsable du projet. Il souhaiterait recevoir une alerte lorsqu'un nouveau projet entre dans SageX afin de pouvoir aller à la rencontre des équipes de recherche pour les sensibiliser sur ces questions.
Problématique	Dès 2023 avec l'entrée en vigueur de la nLPD, obligation de lui signaler tout traitement de données personnelles.
Outils	Tenue d'un registre de traitement. Il utilise une application « Smart global privacy » qui lui permet entre autres de tenir le registre des traitements. Mais ce dernier pourrait tout à fait être réalisé sur un fichier Excel.
Métadonnées	Présence de données personnelles A quel moment se termine le projet Durée de conservation des données Date de collecte des données Présence de données anonymisées / pseudonymisées Lieu et méthode de stockage

La collaboratrice administrative (CA)

La collaboratrice administrative de l'IS s'occupe du suivi administratif et financier des projets de recherche ainsi que des mandats de prestations de service. Elle apporte un soutien administratif aux chercheur-euse-s de l'IS.

Intervention	Planification : avant le dépôt du projet sur les questions liées au montage financier. Elle aide à l'établissement de budget, s'assure qu'il correspond aux règles établies par les bailleurs de fonds, vérifie notamment que la partie ressources humaines (données salariales, heures prévues) ainsi que le matériel et la sous-traitance aient été correctement budgétisés. Tout au long du projet : elle gère la partie administrative (gérer les subventions, facturer les acomptes, garantir la disponibilité des fonds, établir les rapports financiers) ainsi que le budget. Fin de projet : clôture administrative du projet
Besoins	Elle a le sentiment que les équipes de recherche ne savent pas qu'elle peut apporter son aide pour la relecture du plan financier. Elle souhaiterait avoir une rencontre avec chacun-e avant le dépôt du projet pour éviter les mauvaises surprises liées à un manque de budget.
Problématique	Pas d'obligation de la contacter pour vérifier que le budget et la partie RH correspondent aux exigences des bailleurs de fonds et cela peut poser des problèmes pour la suite du projet.
Outils	Outil SageX utilisé pour la gestion des projets de recherche, dès le moment où il a été accepté.

	Au sein du secrétariat un fichier Excel est utilisé pour assurer le suivi de tous les projets, qu'ils soient en préparation, en cours ou terminés.
Métadonnées	Elle n'a pas forcément de besoins en matière de métadonnées car elle collecte déjà toutes les informations nécessaires dans son fichier. A savoir : Chef de projet No du projet Acronyme Titre du projet Financement Etat d'avancement Date de début et date de fin Filière

La juriste (JU)

La juriste de la HES-SO Valais-Wallis intervient dans un projet pour relire et apporter des conseils sur la rédaction des contrats et ainsi que pour toutes autres questions juridiques.

Intervention	Planification : intervient pour relire les contrats de recherche (confidentialité, droits de propriété intellectuelle) et participe parfois aux négociations pour ces contrats. Modèles de contrats disponibles mais en cours de révision (confidentialité, transfert de matériel, propriété intellectuelle)
Besoins	Dans l'idéal contrats écrit même si cela n'est pas une obligation = liberté contractuelle (moins compliqué en cas de problème pour retrouver quelle était la volonté des parties). Les rédiger avant le début de la recherche lors de la prise de contact avec les partenaires Lui transmettre les contrats avant signature pour relecture Être informée si des contrats sont conclus Recevoir les contrats afin qu'ils soient enregistrés sur SageX.
Problématique	La partie juridique est souvent oubliée par les scientifiques, pas toujours de contrat signé, souvent oraux ou seulement un courriel. N'a pas connaissance des contrats signés
Outils	Pas mentionné
Métadonnées	Existence d'un contrat

La spécialiste en information documentaire (SID)

La spécialiste ID intervient sur les questions liées à l'Open Science et plus particulièrement sur l'Open Access. Elle s'occupe de l'archivage des activités de recherche de l'IS sur la plateforme d'archive ouverte institutionnelle de la HES-SO, ArODES⁴⁷ ainsi que sur le serveur interne de l'IS. La spécialiste ID a également le rôle de data manager.

Intervention	Phase publication et partage : obligation pour les chercheuses et chercheurs de transmettre leurs publications et interventions pour archivage sur ArODES ainsi que sur le serveur interne. En règle générale dès leur parution elles-ils lui font parvenir un exemplaire de leur publication, parfois le contrat d'édition s'il est en leur possession. En début d'année un rappel est envoyé à tout l'institut pour récolter les publications qui n'auraient pas encore été transmises. Conseils sur les questions Open Access
Besoins	Être informée des nouveaux projets de recherche et de leur fin Être informée si des données sont déposées sur un dépôt ouvert afin de faire le lien avec les publications
Problématique	N'a pas connaissance des nouveaux projets de recherche ou de la fin d'un projet, ni si des données sont partagées sur des dépôts de données Pouvoir présenter les services liés à la recherche en début d'activité d'une nouvelle collaboratrice ou d'un nouveau collaborateur
Outils	Une procédure d'archivage et de valorisation des activités de recherche ⁴⁸ Fichier Excel pour assurer le suivi des publications archives hors et dans ArODES.
Métadonnées	Début et fin du projet de recherche Lien sur la publication archivée dans ArODES Lien sur la plateforme de partage

Data manager (DM)

Comme nous l'avons vu précédemment, un service de data management a vu le jour depuis quelques mois et une data manager a été nommée pour la HEdS. Elle apporte son soutien sur les questions liées à l'Open Science et plus particulièrement sur la gestion des données de recherche.

Intervention	Tout au long du projet : Peut apporter son soutien sur les questions liées à l'Open Science et sur la gestion des données de recherche Rédige des guidelines, propose et réalise des formations
---------------------	--

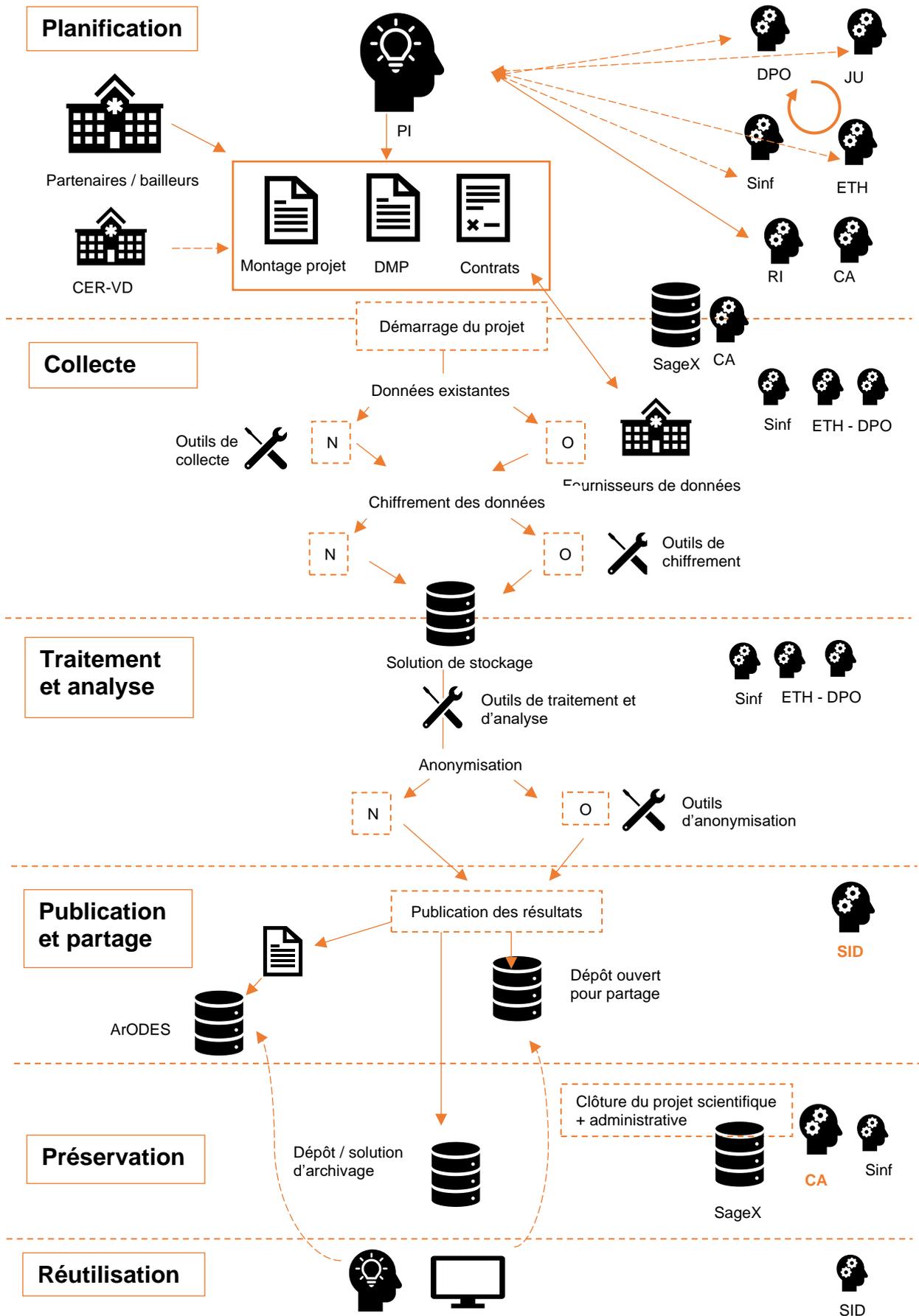
⁴⁷ <https://arodes.hes-so.ch/>

⁴⁸

https://biblioheds.hevs.ch/Portals/56/Documents/DocumentsRaD/Procedure_ARODES_version%20definitive.pdf

	Fait le lien entre les équipes de recherche et les personnes ressources si besoin.
Besoins	Avoir connaissance de tout nouveau projet en préparation afin de pouvoir solliciter une rencontre avec la chercheuse ou le chercheur et pouvoir si besoin transmettre la demande aux autres personnes ressources. Obtenir le DMP s'il est rédigé
Problématique	Toutes les équipes de recherche ne savent pas qu'il existe désormais une DM au sein de l'IS. Elle n'a pas connaissance des nouveaux projets ni de ceux qui sont terminés N'a pour l'instant pas (peu) de contact avec les personnes ressources Pas d'obligation de la contacter
Métadonnées	Nom du projet de recherche Personne responsable des données Date de début et de fin du projet Type de données Présence de données personnelles ou sensibles Présence de données anonymisées / pseudonymisées Présence d'un DMP, et si oui en obtenir une copie Lieu et méthode de stockage Durée de conservation des données + date de destruction Données partagées et si oui sur quel(s) dépôt(s) Financeurs (FNS, H2020...) qui ont des exigences en termes de partage Etat du projet

Schéma interventions des personnes ressources



(Inspiré de Delafontaine, 2020)

Annexe 4 : Guide d'entretien institutions

Organisation :

Date de la séance :

QUESTIONS GÉNÉRALES SUR L'OPEN DATA

- Avez-vous une politique institutionnelle sur l'Open data ?
- Avez-vous un service ou une personne qui répond aux questions liées à la gestion des données de la recherche ?
- Avez-vous une procédure, workflow ou des guidelines pour accompagner les chercheuses et chercheurs sur les tâches à réaliser et sur les bonnes pratiques liées à la GDR ?

MISE EN PLACE DE L'INVENTAIRE

- Comment avez-vous procédé pour mettre en place cet inventaire ? [Entretien, questionnaire, groupe pilote].
- Quelles métadonnées collectez-vous pour chaque jeu de données ?
- Quels jeux de données sont enregistrés dans l'inventaire ?
- A quel niveau de profondeur enregistrez-vous les jeux de données ? au niveau du projet, collection ou au niveau fichier ?
- A quel moment les données sont-elles entrées dans l'inventaire et par qui ?
- Les chercheuses et chercheurs ont-ils-elles l'obligation de transmettre les données / de remplir l'inventaire ?
- Qui s'occupe de la gestion, du suivi et de la mise à jour de l'inventaire ?
- Quel(s) outil(s) utilisez-vous pour la gestion de l'inventaire ?
- Comment êtes-vous informé-e qu'un nouveau jeu de données peut-être / a été rajouté dans l'inventaire ?
- Avez-vous connaissance d'autre(s) institution(s) qui réalise(nt) un inventaire de leurs données ?

ARCHIVAGE ET PARTAGE DES DONNÉES

- Avez-vous une plateforme d'archivage institutionnel ?
- Le serveur de la plateforme est archivé dans votre institution ?
- Tous les jeux sont répertoriés sur votre plateforme d'archivage ou uniquement ceux qui peuvent être partagés ?
- Qui décide qu'un jeu de données possède un intérêt scientifique et pourrait donc être partagé à la communauté ?
- Les jeux de données qui ont des contraintes légales ou contractuelles sont-ils répertoriés et archivés sur cette plateforme ?
- Avez-vous établi des règles de conservation pour les jeux de données en termes de temps et de critères de conservation ?
- Avez-vous des alertes qui vous signalent lorsqu'un jeu de données doit être détruit ?
- Quelle(s) dépôt(s) de données ouvert(s) conseillez-vous pour le partage des données ?

Annexe 5 : Description des inventaires

Data Asset Framework (DAF)

Le Data Asset Framework (DAF) est une méthodologie développée par le Humanities Advanced Technology and information Institute (HATII), University of Glasgow en collaboration avec le Digital Curation Center (2009) afin de permettre aux institutions d'identifier et de localiser les actifs de données existants et d'obtenir des informations sur la manière dont elles sont stockées, gérées, partagées et réutilisées afin de réduire le risque de perte ou de mauvaise utilisation des données. Cette méthode a été développée puis testée auprès de 6 institutions.

Quatre étapes sont proposées dans le DAF : la planification de l'audit, l'identification et la classification des actifs de données, l'évaluation de la gestion des actifs de données et la communication des résultats et recommandations.

Catégories et métadonnées

- 16 métadonnées minimum sont requises mais une 2^e liste permet de collecter jusqu'à 50 métadonnées.
- 6 catégories : description, provenance, propriété, localisation, préservation/conservation et gestion

CABI – Data sharing toolkit

Le document « How to create a data inventory » de Beale et al. (2021) propose un guide en cinq étapes qui a pour objectif de créer un inventaire pour aider à la localisation, à la gestion, à l'utilisation et au partage des données financées par la Bill & Melinda Gates Fondation. Elle comprend : la planification de l'inventaire, la sélection des attributs de jeux de données à collecter, le remplissage de l'inventaire, la publication de l'inventaire et pour finir sa mise à jour.

Catégories et métadonnées

- 14 métadonnées
- Pas de catégorie

CIRAD

Le Centre de coopération internationale en recherche agronomique pour le développement (CIRAD 2022) est « l'organisme français de recherche agronomique et de coopération internationale pour le développement durable des régions tropicales et méditerranéennes ». Il a pour mission le développement de solutions pour une agriculture durable par le biais de la recherche, de la formation et du partage des connaissances. Le Cirad comprend 29 unités de recherche répartis dans trois départements et comprend plus de 1140 scientifiques (CIRAD 2021; 2022).

En 2015, dans le cadre du projet Patrimoine numérique scientifique (PNS) qui vise « à examiner les sujets liés à la gestion, la valorisation et la diffusion des données de recherche produites par l'établissement dans le cadre de son activité. » (Fortuno 2018, p. 2), un groupe de travail a été formé et a officié durant trois ans pour réaliser un inventaire des données de

recherche du Cirad⁴⁹. Il s'agissait d'obtenir un panorama des données de chaque unité ; de « gérer, valoriser et sécuriser ce capital », mais également d'anticiper les exigences des bailleurs de fonds et d'aider à la prise de décision pour l'ouverture des données (Fortuno 2015a, p. 3).

