Rapport interne FGSE
Résultats de l'enquête « Pratiques numériques de recherche »

Table des matières

1. Contexte et objectifs de l'enquête

Les données de recherche soulèvent aujourd'hui des enjeux complexes et délicats – disciplinaires, juridiques, méthodologiques, techniques. La gestion de grands volumes de données, l'application des principes FAIR[1] – désormais considérés comme une des priorités pour l'agenda de la Science Ouverte – la résolution des questions éthiques et de sécurité, sont autant de défis qui exigent des chercheur·es d'adapter continuellement leurs pratiques.

Afin de fournir un soutien axé sur les besoins réels des chercheuses et chercheurs de la FGSE, l'enquête « pratiques numériques de recherche » a été menée au printemps 2023. Elle avait pour objectifs de :

- faire le point sur les pratiques numériques de gestion des données de recherche ;
- identifier les besoins des chercheur·es en termes de soutien, d'information et de formation dans ce domaine ;
- identifier des manques au regard des infrastructures et outils numériques nécessaires pour mener à bien la recherche scientifique.

Le présent rapport résume les principaux résultats et constats de l'enquête, présente clarifications et solutions aux problèmes et questions spécifiques soulevés par les répondant·es. Il identifie également les points de préoccupation et donne quelques recommandations pour améliorer la gestion des données de recherche.

L'enquête et le présent rapport ont été réalisées par un groupe de travail (GT) constitué notamment de personnels de soutien à la recherche de la FGSE : la consultante recherche Amélie Dreiss, la spécialiste données de recherche Zhargalma Dandarova (data steward), les ingénieur·es recherche Margot Sirdey et Flavio Calvo, ainsi que de Carmen Jambé de l'UNIRIS et de Thierry Lombardot de la DCSR.

2. Méthodologie de l'enquête

Le questionnaire, élaboré par le GT, abordait les pratiques numériques en fonction des principales phases du cycle de vie des données de recherche (collecte, traitement, stockage, partage et archivage). Le lien vers le questionnaire en ligne, réalisé avec le logiciel LimeSurvey, a été envoyé par courriel à l'ensemble des chercheur·es de la faculté. Le lien est resté actif du 02.03 au 23.04 2023.

3. Participant·e·s à l'enquête

En tout, 60 personnes ont accepté de répondre au questionnaire (17 % de la totalité des membres FGSE interrogés) :

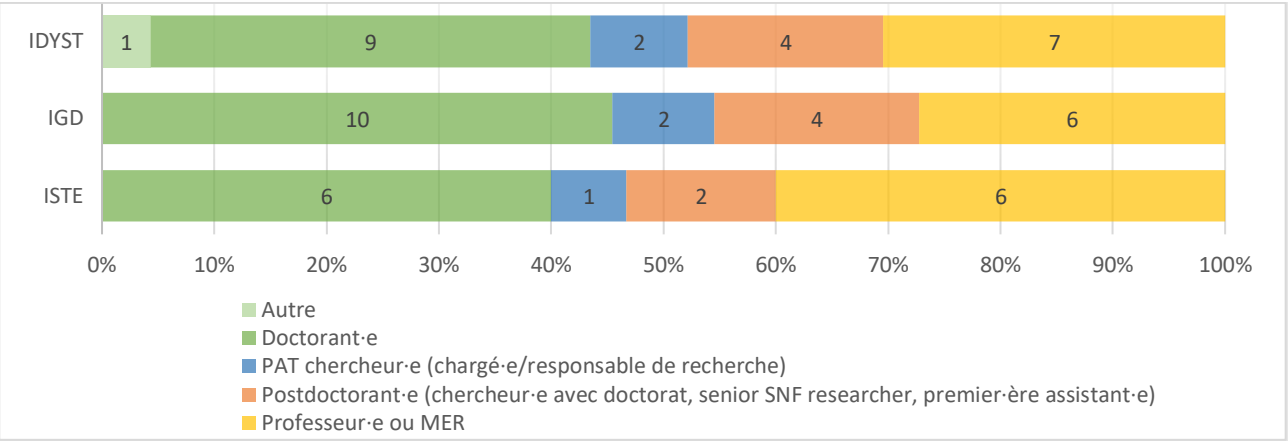
- ISTE (N=15),
- IGD (N=22),
- IDYST (N=23)

L'échantillon couvre l'ensemble des instituts de la FGSE, ainsi qu'une diversité de fonctions académiques : 42 % de doctorant·es, 17 % de postdoctorant·es (senior researcher FNS,

[1] Faciles à trouver, Accessibles, Interopérables et Réutilisables (voir L'Open Science à l'UNIL)

premier·ères assistant·es), 10 % de chercheur·es PAT (chargé·e ou responsables de recherche) et 32 % de professeur·es et de MER ont participé à l'enquête.

Figure 1 - Répartition de l'échantillon par statut professionnel et par institut de rattachement



4. Outils et infrastructures informatiques utilisés

Objectif de cette partie : faire l'état des lieux des outils, des infrastructures et des pratiques numériques utilisés pour la collecte et l'analyse des données de recherche ; identifier les besoins et les manques dans ce domaine.

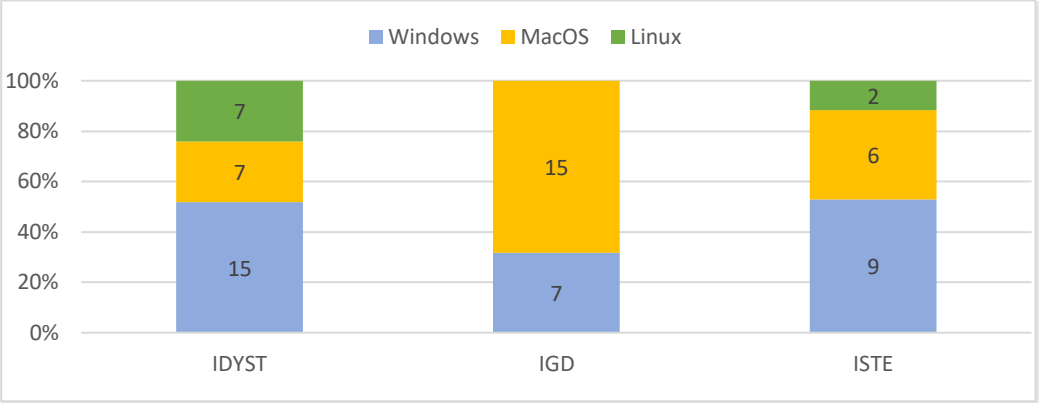
4.1. Ordinateurs et systèmes d'exploitation

Les ordinateurs portables sont de loin les machines professionnelles principales, d'après l'enquête (97 % des sondé·es). 30 % déclarent aussi utiliser un ordinateur de bureau, 17 % des machines virtuelles de l'UNIL et enfin 13 % travaillent sur le cluster de l'UNIL Curnagl. Les autres centres de calcul nationaux et internationaux semblent très peu utilisés.

Ordinateur professionnel	IDYST (N=23)	IGD (N=22)	ISTE (N=15)	Total (N=60)
Ordinateur portable	23	20	15	97% (58)
Ordinateur de bureau	7	3	8	30% (18)
Machines virtuelles de l'UNIL	10	0	0	17% (10)
Curnagl (UNIL HPC cluster)	7	1	0	13% (8)
Infrastructure de l'EPFL	0	0	0	0
Autres centres de calcul HPC nationaux/internationaux	4	0	1	8% (5)
Services cloud privés (Google, NVIDIA, Amazon, Intel..)	2	0	0	3% (2)
Autre (ordinateur de labo, serveur d'institut)	0	0	3	5% (3)

Concernant les systèmes d'exploitation utilisés sur les machines personnelles (portables ou de bureau), l'enquête montre une répartition équitable de l'utilisation de Windows et macOS au sein de toute la faculté (avec des préférences claires de Windows pour l'IDYST et de macOS pour l'IGD). Le système d'exploitation Linux semble peu présent (de l'ordre de 15 % sur l'ensemble de la faculté).

