



---

## **CELLULE DATA STEWARDSHIP RAPPORT ENTRETIENS CHERCHEUR·EUSES DOMAINE INGÉNIERIE ET ARCHITECTURE**

Xavier Brochet, Data steward – Octobre 2024

---

### **TABLE DES MATIÈRES**

I. INTRODUCTION.....	1
II. PRATIQUES EN MATIÈRE DE GESTION DES DONNÉES.....	2
III. PRINCIPAUX DÉFIS.....	7
IV. CONCLUSION ET RECOMMANDATIONS.....	9

### **I. INTRODUCTION**

---

La gestion des données de recherche, en lien avec les principes de l'Open Research Data (ORD), prend une place croissante dans le milieu académique. À la HES-SO, l'ORD constitue un axe stratégique pour assurer la transparence, la reproductibilité et la visibilité des résultats de recherche. Cependant, l'application de ces principes varie selon les disciplines, en fonction des caractéristiques des données et des pratiques des chercheur·euses. Dans ce cadre, la Cellule Data Stewardship s'engage à sensibiliser et accompagner dans la mesure du possible, les chercheur·euses dans la gestion et l'ouverture de leurs données, tout en prenant en compte les spécificités disciplinaires.

Pour mieux comprendre les pratiques et les défis liés à l'ORD dans le domaine Ingénierie et Architecture (IA), 22 entretiens semi-structurés ont été menés par Xavier Brochet, data steward du domaine IA, entre janvier et mai 2024 (voir liste des entretiens, tableau 1). Ces entretiens avaient pour but de recueillir des retours d'expérience et d'identifier les pratiques des chercheur·euses en matière de stockage, de partage et de diffusion des données mais aussi de pointer les défis associés à cette nouvelle culture.

Ce rapport synthétise les résultats de cette enquête, en mettant en lumière la diversité des données produites et les enjeux spécifiques à leur gestion. Anonymisées pour préserver la confidentialité des participant·es, les informations recueillies serviront à formuler des recommandations pour améliorer la gestion des données dans le domaine IA. Ce document vise à nourrir les réflexions internes en vue de l'élaboration de la stratégie et du plan d'action ORD de la HES-SO et peut également servir de support pour alimenter d'autres processus de planification stratégique au sein des Hautes écoles et/ou du domaine.





	Institution	Domaine
Entretien 1	HES-SO Valais	Conception mécanique et hydroélectricité Nanotechnologie, Physique appliquée, Surfaces et matériaux fonctionnels Architecture logiciel et application industrielle Système embarqué et intelligence artificielle Foudre et champs électromagnétiques ; Génie électrique, électronique et communication Informatique embarqué et pharmacologie chimique, Approches mathématiques de la pharmacocinétique Domaine d'application des sols et de l'environnement Biologie moléculaire et végétale Analyse de données Énergie, Valorisation énergétique de la biomasse Technologie de l'éducation Technologie alimentaire et Biotechnologie Conception mécanique et simulation numérique Internet des objets UX, Conception centrée sur l'utilisateur Chimie industrielle, Automatisation, prototypage et conduite de procédé Mécanique des fluides et aérodynamique Ingénierie industrielle et informatique Gestion de la nature, biodiversité spécialisée écosystème aquatique Science du sol, pédologie et environnement
Entretien 2	HEIG-VD	
Entretien 3	HEIG-VD	
Entretien 4	HEIG-VD	
Entretien 5	HEIG-VD	
Entretien 6	HEIG-VD	
Entretien 7	HES-SO Valais	
Entretien 8	Changins	
Entretien 9	Changins	
Entretien 10	HEIG-VD	
Entretien 11	HEIG-VD	
Entretien 12	HEIG-VD	
Entretien 13	HES-SO Valais	
Entretien 14	HEIG-VD	
Entretien 15	HEIG-VD	
Entretien 16	HE-Arcs	
Entretien 17	HEIA-FR	
Entretien 18	HEPIA	
Entretien 19	HEPIA	
Entretien 20	HEPIA	
Entretien 21	HE-Arcs	
Entretien 22	HEIA-FR	

Tableau 1 - Liste des entretiens

## II. PRATIQUES EN MATIÈRE DE GESTION DES DONNÉES

### 2.1 Types de données

Les chercheur-euses du domaine IA manipulent une grande variété de données, principalement quantitatives et électroniques. La diversité des disciplines implique des outils et pratiques variés, rendant la transition vers des formats libres complexe pour certaines catégories de données (tableau 2).