Un processus en six étapes a permis la réalisation de l'inventaire (Fortuno 2015a) :

- « Identifier l'ensemble des jeux de données d'intérêt scientifique et/ou présentant un intérêt patrimonial
- Décrire leurs caractéristiques à l'aide de métadonnées
- Établir des notices descriptives des jeux de données
- Valider les jeux d'intérêt et évaluer leur vulnérabilité et les éventuels besoins de protection
- Établir un premier bilan des jeux de données validés (famille/thématique/temporalité...)
- Généralisation de la démarche » (slide 7)

Le projet s'est ciblé sur la description de jeux de données emblématiques par département, sur une période de 2008 à 2015 où des unités pilotes ont réalisé ce travail d'inventaire. Chaque scientifique était invité-e à réaliser cet exercice mais sans obligation de leur part. De même le choix des jeux de données à décrire était laissé à leurs bons soins. De ce fait, cet inventaire n'était pas exhaustif⁵⁰

Catégories et métadonnées

- 56 métadonnées sont proposées dont 24 doivent obligatoirement être renseignées
- 9 catégories : description du contexte de la ressource, présentation de la ressource, caractérisation de la ressource, stockage et accès à la ressource, dispositif de recherche, auteur(s) contributeur(s) publication, 1^e qualification de la ressource, 2^e qualification de la ressource, qualification juridique de la ressource

LESC

Le Laboratoire d'ethnologie et de sociologie comparative (LESC [sans date]) est « une unité de recherche du Centre national de la recherche scientifique (CNRS) et de l'Université Paris Nanterre » qui regroupe environ 75 chercheuses et chercheurs, enseignantes-chercheuses et enseignants-chercheurs ainsi que 60 doctorant-e-s. En 2020, l'Institut des sciences humaines et sociales (InSHS) du CNRS a souhaité rédiger une politique de données à l'échelon du laboratoire résumant « les choix du laboratoire en matière de science ouverte, de gestion et de partage des données ainsi que les responsabilités des différents acteurs impliqués » (Heintz, Gautier 2022a, p. 4). A cette fin plusieurs unités de recherche ont été contactées dont le LESG.

La première étape de ce travail a été de réaliser un recensement non exhaustif des jeux de données au travers d'un inventaire afin d'avoir un aperçu des types de données, des volumétries, de leur localisation... (p. 5). Cela permettait également de disposer des jeux de données permettant de tester la politique et d'évaluer les moyens techniques et humains à mobiliser pour sa mise en œuvre (Heintz, Gautier 2022b).

⁴⁹ Entretien, S. Fortuno, 06 avril 2022

⁵⁰ Cf. 39

Catégories et métadonnées

- 39 métadonnées dont 16 prioritaires.
- 5 catégories : identification du jeu de données, caractérisation des données produites, informations complémentaires, modes de diffusion/archivage, modalité de partage

GovEx Labs

La Johns Hopkins University a créé un Centre d'excellence gouvernementale GovEx Labs afin d'accompagner les villes dans leurs pratiques liées à la gestion des données. Un guide d'inventaire a notamment été créé et comprend cinq étapes (Johns Hopkins University, Center of Government Excellence 2019) : établir une autorité de surveillance, déterminer la portée et le plan de l'inventaire des données, cataloguer les actifs de données conformément au plan d'inventaire, vérifier la qualité de l'inventaire des données, Initier les efforts de priorisation des données.

TM Mondaini

Dans son travail de Master réalisé pour l'IS, Mondaini (2021) propose une méthode et des indicateurs pour analyser la qualité des jeux de données ainsi qu'un tableau de bord qui comprend un certain nombre de métadonnées qui serviront de base pour ce travail.

Catégories et métadonnées

- 14 métadonnées qui décrivent le jeu de données, ainsi que 3 groupes métriques de qualité : principes FAIR, data quality dimensions et des critères infonomics.

HESAV

En vue de la réalisation d'un travail de Master en Sciences de l'information sur la thématique des données de la recherche mandaté par la Haute école de santé Vaud (HESAV), le responsable Open Science a établi un inventaire des données de recherche de son institution qui servira de base de travail pour le TM. L'inventaire porte sur la période 2018 à 2022 avec comme date de début l'obligation de rédaction de DMP imposée par le FNS. Il doit permettre d'obtenir une première photographie des types de données et de leur importance pour ensuite réaliser des entretiens et proposer des recommandations sur la manière dont l'HESAV doit s'organiser pour répondre au besoin de soutien de ces équipes de recherche⁵¹.

Catégories et métadonnées

- 18 métadonnées
- Pas de catégorie

⁵¹ Entretien, L. Amiotte-Suchet, Teams, 14 avril 2022

Annexe 6 : Analyse des inventaires de données

Tous les champs de l'inventaire du DAF ont été repris et traduits afin de servir comme base de ce tableau (Jones, Ruusalepp, Ross 2009, p. 45-47) . Certains champs tirés des autres inventaires et ayant un intérêt pour le projet ont été rajoutés dans le tableau. Ils sont mentionnés par un * à la fin de leur nom.

Liste des inventaires analysés :

- DAF Data Asset Framework réalisé par University of Glasgow en collaboration avec le Digital Curation Center
- Travail de Master d'Alessandro Mondaini
- CABI – Data sharing toolkit propose par Beale et al. pour les projets financés par Bill & Melinda Gates Fondation
- HESAV
- CIRAD – Centre de coopération internationale en recherche agronomique pour le développement en France
- Lesc – Laboratoire d'ethnologie et de sociologie comparative, une unité de recherche du Centre national de la recherche scientifique (CNRS) et de l'Université Paris Nanterre
- Analyse des besoins des personnes ressources de la HES-SO Valais-Wallis (responsable IS, Secrétariat, DPO, éthique, juridique, service informatique, bibliothèque, data manager)

Les croix correspondent aux champs obligatoires, les zéros aux champs facultatifs.

Organisation des champs :

- DAF propose de classer les champs selon : description, provenance, propriété, localisation, préservation/conservation, et gestion.
- Cirad : description du contexte de la ressource, présentation de la ressource, caractérisation de la ressource, stockage et accès à la ressource, dispositif de recherche, auteur(s) contributeur(s) publication,
- Lesc : caractérisation des données produites, informations complémentaires, modes de diffusion/archivage, modalité de partage.

	Champs	Définition (tirée de DAF)	DAF rk	Mondaini	CABI	HESAV	CIRAD	Lesc	PR – HES-SO VS-WS
Description	ID	Attribué par l'auditeur ou l'organisation	X	X	X			O	X
	Nom du projet de recherche*	Nom du projet de recherche dans lequel est créé le jeu de données				X	O	X	X
	Titre du jeu de données	Nom officiel de la ressource du jeu de données	X	X	X	X	X	X	X
	Forme des données*	Description de la forme du jeu de données (base de données, photos, textes, etc.) <u>Commentaires</u> <i>Le Lesc parle de caractérisation des données. Des exemples sont donnés comme : transcription textuelle, collection d'images, application web personnalisée, photographie de terrain</i> <i>Le Cirad propose la liste suivante : jeu de données (feuille et classeur, tableau de données : texte, word), base de données, carte, photo, image, audio, vidéo, logiciel, service portail web, objet physique, autres</i>	X	X	X	(X)	X	X	X
	Type de données	Description du type de données <u>Commentaires</u> <i>Le Lesc mentionne si des métadonnées sont présentes. Il propose la liste suivante : données textuelles, donnée enquête et sondage, données d'observation, données chiffrées, données audiovisuelles, images fixes, données web, données expérimentales, données 3D, données de simulation,</i>				(X)	X	X	X

	<p>codes informatiques et programmes, données dérivées ou compilées, données de référence, autres</p> <p>Le Cirad pour sa part parle de famille de la ressource et propose une liste multichoix : OMIQ-Données Omiques, RGB-Ressources Génétiques/Collections biologiques, EXP-Essais/Expérimentations/Analyses laboratoire (mesures instruments...), OBS-Observations de terrain (relevés, photos, imagerie satellite, film...), ENQ-Enquêtes (socio éco, épidémiologie, zootechnie, pratiques culturelles...), TMPS-Séries temporelles (cohortes + séries chrono + mesures répétées), TAXO-Taxonomie/Ontologie/Référentiel, LOGI-Informatique scientifique originale (logiciels, modèles, scripts,...), AUTR-Autre (non classifiable)</p> <p>La classification la plus couramment utilisée (UNIL, Doranum) est celle citée dans la partie méthode de ce travail à savoir : données d'observation, expérimentales, computationnelles ou de simulation, données dérivées ou compilées et les données de référence. Il serait peut-être intéressant de proposer deux options pour les données d'observation a) enquête et sondage b) observation de terrain</p>							
Propriétaire(s)	<p>Propriétaire(s) officiel(s) des données en termes de droits intellectuels.</p> <p><u>Commentaires</u></p> <p>Le détenteur de la propriété intellectuelle des données, peut être le chercheur mais également un fournisseur externe</p>	X					O	
Responsable du projet*	Nom de la personne principalement responsable du projet		X		X	O		X
Sujet	<p>Informations sur la couverture du sujet des données</p> <p><u>Commentaires</u></p> <p>Le Cirad propose une liste thématique propre à son domaine</p>	X	X	X		X		X
Langue	Langue(s) du contenu des données.	O		X	X	O		
Nom de la variante	Nom alternatif ou communément utilisé, s'il existe, acronyme	O			X			X

	Niveau	Niveau auquel la description actuelle est appliquée (fichiers, collections, projet).	O							
	Résumé	Texte décrivant le jeu de données	O							
	Résumé du projet*	Texte décrivant le projet				X	O			
	Mots-clés	Mots-clés pertinents qui décrivent le jeu de données <u>Commentaires</u> <i>Le Cirad propose un champ en texte libre</i>	O				X			
	Institutions partenaires*	Nom des institutions partenaires qui participent à la co-production				X	O			
	Ecole qui porte le projet*	Ecole qui est la requérante principale et qui porte la responsabilité du projet <u>Commentaires</u> <i>Le Cirad ne précise pas qui porte le projet mais seulement quel département et unité participe au projet</i>				X	X			
	Type(s) de financement*	Type de financement obtenu				X		O	X	
Provenance	Objectif initial	Description de la raison principale de la création de la ressource de données.	X							
	Description	Description des informations contenues dans le jeu de données <u>Commentaires</u> : <i>Le Cirad a intitulé ce champ résumé sur le contenu intellectuel de la ressource.</i>	X				X			
	Date de début	Date à laquelle le jeu de données a été créé / démarré <u>Commentaires</u> <i>Le Lesc et Beale appellent ce champ « Date de collecte » ce qui permet d'y insérer les dates de début et de fin de collecte</i> <i>Mondaini indique « Date du jeu de données » mais sans préciser s'il s'agit de la date de collecte ou de la date de la dernière modification</i>	X	X	(X)		X	O	X	

Date du projet*	Date de début et de fin de projet		X			O	O	X
Fréquence de mise à jour	Fréquence des mises à jour de cet ensemble de données pour en indiquer l'actualité	X		X				
Description du contexte	Description de l'utilisation initiale et du contexte des données.	X						
Source	<p>La ou les sources des informations trouvées dans le jeu, la description des méthodes de collecte des données ou des ensembles de données de tiers utilisées</p> <p><u>Commentaires</u></p> <p><i>Le Lesc communique des informations sur l'origine des données (départ à la retraite, données achetées avec limitation d'usage, données collectées sur le web...)</i></p> <p><i>Le Cird et Hesav demandent s'il s'agit de données collectées ou de données produites</i></p>	X			X	O	O	
Mode d'acquisition*	<p>Comment ces données ont-elles été acquises ?</p> <p><u>Commentaires</u></p> <p>Le Lesc souhaite savoir comment les données ont été acquises (achat, contrat d'usage...)</p>						O	
Type d'analyse*	<p>Type d'analyse réalisée sur les données</p> <p><u>Commentaires</u></p> <p><i>Hesav propose un champ en texte libre qui pourrait contenir : analyse quantitative, analyse qualitative, analyse descriptive, analyse compréhensive, analyse de contenu, analyse statistique, analyse thématique , analyse historiographique...</i></p>				X			
Logiciels de collecte et de traitement des données*	<p>Logiciels de collecte et de traitement des données</p> <p><u>Commentaires</u></p> <p><i>Hesav propose un champ en texte libre</i></p>				X		O	

	Date d'achèvement	Date à laquelle le jeu de données a été achevé / la collecte de données a cessé	O	X			X	O	
	Date de dernière modification	Date à laquelle le jeu de données a été mis à jour ou modifié pour la dernière fois. <u>Commentaires</u> <i>Mondaini indique la date du jeu de données mais sans préciser s'il s'agit de la date de collecte ou de la date de la dernière modification</i>	O		(X)				
	Gestion à ce jour	Historique de la maintenance et de l'intégrité du jeu de données	O						
	Curation à ce jour	Historique des activités de préservation et de conservation <u>Commentaires</u> <i>Le Lesc parle de contrôle de la qualité technique</i>	O					O	
	Description par des métadonnées riches*	<u>Commentaires</u> <i>Le Lesc propose plusieurs champs sur les métadonnées, identifiants des métadonnées, standards et référentiels documentaires</i>						O	
Propriété	Créateur(s) de données	Personne, groupe ou organisation responsable du contenu intellectuel du jeu de données	X		X		X	O	
	Gestionnaire(s) des données	Nom et coordonnées de la personne responsable de la gestion du jeu de données	X		X			X	X
	Droits	Indication des droits de l'utilisateur à visualiser, copier, redistribuer ou republier tout ou partie <u>Commentaires</u> <i>Pour la plupart cela comprend les licences</i> <i>Beale regroupe en un champ les droits, contraintes d'utilisation et d'accès</i>	X	X	(X)	X	X	X	X

Niveau de partage*	Indication si le jeu de données est actuellement partagé avec d'autres utilisateurs <u>Commentaires</u> <i>Périmètre de partage au Lesc : créateur/collecteur, équipe, laboratoire, employeur/tutelle, chercheurs dans le domaine, toutes communautés scientifiques, tout le monde)</i> <i>Au Cirad propose une liste à choix : pas partagé, laboratoire, unité, communauté interne Cirad, partenaire projet (cours/fin projet), en interne tout cirad, communauté externe au cirad, public (open data), autre</i>					X	X	
Période d'embargo*	Période d'embargo sur le jeu de données					O		
Contraintes d'utilisation / (accès)	Restrictions d'accès appliquées au jeu de données <u>Commentaires</u> <i>Mondaini : Niveau d'accès : privé, public, restreint</i> <i>Cirad : modalité d'accès : libre, contrôlé sur demande, privé pas de partage.</i> <i>Beale regroupe en un champ les droits et contraintes d'utilisation et d'accès</i>	X	X	(X)		X		X
Anciens gestionnaires des données	Historique de la conservation / chaîne de conservation de l'ensemble de données	O						
Autres mentions	Noms de contact des autres chercheurs et co-auteurs qui ont travaillé sur le jeu de données.	O				O		
Fréquence d'utilisation	Fréquence d'utilisation estimée et, si elle est connue	O						