Figure 2 Systèmes d'exploitation par institut de rattachement



Remarque de répondant·e : ***Il manque à l'Unil un engagement sérieux en faveur des logiciels libres. Je regrette en particulier (1) que la suite Microsoft Office soit encore la norme ultra-majoritaire (2) que Microsoft 365 soit proposé malgré les nombreux problèmes que posent les solutions cloud-based (3) que Zoom soit la solution de visioconférence proposée alors qu'existent des logiciels comme Big Blue Button.***
Réponse du GT : Une fraction importante des utilisateur·rices préfère la facilité à un engagement pour des solutions éthiques, garantissant la souveraineté numérique, transparentes, et généralement aussi OpenSource. Les ingénieur·es de recherche (research-computing-fgse@unil.ch) sont disponibles pour aider les personnes souhaitant se familiariser avec ces solutions alternatives.
Pour le point (3) il y a un malentendu : Zoom n'est pas la solution de visioconférence préconisée à l'UNIL, mais celle généralement adoptée par les chercheur·es. L'UNIL préconise en revanche l'usage de Teams (qui n'est certes pas non plus un logiciel libre).

4.2. Logiciels et outils informatiques pour la collecte[2] et le traitement[3] des données

Certains logiciels sont très utilisés, notamment ceux liés à la cartographie, la programmation, les statistiques, mais aussi l'analyse d'image et les entretiens audio et vidéo.

Type d'outils de traitement et de collecte cités	IDYST	IGD	ISTE	Total
analyse qualitative (Atlas.ti)		10		10
cartographie/télédétection (OpenStreetMap, TerrSet, ArcGIS/ArcMap, QGIS/ QField)	6	9	1	16
enregistrement/vidéo (Open Broadcast Software, VLC)		3		3
entretien/enquête (Limesurvey, SenseMaker, Surveymonkey, Teams, Zoom)		7		7
image 3D/nuage (CloudCompare, Blobb3D, logiciels spécifiques)			4	4
photo/image (Agisoft Metashape, Fiji, Photoshop)	2		2	4
logiciel spécifique labo	2		4	6
programmation (Python, Bash, C, Fortran, JavaScript, Jetbrain IDEs, Julia, Matlab)	25	2	18	45
statistiques (R, JMP, SAS, Spss, Stata, Matlab)	12	11	8	31
tableur (libre office calc / Excel, Access)	12	20	11	43
traitement de texte (Word, Google Docs, LaTeX)	2	7	1	10
Transcription (f4transcript, Nvivo, YobiYoba, Vocapia)		6		6

Question de répondant·e : ***Pourquoi limiter l'accès à certains logiciels aux doctorant·es ?***

[2] Collecter/générer = récolter des informations numériques pour ensuite les analyser dans le cadre de votre recherche
[3] Traitement = analyse, nettoyage, structuration, modélisation de données numériques

Réponse du GT : Les logiciels qui ont un accès limité sont des logiciels propriétaires, qui sont dès lors payants, parfois très chers, et pour lesquels le nombre de licences achetées par l'UNIL est dès lors trop restreint pour proposer un accès à tous les doctorant·es. Les doctorant·es qui en ont besoin pour leur projet, avec accord de leur PI, peuvent contacter research-computing-fgse@unil.ch pour voir si une machine virtuelle (avec le logiciel installé) peut être mise en place pour le groupe de recherche.

Remarque de répondant·e : ***La version mobile de Limesurvey mériterait d'être améliorée***
Réponse du GT : Bien qu'en partie sponsorisé par une entreprise, LimeSurvey est aussi un logiciel OpenSource, avec une communauté de développeurs bénévoles et enthousiastes. C'est certainement l'occasion de les remercier et de leur envoyer toutes les idées d'amélioration : https://community.limesurvey.org/feature-request/. L'UNIL compte aussi des développeur·euses compétent·es, mais qui sont déjà très prises par le travail fait en local pour les chercheur·es de l'UNIL.

Plusieurs remarques de chercheur·e concernent l'accès à des logiciel de transcription. En voici une : ***En moyenne, une bonne moitié des membres du corps intermédiaire de l'IGD (env. 40 pers.) consacre 12 journées entières chaque année à la transcription. Une demande d'outil de transcription automatique a été relayée par la direction d'institut au CI qui a répondu qu'il était impossible d'acquérir des logiciels ne pouvant pas être installés en 'local' sur les ordinateurs, pour des raisons de sécurité. La sécurité des données est certes primordiale, mais qu'en est-il des transcriptions sur YouTube, quand le téléversement d'entretiens sur YouTube pose énormément de problèmes en termes de circulation et de sécurité de nos données ? Cette situation est assez paradoxale.***
Réponse du GT : La DCSR et le CI sont en train de travailler pour proposer une solution aussi simple que possible pour la transcription, avec les données traitées en local. N'hésitez pas à partager vos besoins avec research-computing-fgse@unil.ch, ce qui permettrait de prioriser davantage cette thématique et de modeler la solution à vos besoins.
Quant à l'utilisation massive du cloud pour traiter des données personnelles, voire sensibles, elle montre une méconnaissance des risques et du cadre juridique parmi les utilisateur·rices. Afin de connaitre les bonnes pratiques n'hésitez pas à consulter les site UNIL d'UNIRIS (Open research data : https://www.unil.ch/openscience/openscience/openresearchdata). Il est possible également, pour un soutien individuel, ou pour l'organisation d'ateliers en petits groupes, de s'adresser à researchdata@unil.ch.

Demande de répondant·e : ***De bons outils d'analyse d'images, flexibles, contenant des extensions en 3D et utilisant la texture comme base, seraient très utiles.***
Réponse du GT : Hélas le développement de tels logiciels est un travail de longue haleine et qui va bien au-delà des ressources actuelles de l'UNIL. Si des logiciels payants qui s'ajustent à vos besoins existent, l'achat d'une licence est possible, en faisant la demande au PI de votre groupe. L'UNIL ne peut acheter de licences à un niveau institutionnel que pour les logiciels pour lesquels il y a une demande avérée sur l'ensemble de la communauté universitaire.

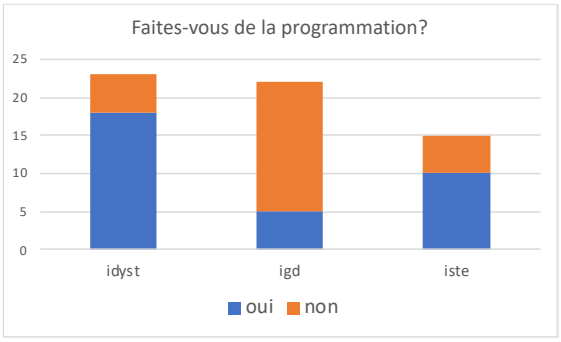
Quelques remarques de répondant·es ont porté sur le ***manque d'information sur le catalogue d'outils offerts à l'UNIL et de cours et tutoriels sur certains outils comme Macro Excel, Python, Adobe Illustrator, QGIS/ARGIS, Github.***
Réponse du GT : Un catalogue des outils proposés/recommandés par l'UNIL avec une documentation basique en français se trouve sur https://www.unil.ch/ci/fr/home/menuinst/catalogue-de-services/materiel-et-logiciel/distribution-de-logiciels.html. Compte tenu de la large palette d'outils utilisée par notre communauté universitaire, une réécriture systématique de documentation et de tutoriels en interne est hélas irréaliste. Ces outils (en particulier Excel, Python et autres langages de programmation) sont largement documentés en anglais et en détail par leurs développeurs, il existe donc une immense offre de tutoriels gratuits et d'exercices sur internet.
Pour ce qui concerne la formation, les ingénieur·es recherche soulignent par contre que la suite Office est une suite bureautique. Au lieu des macros Excel, pour des applications scientifiques, l'apprentissage de langages de programmation tels que Python ou R (pour lesquels des cours sont offerts par la DCSR) est encouragé et rentable sur le long terme. Indépendamment des outils choisis, une aide ponctuelle est aussi offerte lors des permanences (mardi matin GEO-4631) ou par e-mail (research-computing-fgse@unil.ch).
Pour les cours **Illustrator** et **Excel** (bureautique), voir https://courses.unil.ch/ci.
Pour **Python** et **Github**, la DCSR annonce périodiquement des cours. Le CI UNIL offre aussi accès à une offre de cours e-leaning (proposée par le produit *LinkedIn Learning*). Une fois les concepts de base acquis, il est possible de créer des cours supplémentaires pour de petits groupes (research-computing-fgse@unil.ch).
Pour **QGIS/ARGIS**, si intérêt avéré par de petits groupes, prendre contact avec les ingénieur·es recherche : research-computing-fgse@unil.ch.