Catégorie	Données	Outils Propriétaires et Formats	Problèmes Potentiels (Transition vers formats FAIR)	Complexité de la Transition
<b>Données quantitatives</b>	Mesures, Statistiques, Données tabulaires, Signaux de longueur d'onde de laser, Données de spectre	<b>Excel (XLSX)</b> , MATLAB, logiciels de spectroscopie spécifiques	Conversion vers formats ouverts comme CSV, risque de perte de structure ou précision	<b>Faible</b> : Peu de perte de données avec CSV ou TXT
<b>Données qualitatives</b>	Entretiens, Observations, Description des terrains (structure, texture)	<b>Word (DOCX)</b> , PDF, logiciels de traitement de texte propriétaires	Conversion facile vers des formats ouverts (TXT, JSON), perte de mise en page mais peu d'impact sur les données	<b>Très faible</b> : Peu de risques
<b>Modèles et simulations</b>	Modèles numériques	<b>Autocad (DWG)</b> , SolidWorks, Ansys	Perte de détails ou fonctionnalités	<b>Élevée</b> : Logiciels





	(CAO/DAO), Simulations et résultats d'expériences, Modèle physique		spécifiques lors de la conversion vers STEP, DXF	complexes, peu de solutions libres viables
<b>Données géospatiales</b>	Cartes, SIG (Systèmes d'Information Géographique)	<b>ArcGIS, Google Earth</b>	Difficulté à convertir des formats vers GeoJSON ou KML, perte de métadonnées spatiales	<b>Moyenne</b> : Solutions libres (QGIS) disponibles mais conversion complexe
<b>Imagerie et échantillons</b>	Images en haute définition, Échantillons de sol, Scanner du sol pour obtenir des images 3D	<b>Photoshop (PSD), RAW, TIFF</b>	Conversion des formats propriétaires (RAW, PSD) vers PNG ou JPEG peut entraîner une perte de qualité	<b>Faible</b> : Perte de qualité possible mais outils disponibles
<b>Données de conception et plans</b>	Données CAO, Plans mécaniques 3D, Fichiers d'impression 3D, Cartes électroniques, Listes de matériels, Analyse chimique	<b>Autocad (DWG), SolidWorks (SLDPRT), EagleCAD</b>	Conversion vers STEP ou STL entraîne souvent une perte de détails ou de fonctionnalités	<b>Élevée</b> : Logiciels complexes, peu d'alternatives libres
<b>Données biologiques</b>	Séquences biologiques	<b>Propriétaires des laboratoires de séquençage</b>	Conversion vers FASTA possible mais risques de perte de métadonnées	<b>Moyenne</b> : Formats ouverts disponibles mais conversion délicate
<b>Documents textuels</b>	Questionnaires, Rapports	<b>Word (DOCX), PDF, Excel (XLSX)</b>	Conversion facile vers des formats ouverts (TXT, CSV), risque minime pour les données	<b>Très faible</b> : Conversion simple et peu de perte
<b>Images</b>	Photos, Dessins, Scans	<b>Photoshop (PSD), RAW, TIFF, JPEG</b>	Compression et perte de qualité possibles lors de la conversion vers PNG, JPEG	<b>Faible</b> : Impact limité sur les données
<b>Données 3D</b>	Schémas, Conception assistée par ordinateur	<b>Autocad (DWG), SolidWorks (SLDPRT)</b>	Conversion vers STL ou DXF peut entraîner une perte de détails et de précision	<b>Élevée</b> : Transition difficile avec peu d'outils libres
<b>Audio</b>	Enregistrements sonores	<b>Audacity (AUP), MP3, WAV</b>	Conversion des fichiers MP3 en formats libres peut réduire la qualité sonore	<b>Faible</b> : Risque de qualité sonore, mais outils disponibles
<b>Vidéo</b>	Enregistrements vidéo	<b>Adobe Premiere (PRPROJ), MP4</b>	Conversion vers des formats ouverts entraîne parfois des pertes de qualité vidéo	<b>Faible</b> : Problèmes minimes avec des outils standards
<b>Mesures</b>	Microscopes, spectromètre, capteur de mouvement, température, hydrométrie	<b>LabVIEW, MATLAB,</b> logiciels spécifiques propriétaires	Conversion vers des formats ouverts est possible mais perte de précision et métadonnées	<b>Moyenne</b> : Conversion faisable, mais perte de détails
<b>Code informatique</b>	Scripts, Algorithmes	<b>MATLAB, RStudio, Eclipse (JAVA)</b>	Conversion possible vers des formats	<b>Faible à Moyenne</b> :





et algorithmes			libres comme Python ou R, mais difficulté de compatibilité	Outils libres disponibles mais compatibilité limitée
----------------	--	--	--	--