	Ethique, protection des données et à la vie privée	Description de toute question potentielle relative à la protection des données ou à l'éthique liée au contenu de la ressource de données et si des restrictions sont appliquées <u>Commentaires</u> <i>Mondaini : sensibilité oui/non</i> <i>Personnes ressources : données personnelles ou sensibles, données soumises à la LRH, comité d'être oui/non, + date du passage + données confidentielles (qu'on ne peut pas rendre public, secrets recherche, contractuel)</i> <i>Hesav : LRH</i>	O			X			X
	Réutilisations potentielles	Description des réutilisations qui peuvent être envisagées. <u>Commentaires</u> <i>Le Cirad propose une liste à choix : ENS-Enseignement/Etudiants, PAR-Partenariat institutionnel, REC-Recherche/Chercheurs, EXPT-Expertise / Privé, PUB-Grand public/Citoyen</i> <i>Le Lesc propose un champ Réutilisation effective (O/N)</i>	O				O	X	
	Traitement sur les données pour partage*	Données anonymisées, pseudonymisées							X
Localisation	Lieu de stockage	Lieu de stockage et chemin ou adresse www où le jeu de données peut être trouvé <u>Commentaires</u> <i>Le Cirad propose un champ lien d'accès à la ressource et un autre support de stockage actuel. Liste à choix : Poste travail Cirad, Serveur interne géré par l'unité, Entrepôt externe – Editeur scientifique, Ddur externe/Cdrom/Dvd/Clé Usb, Serveur interne géré par le Cirad Entrepôt externe-Scientifique généraliste ou thématique, Poste travail partenaire, Serveur externe partenaire projet, Service externe Privé (Cloud, Drop Box, Amazon..). Autre...</i> <i>DAF : nom du champ Emplacement actuel</i>	X	X	X	X	X	X	X

Plateforme de partage des données*	Nom et adresse URL du dépôt de données ouvert où est partagé le jeu de données pour réutilisation				X		X	X
Couverture	Domaine intellectuel ou domaine couvert par les informations du jeu de données. <i>Commentaires :</i> <i>Le Cirad et le Lesc y intègrent également des informations sur la zone géographique et la couverture temporelle</i>	X				X	O	
Publication liée au jeu de données	Description des relations que l'actif de données entretient avec d'autres données (indiquer DOI, ISBN...). → ARODES <u>Commentaires</u> <i>DAF intitulé peu parlant : Relation</i>	X				O	O	X
Version	Version actuelle de l'ensemble de données	O						
Responsabilité du bien à long terme	Description de la politique de conservation et de gestion du bien à long terme.	O						
Peut-on / doit-on confier la conservation à un prestataire de services ?	L'organisation peut-elle prendre en charge la conservation du bien à long terme ou devra-t-elle faire appel à des experts ou à des services pour cela ?	O						
Sécurisation des données*	Evaluation du niveau de sécurisation de la ressource (quels moyens humains, techniques et organisationnels attribués à la gestion technique de la ressource) <u>Commentaires</u> <i>Cirad : Niveau de sécurisation : 0- aucun 1-faible 2-moyenne 3- forte</i>					X		

Préservation et conservation	Valeur à long terme	Description de la valeur que l'actif de données pourrait avoir à long terme. <u>Commentaires</u> <i>Cirad : importance stratégique , 0- aucun 1-faible 2-moyenne 3- forte</i>	X				X		X
	Politique sauvegarde/d'archivage	Nbre de copies actuellement stockées, fréquence des procédures de sauvegarde et d'archivage.	X						
	Mesures de reprise après sinistre	Description du processus de récupération au cas le jeu de données serait endommagé	X						
	Période de conservation	Date de fin prévue ou période de conservation du bien (si elle existe).	O						X
	Politique de préservation	Description de toute activité de préservation ou de conservation prévue ou appliquée	O						
Gestion	Format(s) de fichier	Format(s) de fichier et leur(s) version(s) utilisés	X		X	X	X	X	
	Structure du jeu de données	Détails de la structure du jeu de données <u>Commentaires</u> <i>Cirad si existence ou non d'une structure de données</i>	X				X		
	Documentation disponible	Documentation disponible (par exemple, manuels d'utilisation, tables de codage) <u>Commentaires</u> <i>Le Cirad questionne si la documentation est nécessaire pour comprendre et réutiliser la ressource ? pas nécessaire, oui nécessaire et disponible, oui mais plus disponible</i>	X				O	O	
	Piste d'audit et fixité	Description de toute mesure utilisée pour garantir l'authenticité des données (contrôle, piste d'audit).	X						
	Coût actuel	Coûts de maintenance actuels (par an) du support de données.	X						

Base de financement	Source de financement disponible pour le jeu de données actuellement et maintien à l'avenir	X						
Coût initial de création du jeu de données	Coût initial de création du jeu de données	O						
Coûts prévus de maintenance	Plans et coûts existants pour l'amélioration du jeu ou sa maintenance	O						
Taille	Taille du jeu de données <u>Commentaires</u> <i>Cirad propose en volumétrie les tranches suivantes : < 1 Mo, [1Mo ; 100 Mo], [101 Mo ; 500 Mo], [500 Mo ; 1Go], [1 Go ; 200Go], [200 Go ; 1 To], > 1To</i> <i>Mondaini donne le nombre de fichiers</i>	O	X		X	O	X	X
Exigences matérielles/logicielles	Description de toutes les exigences matérielles ou logicielles spécialisées	O						
Présence d'un DMP*	Un DMP a été ou va être réalisé						X	X
Contrat*	<u>Commentaires</u> <i>Liste à choix : Non, contrat de confidentialité, contrat de propriété intellectuelle, accord de transfert et d'utilisation des données, autres</i>							X
Analyse de la qualité*	<u>Commentaires</u> <i>Mondaini : Principes FAIR : pour chaque principe une évaluation (non-conforme, vague, conforme) ; Data quality dimension : exhaustivité, unicité, validité (en termes de pourcentage ainsi qu'une évaluation (haut, ...)) ; Critères infonomics : intrinsic value of information et business value of information (p. 69).</i> <i>Lesc : Contrôle de qualité technique oui-non</i>		X				O	

	Etat de finalisation du projet*	<u>Commentaires</u> <i>Proposé par les personnes ressources : En cours, En préparation, terminé scientifiquement, terminé administrativement, archivé</i>						X	X
	Date de mise à jour de la fiche*					X			

Annexe 7 : Choix des métadonnées de l'inventaire

Bleu : Description du projet, Jaune : Description du jeu de données, Orange : Provenance des données, Vert : Modalités de partage, Gris : modalité de stockage et archivage, Violet : gestion, Rose : qualité

	Champs	Définition	Type de champ	Liste des options	Remarque
1	ID	Numéro SageX	texte libre		
2	Acronyme	Nom alternatif ou communément utilisé, s'il existe, acronyme	texte libre		
3	Nom du projet de recherche	Nom du projet de recherche dans lequel est créé le jeu de données	texte libre		
4	Responsable du projet	Nom Prénom	texte libre		Nom Prénom
5	Filière	Informations sur la couverture du sujet des données	case à cocher	SI PHY	
6	Dates de début projet	AAAA-AAAA	texte libre		Format : AAAA-AAAA
7	Dates de fin projet	AAAA-AAAA	texte libre		Format : AAAA-AAAA
8	Types de financement	Type de financement obtenu	case à cocher	FNS, Fondation/Association, Recherche et impulsions, Ecole/établissement, Recherche et mandat, InnoSuisse, Organisme public, PNR, SEFRI, Socle RA&D, Autre (à préciser)	
9	Titre du jeu de données	Nom officiel du jeu de données	texte libre		
10	Responsable des données	Nom prénom de la personne responsable de la gestion du jeu de données et si externe le nom de l'institution	texte libre		Nom Prénom
11	Langue	Langue(s) du contenu des données	choix multiple	FR DE	

				EN PR autre	
12	Type de données	Type(s) de données du jeu de données	choix multiple	Données d'entretien Données de questionnaire Données d'observation de terrain Données expérimentales Données de simulation, Données dérivées, Données de référence,	<p>Données d'observation : sont capturées en temps réel, la plupart du temps en dehors du laboratoire. Elles sont habituellement uniques et donc impossibles à reproduire. Elles sont ici séparées en deux catégories :</p> <p>a) données d'entretiens b) données de questionnaires c) données d'observation de terrain</p> <p>Données expérimentales : produites en laboratoire, dans un contexte contrôlé et donc reproductibles, ex. Essais, expérimentation, analyses laboratoire</p> <p>Données de simulation ou computationnelle : générées par des modèles informatiques ou de simulation. Elles sont souvent reproductibles à condition que le modèle soit correctement documenté.</p> <p>Données dérivées ou compilées : résultant du traitement, de la sélection, de la compilation ou de l'agrégation de données brutes. Elles sont souvent reproductibles mais coûteuses. Ex. fouille de texte, bases de données compilées, modèles 3D.</p> <p>Données de référence : données publiées antérieurement qui sont collectées, triées et agrégées ; ex. sources textuelles primaires tiré de (Doranum, 2022)</p>
13	Forme des données	Description de la forme du jeu de données	choix multiple	Matérielles et physiques, Textuelles, Tableurs, Base de données, Audio , Vidéo, Images,	si autre préciser sous le champ 30 Remarques

				Géospatiales, 3D, Modèles, Visualisations, Codes sources, Logiciels, Programmes informatiques, Autre (à préciser)	
14	Format(s) de fichier	format(s) de fichier	texte libre		Liste des formats si nécessaire
15	Taille	Taille du jeu de données en b/Gb	case à cocher	< 1 Mo [1Mo ; 100 Mo] [101 Mo ; 500 Mo] [500 Mo ; 1Go] [1 Go ; 200Go] [200 Go ; 1 To] > 1To	
16	Description	Courte description des informations contenues dans le jeu de données.	Texte libre		
17	Date de début – fin de collecte	Date de début et de fin de la collecte des données	texte libre		Format date : jj-mm-aaaa – jj-mm-aaaa ou mm-aaaa – mm-aaaa
18	Source des données	Le jeu de données contient-il des données produites, réutilisées ou une combinaison des deux ?	Choix multiple	Données produites, Données réutilisées	Si le jeu de données contient des données réutilisées, ajouter des informations sur le nom du fournisseur et le mode d'acquisition sous le champ 30 Remarques
19	Licence	Indication sur la licence utilisée pour le jeu de données	case à cocher	Fermé, CC0, CC-BY, CC-BY-SA, CC-BY-NC-SA, CC-BY-NC, CC-BY-NC-ND, CC-ND, Pas encore définie, Autre (à préciser)	Si autre à préciser sous le champ 30 Remarques
20	Niveau d'accès	Restrictions d'accès appliquées au jeu de données	case à cocher	Privé, Institution, Public,	

				Restreint sur demande	
21	Niveau de sensibilité des données	Description du niveau de sensibilité des données	choix multiple	Sans sensibilité, Confidentielles, Personnelles, Sensibles, Soumises à la LRH	Si le jeu de données contient des données soumises à la LRH, ajouter dans le champ Remarques la date de la validation par la commission d'éthique
22	Réutilisations potentielles	Description des réutilisations potentielles qui pourraient être envisagées	choix multiple	Enseignement/Etudiants, Partenariat institutionnel, Recherche/Chercheurs, Expertise / Privé, Grand public/Citoyen, Pas de réutilisation possible	
23	Emplacement(s) actuel(s)	Lieu(x) de stockage, archivage, partage des données	choix multiple	Ordinateur portable personnel, Ordinateur portable institutionnel, Disque dur externe/Clé USB, Serveur interne de l'école, SWITCHdrive , Service externe privé (Dropbox, Google drive), Serveur externe partenaire projet (REDCap), Dépôt externe de l'éditeur scientifique, Dépôt ouvert généraliste ou thématique (Olos, SwissUbase, Zenodo...), Autre (à préciser)	Si autre à préciser sous le champ 30 Remarques
24	URL de l'emplacement	Chemin d'accès ou adresse URL de(s) emplacements de données	texte libre		
25	Publication associées	Lien sur la publication ArODES liées au jeu de données	texte libre		
26	Période de conservation	Date de fin prévue ou période de conservation du bien (si elle existe).	Texte libre		Format date : jj-mm-aaaa

27	Documentation disponible	Une documentation accompagne-t-elle vos données ?	Choix multiple	Aucune Readme Codebook (dictionnaire de variables) Cahier de laboratoire électronique (ELN) Carnet de terrain Manuels d'utilisation Intégrée directement dans le fichier de données Protocole Autre (à préciser)	La documentation va fournir des informations contextuelles et explicatives (métadonnées) qui vont donner du sens aux données, afin d'en faciliter l'utilisation, la récupération et la gestion. On y trouve généralement des informations sur le projet (contexte, instruments de recueil de données, échantillonnage, protection données) ainsi que sur les données elles-mêmes (contexte récolte, préparation et nettoyage, procédures méthodes et outils d'analyses...) (FORS, 2018) Si autre à préciser sous le champ 30 Remarques
28	Présence d'un DMP	Un DMP (data management plan) a/va être réalisé	case à cocher	Oui, Non, Prévu	Note : merci de transmettre une copie du DMP pour archivage dans SageX
29	Signature de contrats	Contrats signés dans le cadre du projet	choix multiple	Non, Contrat de confidentialité, Contrat de propriété intellectuelle, Accords de transfert et d'utilisation, Autre (à préciser)	Note : merci de transmettre une copie des contrats pour archivage dans SageX
30	Remarques sur les données		texte libre		
31	Etat de finalisation du projet	Etat de finalisation du projet de recherche	case à cocher	En cours, En préparation, Terminé scientifiquement, Terminé administrativement, Archivé	NE PAS REMPLIR CE CHAMP
32	Date de mise à jour de la fiche	Date de la dernière mise à jour des informations du jeu de données sur le fichier Excel	texte libre		Format date : jj-mm-aaaa NE PAS REMPLIR CE CHAMP

33	Analyse de la qualité	Mondaini : Principes FAIR : pour chaque principe une évaluation (non-conforme, vague, conforme) ; Data quality dimension : exhaustivité, unicité, validité (en termes de pourcentage ainsi qu'une évaluation (haut, ...)) ; Critères infonomics : intrinsic value of information et business value of information (p. 69).			NE PAS REMPLIR CE CHAMP
34	Valeur à long terme	Description de la valeur que pourrait avoir le jeu de données à long terme	choix multiple	Légale, Contractuelle, Commerciale, Scientifique (réutilisation)	NE PAS REMPLIR CE CHAMP

Annexe 8 : Guide d'entretien des chercheuses et chercheurs

ID :

Date de la séance :

MODÉLISATION ET INTERACTIONS

- **Validité des schémas et personnes ressources** [ordre des phases, des tâches effectuées, intervention(s) des personnes ressources dans les phases du projet]

PLANIFICATION

- **Planification préalable** : [étapes de planification, (collecte des données, outils informatiques nécessaires, types de données (sensible – non), volumétrie, stockage, préservation)...]
- **DMP** [obligation, canevas utilisé, relectures, partage du DMP]
- **Signature de contrats liés aux données** [quels types, forme, fréquence (chaque projet, seulement si imposé), transmission au secrétariat/juriste, (contrat confidentialité, de transfert et d'usage, d'appariement, de propriété intellectuelle, de collaboration...)]
- **Connaissance du soutien / personnes ressources** [DMP, contrats, montage financier, préparation dossier éthique, solutions informatiques et analyse de risques, données personnelles, sensibles, LRH]
- **Soutien** [difficultés rencontrées, besoins de soutien pour cette étape, checklist, entretien avec la DM, remplir un DMP]

COLLECTE

- **Catégories données sources** [avant tout traitement et analyse, données collectées (primaires) ou réutilisées (secondaires) fournisseurs de données (hôpital, OFSP, autres), type de contrat, méthode de transfert, contraintes particulières, difficultés rencontrées]
- **Logiciels de collecte**
- **Organisation et partage des données durant le projet** [règles, responsabilité, méthode, convention nommage, organisation des fichiers, contrôle version, conditions accès]
- **Données personnelles – sensibles** [Signalement au DPO, commission d'éthique, gestion (enregistrements), consentement (publication, réutilisation), Anonymisées/pseudonymisées, (quel moment, outils, accès), chiffrement (outils)]
- **Documentation des données** [type de documentation (readme, protocole...), standards de métadonnées utilisé + connu, mode de fonctionnement, étapes qui sont documentées (collecte, traitement/analyse, partage, réutilisation, stockage), Information scientifique (méthodologie d'enquête, plan d'échantillonnage...), Information technique (logiciel utilisé, liste des fichiers de données, format, ...),