Il existe aussi des cours de programmation à la FGSE destinés aux étudiant·es, et il est possible de demander d'y assister en candidat libre (https://www.unil.ch/gse/home/menuinst/formations.html).

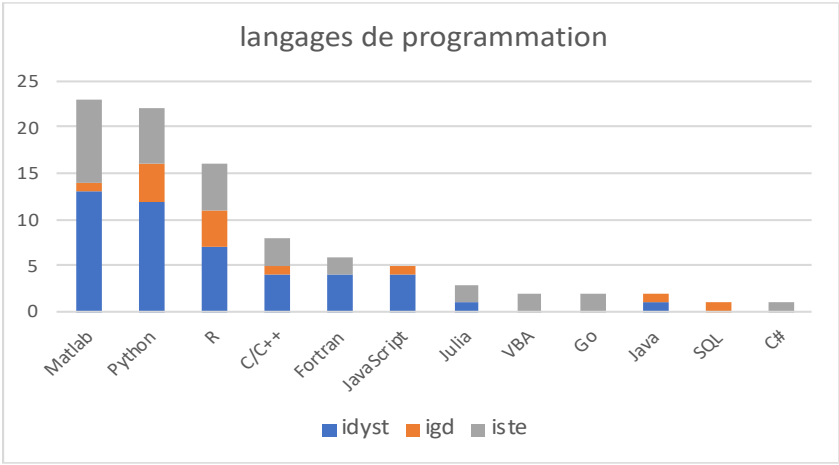
Demande de répondant·e : ***Un soutien pour l'utilisation d'ATLAS.ti serait apprécié.***
Réponse du GT : ATLAS.ti est un outil propriétaire et payant, utilisé par une petite minorité de la communauté universitaire. Compte tenu de la large palette d'outils utilisés dans notre communauté, offrir du support pour tous ces outils est hélas impossible. Lorsque ces outils sont en sus propriétaires et payants, les ingénieur·es recherche n'ont pas la possibilité de se former eux-mêmes pour offrir un tel support.
Les résultats de l'enquête ont montré que plusieurs personnes de la faculté utilisent cet outil. La création d'un réseau d'échanges de bonnes pratiques parmi les utilisateur·rices de cet outil pourrait être une solution à envisager.

4.3. Programmation et HPC[4]

L'IDYST et l'ISTE semblent utiliser plus largement la programmation dans la recherche (plus de 75 % des participant·es au sondage ; avec notamment l'utilisation de Matlab, Python et R) que l'IGD (moins de 25 % d'utilisateur·rices, majoritairement de Python et R).



Le graphique ci-dessous présente les langages de programmation les plus utilisés par les répondant·es de la faculté. Les autres langages mentionnés par les répondant·es sont Bash, SAS, NCL et IDL.



Remarques de répondant·es : ***Il faut un certain temps pour se sentir à l'aise avec certains outils comme Stata, QGIS, R, Python, etc. Bien souvent, les forums ne sont pas d'une grande aide. Avoir un·e expert·e du programme sur le campus ou même à l'Institut à qui l'on pourrait poser des questions sur le code ou le processus serait d'une grande aide.***
Réponse du GT : La DCSR organise des cours pour une sélection de logiciels, souvent liés au calcul de haute performance. Si un nombre suffisant de personnes est intéressé (10 personnes ou plus), des cours

[4] HPC : Calcul Haute Performance

supplémentaires peuvent être organisés. Le CI UNIL offre aussi accès à une offre de cours e-leaning (notamment R, Python, voir le catalogue « LinkedIn Learning »).
En ce qui concerne les logiciels d'analyse statistique, les ingénieur·es recherche vous conseillent plutôt d'utiliser Python ou R, et sont là pour vous aider à optimiser le code et à le porter sur l'infrastructure HPC (contact : research-computing-fgse@unil.ch ou permanence le mardi matin en GEO-4631).
Par ailleurs, si vous souhaitez suivre une formation nécessaire pour votre recherche à l'extérieur de l'UNIL et que vous avez besoin d'un soutien financier, veuillez contacter votre consultante recherche.

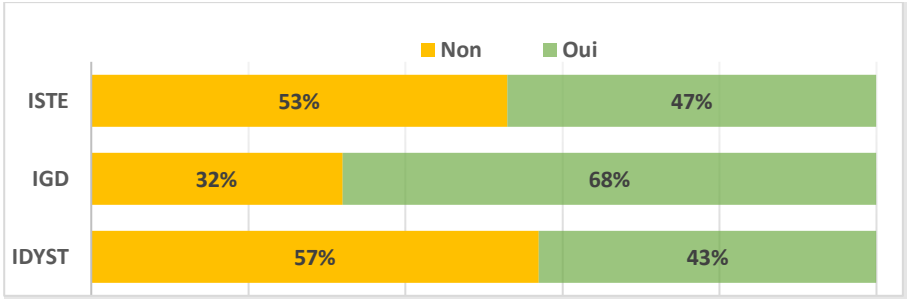
La majorité (63 %) des sondé·es n'ont pas recours à des architectures de Calcul Haute Performance (HPC). Néanmoins, quand c'est le cas, les clusters locaux (de la faculté) sont quasiment autant utilisés que les clusters de l'UNIL. La migration des clusters facultaires aux clusters de l'UNIL semble donc toujours en cours.
D'autre part, à l'échelle des instituts, les deux tiers des sondé·es de l'IDYST utilisent l'infrastructure HPC, pour seulement 20 % sur l'ensemble d'IGD et ISTE.

Solution HPC	IDYST (N=23)	IGD (N=22)	ISTE (N=15)	Total (N=60)
Serveurs d'Institut / de la Faculté	8	0	3	18% (11)
Serveurs de l'UNIL (unité DCSR)	12	2	2	25% (15)
Serveurs fournis par l'EPFL	0	0	0	0
Serveurs fournis par des entreprises privées (Amazon, Intel, etc.)	2	0	0	0
Serveurs fournis par d'autres universités dans le cadre de collaborations avec celles-ci	4	1	1	10% (6)
Mes propres serveurs / workstations	3	0	1	7% (4)
Autre	-	-	1	-
Je n'utilise pas de calcul à haute performance	7	19	11	63% (38)

4.4. Puissance de calcul

Plus de la moitié des participant·es (53 %) ont déclaré s'être heurtés aux limites de calcul de leur ordinateur portable ou de bureau (espace disque insuffisant ou bien pas assez de mémoire vive) et avoir laissé leur ordinateur allumé toute la nuit pour terminer les calculs. Le graphique ci-dessous montre la répartition des répondant·es par institut de rattachement.