Tableau 2 - Types de données ; Outils & Formats

Les types de données et les outils utilisés varient grandement en termes de facilité de conversion vers des formats libres conformes aux principes FAIR. Les outils de traitement de texte, comme **Word (DOCX)** ou **Excel (XLSX)**, sont faciles à convertir en formats libres tels que **TXT** ou **CSV** avec peu de risque de perte significative des données. La principale perte est la mise en page ou le formatage, ce qui n'impacte généralement pas la précision des données elles-mêmes. En revanche, pour des logiciels plus complexes comme **Autocad (DWG)**, **SolidWorks** ou **Ansys**, la conversion pose des défis majeurs. Ces outils propriétaires gèrent des données très spécifiques, comme des modèles 3D ou des simulations numériques, qui peuvent inclure des couches, des annotations ou des détails précis difficilement convertibles dans des formats libres comme **STEP** ou **STL** sans perte de précision. De plus, les solutions libres équivalentes à ces logiciels propriétaires sont limitées et souvent moins performantes, ce qui rend la transition beaucoup plus difficile pour des projets complexes. Pour des données comme celles gérées par des systèmes **SIG (ArcGIS)**, des solutions libres existent (**QGIS** par exemple), mais la conversion entre formats peut entraîner la perte de couches géospatiales ou de métadonnées. Ce qui rend ces transitions plus compliquées que des simples documents textuels. Les données issues des logiciels de CAO ou de simulation nécessitent une attention particulière dans la transition vers des formats libres, car les pertes de détails, de calibrage ou de fonctionnalités peuvent affecter les résultats finaux. En revanche, les données textuelles ou basiques peuvent être converties avec peu de contrainte, et les outils libres de traitement de texte sont largement accessibles.

## 2.2 Stockage des données

Les volumes de données manipulés par les chercheur·euses peuvent atteindre des tailles très importantes, allant jusqu'à plusieurs dizaines de To. Lorsque ces grands volumes dépassent les capacités des infrastructures institutionnelles, les chercheur·euses se tournent souvent vers des solutions personnelles. Celle-ci impliquent fréquemment des pratiques non standardisées et non supervisées ou évaluées par le service informatique de l'institution, ce qui augmente les risques pour l'intégrité des données (voir tableau 3).

Catégorie	Solution	Responsabilité/Utilisation	Remarques
Gestion par équipes ou SI	NAS (Network Attached Storage)	Géré par les équipes de recherche	Partagé entre plusieurs utilisateurs pour stockage localisé
	Serveur de données	Géré par les équipes ou le SI	Utilisé pour héberger des données critiques
	Serveur de calcul	Géré par les instituts ou le SI des écoles	Haute performance pour calcul et stockage temporaire
	Serveurs de stockage et d'archivage de l'école	Géré par l'école	Solution centralisée pour conservation à long terme
Solutions de stockage en ligne	OneDrive, SwitchDrive, Google Drive, Dropbox	Partagé entre les utilisateurs	Accessible depuis n'importe quel appareil connecté
	REDCap	Utilisé pour les données médicales	Conforme aux normes de sécurité médicale
	Git (GitLab, GitHub)	Pour le versioning et la gestion de code	Utilisé principalement par les développeurs
	Google Docs	Collaboratif	Pour l'édition en temps réel





	Teams	Collaboration et stockage de fichiers	Intégré aux outils Microsoft Office
Stockage local	Ordinateur de bureau	Personnel ou de l'école	Risque de perte si non sauvegardé
	Ordinateur portable	Personnel ou de l'école	Mobilité, mais plus fragile
	Disque dur externe	Personnel ou de l'école	Utile pour les sauvegardes, mais risque de panne matérielle
	Site web	Pour le stockage en ligne accessible	Visibilité externe potentielle
	Base de données	Gestion efficace des données	Adapté pour les grandes quantités de données structurées
Autres	Boîtes e-mail	Communication	Risque de surcharge et perte de traçabilité