Information standardisée (titre, auteur, mots-clés, date de création, droits associés sur les accès et les licences, DOI, difficultés rencontrées]

- **Stockage et sauvegarde** [fréquence, lieu(x) de sauvegarde, tout/partie]
- **Soutien** [difficultés rencontrées, besoins de soutien pour cette étape]

TRAITEMENT / ANALYSE

- **Type(s) de traitement/ analyse/nettoyage** [méthodes, outils]
- **Soutien** [difficultés rencontrées, besoins de soutien pour cette étape]

PUBLICATION ET PARTAGE

- **Dépôts de données ouverts** [connaissance de dépôts, types dépôts, avantages/inconvénients]
- **Partage des données** [déjà déposé des données, où, nom(s), pourquoi ce choix, autres dépôts connus, type accès (libre, restreint, embargo – pourquoi ce choix, critères), licence (pourquoi), si non pourquoi, freins liés au partage, souhait de partage de jeux]
- **Soutien** [difficultés rencontrées, besoins de soutien pour cette étape]

PRÉSERVATION

- **Tri et évaluation** [tri, élimination avant de clore, critères de conservation]
- **Archivage** [formats ouverts, lieux pour l'archivage, vérification]
- **Soutien** [difficultés rencontrées, besoins de soutien pour cette étape]

RÉUTILISATION

- **Réutilisation** : [soit propres données, soit données autres personnes, interne école, déposées sur dépôts ouverts, quel(s) dépôts, méthode découverte]
- **Soutien** [difficultés rencontrées, besoins de soutien pour cette étape]

MISE EN PLACE D'UN INVENTAIRE

- **Fichier Excel d'inventaire** [difficultés rencontrées, compréhension des champs, champs manquants, champs non pertinents, propositions améliorations, remarques]
- **Processus de mise à jour de l'inventaire** [personne responsable, moment(s) pour la récolte, comment rappeler aux gens de participer, besoins, intérêt pour checklist, arbre décisionnel, rencontre au départ du projet]

Annexe 9 : Cahier des charges

1. Analyse de l'existant

Différents outils déjà utilisés au sein de l'école pour la gestion des activités de recherche récoltent des informations utiles à la tenue de l'inventaire :

- L'application Sage-X pour la gestion des projets validés
- Un fichier Excel est utilisé par la CA pour assurer le suivi des projets. Le fichier contient un onglet par chercheuse et chercheur pour les projets en préparation, en cours et archivés.
- L'application « Smart global privacy » gère entre autres le registre de traitements du DPO
- Un fichier Excel pour la gestion de l'archivage des publications (sur ArODES et sur le serveur interne) est utilisé par la Spécialiste ID
- Le DMP du projet contient une grande partie des informations nécessaires à l'inventaire
- Les dépôts de données ouverts archivent et partagent les données de recherche des équipes de recherche

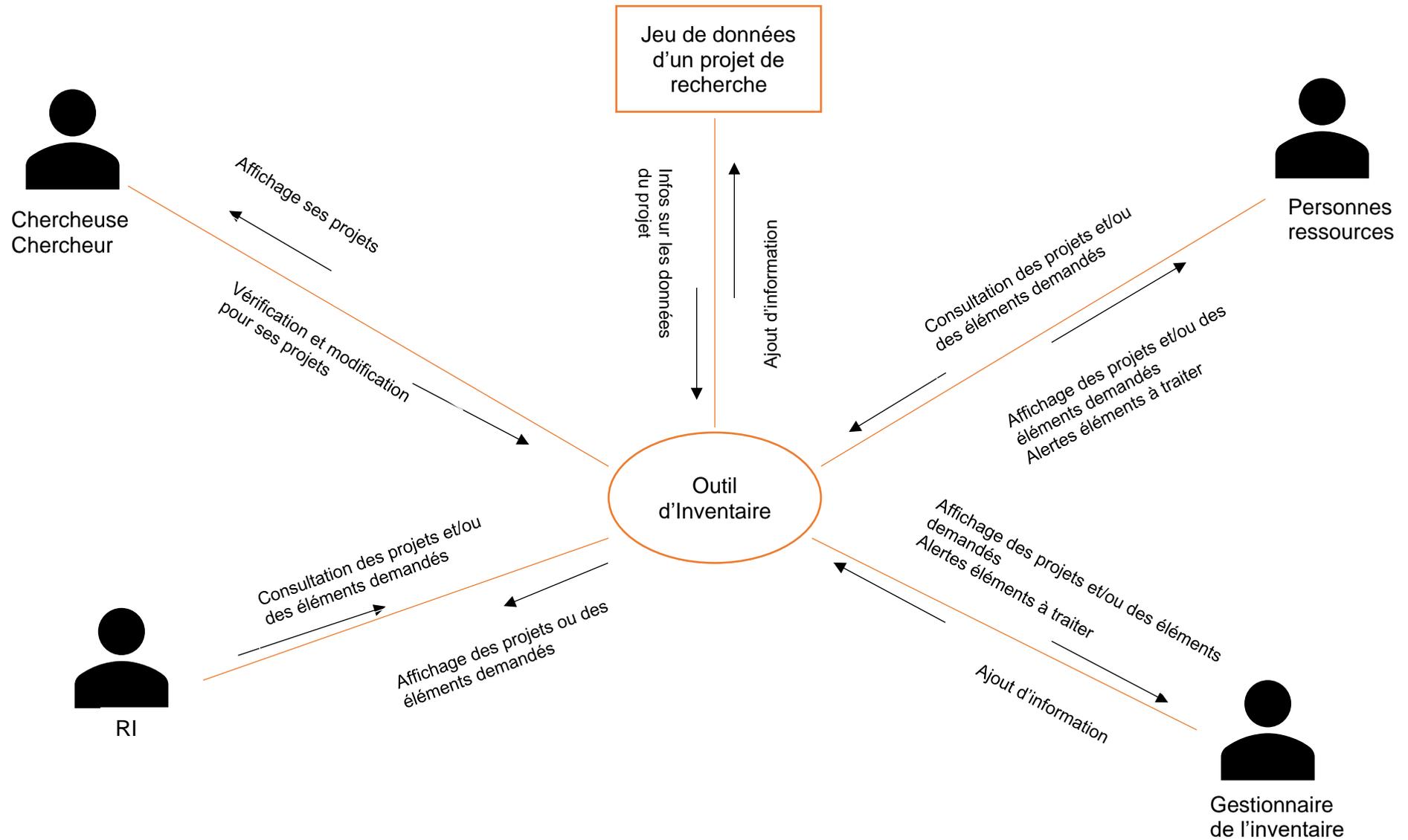
2. Objectifs de la solution

L'objectif de l'inventaire est de répertorier les jeux de données produits dans le cadre des recherches de l'IS et d'en assurer un suivi administratif et scientifique. Il va permettre d'assurer la traçabilité des données de recherche, de veiller au respect des exigences légales, réglementaires et contractuelles, de consacrer les moyens et ressources (financière, sécuritaire, personnel) adéquats et de favoriser la pérennisation, l'accès et la réutilisation des données.

Les indicateurs suivants peuvent être utilisés pour vérifier l'atteinte des objectifs :

- Nombre de nouveaux jeux de données enregistrés dans l'inventaire par rapport au nombre de nouveaux projets créés dans SageX
- Nombre de jeux de données mis à jour dans l'inventaire par rapport au nombre de projets terminés scientifiquement et administrativement dans SageX.

3. Périmètre du projet



4. Acteurs et utilisateurs

Partie prenante	Rôle	Objectifs de l'inventaire	Problématique	Niveau accès	Fréquence
DG	Responsable stratégique		Respect exigences légales, réglementaires, contractuelles		
Responsable institut	Utilisateur direct	Assurer un suivi administratif et scientifique des données de l'IS	Manque de connaissance des données produites Améliorer la valorisation des données Manque de réutilisation données	Consultation	Occasionnel
Gestionnaire de l'inventaire	Gestionnaire de l'inventaire	Permettre de maintenir à jour les informations sur les données de recherche de chaque projet	Obtenir les informations de la part des chercheuses et chercheurs	Utilisation	Régulier
Personnes ressources	Utilisateur occasionnel	Fournir les informations nécessaires pour s'assurer de la bonne gestion des données de recherche	Pas de connaissance de la manière dont sont gérées les données (sécurité, respect exigences éthiques et légales) Pas de suivi par rapport aux données (stockage, accès, destruction)	Consultation	Occasionnel
Chercheuses et chercheurs	Fournisseur d'informations	Enregistrer les informations sur les données de leurs projets de recherche	Appliquer les demandes de l'institution dans le temps imparti Manque de cadres, de guidelines pour les accompagner	Consultation limitée à leurs projets	Une fois par année
Responsable produit	Fournisseur du produit	Création et/ou maintien de la solution informatique		Administration	Occasionnel

5. Fonctionnalités

5.1 Principales

- Intégration de 35 champs répartis en 6 catégories. Liste en page suivante.
- Intégrer différents types de champs : texte libre, liste déroulante, liste à choix multiple
- Ajouter ou supprimer un champ et/ou une catégorie si nécessaire
- Ajouter ou supprimer un ou des jeux de données rattachés à un projet de recherche
- Proposer des bulles informatives contenant une définition du champ.
- Respecter un format normalisé pour certains champs comme pour les dates, jj-mm-aaaa ou aaaa,
- Avoir des niveaux d'accès en fonction du type d'utilisateur : lecture, écriture
 - Accès restreint pour la-le PI, aux jeux de données la-le concernant
- Possibilité de rajouter des documents : DMP, contrats, protocoles qui sont reliés au jeu de données et/ou au projet concerné
- Recevoir des alertes automatiques lorsque des dates de préservation arrivent à échéance
- Effectuer des recherches sur un ou plusieurs champs
- Avoir des options de tri en fonction des champs et des critères de chaque champ
- Réaliser des extractions en fonction des critères de tri sélectionnés
- Réaliser des impressions en fonction des critères de tri sélectionnés

5.2 Secondaires

- Récupérer automatiquement certaines données depuis d'autres plateformes comme Sage-X, l'outil du DPO, le fichier Excel du secrétariat, ArODES, Dépôt de données...
- Envoi de courriels automatiques lorsqu'une action doit être effectuée (suppression d'un jeu de données)
- Envoi de courriels automatiques à la personne ressource concernée lors de l'ajout ou la modification d'un projet nécessitant son attention

5.3 Liste des champs de l'inventaire

Bleu : Description du projet, Jaune : Description du jeu de données, Orange : Provenance des données, Gris : modalité de stockage et archivage, Vert : Modalités de partage, Violet : gestion, Rose : qualité

	Champs	Définition	Type de champ	Liste des options	Remarque
1	ID	Numéro SageX	texte libre		
2	Acronyme	Nom alternatif ou communément utilisé, s'il existe, acronyme	texte libre		
3	Nom du projet de recherche	Nom du projet de recherche dans lequel est créé le jeu de données	texte libre		
4	Responsable du projet	Nom Prénom	texte libre		Nom Prénom
5	Filière	Informations sur la couverture du sujet des données	case à cocher	SI PHY	
6	Dates de début projet	AAAA-AAAA	texte libre		Format : AAAA
7	Dates de fin projet	AAAA-AAAA	texte libre		Format : AAAA
8	Types de financement	Type de financement obtenu	case à cocher	FNS, Fondation/Association, Recherche et impulsions, Ecole/établissement, Recherche et mandat, InnoSuisse, Organisme public, PNR, SEFRI, Socle RA&D, Autre (à préciser)	
9	Titre du jeu de données	Nom officiel du jeu de données	texte libre		
10	Responsable des données	Nom prénom de la personne responsable de la gestion du jeu de données et si externe le nom de l'institution	texte libre		Nom Prénom
11	Langue	Langue(s) du contenu des données	choix multiple	FR DE EN	

				PR Autre (à préciser)	
12	Type de données	Type(s) de données du jeu de données	choix multiple	Données d'entretien Données de questionnaire Données d'observation de terrain Données expérimentales Données de simulation, Données dérivées, Données de référence,	<p>Données d'observation : capturées en temps réel, la plupart du temps en dehors du laboratoire, habituellement uniques et donc impossibles à reproduire. Elles sont séparées en 3 catégories : a) données d'entretien, b) données de questionnaires, c) données d'observation de terrain</p> <p>Données expérimentales : produites en laboratoire, dans un contexte contrôlé et donc reproductibles, ex. Essais, expérimentation, analyses laboratoire</p> <p>Données de simulation : générées par des modèles informatiques ou de simulation. Elles sont souvent reproductibles à condition que le modèle soit correctement documenté . Ex. modélisation, calcul</p> <p>Données dérivées ou compilées : résultant du traitement, de la sélection, de la compilation ou de l'agrégation de données brutes. Elles sont souvent reproductibles mais coûteuses. « Les données dérivées peuvent très justement être décrites comme des données qui ont été générées à partir de données préexistantes. Les données dérivées prennent souvent des données qui ont déjà été collectées et les modifient ou y ajoutent de la valeur afin de créer une toute nouvelle interprétation des données. Un exemple de données dérivées serait lorsqu'un chercheur prend des données de phénotype précédemment collectées et les combine avec des données de génotype nouvellement générées. Cette combinaison crée un nouvel ensemble de données dérivé de données précédemment collectées ». (National Library of Medicine [sans date])Ex. fouille de texte, bases de données compilées</p>

					Données de référence : collection de petits jeux de données revus par les pairs et publiés antérieurement qui sont collectées, triées et agrégées ; ex. corpus de textes, recueils statistiques, cartes et plans (Doranum, 2022)
13	Forme des données	Description de la forme du jeu de données	choix multiple	Matérielles et physiques, Textuelles (word, PDF), Tableurs (csv, excel), Base de données (Access), Audio , Vidéo, Images, Visualisations, Codes sources, Logiciels, Programmes informatiques, Autre (à préciser)	si autre préciser sous le champ 30 Remarques
14	Format(s) de fichier	format(s) de fichier	texte libre		<u>Liste des formats si nécessaire</u>
15	Taille	Taille du jeu de données en b/Gb	case à cocher	< 1 Mo [1Mo ; 100 Mo] [101 Mo ; 500 Mo] [500 Mo ; 1Go] [1 Go ; 200Go] [200 Go ; 1 To] > 1To	
16	Description	Courte description des informations contenues dans le jeu de données.	Texte libre		
17	Date de début – fin de collecte	Date de début et de fin de la collecte des données	texte libre		Format date : jj-mm-aaaa – jj-mm-aaaa ou mm-aaaa – mm-aaaa
18	Source des données	Le jeu de données contient-il des données produites, réutilisées ou une combinaison des deux ?	Choix multiple	Données produites, Données réutilisées	Si le jeu de données contient des données réutilisées, ajouter des informations sur le nom du fournisseur et le mode d'acquisition sous le champ 30 Remarques
19	Niveau de sensibilité des données	Description du niveau de sensibilité des données	choix multiple	Sans sensibilité, Confidentielles,	Données personnelles liées à la santé : les informations concernant une personne

				Personnelles, Sensibles, Soumises à la LRH	déterminée ou déterminable qui ont un lien avec son état de santé ou sa maladie, données génétiques comprises Données sensibles : sont des données personnelles qui portent sur les opinions ou activités religieuses, philosophiques, politiques ou syndicales, la santé, la sphère intime ou l'appartenance à une race, des mesures d'aide sociale, des poursuites ou sanctions pénales et administratives, données génétiques ainsi que les données biométriques identifiant une personne de physique de manière univoque Données personnelles : toutes les informations qui se rapportent à une personne identifiée ou identifiable. On y trouve des identifiants directs comme le nom, le numéro de téléphone, le numéro AVS, les empreintes digitales... ou indirects qui une fois croisés peuvent relever l'identité d'une personne. Données confidentielles Si le jeu de données contient des données soumises à la LRH, ajouter dans le champ Remarques la date de la validation par la commission d'éthique
20	Mesures liées à la protection des données	Description des mesures prises pour assurer la protection des données durant la phase de réalisation du projet	Choix multiple	Chiffrement Anonymisation Codage Autre (à préciser)	
21	Emplacement de stockage et d'archivage	Lieu(x) de stockage et d'archivage des données	choix multiple	Ordinateur portable Disque dur externe/Clé USB, NAS OneDrive Teams Sharepoint, SWITCHdrive personnel SWITCHdrive folders REDCap VM Serveur externe partenaire projet	Si autre à préciser sous le champ 30 Remarques