Figure 3 - Vous arrive-t-il de heurter les limites de calcul de votre ordinateur portable ou de bureau ?



Alors que 85 % des chercheur·es de l'IGD ayant répondu au sondage n'utilisent pas l'infrastructure HPC, 68 % déclarent heurter les limites de calculs de leurs machines. Cela laisse à penser que le besoin en HPC est présent dans cet institut. L'IDYST, de son côté, fait aussi face aux limites des machines personnelles, même si plus de personnes semblent avoir déjà migré vers l'infrastructure HPC (70 % des 23 participant·es de l'IDYST y ont recours).

Outils et infrastructures : Résumé et points clés

On constate que les outils et logiciels utilisés par les chercheur-es de la FGSE sont variés. Un certain nombre de répondant·es ont mentionné le besoin de formation et de l'appui pour leur utilisation. Le GT constate que certaines demandes témoignent d'une méconnaissance du soutien existant. Dans ses réponses aux sondé·es, le GT indique que des formations existent pour certains outils à la DCSR ou au CI (cours pour l'utilisation de Github, du Cluster, de librairies de parallélisation, etc.). De plus, si une demande en ce sens émane d'un groupe de chercheur·e, des ateliers sur mesure peuvent être mis en place par la DCSR et le personnel de soutien FGSE. Les ingénieur·es recherche fournissent également une assistance individuelle en matière de programmation et sur certains outils. En cas de besoin de soutien financier pour suivre des formations externes, vous pouvez contacter la consultante recherche de la FGSE.

Un autre résultat de l'enquête est que certaines personnes atteignent les limites de capacité de leurs ordinateurs sans pourtant faire appel à l'utilisation du cluster (particulièrement à l'IGD et l'ISTE). Le GT rappelle que des ressources de stockage des données (le NAS) et de calcul scientifique (les clusters Curnagl et Urblauna) sont à la disposition des chercheurs. Au vu des résultats du l'enquête, le GT recommande de s'adresser davantage aux ingénieur·es recherche pour mettre en place ces solutions de calcul scientifique.

5. Gestion de données numériques de recherche

Objectif de cette partie : faire l'état des lieux des pratiques des chercheur·es en matière de stockage, partage et archivage des données de recherche, y compris les données sensibles ou à caractère personnel, identifier les besoins et les manques.

5.1. Sources de données préexistantes

80 % des répondant·es déclarent acquérir des données numériques préexistantes pour leur recherche. On constate qu'il s'agit tout d'abord des bases de données spécifiques à certains domaines tandis que les données ouvertes stockées sur les dépôts génériques (comme Zenodo, DRYAD, re3data etc.) semble encore très peu réutilisées dans le cadre de la recherche à la FGSE.

Sources de données préexistantes	IDYST (N=23)	IGD (N=22)	ISTE (N=15)	Total (N=60)
Bases de données existantes[1]	18	8	6	53% (32)
Bibliothèques ou centres de documentation[2]	2	5	1	13 % (8)
Entreprises ou autres organisations privées[3]	1	3	0	7% (4)
Administrations publiques[4]	8	6	3	28% (17)
Autres universités (dans le cadre de collaborations)]	7	2	2	18% (11)
Dépôts de données[5]	5	2	0	12% (7)
Données d'un précédent projet de recherche propre	9	4	5	30% (18)
Autre[6]	7	2	4	22% (13)

Exemples de sources de données citées par les sondé·es :
[1] Copernicus, UnilGIS, Goodle Earth Engine, HydroSHEDS**,** Data lakes**,** Climate Data store, Australian research council, scopus**,** SHIPS, ERA5, Merra-2, CMORPH, OSM, InfoSpecies, Protected Planet, IBAT, IUCN Red List, World Bank, UN, Georock, ALTIDEM, CRUST1.0, meteosuisse, Africapolis
[2] BCUL, secondary data published in various sources or media articles
[3] t-I (TP Lausanne)
[4] OFEV, swisstopo, OFS, CDC, Hydroatlas.ch, IDAWEB, SLF snow data
[5] Zenodo
[6] articles scientifiques, données publiées, données de suppl. mat., collègues, précédent projet dans le même domaine, articles de presse numérisés accessibles sur internet, associations, GIS database from institutions.

Remarque de répondant·e : ***L'UnilGIS contient beaucoup de données, mais je ne sais pas lesquelles. Il existe apparemment un logiciel qui peut afficher les métadonnées dans chaque dossier, mais il n'est pas facile à utiliser. Ce serait formidable s'il existait une documentation sur UnilGIS permettant de rechercher des mots-clés et de savoir exactement ce qui est disponible.***
Réponse du GT : Il existe deux logiciels, ArcGIS (un logiciel propriétaire) et GeoManager (développé à l'UNIL par Alexandre Hirzel), comme indiqué dans le fichier ReadMe.txt du dossier d'installation disponible sur la page https://www.unil.ch/gis/fr/home/menuinst/geodonnees/geomanager.html. Ce logiciel est encore en phase de développement bêta (ce qui signifie qu'il n'a pas atteint sa maturité), et malheureusement le CI a des ressources limitées pour le développement. Les suggestions peuvent être envoyées directement à Alexandre Hirzel (Alexandre.Hirzel@unil.ch) qui pourra peut-être apporter des améliorations. N'hésitez pas à contacter les ingénieur·es recherche si vous avez besoin d'une aide immédiate pour analyser les métadonnées (research-computing-fgse@unil.ch).

5.2. Solutions de stockage[5] des données de recherche durant le projet

Parmi les différents supports, l'ordinateur personnel (87 %) et le disque dur externe ou la clé USB restent largement utilisés. L'utilisation des infrastructures de l'UNIL dédiées au stockage des données de recherche n'est pas encore une pratique courante : 30 % des sondée·es recourent au NAS DCSR et 22 % stockent leurs données sur les serveurs mis à disposition dans leurs instituts. En revanche, le SWITCHdrive est largement utilisé (57 %). L'utilisation des clouds commerciaux (Dropbox, Google Drive, etc.) reste très répandue (28 %).