Tableau 3 - Cartographie des Différentes Options de Stockage de Données

## 2.3 Archivage à long terme

Les pratiques d'archivage sont également hétérogènes, variant selon les écoles et les chercheur·euses. Il n'existe pas de procédure commune entre toutes les écoles pour l'archivage des données de recherche, ni de directive pour aider les chercheur·euses à sélectionner les données à archiver. En l'absence de directives claires, les chercheur·euses manquent de connaissances sur les bonnes pratiques d'archivage. Ce qui entraîne un manque de repère concernant la sélection des données à archiver, la durée de conservation, ou les procédures de destruction des données après une certaine période. De plus, les infrastructures adaptées à ces besoins sont souvent insuffisantes, rendant la gestion de l'archivage parfois inefficace et peu coordonnée.

## 2.4 Partage des données

Sur les 22 chercheurs·euses interrogés·es, 9 ont partagé des données sur des dépôts FAIR non commerciaux. Cependant, ce chiffre ne reflète qu'une partie des pratiques au sein de la HES-SO, l'échantillon étant limité. Le contexte de la recherche souvent réalisé en collaboration avec des entreprises, complique grandement le partage de données, celles-ci étant souvent considérées comme propriété privée des partenaires industriels. Ces derniers préfèrent généralement garder ces informations confidentielles pour des raisons stratégiques.

Le faible taux de partage des données ne s'explique pas uniquement par des contraintes techniques ou juridiques. Les chercheur·euses évoquent également un manque de motivation. Beaucoup estiment que le partage des données représente un coût en temps et en ressources, sans réel bénéfice immédiat. Ils·elles peinent à percevoir les avantages, en particulier lorsque le champ d'application de leurs données est limité à une communauté restreinte.

Les chercheur·euses manquent aussi de reconnaissance institutionnelle pour ces efforts, ce qui contribue à la réticence. De plus, ils·elles perçoivent souvent leurs données comme trop contextuelles ou spécialisées pour être utiles à d'autres. Le choix d'une licence appropriée pour le partage et la crainte de perdre un avantage concurrentiel sont également des obstacles récurrents. Les chercheur·euses interrogés ont soulevé plusieurs interrogations liées au partage des données:

- **Connaissance des dépôts** : La plupart des chercheur·euses connaissent peu les dépôts de données disponibles, qu'ils soient spécialisés ou généralistes.
- **Sélection des données à partager** : Les chercheur·euses hésitent souvent sur le type de données à partager (données brutes, nettoyées, ou seulement les métadonnées).





- **Compréhension des données partagées** : Il est crucial pour eux de s'assurer que les données partagées soient compréhensibles et utilisables par d'autres, ce qui nécessite un effort de documentation et de contextualisation.

## 2.5 Réutilisation des données

La réutilisation des jeux de données déposés sur des plateformes FAIR comme Zenodo reste une pratique peu courante. Aucun exemple concret de réutilisation de données provenant de ces dépôts n'a été identifié lors de cette enquête. Bien que certains chercheur·euses consultent des informations dans des bases de données publiques, ils·elles ne considèrent généralement pas cette démarche comme une réutilisation de données, révélant une certaine confusion ou une incompréhension du terme « réutilisation de données ». Il semble également exister une incompréhension concernant la distinction entre les bases de données traditionnelles, reconnues et bien établies au sein de la communauté scientifique, et les dépôts de données qui prennent de plus en plus d'importance avec l'adoption des procédures liées à l'ORD. Alors que les premières sont perçues comme des sources classiques et fiables, les dépôts de données ORD sont encore mal compris, et leur rôle dans la réutilisation des données est souvent sous-estimé. En théorie, les chercheur·euses sont ouverts à l'idée d'utiliser des données externes, à condition qu'elles soient pertinentes, fiables et bien documentées. Toutefois, ils·elles se montrent réticents à réutiliser des données sans en évaluer rigoureusement la qualité. Cette méfiance s'explique par divers facteurs: la fiabilité des sources, des méthodologies de collecte potentiellement biaisées, le manque de standardisation, des erreurs humaines possibles, ou encore l'obsolescence des données. L'absence de transparence sur les processus de collecte et de traitement renforce également ces doutes, amenant les chercheur·euses à remettre en question la pertinence des données partagées pour leur propre recherche. Souvent, ils·elles préfèrent obtenir directement les données auprès de leurs pairs, via leurs réseaux de connaissances. Une exception notable concerne les codes informatiques, qui sont fréquemment partagés et réutilisés via des plateformes de gestion de versions comme GitHub ou GitLab, ainsi que par des Gits institutionnels. Contrairement à Zenodo, Git n'est pas un dépôt de données, mais un système de contrôle de version utilisé pour la gestion des logiciels. La réutilisation des données varie également selon les domaines d'application :