				Autre (à préciser)	
22	Période de conservation	Date de fin prévue ou période de conservation (si elle existe).	Texte libre		Format date : jj-mm-aaaa
23	Publication associées	Lien sur la publication ArODES liées au jeu de données	texte libre		
24	Partage des données	Nom du dépôt sur lequel sont partagés les données	Texte libre		dépôt externe de l'éditeur scientifique ou dépôt ouvert généraliste ou thématique (Olos, SwissUbase, Zenodo, OSF...),
25	Licence	Indication sur la licence utilisée pour le partage du jeu de données	case à cocher	Fermé, CC0, CC-BY, CC-BY-SA, CC-BY-NC-SA, CC-BY-NC, CC-BY-NC-ND, CC-ND, Pas encore définie, Autre (à préciser)	Si autre à préciser sous le champ 30 Remarques <u>Liste des licences creatives commons</u>
26	Niveau d'accès	Restrictions d'accès appliquées au jeu de données	case à cocher	Privé Sous embargo Restreint sur demande	
27	Réutilisations potentielles	Description des réutilisations potentielles qui pourraient être envisagées	choix multiple	Enseignement/Etudiants, Partenariat institutionnel, Recherche/Chercheurs, Expertise / Privé, Grand public/Citoyen, Pas de réutilisation possible	
28	Documentation disponible	Une documentation accompagne-t-elle vos données ?	Choix multiple	Aucune Readme Codebook (dictionnaire de variables) Cahier de laboratoire électronique (ELN) Carnet de terrain Manuels d'utilisation	La documentation va fournir des informations contextuelles et explicatives (métadonnées) qui vont donner du sens aux données, afin d'en faciliter l'utilisation, la récupération et la gestion. On y trouve généralement des informations sur le projet (contexte, instruments de recueil de données, échantillonnage, protection données) ainsi que sur les données elles-mêmes (contexte)

				Intégrée directement dans le fichier de données Protocole Autre (à préciser)s	récolte, préparation et nettoyage, procédures méthodes et outils d'analyses...) (FORS, 2018) Si autre à préciser sous le champ 30 Remarques
29	Présence d'un DMP	Un DMP (data management plan) a/France être réalisé	case à cocher	Oui, Non, Prévu	Note : merci de transmettre une copie du DMP pour archivage dans SageX
30	Signature de contrats	Est-ce que des contrats ont été signés dans le cadre du projet	choix multiple	Non, Contrat de confidentialité, Contrat de propriété intellectuelle, Accords de transfert et d'utilisation, Contrat de prestations Autre (à préciser)	Note : merci de transmettre une copie des contrats pour archivage dans SageX
31	Remarques sur les données		texte libre		
32	Etat de finalisation du projet	Etat de finalisation du projet de recherche	case à cocher	En cours, En préparation, Terminé scientifiquement, Terminé administrativement, Archivé	
33	Date de mise à jour de la fiche	Date de la dernière mise à jour des informations du jeu de données sur le fichier Excel	texte libre		Format date : jj-mm-aaaa
34	Analyse de la qualité	Analyse de la valeur des jeux de données			Mondaini : Principes FAIR : pour chaque principe une évaluation (non-conforme, vague, conforme) ; Data quality dimension : exhaustivité, unicité, validité (en termes de pourcentage ainsi qu'une évaluation (haut, ...)) ; Critères infonomics : intrinsic value of information et business value of information
35	Valeur à long terme	Description de la valeur que pourrait avoir le jeu de données à long terme	choix multiple	Légale, Contractuelle, Commerciale, Scientifique (réutilisation)	

Définitions

But de l'inventaire : Assurer un suivi administratif et scientifique des données de recherche de l'Institut santé afin de veiller au respect des exigences légales, réglementaires et contractuelles, et de consacrer les moyens et ressources adéquats pour favoriser la sécurité, la pérennisation, l'accès et la réutilisation des données.

Qu'entend-on par données de recherche : « Les données de recherche sont les preuves qui sous-tendent la réponse à la question de recherche, et peuvent être utilisées pour valider les résultats, quelle que soit leur forme (par exemple, imprimée, numérique ou physique). Il peut s'agir d'informations quantitatives ou d'énoncés qualitatifs recueillis par les chercheurs au cours de leur travail par l'expérimentation, l'observation, la modélisation, l'entretien ou d'autres méthodes, ou d'informations dérivées de preuves existantes. Les données peuvent être brutes ou primaires (par exemple, directement issues de mesures ou de collectes) ou dérivées de données primaires en vue d'une analyse ou d'une interprétation ultérieure (par exemple, nettoyées ou extraites d'un ensemble de données plus vaste), ou encore dérivées de sources existantes dont les droits peuvent être détenus par d'autres. Les données peuvent être définies comme des éléments « relationnels » ou « fonctionnels » de la recherche, ce qui signifie que leur identification et leur valeur dépendent de la manière dont les chercheurs les utilisent pour étayer leurs affirmations [comme preuve pour valider des résultats/leurs affirmations]. Il peut s'agir, par exemple, de statistiques, de collections d'images numériques, d'enregistrements sonores, de transcriptions d'entretiens, de données d'enquêtes et d'observations de terrain avec des annotations appropriées, d'une interprétation, d'une œuvre d'art, d'archives, d'objets trouvés, de textes publiés ou d'un manuscrit. Le but premier des données de recherche est de fournir l'information nécessaire pour appuyer ou valider les observations, les résultats ou les conclusions d'un projet de recherche. » (High Education Funding Council for England et al. 2016)

Qu'entend-on par jeu de données : Un jeu de données ou dataset est un ensemble ou une collection de données cohérentes, « produites dans le cadre d'un même projet, sur un même objet d'étude et/ou recueillies sur un même lieu. Toutes les données d'un dataset peuvent donc être décrites avec une majorité de métadonnées communes. Il comprend des datafiles (fichiers de données), mais aussi le code, la documentation et les métadonnées associées » (IRD data, 2022).

Niveau de profondeur de description (granularité) : Un projet de recherche peut comprendre plusieurs jeux de données. Ex. un jeu de données d'entretiens ainsi que des questionnaires, chacun avec un niveau de sensibilité, un lieu de stockage, des possibilités de partage différents d'où l'intérêt de les décrire séparément en rajoutant autant de lignes que souhaité dans le fichier Excel.

Références

DORANUM, 2022. L'origine et la description des données de la recherche. DoRANum. [en ligne]. 15.03.2022. [Consulté le 6 mai 2022]. Disponible à l'adresse : <https://doranum.fr/plan-gestion-donnees-dmp/origine-description-donnees-recherche>

HIGH EDUCATION FUNDING COUNCIL FOR ENGLAND, RESEARCH COUNCILS UK, UNIVERSITIES UK, et WELLCOME, 2016. *Concordat on Open Research Data*. [en ligne]. 28 août 2016. [Consulté le 26 mai 2022]. Disponible à l'adresse : <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>

INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT, 2021. Définitions. IRD data. [en ligne]. 2021. [Consulté le 21 novembre 2021]. Disponible à l'adresse : <https://data.ird.fr/definitions/>

6. Exigences non fonctionnelles

6.1 Sécurité

Seules quelques personnes mentionnées sous la partie 4 de l'annexe 9 peuvent accéder au fichier. La-le gestionnaire de l'inventaire est la seule personne qui peut avoir un droit d'écriture pour l'ensemble de la base. Elle-il peut ajouter, modifier, supprimer un jeu de données.

Les chercheuses et chercheurs peuvent consulter uniquement les jeux de données de leurs projets. Une fois par année, elles-ils ont la possibilité d'ajouter, modifier ou supprimer des informations concernant leurs jeux de données, lors de la vérification de fin d'année (voir 3.6.1).

Les personnes ressources et le RI ont uniquement le droit de consulter les informations.

7. Contraintes

7.1 Hébergement, sauvegarde, accessibilité

La solution devra être hébergée sur le serveur interne de l'IS, au même endroit que les activités de recherche ainsi une cohérence est conservée. Une sauvegarde journalière réalisée par le SINP est assurée pour cet emplacement. De plus il peut être accessible hors de l'institution via le VPN.

Annexe 10 : Descriptif des outils d'inventaire

Logiciel Tableur Excel

But	Le logiciel «intègre des fonctions de calcul numérique, de représentation graphique, d'analyse de données (notamment de tableau croisé dynamique) et de programmation » (<i>Microsoft Excel 2022</i>)	
Coût :	Payant par abonnement ou achat définitif, inclus dans Microsoft 365. Déjà pris en charge par la HEdS. Pas de coût supplémentaire.	
Données techniques :	Système d'exploitation : SE Windows : Windows 11, Windows 10, Windows 10 LTSC 2021, Windows 10 LTSC 2019, Windows Server 2022 ou Windows Server 2019 Logiciel qui provient de la suite bureautique office 365 Dernière version : Excel 2021 (octobre 2021)	
Fonctionnalités	Intégration de 34 champs répartis en 6 catégories. Liste en annexe 9.	X
	Intégrer différents types de champs : texte libre, liste déroulante, liste à choix multiple	X
	Ajouter ou supprimer un champ et/ou une catégorie si nécessaire	X
	Ajouter ou supprimer un ou des jeux de données rattachés à un projet de recherche	X
	Proposer des bulles informatives contenant une définition du champ	X
	Respecter un format normalisé pour certains champs comme pour les dates, jj-mm-aaaa ou aaaa,	(X)
	Avoir des niveaux d'accès en fonction du type d'utilisateur : lecture, écriture	X
	Possibilité de rajouter des documents : DMP, contrats, protocoles	(X) ajout lien
	Recevoir des alertes automatiques lorsque des dates de préservation arrivent à échéance	X
	Effectuer des recherches sur un ou plusieurs champs	X
	Avoir des options de tri en fonction des champs et des critères de chaque champ	X
	Réaliser des extractions en fonction des critères de tri sélectionnés	X
	Réaliser des impressions en fonction des critères de tri sélectionnés	X
Récupérer automatiquement certaines données : depuis d'autres plateformes comme Sage-X, l'outil du DPO, le fichier Excel du secrétariat, ArODES, Dépôt de données...		

	Envoyer des courriels automatiques lorsqu'une action doit être effectuée (suppression d'un jeu de données)	X
	Envoyer des courriels automatiques à la personne ressource concernée lors de l'ajout ou la modification d'un projet nécessitant son attention	X
Maintenance :	Pris en charge jusqu'en octobre 2026	
Lien	https://www.microsoft.com/fr-ch/microsoft-365/excel	
Propriétaire	Microsoft	
Exemples d'institutions utilisant cette solution	Il a été utilisé notamment par le Cirad, le Lesc, HESAV, le NAA ⁵² ainsi que par plusieurs autres institutions ayant réalisés des registres d'actifs informationnels. On peut citer notamment : NHS East Staffordshire Clinical Commissioning Group ⁵³ , State of Victoria, Department of Health and Human Services ⁵⁴ , Organisation des Nations Unies ⁵⁵ , UK Government ⁵⁶ (Bischoff et al. 2021).	
Avantages	<p><u>Alertes et annotations peuvent être programmées</u> soit pour apparaître dans le fichier soit pour recevoir un courriel</p> <p>Pas de coût supplémentaire</p> <p>Accès anciennes versions du fichier</p> <p>Simplicité d'utilisation, connaissance des utilisateurs</p> <p>L'utilisation de macros permet de créer des listes à choix multiple et permet de développer les fonctionnalités de l'outil</p> <p>Création d'alertes, de notification et d'envoi de courriels automatiques</p>	
Contraintes	<p>Risque d'erreurs élevé dans la saisie</p> <p>N'est pas possible de rendre seulement visible certaines lignes ou certaines colonnes en fonction de l'utilisateur – niveau accès selon le type d'utilisateur</p> <p>Pas possible d'ajouter des pièces jointes</p> <p>Interface limitée (si beaucoup de colonne, perte de lisibilité)</p> <p>Pas de synchronisation et d'interopérabilité avec d'autres logiciels</p>	
Autres versions	Google Sheets, LibreOffice Calc	
Sources	(Microsoft 2022a; <i>Microsoft Excel 2022</i> ; ExtendOffice 2022; AGNES Tutos et Formations 2021)	

⁵² <https://www.naa.gov.au/information-management/information-governance/conducting-information-review/designing-and-maintaining-information-asset-register#information-asset-register-template>

⁵³ <https://eaststaffscg.nhs.uk/publications/information-assets-register/>

⁵⁴ <https://www.dhhs.vic.gov.au/department-health-and-human-services-information-asset-register-public>

⁵⁵ <https://archives.un.org/fr/content/field-information-management>

⁵⁶ <https://www.gov.uk/government/publications/information-asset-register>

Système de gestion de base de données relationnelles (type Access)

But	Application qui permet de créer, éditer et mettre à jour une base de données contenant des tables, des formulaires ainsi qu'une interface visuelle (WikiHow [sans date])	
Coût :	Payant par abonnement ou achat définitif, inclus dans Microsoft 365. Déjà pris en charge par la HEdS. Pas de coût supplémentaire.	
Données techniques :	<p>Fait partie de la suite Microsoft Office</p> <p>Composé de plusieurs programmes : « le moteur de base de données Microsoft Jet, un éditeur graphique, une interface de type Query by Example pour interroger les bases de données, et le langage de programmation Visual Basic for Applications » (<i>Microsoft Access 2022</i>)</p> <p>Dernière version disponible avec l'abonnement Microsoft 365. 2019 pour la version achat seul</p> <p>Disponible sur PC uniquement (Microsoft 2022b)</p>	
Fonctionnalités	Intégration de 34 champs répartis en 6 catégories. Liste en annexe 9.	X
	Intégrer différents types de champs : texte libre, liste déroulante, liste à choix multiple	X
	Ajouter ou supprimer un champ et/ou une catégorie si nécessaire	X
	Ajouter ou supprimer un ou des jeux de données rattachés à un projet de recherche	X
	Proposer des bulles informatives contenant une définition du champ	X
	Respecter un format normalisé pour certains champs comme pour les dates, jj-mm-aaaa ou aaaa,	X
	Avoir des niveaux d'accès en fonction du type d'utilisateur : lecture, écriture	X
	Possibilité de rajouter des documents : DMP, contrats, protocoles	X
	Recevoir des alertes automatiques lorsque des dates de préservation arrivent à échéance	X
	Effectuer des recherches sur un ou plusieurs champs	X
	Avoir des options de tri en fonction des champs et des critères de chaque champ	X
	Réaliser des extractions en fonction des critères de tri sélectionnés	X
	Réaliser des impressions en fonction des critères de tri sélectionnés	X
Récupérer automatiquement certaines données : depuis d'autres plateformes comme Sage-X, l'outil du DPO, le fichier Excel du secrétariat, ArODES, Dépôt de données...	X	