Figure 4 - Solution de stockage des données de recherche

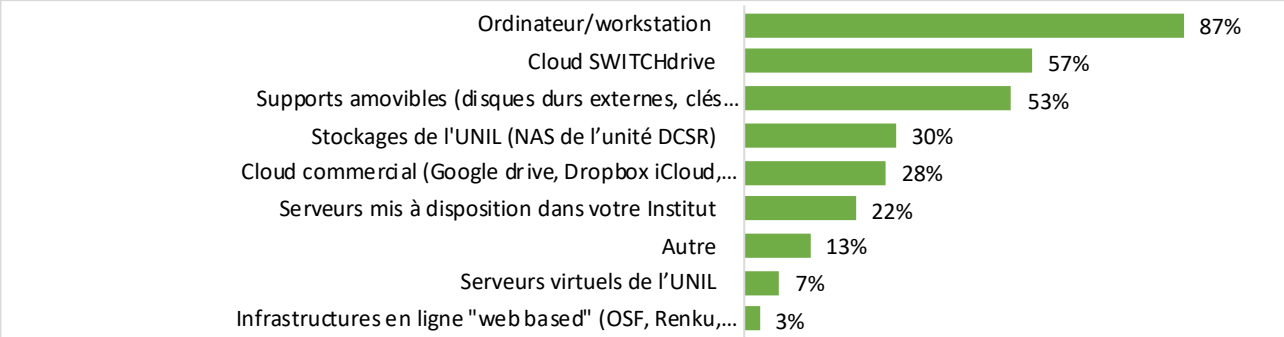
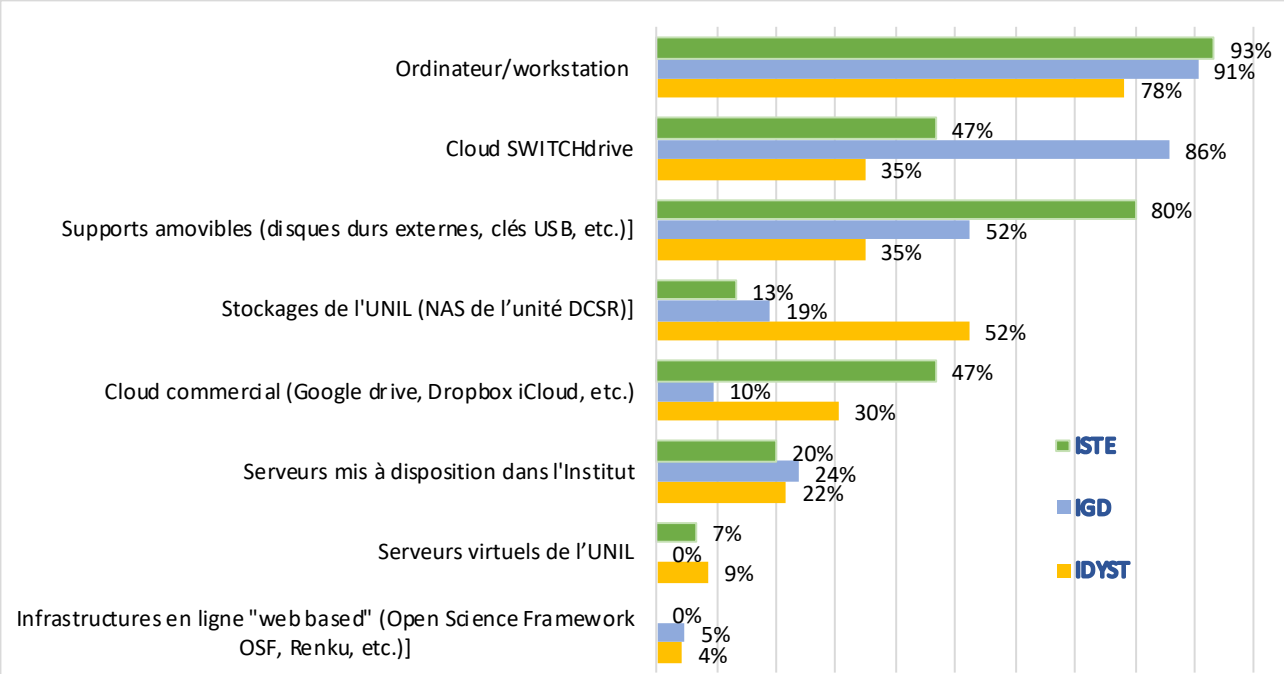


Figure 5 - Répartition solutions de stockage par instituts de rattachement



Remarque de répondant·e : ***La création de "data repositories" avec accès à l'utilisation pour chaque groupe de recherche afin d'éviter la duplication et la redondance serait utile.***
Réponse du GT : C'est tout à fait comme cela que le NAS fonctionne ! Plus d'information sur https://unil.ch/ci/home/menuinst/catalogue-de-services/recherche/stockage-de-donnees-de-recherche.html.

[5] stockage = sauvegarde courante des données de recherche pendant la phase de réalisation du projet

Remarques de répondant·es : ***Je ne connais pas les options de stockage qui existent. Suis-je obligé d'utiliser l'une d'entre elles en tant que membre du personnel de l'UNIL ? Il faudrait informer les personnes au moment de leur arrivée.***
Réponse du GT : Les informations sur le stockage sont sur le site UNIL open science https://www.unil.ch/openscience/home/menuinst/open-research-data/research-data-management/stockage--securite.html. Vous pouvez aussi prendre rendez-vous avec notre spécialiste données de recherche pour identifier des solutions pour votre recherche. Les chercheur·es n'ont aucune obligation de la part de l'UNIL, seulement des recommandations (Directive 4.5). Ils ou elles doivent néanmoins respecter la loi, par exemple lorsqu'ils traitent des données personnelles et sensibles qui sont soumis à la Loi fédérale sur la Protection des Données (LPD) ainsi qu'à la Loi cantonale vaudoise sur la protection des données personnelles (LPrD). Il convient de rappeler que, conformément à la directive 4.5 de l'UNIL, la sécurité des données de recherche doit être garantie et contrôlée sous la responsabilité du ou de la chercheur·e principal·e.

Remarque de répondant·e : ***Quelques adaptations mineures pourraient améliorer l'utilisation du cluster Curnagl de la DCSR. Il y a des limites assez strictes à l'utilisation de la mémoire des shells interactifs, et bien que je sois conscient que sur demande plus de mémoire peut être allouée, je me demande si les contraintes de mémoire générales peuvent être assouplies.***
Réponse du GT : La DCSR impose des règles parfois strictes sur le cluster mais ces restrictions visent un partage équitable et efficient des ressources entre les utilisateur·rices. C'est aussi la raison pour laquelle les utilisateur·rices sont très rarement sur file d'attente pour les ressources interactives. Notez que l'UNIL fait déjà une exception en permettant un usage en interactif (ce qui est proscrit sur la plupart des clusters, car les ressources attendent alors l'utilisateur·rice et l'usage devient extrêmement inefficient). En production, il faut donc créer un script SLURM pour faire tourner des travaux en arrière-plan, et la DCSR vous accompagne avec plaisir dans la tâche. Dans de très rares cas, l'utilisation des ressources en interactif plutôt que via des scripts peut être justifiée, et des exceptions peuvent être faites. N'hésitez pas à contacter les ingémieur·es recherche : research-computing-fgse@unil.ch.

5.3. Solutions de partage des données lors du travail collaboratif

Le SWITCHdrive reste l'un des espaces les plus utilisés lors d'un besoin d'espace virtuel pour partager et travailler sur des données communes. L'utilisation de nuages commerciaux, principalement Dropbox et Google Drive, pour le partage des données reste très répandue.

Solutions de partage des données	IDYST (N=23)	IGD (N=22)	ISTE (N=15)	Total (N=60)
Cloud SWITCHdrive	9	18	9	60% (36)
Service de transfert de fichier (Wetransfer, SwissTransfer, etc.)	12	11	8	52% (31)
Cloud commercial (Google drive, Dropbox iCloud, etc.)	8	6	6	33% (20)
Stockages de l'UNIL (NAS)	10	1	2	22% (13)
Serveurs mis à disposition dans votre Institut	4	3	2	15% (9)
Serveurs virtuels de l'UNIL	2	0	1	5% (3)
Infrastructures en ligne "web based" (Open Science Framework OSF, Renku, etc.)	1	0	0	2% (1)
Je n'ai pas besoin de partager mes données	2	2	0	7% (4)

5.4. Gestion de données personnelles[6] et sensibles[7]

L'enquête montre que 35 % (N=21) des sondé·es collectent des données personnelles et 20 % (N=12) des interrogé·es collectent des données sensibles. À l'IGD, ce chiffre monte à 68 % et 45 % (Fig. ci-dessous).
Parmi tous les sondé·es ayant collecté des données à caractère personnel (N=21), seulement 2 personnes (9,5 %) utilisent le NAS pour le stockage. Pour ce qui concerne les chercheur·es ayant des données sensibles (N=12), 4 personnes recourent au NAS pour le stockage.