- **Biologie** : Utilisation fréquente de bases de données en ligne telles que NCBI et UniProt.
- **Data Science** : Utilisation de jeux de données disponibles sur des plateformes comme Kaggle.
- **Imagerie** : Usage d'images et de vidéos provenant de sources telles que la NASA ou l'ESA.
- **Environnement** : Exploitation de données provenant de MétéoSuisse, InfoSpécies, SITG (Système d'Information du Territoire Genevois) ou de l'Office fédéral de la topographie.

Les chercheur·euses peuvent également contribuer à enrichir ces bases de données en y ajoutant les données collectées lors de leurs recherches. Cependant, la réutilisation des jeux de données FAIR reste, pour l'instant, limitée et dépend fortement du domaine d'application et des besoins spécifiques des chercheur·euses.





## III. PRINCIPAUX DÉFIS

---

### 3.1 La propriété des données

La question de la propriété des données est un enjeu central dans l'ORD. La majorité des personnes interrogées sont conscientes que les données collectées et ou produites appartiennent à l'institution qui les emploie. Toutefois, des ambiguïtés surgissent souvent, notamment dans le cadre de projets collaboratifs impliquant plusieurs institutions publiques. Il devient alors essentiel d'établir des accords clairs dès le début des projets pour éviter les conflits, assurer une gestion efficace des données et garantir le respect des réglementations en vigueur. Un autre point de confusion fréquent concerne la distinction entre la propriété des données et la propriété intellectuelle. Si les chercheur·euses reconnaissent généralement que les données appartiennent à leur employeur, ils-elles peuvent néanmoins percevoir ces données comme faisant partie de leur propre travail intellectuel, d'où une confusion entre ces deux notions. Deux exemples concrets illustrent l'impact que cette confusion peut entraîner :

- **Absence de contrat spécifique** : contrairement aux projets avec des partenaires industriels, les projets impliquant plusieurs chercheur·euses ou institutions ne font généralement pas l'objet de contrats spécifiques, laissant floue la question de la propriété des données. Cela peut générer des tensions quant à l'accès, l'utilisation ou le partage des données entre les parties.
- **Projets menés par des étudiant·es** : la propriété des données issues des projets étudiant·es (master, doctorat) demeure un sujet sensible. Les personnes interrogées s'interrogent sur la titularité des droits de propriété intellectuelle pour le corps étudiantin·e qui n'est pas contractuellement lié à une Haute école.

Ces problématiques soulignent l'importance d'une clarification des rôles et des responsabilités dès le début des projets, ainsi que d'une meilleure compréhension des distinctions entre propriété des données et propriété intellectuelle.

### 3.2 Les aspects légaux

- **Utilisation des logiciels sous licence** : des questions se posent sur l'utilisation de logiciels sous licence universitaire pour des projets facturés (mandats) ou de recherche (FNS, InnoSuisse). Si l'utilisation est restreinte à l'enseignement, les institutions doivent proposer des alternatives pour des logiciels non libres comme Matlab ou des logiciels de CAO (exemples : Ansys, COMSOL, AutoCAD, SolidWorks).
- **Problèmes juridiques avec des partenaires étrangers** : lors de projets internationaux, des questions sur la législation applicable (suisse ou du pays tiers) se posent, notamment pour la collecte de données et l'utilisation des licences. Il est crucial de clarifier quelle loi s'applique et comment gérer les contradictions éventuelles entre les législations suisses et étrangères, en particulier lorsque des partenaires locaux sont impliqués.

Une meilleure sensibilisation et des clarifications sur ces questions juridiques sont nécessaires pour garantir la conformité éthique et légale des projets internationaux.

### 3.3 Transfert et partage des données

Le transfert de données est une pratique courante dans les projets collaboratifs, notamment lorsque plusieurs partenaires sont impliqués. Cependant, ce processus présente des défis importants, en particulier lorsque les données sont sensibles ou contiennent des informations personnelles. Bien que des mesures de sécurité soient souvent mises en place, l'absence de protocoles standardisés et de bonnes pratiques partagées pousse fréquemment les chercheur·euses à développer des solutions internes. Ces systèmes, conçus au cas par cas, varient considérablement d'une équipe à l'autre.