	Envoyer des courriels automatiques lorsqu'une action doit être effectuée (suppression d'un jeu de données)	X
	Envoyer des courriels automatiques à la personne ressource concernée lors de l'ajout ou la modification d'un projet nécessitant son attention	X
Maintenance :	Pris en charge jusqu'en octobre 2026	
Lien	https://www.microsoft.com/fr-ch/microsoft-365/access	
Propriétaire	Microsoft	
Exemples d'institutions utilisant cette solution	NCCR, UK Data Service (propose un canevas d'inventaire pour les données de recherche)	
Avantages	<p>Permet d'extraire des données, de créer des rapports</p> <p>Nombreuses options d'import/export vers d'autres systèmes</p> <p>Meilleure ergonomie d'affichage</p> <p>Peut être utilisé aussi bien par des débutants que des utilisatrices-utilisateurs avancés</p> <p>Possibilité d'automatiser de nombreuses tâches grâce à l'utilisation de macros.</p>	
Contraintes	<p>Limitation à 2 GB l'espace de la base de données</p> <p>Disponible que pour le système fonctionnant sous Windows</p> <p>Si le fichier est lourd le programme peut devenir lent et instable.</p> <p>Utilisation simultanée assez limité</p>	
Autres versions	Libre Office Base, OpenOffice Base	
Sources	(Microsoft 2022b; 2022a; <i>Microsoft Access</i> 2022; IONOS 2022; UK Data Service 2012; T Golden Eye 2014), Entretien A. Perret Lundberg, Teams, 26.04 2022)	

AGP / Sage-X

But	Application de gestion de projet pour la HES-SO qui permet un suivi administratif, scientifique et financier des projets de la HES-SO
Coût :	Coût déjà pris en charge actuellement
Données techniques :	<p>AGP/Sage-X v.1.7.1</p> <p>« L'application fonctionne sur des serveurs centralisés installés au service informatique de l'Ecole d'ingénieurs et d'architectes de Fribourg (EIAFR). Celui-ci assure la maintenance et l'exploitation des systèmes. Les serveurs sont installés dans un environnement sécurisé tant au niveau de l'accès physique que de l'alimentation électrique des systèmes » (Haute école spécialisée de Suisse occidentale 2014b, p. 2)</p>

	Logiciels utilisés : AGP (application de gestion de projets), Oracle (base de données), Discoverer (Oracle) « outil d'infocentre permettant d'extraire de la base de données Oracle les informations nécessaires à la création de tous documents, listes, statistiques et rapports utilisés dans les différentes écoles. » (p. 3) Prérequis : navigateurs Internet Explorer (jusqu'à la version 11), Mozilla Firefox (jusqu'à la version 60), Google Chrome (jusqu'à la version 70) (Définition 2021)	
Fonctionnalités	Intégration de 34 champs répartis en 6 catégories. Liste en annexe 9.	
	Intégrer différents types de champs : texte libre, liste déroulante, liste à choix multiple	X
	Ajouter ou supprimer un champ et/ou une catégorie si nécessaire	
	Ajouter ou supprimer un ou des jeux de données rattachés à un projet de recherche	
	Proposer des bulles informatives contenant une définition du champ	X
	Respecter un format normalisé pour certains champs comme pour les dates, jj-mm-aaaa ou aaaa,	X
	Avoir des niveaux d'accès en fonction du type d'utilisateur : lecture, écriture	X
	Possibilité de rajouter des documents : DMP, contrats, protocoles	X
	Recevoir des alertes automatiques lorsque des dates de préservation arrivent à échéance	
	Effectuer des recherches sur un ou plusieurs champs	(X)
	Avoir des options de tri en fonction des champs et des critères de chaque champ	(X)
	Réaliser des extractions en fonction des critères de tri sélectionnés	(X)
	Réaliser des impressions en fonction des critères de tri sélectionnés	(X)
	Récupérer automatiquement certaines données : depuis d'autres plateformes comme Sage-X, l'outil du DPO, le fichier Excel du secrétariat, ArODES, Dépôt de données...	
Envoyer des courriels automatiques lorsqu'une action doit être effectuée (suppression d'un jeu de données)		
Maintenance :	Cet outil n'est plus développé et est amené à disparaître. Maintenances régulières (chaque mois) Migration de la base de données Oracle de la version 12c à la version 19c prévu le 23.09.2022	
Lien	https://agp.hes-so.ch/	

	https://agp-wiki.hes-so.ch/
Propriétaire	HES-SO
Exemples d'institutions utilisant cette solution	Les données sont saisies principalement par la-le chef-fe de projet assisté par la collaboratrice administrative.
Avantages	<p>Déjà une partie des informations sont présentes sur le projet (no, acronyme, nom du projet, responsable du projet, financeurs).</p> <p>Définition de groupes d'utilisatrices-utilisateurs et gestion des accès et de la visibilité des écrans et des projets en fonction de ces derniers</p> <p>Possibilité d'exporter des données</p> <p>Authentification avec Switch edu-ID</p> <p>Possibilité d'ajouter sous Documents- Publications : soit un fichier soit un lien sur les publications ArODES ainsi que d'autres fichiers comme le DMP ou les contrats</p> <p>Espace valorisation prévue ou pourrait être ajouté des liens sur les publications ArODES, ainsi que sur les données déposées sur des dépôts ouverts</p> <p>Interopérabilité entre AGP-SageX et People@HES-SO. Les données des projets dans SageX sont extraites et intégrées dans People@HES-SO</p>
Contraintes	<p>N'est plus développé .Une nouvelle solution va être proposée prochainement mais pas de date ni de solution encore envisagée.</p> <p>Peu d'autonomie pour les utilisatrices et utilisateurs (pas possible de faire des extractions soi-même, ni des requêtes)</p> <p>N'est pas exclusivement lié à la HES-SO Valais-Wallis mais est utilisé par l'ensemble de la HES-SO.</p>
Sources	(Spada 2021; Haute école spécialisée de Suisse occidentale 2014b; <i>Définition</i> 2021)

Annexe 11 : Analyse des outils de stockage

Local

	Ordinateur portable	Disque dur externe / Clé USB
But	Gestion quotidienne de son activité	Pour la sauvegarde de données
Gouvernance	Personnel ou fourni par la HES-SO VS-WS	Personnel ou proposé par la HES-SO VS-WS
Hébergement	Interne à l'appareil	Externe à l'appareil
Accès, sécurité	Utilisation de mot de passe de l'ordinateur	Utilisation de mot de passe possible
Partage	Partage de main à main	Partage de main à main
Sauvegardes	Pas de sauvegarde automatique. A faire manuellement sur un autre support	Pas de sauvegarde automatique. A faire manuellement sur un autre support
Espace stockage	Dépend du modèle choisi	Dépend du modèle choisi
Avantages	Facilité de transport Pas besoin d'internet Contrôle total des fichiers Facile à protéger contre l'accès non autorisé	Facilité de transport Faible coût Pas besoin d'internet Contrôle total des fichiers Facile à protéger contre l'accès non autorisé
Contraintes	Facile de perdre ou d'endommager Pas adapté pour le stockage à long terme	Facile de perdre ou d'endommager Pas adapté pour le stockage à long terme
Remarques	Utiliser un mot de passe fort, voir chiffre le disque dur	« Utiliser pour le stockage temporaire à court terme de données non sensibles, par exemple sur le terrain ou pour transporter

	<p>Utiliser seulement si le projet implique peu idéalement une seule personne et si les données ne sont pas fréquemment déplacées. (Recker, CESSDA Training Team 2017)</p> <p>S'assurer de toujours travailler sur la version la plus récente.</p> <p>S'assure que la version la plus récente est toujours sauvegardée (Recker, CESSDA Training Team 2017)</p>	<p>des données et des fichiers lorsque la transmission en ligne n'est pas possible » (Recker, CESSDA Training Team 2017)</p> <p>« Effectuer des vérifications régulières pour s'assurer que l'appareil fonctionne et que les fichiers sont accessibles » (Recker, CESSDA Training Team 2017)</p> <p>Ne pas utiliser pour un stockage à long terme ou pour réaliser des copies maitresses» (Recker, CESSDA Training Team 2017)</p>
Sources	(Recker, CESSDA Training Team 2017; Summers, Corti 2020; Diaz et al. 2022)	(Recker, CESSDA Training Team 2017; Summers, Corti 2020; Diaz et al. 2022)

Cloud

	OneDrive Enterprise	Teams Sharepoint Microsoft 365
But	<p>Service de stockage personnel et de synchronisation de fichiers en ligne</p> <p>Comprend un espace personnel</p>	<p>Solution de partage et de collaboration</p> <p>Comprend les dossiers liés aux équipes et canaux créés dans Microsoft Teams</p>
Gouvernance	HES-SO et HES-SO Valais-Wallis	HES-SO et HES-SO Valais-Wallis
Hébergement	Serveurs localisés en Suisse depuis 2022	Serveurs localisés en Suisse depuis 2022
Accès / sécurité	<p>Accès en ligne et en local via l'installation d'un client OneDrive</p> <p>Accès par l'institution : l'institution n'a légalement pas accès aux données</p> <p>Plus accessible au départ du personnel</p> <p>Droits d'accès : gestion des droits d'accès pour chaque fichier/dossier (modification, affichage)</p> <p>Chiffrement : Les données ne sont pas chiffrées mais la transmission des données entre l'ordinateur et les serveurs est chiffrée</p>	<p>Accès en ligne et en local via l'installation Teams ou avec le client OneDrive</p> <p>Accès par l'institution : « Comme il ne s'agit pas d'un dossier personnel, l'institution pourrait y accéder légalement sur demande à la HES-SO » (P. Schneider, 2022)</p> <p>Au départ du personnel transfert de la propriété à un autre membre de l'équipe</p> <p>Droits d'accès : gestion des droits d'accès pour chaque fichier/dossier (modification, affichage)</p>

		Chiffrement : Les données ne sont pas chiffrées mais la transmission des données entre l'ordinateur et les serveurs est chiffrée
Partage	Internes HES-SO (donc aussi VS), externes à la HES-SO Partage de lien par mot de passe avec définition d'une durée de validé du lien	Internes HES-SO (donc aussi VS), externes à la HES-SO « Le partage par un lien est désactivé pour des raisons de sécurité. Il est cependant possible pour un externe de se créer un compte invité et d'être ajouté comme ayant droit de lecture ou écriture sur un dossier/fichier. Cet invité recevra un email lorsqu'il veut y accéder avec un mot de passe aléatoire à chaque connexion. Cette solution garanti un accès nominatif et que seul l'utilisateur qui a été autorisé a bien accès au document et qu'il ne peut pas partager cet accès » (P. Schneider, 2022)
Sauvegardes	Pas de sauvegarde faite dans un autre système permettant de restaurer l'intégralité du OneDrive à un état précédent Historique de versions qui permet de revenir à une version antérieure d'un fichier Corbeille permettant de restaurer un fichier supprimé durant 30 jours qui suivent la suppression (HES-SO Fribourg, Service informatique 2020)	Pas de sauvegarde faite dans un autre système permettant de restaurer l'intégralité de Sharepoint à un état précédent. Une demande au SINP peut être faite pour réaliser une sauvegarde spécifique sur une équipe Teams. Dans ce cas les sauvegardes sont faites toutes les nuits. Historique de versions qui permet de revenir à une version antérieure d'un fichier Corbeille permettant de restaurer un fichier supprimé durant 30 jours qui suivent la suppression (HES-SO Fribourg, Service informatique 2020)
Espace st.	< 1 Tb	> 1 Tb
Avantages	Possibilité de partager avec des externes de HES-SO Simple d'utilisation Accès facilité	Possibilité de partager avec des externes de HES-SO Simple d'utilisation, Accès facilité En cas de départ du propriétaire les droits peuvent être transmis à un-e autre membre de l'équipe.

Contraintes	<p>Pas de sauvegarde donc risque de perte des données en cas d'attaque</p> <p>En cas de départ du propriétaire personne n'a accès aux dossiers sauf si ce dernier en donne l'autorisation</p>	<p>Pas de sauvegarde donc risque de perte des données en cas d'attaque</p>
Remarques	<p>Possible de stocker des données personnelles comme « le tenant se trouve en Suisse ». Pour les données sensibles des mesures complémentaires peuvent être nécessaires (P. Schneider, 2022).</p> <p>Utiliser une solution de chiffrement pour les données personnelles et sensibles</p> <p>Plateforme à utiliser dans la phase de réalisation du projet, mais n'est pas adapté ni pour la conservation à moyen terme (5 ans) ni pour l'archivage à long terme.</p> <p>Réaliser des sauvegardes régulières de son côté, dans l'idéal sur le serveur de l'école</p>	<p>Possible de stocker des données personnelles comme « le tenant se trouve en Suisse ». Pour les données sensibles des mesures complémentaires peuvent être nécessaires (P. Schneider, 2022).</p> <p>Utiliser une solution de chiffrement pour les données personnelles et sensibles</p> <p>Plateforme à utiliser dans la phase de réalisation du projet, mais n'est pas adapté ni pour la conservation à moyen terme (5 ans) ni pour l'archivage à long terme</p> <p>Réaliser des sauvegardes régulières de son côté, dans l'idéal sur le serveur de l'école</p>
Sources	<p>(HES-SO Fribourg, Service informatique 2020; HES-SO Valais-Wallis [sans date]; Microsoft 2022c) Entretien, P. Schneider, Teams, 14.07.2022 et 19.07.2022</p>	<p>(HES-SO Fribourg, Service informatique 2020; HES-SO Valais-Wallis [sans date]; Microsoft 2022c), Entretien, P. Schneider, Teams, 14.07.2022 et 19.07.2022</p>

Cloud suite

	SWITCHdrive (espace personnel)	SWITCHdrive Projects Folders
But	Service SaaS de stockage et de partage de fichiers Comprend un stockage personnel	Service SaaS de stockage et de partage de fichiers Comprend un stockage pour les projets
Gouvernance	SWITCH	SWITCH
Hébergement	Serveurs localisés en Suisse	Serveurs localisés en Suisse
Accès / sécurité	<p>Accès en ligne et possibilité d'installer le client afin d'accéder aux dossiers en local</p> <p>Accès par l'institution : l'institution n'a légalement pas accès aux données</p> <p>Plus accessible au départ du personnel</p> <p>Droits d'accès : possibilité de donner des droits d'accès (ajout, modification suppression)</p> <p>Possibilité d'attribuer un mot de passe aux dossiers et fichiers</p> <p>Chiffrement : Les données ne sont pas chiffrées mais la transmission des données entre l'ordinateur et les serveurs est chiffrée.</p> <p>Il est possible de chiffrer les données en utilisant un outil comme Cryptomator. Attention cependant, car cela rend certaines fonctionnalités inutilisables (lecture et édition en ligne d'un document)</p>	<p>Accès en ligne et possibilité d'installer le client afin d'accéder aux dossiers en local</p> <p>Accès par l'institution : espace collaboratif appartenant à l'institution, de ce fait l'institution pourrait y accéder légalement sur demande à la HES-SO » (P. Schneider, 2022)</p> <p>Au départ du personnel les accès restent actifs pour les autres membres de l'équipe</p> <p>Droits d'accès : possibilité de donner des droits d'accès (ajout, modification suppression)</p> <p>Possibilité d'attribuer un mot de passe aux dossiers et fichiers</p> <p>Chiffrement : Les données ne sont pas chiffrées mais la transmission des données entre l'ordinateur et les serveurs est chiffrée</p> <p>Il est possible de chiffrer les données en utilisant un outil comme Cryptomator. Attention cependant, car cela rend certaines fonctionnalités inutilisables (lecture et édition en ligne d'un document)</p>
Partage	Internes HES-SO (donc aussi VS), externes à la HES-SO Partage de lien avec mot de passe et définition d'une durée de validé du lien. Solution permet le partage sans avoir un compte SWITCHdrive -> A privilégier pour les externes	Internes HES-SO (donc aussi VS), externes à la HES-SO Partage de lien avec mot de passe et définition d'une durée de validé du lien. Solution permet le partage sans avoir un compte SWITCHdrive -> A privilégier pour les externes