Figure 6 - Collecte des données personnelles par institut de rattachement

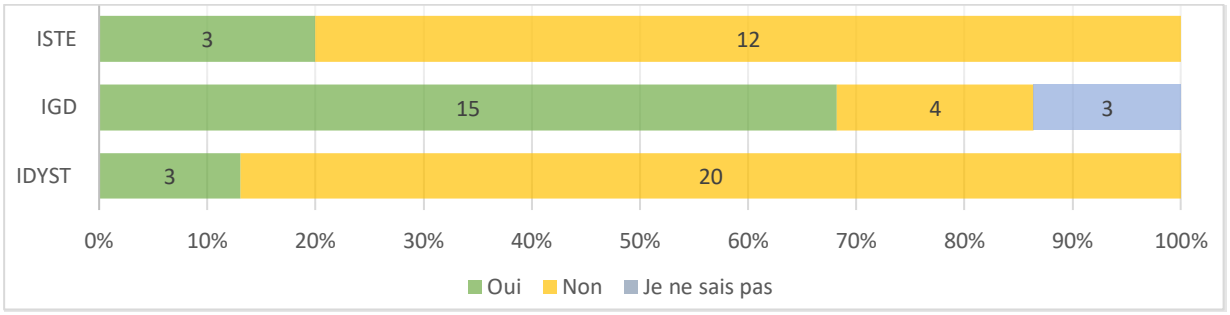
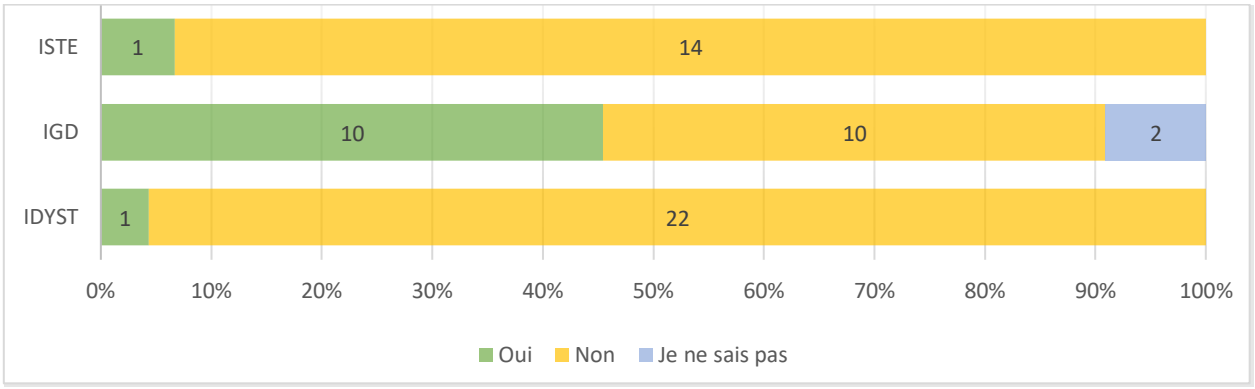


Figure 7 - Collecte des données sensibles par institut de rattachement



Les outils d'anonymisation et de chiffrement des données sont peu utilisés. Les sondé·es leur préfèrent une suppression ou chiffrement manuel des données sensibles ou un contrôle d'accès par contrats de confidentialité. Une proportion importante (10/23) ne sait pas quelle solution utiliser.

[6] C'est-à-dire contenant des informations qui se rapportent à une personne identifiée ou identifiable.
[7] C'est-à-dire des données se rapportant à : des opinions ou activités religieuses, philosophiques, politiques ou syndicales, une origine ethnique, la sphère intime de la personne, en particulier à son état psychique, mental ou physique, des mesures et aides individuelles découlant des législations sociales des poursuites ou sanctions pénales et administratives.

Tableau 1 - Solutions utilisées pour la protection des données personnelles ou sensibles parmi les chercheur·es qui ont ce type des données

Solutions et mesures de protection	IDYST (N=3)	IGD (N=17)	ISTE (N=3)	Total (N=23)
Outils d'anonymisation (*Amnesia, ARX Data Anonymization Tool, Deface*, etc.)	0	1	0	4% (1)
Logiciels pour le chiffrement des données (par ex. *Boxcryptor*)	0	0	0	0
Suppression/chiffrement manuel des données	0	9	0	39% (9)
Logiciels pour la destruction de données	0	0	0	0
Contrôle d'accès aux données (contrats de confidentialité)	1	4	0	22% (5)
Je ne sais pas	1	6	3	43% (10)

Remarques de répondant·es : ***Des vadémécum et procédure-type seraient utiles pour nous aider à gérer les données sensibles et pour capitaliser les connaissances.***
Questions de répondant·es : ***Comment anonymiser sans perdre les informations utiles et conserver une qualité de données suffisante ? Comment gérer ces données dans des recherches qualitatives où l'anonymisation n'est pas possible ou pas désirable pour des raisons scientifiques ?***
Réponse du GT : Actuellement UNIRIS travaille sur la mise à jour du site Open Research data UNIL sur la gestion de données, y compris des données sensibles et personnelles. UNIRIS prévoit également de produire une brochure d'information (numérique) pour la gestion de données (dont données sensibles et personnelles). Un cycle de formation sur la gestion de données de recherche au niveau UNIL, y compris sur les données sensibles et personnelles, sera proposé durant l'année académique 2023-2024 https://unil.ch/openscience/home/menuinst/formations.html. Si intérêt avéré par de petits groupes, contactez la consultante recherche pour l'organisation d'ateliers sur les thématiques souhaitées.
Vous pouvez également trouver des informations sur la gestion des données sensibles et personnelles sur le site https://www.unil.ch/openscience/home/menuinst/open-research-data/conformite--exigences/donnees-personnelles--sensibles.html de l'UNIL. N'hésitez pas non plus à demander un soutien pour la gestion des données sensibles et personnelles une à titre individuel. Pour ce faire, contactez : researchdata@unil.ch.

Remarque de répondant·e : ***Où stocker les données sensibles à la fin de son contrat avec l'UNIL ?***
Réponse du GT : De manière générale, lorsque qu'un PI quitte l'UNIL, il doit transmettre, dans des formats répondant aux standards internationaux, au décanat de sa faculté les données de recherche de son groupe de recherche qu'il désire ou doit conserver, ainsi que toutes les informations nécessaires à leur compréhension et à leur gestion (Directive 4.5. Article 19). Concernant les doctorant·es, il est sous-entendu que leurs directrices ou directeurs de thèse sont considéré·es comme PI. Par conséquent, lorsque un·e doctorant·e quitte l'UNIL, les données sont transmises à son PI. Concernant les postdocs, leurs statut de PI est définit au cas par cas.
En ce qui concerne l'archivage des données sensibles, actuellement la DCSR propose une solution d'archivage des données de recherche sur les bandes magnétiques (LTS, Long Term Storage) : https://unil.ch/openscience/home/menuinst/open-research-data/research-data-management/stockage--securite.html.

5.5. Solution d'archivage[8] des données numériques de recherche

La moitié des sondé·es semble suivre les bonnes pratiques en archivant des données (ou en prévoyant de le faire) sur les infrastructures NAS et LTS[9] de l'UNIL ou sur un dépôt. En même temps les résultats de l'enquête montrent que la majorité des personnes interrogées (73,4 %) conservent leurs données de recherche sur un ordinateur ou un disque dur une fois le projet de recherche terminé (73,4 % des personnes interrogées le font ou prévoient de le faire ; Tableau ci-dessous). Une analyse plus approfondie révèle que la moitié de ces répondant·es (37 %) ne prévoient pas d'utiliser d'autres solutions d'archivage et conservent

[8] Archivage = conservation à long terme des données numériques, une fois le projet terminé
[9] LTS (Long Term Storage) ou l'hébergement des données de recherche pour du stockage à long terme. IL s'agit de l'archivage de données de recherche sur les bandes magnétiques à la DCSR, sur le site de l'UNIL.

leurs données (ou envisagent de le faire) uniquement sur leurs ordinateurs et/ou disques durs.
Pour ce qui concerne de l'archivage des données par le biais de divers dépôts tels SWISSUbase ou Zenodo (conformément les principes de la Science Ouverte), seulement un tiers de répondant·es (32 %) le font ou envisagent de le faire. Une analyse des réponses selon la catégorie/statut a montré que ce sont surtout de doctorant·es qui n'envisagent pas cette possibilité. Seulement 12 % des doctorant·es pensent archiver les données dans un dépôt contre 47 % des professeur·es/MER.