La diversité des méthodes de transfert complique la collaboration entre les partenaires et soulève des enjeux de sécurité, de confidentialité et de gestion cohérente des données. En l'absence de standards communs, les risques de failles de sécurité, d'utilisation non autorisée ou de mauvaise manipulation augmentent. De plus, le non-respect de réglementations strictes telles que le RGPD en Europe ou la LPD en Suisse (Loi fédérale sur la protection des données) peut entraîner des conséquences juridiques importantes.

Pourtant, la mise en place d'infrastructures sécurisées et le développement de normes communes pour le transfert et le partage des données apporteraient de nombreux bénéfices. Non seulement cela garantirait une meilleure sécurité et confidentialité, mais cela faciliterait également la collaboration, la reproduction des résultats, et l'innovation scientifique à grande échelle.

Plusieurs obstacles rendent le partage des données difficile :

- **Temps et coût** : préparer des données pour le partage est chronophage et coûteux. Aucun budget spécifique n'est souvent prévu pour cette tâche.
- **Projets avec partenaires industriels** : La propriété intellectuelle et les données générées sont souvent protégées par des contrats, limitant ainsi la possibilité de partager ces informations.
- **Manque d'intérêt** : Certains chercheur·euses estiment que partager leurs données n'apporte pas de réelle valeur, surtout si leur domaine de recherche est très spécialisé.
- **Risque de concurrence** : la crainte de perdre un avantage concurrentiel en partageant des données est un frein important.
- **Manque d'expérience** : les chercheur·euses manquent souvent de pratique dans le partage de données et connaissent peu les procédures pour déposer ou accéder à des datasets dans les dépôts existants.
- **Documentation insuffisante** : le manque de documentation adéquate (métadonnées, fichiers explicatifs) rend souvent difficile l'utilisation des données partagées.
- **Propriété des données** : il y a une confusion fréquente sur la propriété des données publiquement partagées. Les chercheur·euses pensent parfois à tort que ces données ne sont plus sous la propriété de l'institution une fois publiées.
- **Confusion conceptuelle** : les chercheur·euses mélangent souvent la valorisation et le partage des données, et d'une manière plus large l'ORD avec l'Open Access.

### 3.4 La réutilisation des données existantes

**Utilisation des données issues de forums ou outils numériques** : les chercheur·euses se posent des questions sur les réglementations légales entourant l'extraction et l'utilisation de données provenant de forums en ligne ou d'autres sources numériques. Il est essentiel de connaître les lois applicables pour garantir une utilisation éthique et légale de ces données.

### 3.5 Soutien

**Le rôle des bibliothécaires** : le rôle des bibliothécaires dans le soutien à l'Open Science, en particulier pour la relecture des DMP, est apprécié. Toutefois, il a été mentionné à plusieurs reprises, qu'un appui supplémentaire de personnes ayant une expérience en recherche, renforcerait la pertinence de cet accompagnement. Une collaboration entre bibliothécaires et chercheur·euses expérimenté·es permettrait de mieux répondre aux besoins spécifiques des différents domaines.

### 3.6 Divers

**Impact environnemental** : le stockage inutile de données est un point de préoccupation, les chercheur·euses interrogé·es soulignant l'impact environnemental d'une telle pratique.







## IV. CONCLUSION ET RECOMMANDATIONS

---

### 4.1 Conclusion

Les pratiques liées à l'ORD sont encore relativement nouvelles et doivent être progressivement intégrées dans les processus de travail des chercheur·euses. Ces changements peuvent être perçus comme complexes, voire perturbants, mais il est essentiel de préparer les équipes de recherche à adopter ces pratiques, qui deviendront progressivement la norme. L'enquête révèle que la gestion des données de recherche n'est souvent pas optimisée. Les équipes définissent leur propre méthode, mais peu d'entre elles suivent des procédures formelles, notamment pour la réutilisation et la sécurité des données. Bien que la nécessité d'une bonne gestion soit reconnue, elle n'est pas systématiquement mise en œuvre, et les pratiques varient d'une équipe à l'autre. Ce manque de standardisation et de procédures formalisées complique la gestion et le partage futur des données, principalement pour les données sensibles. Pour encourager une gestion efficace des données de recherche selon les principes FAIR, un soutien renforcé est nécessaire sous forme de formations, de ressources humaines dédiées, d'infrastructures adaptées et de guides pratiques. Bien que certaines initiatives soient déjà en place au sein des écoles et du rectorat, il est impératif d'intensifier les efforts pour sensibiliser les chercheur·euses aux bonnes pratiques de gestion des données, à l'utilisation des différentes licences, et aux avantages d'un partage sécurisé. Une approche coordonnée et bien financée est cruciale pour répondre à ces besoins. Une meilleure coopération entre les différentes entités de la HES-SO, ainsi qu'une mutualisation des ressources et une communication efficace, sont essentielles pour sensibiliser les chercheur·euses et améliorer les pratiques de gestion des données.