	Possibilité d'accès nominatif pour les externes n'ayant pas de compte edu-ID « académique » avec la création d'un compte edu-ID + voucher à demander au service informatique.	Possibilité d'accès nominatif pour les externes n'ayant pas de compte edu-ID « académique » avec la création d'un compte edu-ID + voucher à demander au service informatique.
Sauvegardes	<p>Sauvegardes : contractuellement SWITCH ne garantit pas la sauvegarde. Seules des sauvegardes à l'échelle du système sont réalisées.</p> <p>Restauration : des restaurations du système peuvent être faites en cas de sinistre mais pas de restauration individuelle. Mais corbeille personnelle où il est possible de restaurer des éléments durant 90 jours après leur suppression.</p> <p>Historique des versions : dispose d'un système de gestion des versions jusqu'à 30 jours</p>	<p>Sauvegardes : contractuellement SWITCH ne garantit pas la sauvegarde. Seules des sauvegardes à l'échelle du système sont réalisées.</p> <p>Restauration : des restaurations du système peuvent être faites en cas de sinistre mais pas de restauration individuelle. Mais corbeille personnelle où il est possible de restaurer des éléments durant 90 jours après leur suppression.</p> <p>Historique des versions : dispose d'un système de gestion des versions jusqu'à 30 jours</p>
Espace stockage	<p>100 Go jusqu'à Tb.</p> <p>Possibilité d'obtenir du stockage supplémentaire en faisant une demande au SINF</p>	<p>100 Go dédié au projet jusqu'à 1 Tb. Si plus de 1 Tb, créer plusieurs Projets Folders</p> <p>Faire une demande au SINF pour obtenir un « Projet Folder » avec un espace de stockage dédié .</p> <p>Possibilité d'obtenir du stockage supplémentaire en faisant une demande au SINF</p>
Avantages	<p>Possibilité de travailler avec des personnes hors HES</p> <p>Accès facilité</p>	<p>Possibilité de travailler avec des personnes hors HES</p> <p>Accès facilité</p>
Contraintes	<p>Problèmes rencontrés pour le partage avec certaines écoles (ex. CHUV)</p> <p>Souci de synchronisation si gros volumes de fichiers ou si VM</p> <p>En cas de départ du propriétaire personne n'a accès aux dossiers sauf si ce dernier en donne l'autorisation</p> <p>Pas de sauvegarde donc risque de perte des données en cas d'attaque</p>	<p>Problèmes rencontrés pour le partage avec certaines écoles (ex. CHUV)</p> <p>Souci de synchronisation si gros volumes de fichiers ou si VM</p> <p>En cas de départ du propriétaire personne n'a accès aux dossiers sauf si ce dernier en donne l'autorisation</p> <p>Une demande doit être faite au SINF pour ouvrir un espace</p> <p>Pas de sauvegarde donc risque de perte des données en cas d'attaque</p>

Remarques	Utiliser une solution de chiffrement pour les données personnelles et sensibles Plateforme à utiliser dans la phase de réalisation du projet, mais n'est pas adapté pour la conservation à moyen terme (5 ans) ni pour l'archivage à long terme (UNIL sans date, p. 4) Réaliser des sauvegardes régulières de son côté, dans l'idéal sur le serveur de l'école	Utiliser une solution de chiffrement pour les données personnelles et sensibles Plateforme à utiliser dans la phase de réalisation du projet, mais n'est pas adapté pour la conservation à moyen terme (5 ans) ni pour l'archivage à long terme (UNIL sans date, p. 4) Réaliser des sauvegardes régulières de son côté, dans l'idéal sur le serveur de l'école
Sources	(UNIL, Faculté des sciences sociales et politiques [sans date]; SWITCH 2022b; Schneider 2022), Entretien, P. Schneider, Teams, 14.07.2022 et 19.07.2022	(UNIL, Faculté des sciences sociales et politiques [sans date]; SWITCH 2022b; Schneider 2022), Entretien, P. Schneider, Teams, 14.07.2022 et 19.07.2022

Cloud (suite)

	REDCap
But	« REDCap (Recherche Electronic Data Capture), un service en ligne avec application mobile conçu pour développer et gérer des enquêtes et sondages en ligne » (DAUnit 2022). Il est utilisé principalement pour l'acquisition de données personnelles et/ou sensibles.
Gouvernance	HES-SO Fribourg
Hébergement	Stockées sur un serveur sécurisé de la HES-SO à Fribourg. La DAUnit gère les aspects techniques liés à la création des comptes, méthodologiques et le support en général
Accès / sécurité	Accès est fait avec Switch edu-ID, aussi accès pour des externes (création compte AAI avec une date d'expiration). Accès à distance plus via l'application mobile. Accès par l'institution : uniquement le PI et les personnes à qui il donne des droits Plus accessible au départ du personnel : Droits d'accès : limités au projet , possibilité d'accorder différents droits d'accès selon les utilisateurs Dispositif d'anonymisation des données, respect de la confidentialité des données – toutes les données peuvent être mises dans REDCap. Possible de définir quelles données sont identifiantes. Cependant les données identifiantes ne devraient pas être mises car il est possible de les retrouver via le journal de suivi.

	Chiffrement : Les données ne sont pas chiffrées mais la transmission des données entre l'ordinateur et les serveurs est chiffrée Authentification à deux facteurs
Partage	Internes HES-SO (donc aussi VS), externes à la HES-SO
Sauvegardes	Sauvegardes : des sauvegardes sont effectuées par le SINF chaque jour Historique des versions : oui, journal de suivi du point de vue admin mais également dans son projet. Possibilité d'avoir des doubles vérifications. Restauration : deux restaurations qui peuvent être faites, au niveau de la base complète et pour le projet mais risque de perte de données
Espace stockage	> 1Tb Jusqu'à maintenant pas de nettoyage qui était fait au niveau de la base. Mais c'est en cours de discussion
Avantages	Sécurisé Validation des données Précision au niveau du formatage des champs Lier à une ontologie – vocabulaire contrôlé Forum, communauté, avec beaucoup de vidéos Droit d'accès très précis par type d'utilisateur Combiner des questionnaires avec d'autres (rappel automatique, invitation) Télécharger des documents dans la base de données et les lier aux participants, à l'instrument Possibilité d'ajouter des commentaires pour chaque variable
Contraintes	Multilinguisme est difficile à gérer Tout n'est pas traduit dans REDCap (bouton submit) Temps pour apprentissage Temps pour préparer Possible de copier et de modifier une variable. Une fois en production seul-e-s les administratrices-teurs peuvent effectuer ces changements. Obligation de le faire en ligne. Pour le hors ligne il faut l'application

Remarques	<p>Sensibilité des données : données personnelles ou sensibles</p> <p>Plateforme à utiliser dans la phase de réalisation du projet ou pour la conservation à moyen terme (5 ans) mais n'est pas une plateforme d'archivage à long terme</p> <p>Chaque personne doit réaliser des sauvegardes régulières de son côté dans l'idéal sur le serveur de l'école étant donné le risque de perte en cas de récupération</p>
Sources	Entretiens : C. Martinez, Teams, 28.06.2022, A. Cotting, Teams, 13.07.2022

Lecteur réseau

	Network Attached Storage (NAS)	Machine virtuelle (VM)
But	Espace de stockage interne dédié pour les projets de recherche à la HES-SO VS-WS	Machine virtuelle qui contient aussi bien les fichiers de données et les applications utilisées pour l'analyse et le traitement afin d'avoir un espace de travail et de stockage hautement sécurisé
Gouvernance	HES-SO Valais-Wallis	HES-SO Valais-Wallis
Hébergement	Serveurs localisés en Valais	Serveurs localisés en Valais
Accès / sécurité	<p>Accès en local ou à distance via le VPN</p> <p>Accès par l'institution : l'institution a accès à ces données sur demande au SINF</p> <p>« Un utilisateur qui quitte la HES-SO Valais-Wallis n'aura plus accès, mais l'institution pourra toujours demander l'attribution des accès à d'autres utilisateurs sur demande » (P. Schneider, 2022).</p> <p>Droits d'accès : possibilité de donner des droits d'accès différents pour chaque personne, est géré par le SINF</p> <p>Chiffrement : Les données ne sont pas chiffrées tout comme les sauvegardes car risque de pertes</p>	<p>Machine virtuelle Windows accessible en remote desktop</p> <p>Connexion avec un VPN</p> <p>Accès par l'institution : L'institution a accès à ces données</p> <p>« Un utilisateur qui quitte la HES-SO Valais-Wallis n'aura plus accès, mais l'institution pourra toujours demander l'attribution des accès à d'autres utilisateurs sur demande » (P. Schneider, 2022).</p> <p>Droits d'accès : possibilité de donner des droits d'accès différents pour chaque personne, est géré par le SINF</p> <p>Les fichiers contenant les données sensibles ainsi que les applications nécessaires à leur traitement se trouvent dans la VM. Le stockage et le traitement des données s'effectuent dans cet environnement sécurisé</p>

		Chiffrement : disques de la machine virtuelle chiffrés, Authentification à multi-facteurs
Partage	Partage possible uniquement pour les collaboratrices et collaborateurs de la HES-SO Valais-Wallis	Personnel de la HES-SO VS-WS. Possible pour des personnes externes via VPN spécifique
Sauvegardes	Sauvegardes : réalisées chaque jour par le SINF Historique des versions : oui Restauration : possible durant 1 année	Sauvegardes : sauvegardes réalisées chaque jour voire encore plus régulières selon le projet par le SINF Historique des versions : oui Restauration : restauration de l'environnement complet seulement durant 1 an. Il n'est pas possible de restaurer un seul fichier comme il est en accès fermé
Espace st.	> 1 Tb Faire une demande pour la création d'un espace de stockage au SINF	>1 Tb Faire une demande pour la création d'une VM au SINF
Avantages	Espace sécurisé pour données personnelles ou sensibles, Sauvegardes quotidiennes, Possibilité de donner des droits accès pour chaque participant Solution de stockage pour le moyen terme également	Espace de stockage très sécurisé Sauvegardes très régulières Droits d'accès pour chaque participant Possibilité de travailler avec des externes de l'école Solution de stockage pour le moyen terme également
Contraintes	Limité aux internes à la HES-SO VS-WS Une demande doit être faite au SINF pour ouvrir un espace	Système complexe, Nécessite une formation Une demande doit être faite au SINF pour ouvrir un espace
Remarques	Sensibilité des données : peut être utilisé pour le stockage de données personnelles ou sensibles, dans l'idéal en chiffrant les données Plateforme pour la conservation durant la phase de projet et à moyen terme (5 ans) mais n'est pas une plateforme d'archivage à long terme	Sensibilité des données : peut être utilisé pour le stockage de données personnelles, sensibles et hautement sensibles, protégées avec un niveau de sécurité semblable à l'environnement source (hôpital, OFS)

	Utiliser les restrictions d'accès pour s'assurer que tout le monde n'a pas accès à tout si ce n'est pas nécessaire	Plateforme pour la conservation durant la phase de projet et à moyen terme (5 ans) mais n'est pas une plateforme d'archivage à long terme Utiliser les restrictions d'accès pour s'assurer que tout le monde n'a pas accès à tout si ce n'est pas nécessaire
Sources	Entretiens chercheuses et chercheurs, Entretien, P. Schneider, Teams, 14.07.2022	(HES-SO Valais-Wallis Service informatique, Communication personnelle, 19.06.2020), Entretien, P. Schneider, Teams, 14.07.2022

Annexe 12 : Evaluation et sélection des données à conserver

L'augmentation des données collectées lors des projets de recherche, associés aux coûts liés au stockage et au maintien de la lisibilité des formats nécessitent de faire un choix quant aux données à conserver. Ce choix doit être réalisé en gardant à l'esprit le potentiel de partage et de réutilisation (Laplante 2019).

On peut s'appuyer sur l'archivistique et l'approche de gestion des données pour les transposer aux données de recherche afin d'évaluer et sélectionner les données à conserver (Maday, Moysan 2014; Maday 2015; Grant 2017; Laplante 2019; Melly 2021). L'évaluation est selon Couture (1996-1997) « l'acte de juger des valeurs que présentent les documents d'archives » (p. 3) valeur qui peut être primaire ou secondaire et va permettre de définir si les données doivent être conservées ou éliminées et la période durant laquelle elles seront conservées.

La notion de valeur a été définie dans les années 60 par l'archiviste américain Schellenberg lorsqu'il a abordé la question de l'évaluation des archives (Ott 2015; Melly 2021). Il propose alors une distinction entre la valeur primaire et secondaire, la première faisant référence « aux aspects légaux, financiers et administratifs » et la seconde à la valeur de « preuve au sens historique » et de recherche au sens de valeur informationnelle (Couture 1996, p. 8). Ces critères sont toujours pertinents et peuvent aider à fixer une valeur aux données de recherche et ainsi « déterminer leur conservation ou leur élimination » (Laplante 2019, p. 7).

Laplante (2019) propose trois lignes directrices pour l'évaluation des données de recherche :

- « Adapter les critères d'évaluation en fonction de la mission de l'organisme et du type de données à conserver.
- Collaborer avec le chercheur dès la création des données afin d'assurer une gestion efficace des données tout au long du processus et une documentation complète et conforme
- Élaborer les critères d'évaluation en tenant compte des contraintes de la conservation à long terme, particulièrement celles d'ordre pécuniaire » (p. 12)

Un certain nombre de critères de base pour la sélection des données de recherche à archiver sont proposés dans la littérature (Whyte, Wilson 2010; Beagrie 2019; Gastl-Hartmann 2022; Inter-university consortium for political and social research, University of Michigan 2022) :

- Pertinence par rapport à la mission
- Conformité juridique ou légale : restreindre l'accès si nécessaire ou au contraire rendre les données publiques si exigé par les bailleurs de fonds
- Valeur scientifique ou historique à long terme
- Caractère unique du jeu de données
- Potentiel de réutilisation : en lien avec la propriété intellectuelle et les questions éthiques
- Non-répliquabilité
- Evaluer les coûts
- Documentation complète
- Niveau, type, format des données

Kung et Campbell (2016) ont fait l'exercice inverse et propose une liste de données qu'il n'est pas souhaitable de conserver :

- « Données de mauvaise qualité
- Données qui ne peuvent pas être utilisées par d'autres
- Données facilement reproductibles
- Données sans bonnes métadonnées
- Données anciennes qui ne sont pas utilisées et n'ont pas de valeur culturelle ou historique évidente
- Données pilotes, de test ou intermédiaires
- Données propriétaires
- Données sensibles ou confidentielles » (traduit de Kung, Campbell 2016, p. 53)

Durée de conservation

La durée de conservation des données est une question importante

Tableau 11 : Recommandations des institutions sur la durée de conservation

Institution	Recommandations
FNS	Recommandation 10 ans se base sur le 8 ^e principe du Concordat on Open Research (Fonds national suisse 2022b)
ETHZ	Données déposées dans un dépôt doivent être conservées au minimum 10 ans. Pour celles qui ne sont pas déposées dans un dépôt, conservation 10 ans si techniquement et économiquement réalisable (Eidgenössische Technische Hochschule Zürich 2022).
Concordat on Open Research Data	Toutes les données ouvertes ne doivent pas être conservées de manière illimitée. Seules « les données sous-jacentes aux publications doivent être conservées pendant 10 ans à compter de la date de publication » sauf indication contraire des bailleurs (traduit de HEFCE et al. 2016, p. 16).
UNIL	« Selon la directive 4.2 « Intégrité scientifique dans le domaine de la recherche », les données de base produites dans le cadre de la recherche doivent être conservées en sécurité pendant au moins dix ans après l'achèvement de l'étude. Cette durée de 10 ans correspondant au stockage long terme (Long Term Storage – LTS) proposée par le Ci. La directive 4.5 « Traitement et gestion des données de recherche » précise que l'offre de stockage courant des données (Current Storage – CT) est en principe limité à une durée de 5 ans à compter de la date de fin du Projet. Passé ce délai, les données doivent être transférées sur le LTS, et 20 ans pour des données liées à des brevets.