Tableau 2 - Solutions d'archivage par institut de rattachement

Solutions d'archivage citées	**IDYST (N=23)**	**IGD (N=22)**	**ISTE (N=15)**	**Total (N=60)**
Dépôt (Zenodo, Dryad, DaSCH, SWISSUbase, FORSBase etc.)	8	6	5	32% (19)
Ordinateur perso/disque dur	13	18	12	72% (43)
Ordinateur perso/disque dur sans autre solution	*5*	*11*	*6*	*37% (22)*
Serveur de stockages de l'UNIL (NAS de l'unité DCSR, Long Term Storage UNIL)	15	3	3	35% (21)
Serveurs d'Institut	6	2	2	17% (10)
Je n'envisage pas l'archivage de données	0	2	0	3% (2)
Je ne sais pas	1	3	0	7% (4)
Autre (« External harddrive, data included in publications (article + suppl. mat.), transmission des données à mes supérieurs hiérarchiques, computer backed to Dropbox, OneDrive, Comet backup, MinPet, IgPET »)	2	4	3	15% (9)

Plusieurs remarques de répondant·e concernent : ***le besoin d'aide à l'archivage des données (notamment sur bande) et la difficulté de trouver les informations sur le site de l'UNIL.***
Réponse du GT : Il existe la possibilité d'archiver des données de recherche sur des bandes magnétiques (autrement Long Term Storage, LTS). Les demandes de soutien pour l'archivage (sur bande ou autre) sont à adresser à : research-computing-fgse@unil.ch ou en ligne : https://wiki.unil.ch/ci/books/research-data-storage/page/general-information.
Vous pouvez trouver l'information sur les solutions de l'archivage de données de recherche sur les pages Open research data de l'UNIL https://unil.ch/openscience/home/menuinst/open-research-data/research-data-management/stockage--securite.html.

Plusieurs remarques de répondant·e ***concernent la formation et le besoin d'informations sur les dépôts de données, notamment SWISSUbase. Un·e répondant·e note par ailleurs: « il faudrait que les professurs titulaires soient informés de cela en particulier étant donné que le personnel non titulaire ne reste que quelques années à l'Unil »*****.**
Réponse du GT : Un atelier sur l'archivage et le partage des données de recherche via les dépôts, notamment le SWISSUbase) sera proposé dans le cadre du cycle d'ateliers sur la gestion de données de recherche durant l'année académique 2023-2024 https://unil.ch/openscience/home/menuinst/formations.html.

En réponse à vos commentaires, une brochure d'information (numérique) sur la gestion des données de recherche et le soutien à la recherche est prévu prochainement. Par ailleurs, l'équipe UNIRIS/DCSR participe aux journées d'accueil des professeur·es et MER, au cours desquelles des informations sur la gestion des données de recherche sont fournies.

Vous pouvez également demander une consultation à titre individuel pour le dépôt de données researchdata@unil.ch ou demander l'organisation d'atelier en petits groupes à votre consultante recherche.

En outre, à l'automne 2023, la FGSE participera au projet pilote SwissUbase afin de tester la cohérence de la plateforme pour le dépôt des données prévenantes des disciplines liées à cette faculté et d'identifier les principaux points d'amélioration. Si vous souhaitez participer à ce projet pilote, merci à contacter la spécialiste données de recherche : researchdata@unil.ch.

Gestion des données : Résumé et points clés

Certaines questions et remarques des participant·es à l'enquête témoignent d'une méconnaissance des outils et du soutien existants à l'UNIL. Le site FGSE « Vous-êtes chercheur·e[10] » fournit un résumé de ces ressources et des personnes de contact.

En particulier, les réponses laissent penser qu'une partie des sondé·es ignore l'existence des infrastructures de l'UNIL destinées au stockage et l'archivage de données de recherche (le NAS de la DCSR). Il serait utile de rappeler que les ordinateurs et disques durs sont des solutions insécures et sujettes à des défaillances matérielles. L'UNIL encourage les chercheur·es à sauvegarder les données de recherche sur les infrastructures institutionnelles (NAS de la DCSR) (Directive 4.5, art. 8).
L'usage des clouds commerciaux (DropBox, Google Drive, etc.) pour stocker des données de recherche est aussi un point inquiétant. En effet, l'UNIL autorise l'utilisation de services de stockage tiers à condition que les données soient stockées en Suisse (Directive 6.9, art. 5). Si les données sont stockées à l'étranger, les données doivent être chiffrées et la clé de chiffrement doit être stockée en Suisse. Malheureusement, la plupart des services connus ne satisfont pas ces critères et toute utilisation de ces services se fait sous l'unique responsabilité de l'utilisateur·trice qui en assume pleinement les risques[11].
Pour le partage et le stockage, beaucoup de chercheur·es font appel à SWITCHdrive. Un passage à OneDrive devra se faire (pour les données de recherche non sensibles), étant donné l'arrêt du service de SWITCH prévu en 2024. Le Filetransfer[12] de la DCSR est une alternative à l'envoi de fichier volumineux vers et depuis des collaborateur·rices externes.

L'enquête met en évidence l'insuffisance des mesures adoptées pour la gestion des données personnelles et sensibles. Une partie de répondant·es ne sait pas quelles solutions à adopter tandis que les mesures prises par d'autres·es ne semblent pas suffisantes pour assurer la sécurité et la confidentialité de données. Pour exemple, le cryptage des données à l'aide de Boxcryptor (outil proposé par la DCSR) n'a été demandé par aucune·e chercheur·es. Un travail de sensibilisation est donc à mener sur la gestion des données personnelles et sensibles.

Si la moitié des répondant·es semble suivre les pratiques recommandées par l'UNIL pour archiver les données, plus d'un tiers ne prévoit pas d'utiliser d'autres solutions que leurs ordinateurs ou disques durs. Parmi les sondé·es, seul un tiers partage les données sur un dépôt, conformément aux principes de la Science Ouverte. Les doctorant·es en particulier sont encore très peu sensibilisé·es à cette pratique. Des efforts de promotion des nouvelles pratiques d'archivage et de Science Ouverte sont donc nécessaires.

[10] https://unil.ch/gse/home/menuguid/chercheureuses.html
[11] Voir le FAQ « Puis-je utiliser des services dans le Cloud (Dropbox, Google Drive, iCloud, etc.) pour traiter et stocker mes données ? » https://www.unil.ch/openscience/home/menuinst/open-research-data/faq.html
[12] https://wiki.unil.ch/ci/books/high-performance-computing-hpc/page/filetransfer-from-the-cluster

6. Valorisation du profil de recherche sur internet

Une partie importante des sondé·es possède un profil de recherche sur internet.
Les pages institutionnelles de l'UNIL vont faire peau neuve dans les années à venir (projet IRIS). Il sera sans doute plus attractif pour les chercheur·es qui ne sont que 52 % à présent à l'utiliser pour valoriser leur profil.

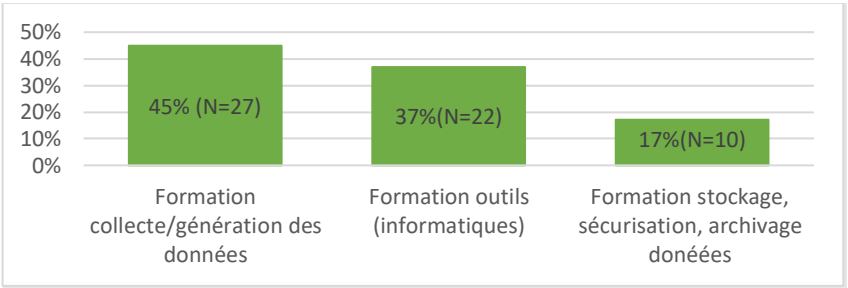
Tableau 3 - Utilisez-vous des pages/plateformes internet pour valoriser vos recherches ?