### 4.2 Recommandations

Dans le cadre de l'amélioration de la gestion des données de recherche et de l'adoption des principes FAIR, nous proposons cinq axes de recommandations à court et moyen terme, visant à renforcer les outils, les ressources humaines, les documents cadres, la formation et sensibilisation, ainsi que les incitations au partage des données. Ces actions permettront de structurer et soutenir efficacement les pratiques de gestion des données au sein des Hautes écoles.

#### Guides et outils

- Créer une check-list pour la gestion des données de recherche adaptée aux exigences et spécificités du domaine IA.
- Créer un guide sur les bonnes pratiques pour l'ouverture des logiciels et codes sources, en réponse aux principes FAIR.
- Promouvoir et encourager l'utilisation d'outils dédiés pour aider les chercheur·euses à sélectionner la licence la plus appropriée pour leurs codes et logiciels, garantissant une gestion claire des droits d'auteur, des conditions de partage et d'utilisation, facilitant la collaboration et favorisant l'adoption du logiciel par un public plus large.

#### Ressources humaines et réseau

- Renforcer et valoriser le réseau des référents en Open Research Data (ORD) en favorisant les collaborations intercantoniales pour faciliter le partage des bonnes pratiques et améliorer l'accès aux ressources, notamment dans les cantons où ces services sont moins développés.
- Soutenir l'interaction entre chercheur·euses et bibliothécaires spécialisés en open science, en particulier sur les questions d'ORD, pour enrichir et renforcer l'aide existante. Le binôme chercheur·euse - bibliothécaire permet de compléter les connaissances et expertises, et d'offrir un accompagnement plus ciblé et efficace dans la gestion des données.





## Documents cadres et infrastructure

- Encourager les Hautes écoles du domaine à adopter des documents-cadres pour la gestion et l'ouverture des données de recherche, précisant clairement les processus de stockage, de transfert, et d'archivage des données, ainsi que les rôles et responsabilités associés.
- Renforcer la collaboration entre les équipes de recherche des Hautes écoles et les services informatiques pour mieux répondre aux exigences de sécurité et de protection des données de recherche.

## Formation et sensibilisation

- Proposer un coaching sur la gestion des données de recherche pour tout nouveau projet ou nouvel arrivant impliqué dans la Ra&D.
- Renforcer les ateliers de sensibilisation sur l'ORD au sein des Hautes écoles.
- Améliorer la communication au sein des hautes écoles et du Rectorat pour mieux promouvoir les ressources en ligne et les services de soutien disponibles, tels que les outils et documentations offerts par la HEIG-VD, la HES-SO valais et les pages ORD du Rectorat.
- Favoriser l'interconnexion des sites web des hautes écoles et du Rectorat afin de rendre les ressources plus accessibles et complémentaires. Cela permettrait de capitaliser sur l'expertise des établissements plus avancés et de renforcer l'accès aux services de soutien à travers un réseau intégré.

## Incitations au partage des données

- **Aide financière** : Allouer un budget spécifique pour soutenir l'ouverture des données de recherche dans les projets non financés par le FNS ou Horizon-Europe, qui incluent déjà des fonds pour l'ORD.
- **Prix ORD du domaine IA** : Créer un fonds dédié à récompenser les chercheur·euses qui partagent activement leurs données et s'engagent dans la promotion de l'Open Research Data.
- **Témoignages et webinaires** : Organiser des présentations de "success stories" ou de "data champions" pour démontrer concrètement les avantages du partage des données, notamment en termes de collaborations et de visibilité accrue.
- **Affichage des DOI** : Intégrer les DOI des jeux de données dans les systèmes internes (comme ArODES) pour donner visibilité et reconnaissance aux chercheur·euses.