	Attention toutefois, selon le projet, les données doivent être détruites plus rapidement, conformément aux consentements signés par les personnes concernées. Dans tous les cas, UNIRIS recommande la destruction des données qui ne sont pas utilisées ou sollicitées au bout de 3 ans ». (UNIL, UNIRIS 2019)
UNIGE	Les « des données de base » doivent être conservées «pendant 5 ans au moins après le terme de la recherche» (UNIGE 2019b)
UNIBE	5 ans de conservation puis archivage d'une sélection de données. Les critères de conservation sont fixés par les Archives de l'Etat (intérêt historique) et non par l'Université (intérêt pour le domaine d'activité) (2022)
HES-SO Valais-Wallis	Actuellement pas de critères mais le SINP préconise de les conserver 5 ans au maximum sur le NAS. (Entretien P. Schneider, Teams, 14.07.2022).

Guides pour aller plus loin

ASSOCIATION DES ARCHIVISTES FRANÇAIS. SECTION AURORE, 2018. Référentiel de gestion des archives de la recherche. *Doranum* [en ligne]. 2018. [Consulté le 06 août 2022]. Disponible à l'adresse : https://doranum.fr/wp-content/uploads/Referentiel_2.pdf

CENTRE DE COOPÉRATION INTERNATIONALE EN RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT et SERVICE INTERMINISTÉRIEL DES ARCHIVES DE France, 2020. Les durées de conservation : guide pratique. *CNIL* [en ligne]. 2020. [Consulté le 06 août 2022]. Disponible à l'adresse : https://www.cnil.fr/sites/default/files/atoms/files/guide_durees_de_conservation.pdf

GASTL-HARTMANN, Angela, 2022. Forschen für die Ewigkeit: Bewertung und Archivierung von Forschungsdaten als neues Arbeitsgebiet für Hochschularchive. *Informationswissenschaft: Theorie, Methode und Praxis*. 22 juin 2022. Vol. 7, no. 1, pp. 396-429. [Consulté le 9 août 2022]. Disponible à l'adresse: <https://doi.org/10.18755/iw.2022.22>.

LAPLANTE, Catherine, 2019. Les données de recherche : comment évaluer pour mieux conserver? *Papyrus*. [en ligne]. 2019. [Consulté le 6 août 2022]. Disponible à l'adresse: https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/22340/laplane_catherine_2019_donnees_recherche_comment_evaluer.pdf?sequence=1&isAllowed=y

Annexe 13 : Analyse de solutions d'archivage

	OLOS	SWISSUbase
Description		
Type de dépôt	Généraliste	Généraliste auparavant disciplinaire
Discipline de recherche	Toutes disciplines	Sciences sociales, linguistiques mais développement pour intégrer toutes les disciplines
Coût (accès, hébergement)	Trois options de tarification : essential, advanced, premium ⁵⁷ Des réductions sont proposées pour les membres	Gratuit pour les chercheuses et chercheurs
Certification	Non	CTS
Conformité aux principes FAIR	Oui	Oui
Types de données acceptées	Tous	Données quantitatives et qualitatives. Critères acceptation : « Les chercheurs affiliés à une institution de recherche suisse ; Les équipes de recherche internationales ayant une composante suisse ; Les services gouvernementaux ; Les offices de statistiques nationales et cantonales ; Les institutions publiques et privées. » (traduit de FORS Data Service 2022a, p. 3)
Localisation des serveurs	Suisse	Suisse
Lien	https://olos.swiss	https://www.swissubase.ch/fr/
Découverte		
Métadonnées de qualité et vocabulaires contrôlés	DataCite. Métadonnées de préservation PREMIS, Métadonnées administratives METS. Métadonnées obligatoires et recommandées. Métadonnées administratives collectées automatiquement. Ajout de métadonnées possible selon son propre schéma	DDI (sciences sociales) et META SHARE pour la linguistique. Développements futurs d'autres schémas spécifiques par disciplines. CMM – CESSDA Metadata Model. Métadonnées obligatoire et recommandées

⁵⁷ A la fin du tableau des informations complémentaires sont données par rapport aux tarifs pour l'archivage sur OLOS

Moissonnage des métadonnées avec OAI-PMH	Oui	Information non trouvée
Métadonnées toujours disponibles (même si données pas/plus disponibles)	Oui si création d'une unité organisationnelle	Oui
Attribution d'un PID	DOI	DOI
Moteur de recherche intégré	Oui, avec recherche avancée	Oui, avec recherche sur les données et les métadonnées
Signalement dans des catalogues	Oui (Re3data)	Oui (Re3data, CESSDA)
Métadonnées indexées par des services de recherche (moteurs recherche)	Oui	Non
Métadonnées disponibles dans formats lisibles par homme/machine	Oui selon le schéma de métadonnées DataCite en xml	Information non trouvée
PID pour les données avec plusieurs niveaux de granularité	DOI au niveau de l'archive Identifiants pour chaque fichier	Pour chaque jeu de données d'un projet un DOI différent
Liaison entre ressources connexes (articles, logiciels)	Oui lien sur des archives similaires. Les références des articles sont rajoutées manuellement au niveau de la description de l'archive	Oui sur les publications
Prise en charge des PID pour auteurs, financeurs...	ORCID	ORCID, financeurs,
Accès		
Conditions d'accès consultation, récupération, versement	Versement : Login institutionnel (SWITCH edu-ID) Consultation : sans condition Récupération : sans condition	Versement : Login Switch edu-ID Consultation : sans condition Récupération : être affilié à une institution académique. Accès à des fins de recherche ou d'enseignement uniquement. Valider le contrat d'utilisation
Accès gratuit pour l'utilisateur	Oui	Oui

Durée garantie de conservation des données	5, 10, 15, 20 ans ou indéfiniment	Pas de durée définie. Mention d'archivage pérenne
Accès pour personnes porteuses de handicap	Non	Non
Mécanisme permettant de mettre des fichiers très volumineux à disposition	Jusqu'à 6 Go en une fois. Si plus contacter le support	Information non trouvée
Moyen alternatif d'accéder à une ressource restreinte	Si restreint ou fermé on peut malgré tout demander des accès en plus	Autorisation spéciale : Aucune, Avec l'accord préalable de l'auteur
Réutilisation		
Information sur les licences et conditions de réutilisation	CC BY, possible de changer la licence par défaut Choix licences ouvertes (8 possibilités)	Utilisation de licences CC Possibilité d'ajouter des restrictions ("Aucune" ; "Recherche académique et enseignement" ; et "Recherche académique uniquement") et des autorisations spéciales
PID citables	Oui	Oui
Liste des formats privilégiés et non propriétaires	Tous les formats	Liste définie de formats : tabulaire, textuel, image, audio, vidéo
Page d'accueil avec les métadonnées relatives à la ressource, y compris citation lisible par machines et humains	Oui	Non depuis la page du jeu de données, pas disponible depuis le projet.
Accessibilité au contenu avec des APIs ouvertes	Oui	Rest-API
Ressources stockées dans des formats standard de la communauté, lisibles par les machines	Oui	Information non trouvée
Intégrité et authenticité		
Politique ou documentation décrivant les pratiques pour éviter l'accès/ la	Différents rôles et permissions en fonction de ces rôles.	Différents rôles et permissions en fonction de ces rôles.

manipulation non autorisées		
Trace des modifications apportées aux métadonnées et aux ressources	Checksums, audit trails	Checksums
Prise en charge du versionnage des (méta)données	Oui	Oui
Informations sur les fournisseurs de métadonnées y compris le nom des personnes/institutions responsables	Oui	Oui
Qualité des données		
Vérification et validation des données et des métadonnées de base	Détection des virus, identification des formats et du niveau de conformité, étape de validation configurable, prévisualisation des données importées,	Contrôle de la qualité des données, de la documentation et des métadonnées par des experts
Documentation disponible sur le processus de curation	Oui	Oui
Annotation, commentaire ou évaluation des (méta)données	Modération qui pourrait être faite avant publication = optionnel Annotation qualité et utilité qui peut être faite par les utilisatrices-utilisateurs	Vérification par des experts en archivage des données
Confidentialité des données sensibles		
Mécanismes qui limitent l'accès aux données aux utilisateurs autorisés	Public, restreint ou fermé, embargo	Ouvert, restreint, embargo, sur demande
Niveaux adaptés aux différents niveaux d'exigence de sécurité des données	DataTags levels « permet d'explicitier à l'aide d'une étiquette le niveau de sensibilité d'un jeu de données» (UNIGE 2018). Pour les tags orange, rouge et cramoisi, la cryptographie est de la responsabilité du déposant. Les clés de cryptographie ne sont pas gérées par OLOS. Pour les	Données toujours anonymisées sauf exception Utilisation de contrat de dépôt entre le producteur et FORS pour « garantir le respect des droits d'utilisation, de la protection des données et des droits de propriété intellectuelle » (FORS Data Service 2022b).

	deux derniers tags, l'authentification à deux facteurs sera implémentée dans une version future. Ces 3 niveaux seront développés prochainement avec l'aide d'un expert en cybersécurité ⁵⁸	Un contrat d'utilisation fixe les conditions de réutilisation. Il doit être lu et accepté par l'utilisatrice-utilisateur en mentionnant l'utilisation qui en sera faite, la durée du contrat et si une publication est prévue. Critères d'accès (cf. voir point Accès)
Préservation		
Plan de préservation à long terme	« Préservation configurable (définition, implémentation, veille, durée...), mesures de continuité d'accès, planification de migrations des formats si détection d'obsolescence » (Guirlet, 2020)	« Environnement de confiance certifié pour la préservation à long terme des données. Continuité d'accès selon modèle OAIS » (Guirlet, 2020)
Documentation ou politique sur les durées et pratiques de préservation	Informations données sur le site	Politique de préservation sur le site de FORS
Approche reconnue de préservation	OAIS	OAIS
Actions nécessaires pour assurer les responsabilités de préservation (ex. droits de copie, de transformation et de stockage)	Gestion de la détérioration des supports, plusieurs systèmes de stockage différents, nombre de copies et distribution géographique configurables Stockage non chiffré. Cependant des données chiffrées peuvent être déposées par la-le PI. La gestion des clés est de leur responsabilité ⁵⁹	Stratégie de backups, répliques, contrôle régulier des médias, plan de recouvrement des désastres, processus de gestion des versions, serveurs distants de 200 km, Chiffrement côté serveur lors du stockage sur le disque et déchiffrement lors du téléchargement
Pérennité et gouvernance		
Organisation responsable de la gestion du dépôt	Association OLOS	FORS
Commercial ou non	Non commercial	Non commercial
Autres caractéristiques		
Documentation ou politique qui définit le périmètre des ressources	Guides de démarrage, outils d'aide, documentation technique, sur le site FAQ,	Guides pas à pas, glossaires, FAQ.

⁵⁸ Courriel, P.-Y. Burgi, Directeur du projet DLCM P5, 03.08.2022.

⁵⁹ Cf. 50.

	bonnes pratiques. Plateforme de test	
Point de contact, service d'assistance	Formulaire de contact, coaching par téléphone, formations proposées Soutien payant sur la GDR (ou inclus dans le forfait)	Service d'assistance, contact par courriel, soutien personnel. Soutien gratuit sur la GDR
Politique, document sur les conditions d'acceptation et d'utilisation	Conditions sur le site OLOS	Conditions sur le site de FORS
Métriques, statistiques	Accessibles dans son espace personnel de préservation pour ses données versées. Peut suivre, contrôler les demandes, les consultations et les téléchargements	Statistiques sur les jeux de données publiées, téléchargements, sur qui a utilisé les données et pour quel usage. Alerte pour les utilisateurs si mise à jour de données téléchargées
Plateforme, technologies	Technologies et applications Open Source	SWITCHengines, Openshift
Sources		
	(Re3data.org 2021; Guirlet 2020; OLOS 2022b; Krug, Noirjean 2022)	(Re3data.org 2022; Guirlet 2020; SWISSUbase 2022a; FORS [sans date]; SWISSUbase 2022b; 2022c; FORS Data Service 2022b; FORS 2021; FORS Data Service 2022a)

Estimation du coût pour l'archivage des données de recherche sur OLOS

Trois niveaux sont proposés : essential, advanced et premium

Si l'on choisit le niveau médium en y incluant les trois durées de conservation (5, 10, 15 ans) il faudra compter 784 CHF pour l'archivage de 605.7 Go de données ou 990 CHF si l'on souhaite une extension jusqu'à 20 ans.

Si l'on choisit le niveau basic avec trois durées de conservation mais sans extension possible jusqu'à 20 ans, on obtient un prix de 653 CHF.

Un service d'assistance et de formation pour un tarif de 150 CHF/heure est proposé. Les institutions qui adhèrent à l'association obtiennent des avantages en fonction de la catégorie choisie. (OLOS, UNIGE 2022).

Annexe 14 : Glossaire

Archivage : L'archivage est une notion mal comprise. Bien souvent, les équipes de recherche pensent qu'il suffit de stocker et de sauvegarder les données sur le serveur de l'institution ou sur un disque dur externe pour les archiver mais il n'en est rien (Sbeih, Dedet, et al. 2020; Hadrossek et al. 2021). Il s'agit dans ce cas de stockage à long terme mais non d'archivage. Le processus commence dès la création des données en s'attachant à « décrire, documenter et contextualiser les données » et va se poursuivre tout au long du projet (Hadrossek et al. 2021, p. 14). Sbeih et al. (2020) parle de « démarche qui vise à définir et mettre en œuvre des méthodes et des outils pour gérer et conserver des documents et des données... [afin de les rendre] disponibles et accessibles pour répondre aux besoins métiers et scientifiques ainsi qu'aux obligations légales et sauvegarder les documents et données ayant un intérêt patrimonial » (p. 2)

Conservation : En archivistique, la norme ISO 15489 sur le Records management définit la conservation comme les « actions et tâches concourant à la pérennité technique et intellectuelle des documents authentiques » (Organisation internationale de normalisation 2001).

Partage : Partager ses données consister à « divulguer publiquement les données de recherche que vous avez collectées, en les rendant trouvables, accessibles et réutilisables » (traduit de Bezjak, Vipavc Brvar, CESSDA Training Team 2017, p. 157).

Pérennisation : Souvent considérée comme un synonyme de préservation (Hadrossek et al. 2021; Sbeih, Dedet, et al. 2020), la pérennisation « permet de faire face à la perte d'informations d'identification ainsi qu'à l'obsolescence des supports et des logiciels. Elle consiste en effet à identifier et à conserver des documents et des données pour les rendre accessibles sur le moyen (10 ans et plus) et le long terme (50 ans et plus) » (Sbeih, Dedet, et al. 2020, p. 4). Selon l'OAIS la pérennisation est « le fait de maintenir sur le long terme l'information immédiatement compréhensible... avec les preuves de son authenticité » (Comité consultatif pour les systèmes de données spatiales 2017, p. 7).

Préservation : La préservation est le fait « de garantir, protéger, mettre à l'abri, sauver d'un dommage ou d'une destruction (cf. notion de sauvegarde) et au fait de tenir dans le même état, en bon état (intelligible). Elle fait aussi référence à la notion de permanence dans le temps (cf. notion d'archivage) » (Hadrossek et al. 2021, p. 56).

Sauvegarde : Appelée aussi back-up, la sauvegarde consiste à réaliser une copie des données sur un autre support que celui utilisé pour le stockage afin de pouvoir les récupérer « en cas de perte ou de dégradation » (Hadrossek et al. 2021, p. 55).

Stockage : Consiste à enregistrer les données sur un support (ordinateur, serveur, disque dur externe...) pour les rendre accessibles sur le court terme, généralement jusqu'à la fin du projet. Les données ne sont ni sauvegardées, ni sécurisées et cette solution n'offre pas de garantie de pérennisation sur le long terme (Hadrossek et al. 2021; Sbeih, Dedet, et al. 2020; Université de Lausanne. Uniris 2014)