Valorisation internet	IDYST (N=23)	IGD (N=22)	ISTE (N=15)	Total (N=60)	Exemples
oui	78% (18)	91% (20)	47% (7)	75% (45)	
Profil institutionnel UNIL	9	18	4	52% (31)	Unisciences, page IGD, Serval
Profil hors UNIL	15	16	4	58% (35)	Academia, ORCID, Researchgate, HALSHS
Site ou blog de recherche personnel	11	5	4	33% (20)	Github, wordpress, group page, RPubs
Publications sur les réseaux sociaux	3	9	2	23% (14)	Twitter, LinkedIn, Facebook, groupe Whatsapp

7. Formations suivies et demandées

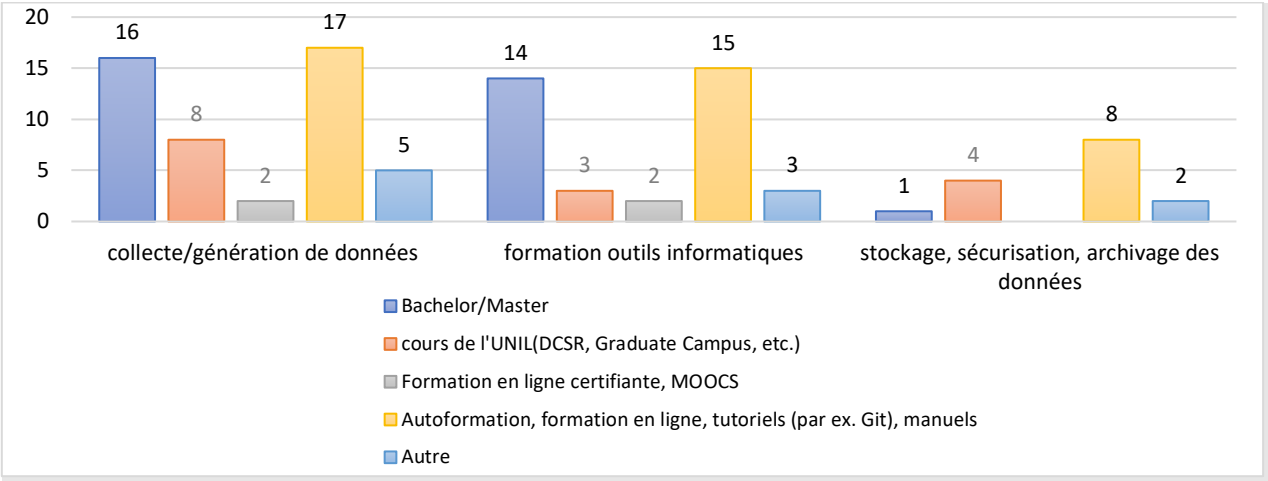
Les résultats de cette section montrent que très peu de répondant·es (17 %) ont suivi une formation à la gestion de données de recherche (GDR), contrairement à deux autres domaines tels que la formation à la collecte/génération des données (45 %) et aux outils informatiques (37 %) (Fig. ci-dessous).

Figure 8 - Formations suivies (%)



Une analyse plus détaillée montre que les chercheur·es développent le plus souvent leurs compétences professionnelles à travers l'auto-apprentissage via l'internet, les manuels, etc. (figure ci-dessous). On constate également que si la formation à la collecte et aux outils fait partie du cursus de premier cycle (Bacchelor/Master) ce n'est pas le cas de la formation en gestion de données de recherche.

Figure 9 - Type de formations suivies par institut de rattachement (N de personnes)



Points clés

Comme le montrent les sections précédentes, des répondant·es ont signalé leur besoin de formation sur divers outils et en gestion de données de recherche. Certaines remarques témoignent d'une méconnaissance de l'offre de formation disponible à l'université. Le GT invite des chercheur·es intéressé·es à s'adresser plus largement aux personnes de soutien. Elles sont disponibles pour des soutiens individuels ou pour organiser des ateliers à un petit groupe de personnes intéressées.

8. Synthèse et recommandations

Besoins et difficultés exprimés, points de préoccupation identifiés par le GT	Mesures à déployer d'après le GT
Méconnaissance du soutien individuel existant à la FGSE. En effet, certaines questions soulevées par les répondant·es pourraient être résolues facilement par une entrevue individuelle.	**Rappeler l'existence des ressources à disposition des chercheur·es**, qui se sont étendues ces dernières années (voir ci-dessous, les personnes travaillant au soutien « données de la recherche » à la FGSE).
Difficulté à trouver les informations sur les bonnes pratiques.	**Rappeler les sites d'information existants** : Vous-êtes-chercheur·es présente les ressources et personnes clés du soutien à la recherche. Le personnel de soutien est aussi disposé à pallier le manque d'informations sur les diverses questions liées à la gestion des données, en **organisant de courtes séances d'information**, dans les conseils d'institut, les groupes, les instituts, sur demande !
Demande de formations sur certains outils et logiciels. Un besoin de cours d'initiation à certains langages informatiques ou bien même à la programmation informatique est notamment apparu.	**Promouvoir les formations** (DCSR, CI, FGSE) **existantes**. La DCSR et le personnel de soutien sont aussi disposés à **mettre en place de nouvelles formations** ou des ateliers ponctuels à la demande des équipes de recherche.
Besoins en calcul scientifique. La moitié des sondé·es se heurte aux limites de calcul sur leurs machines, sans pour autant faire appel à aux solutions de HPC (calcul haute performance).	**Promouvoir les ressources de calcul haute performance** (cluster/machines virtuelles) de l'UNIL.
Méconnaissance des mesures de sécurité pour la collecte et de la gestion des données personnelles et sensibles. Les clouds commerciaux sont encore largement utilisés tandis qu'un faible pourcentage de chercheurs utilise les outils institutionnels sécurisés.	**Rappeler l'existence des infrastructures institutionnelles de stockage et d'archivage de données de recherche** et **le soutien de la spécialiste données de recherche**, qui peut aguiller les chercheur·es sur les outils lors d'entretiens individuels. Sur demande, le personnel de soutien est aussi disposé à mettre en place des **séances d'information** ou des ateliers, dans les instituts et les groupes.
Faible adoption des pratiques de partage de données via des dépôts de données ouvertes. Des demandes et des commentaires de chercheur-e-s concernant la formation et les informations sur ces dépôts (entre autres SWISSUbase) suggèrent qu'il y a cependant un intérêt pour les pratiques de la Science Ouverte.	**Promouvoir la stratégie de l'Open Research Data et le partage** des données de recherche. Le déploiement de SWISSUbase et son adaptation prévue aux géosciences faciliteront aussi l'adoption de ces pratiques.

La gestion des données et sa mutation à l'ère numérique et avec les politiques de la Science Ouverte (Open Science) requièrent une adaptation des pratiques et une formation permanente de la part des chercheur·es.
Le personnel responsable du soutien aux chercheur·es est à disposition pour faciliter cette adaptation. À la FGSE, plusieurs personnes sont répondantes sur la question des données de la recherche. Elles communiquent entre elles régulièrement, et avec les services centraux ; donc sonner à la « mauvaise porte » n'est pas un problème. Cela dit, elles se répartissent le travail comme suit :

- Amélie Dreiss (contact : amelie.dreiss@unil.ch) : questions éthiques, de financement ou en lien avec une postulation,
- Zhargalma Dandarova (contact : researchdata@unil.ch) : questions relatives à la gestion des données,
- Flavio Calvo et Margot Sirdey (contact : research-computing-fgse@unil.ch) : questions relatives à la programmation informatique.