

CERTIFICAT EN GESTION DE DOCUMENTATION ET DE BIBLIOTHÈQUE
2016 – 2017

One size fits none

Quels services pour la gestion des données de la recherche
à la bibliothèque de l'IHEID ?

Guillaume Antoine Pasquier

Yverdon-les-Bains

Automne 2017

TRAVAIL FINAL DE CERTIFICAT

Déposé auprès de

Michèle Furer-Benedetti (Division de l'information scientifique, Université de Genève), responsable scientifique du module « Environnements numériques »

Jean-Blaise Claivaz (Division de l'information scientifique, Université de Genève),
personne de référence pour le suivi pédagogique du travail

Résumé

Les données de la recherche sont devenues depuis quelques années une préoccupation majeure du monde académique. Les chercheurs sont soumis à de nouvelles obligations, sans forcément avoir les connaissances ou pratiques nécessaires pour y répondre. Les bibliothèques académiques développent en conséquence des services pour les soutenir dans ces démarches.

Après une présentation du contexte et un aperçu des pratiques actuelles de quelques chercheurs de l'IHEID, ce travail étudie les services existant dans les bibliothèques d'autres institutions académiques proches, ainsi que les compétences qui devraient être développées pour proposer de tels services.

La conclusion de l'auteur du travail est qu'il faudra des services adaptés aux besoins spécifiques des chercheurs des différentes disciplines, ce qui paraît possible malgré des ressources limitées.

Remerciements

Je souhaite remercier les groupes et personnes suivantes pour leur soutien et leurs encouragements dans la rédaction de ce travail :

- mes collègues de la bibliothèque de l'IHEID, notamment son responsable Yves Corpataux, qui a cru et croit encore en moi,
- les bibliothécaires et spécialistes de l'EPFL, de la HEG, et de l'Université de Genève, mais aussi d'autres institutions (EUI, LSE), pour m'avoir accueilli et pour leurs réponses à mes questions,
- mes camarades de formation à l'Université de Fribourg et nos encadrants, car cette année fut extrêmement riche en souvenirs et en apprentissages,
- Christelle Donius pour sa relecture attentive,
- Giselle Castelo Branco, sans qui ce travail aurait été bien moins bon.

Table des matières

Résumé	2
Remerciements	2
Table des matières	3
1. Contexte	4
1.1. L'IHEID	4
1.2. Les données de la recherche	4
1.3. Exigences des organismes de financement	5
2. Les chercheurs de l'IHEID et leurs données	7
2.1. Type de données	7
2.2. Pratiques de stockage	8
2.3. Pratiques de documentation	8
2.4. Ouverture des données	9
2.5. Obstacles à la publication des données	9
2.6. Bilan des entretiens	11
3. Services pour les données de la recherche	12
3.1. Réseau	12
3.2. Information	13
3.3. Formation	14
3.4. Participation	15
4. La bibliothèque IHEID et ses ressources	17
4.1. Personnel	17
4.2. Identifier les besoins en compétences	18
4.3. Qui, combien, comment ?	19
4.4. Partenaires potentiels (internes et externes)	20
5. Propositions de services et priorités	21
5.1. Sensibilisation	21
5.2. Documentation de référence	21
5.3. Ateliers DMP	22
5.4. Programme de formation	22
5.5. Expertise et consultation	22
5.6. Le cas des services techniques	22
6. Conclusion	24
7. Bibliographie	25
Annexes	27

1. Contexte

1.1. L'IHEID

L'Institut de hautes études internationales et du développement¹ est situé à Genève. Il est issu de la fusion en 2008 de deux instituts universitaires (HEI et IUED) fondés respectivement en 1927 et 1977. Chapeauté par une fondation de droit privé, il est financé par l'État de Genève, par la Confédération, et par des revenus propres, notamment immobiliers (location de locaux dans ses bâtiments et résidences d'étudiants). Depuis 2013, il occupe une partie de la Maison de la paix, bâtiment construit à son initiative à quelques pas de la place des Nations.

C'est un « établissement de recherche et d'enseignement supérieur spécialisé dans l'étude des affaires mondiales » (« L'Institut en 2016-2017 », 2017), qui compte cinq départements académiques². L'approche de chaque sujet est basée sur une vision internationale ou de développement, et les centres de recherche sont généralement interdisciplinaires. En 2016-2017, il comptait 344 doctorants et 167 enseignants et chercheurs pour 553 étudiants en master.

La bibliothèque Kathryn et Shelby Cullom Davis

La nouvelle bibliothèque, qui occupe les deux niveaux inférieurs de la Maison de la Paix, a été financée par un don financier important de madame Kathryn Davis. Elle compte 18 employés permanents (15.65 ETP) et ses collections regroupent plus de 350'000 ouvrages. Pour des raisons historiques, ses collections numériques sont conjointes avec celles de l'Université de Genève. Elle propose par ailleurs des formations aux étudiants (ressources électroniques, Zotero, soutien à des cours disciplinaires de droit).

La bibliothèque a été évaluée au printemps 2015 par deux auditeurs externes (Vanessa Proudman et Jean-Philippe Accart). Parmi leurs recommandations figurait l'importance de développer de nouveaux services à la recherche. Le plan stratégique 2016-2019 développé par la suite a soulevé la question de la gestion des données de la recherche³. C'est dans ce cadre que s'inscrit ce travail.

1.2. Les données de la recherche

Plusieurs institutions ont défini ce qui constitue des « données de la recherche ». Parmi les plus fréquemment citées, on trouve celles de l'OCDE (Organisation de Coopération et de Développement Économiques)⁴ et de l'Office of Management and

¹ Pour plus de simplicité, j'utiliserai les termes « IHEID » ou « l'Institut ». Il dispose également d'une dénomination en anglais : Graduate Institute of International Studies and Development, qui lui donne son logo et marque « The Graduate Institute, Geneva ». Voir <http://graduateinstitute.ch>.

² Anthropologie et sociologie (ANSO), économie internationale (IE), histoire internationale (IH), droit international (LAW), science politique et relations internationales (PS). Il faut ajouter à ce total les masters interdisciplinaires (MINT) en affaires internationales et en développement.

³ La gestion des données de la recherche (GDR) est aussi nommée research data management (RDM). Ces termes ont un sens identique dans l'ensemble du travail.

⁴ « 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. » (Gurria, 2007)

Budget (OMB) des Etats-Unis⁵. La définition retenue par le Fonds national suisse de la recherche scientifique (FNS) sur son site web est issue des travaux d'institutions britanniques :

"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)." (« Concordat on Open Research Data », 2016)

On peut retenir de cette définition que les "données" ne sont pas forcément numériques ou textuelles. C'est avant tout l'usage d'un élément par un chercheur et sa valeur de preuve scientifique qui en font une donnée de la recherche. Par simplicité, on exclura toutefois (comme l'OMB le fait explicitement) les éléments physiques des considérations de ce travail.

Typologies des données de la recherche

Les données de la recherche peuvent être catégorisées de multiples manières : données quantitatives ou qualitatives, données expérimentales, de simulation, d'observation, ou compilées, données primaires ou secondaires, etc. Une typologie complète de ces données échappe à ce travail. D'autres travaux les ont décrites et classées de manière détaillée, et nous nous référerons à eux si ces éléments s'avèrent nécessaires.

Cycle de vie des données de la recherche

Le traitement des données doit être anticipé pour tout leur cycle de vie (research data lifecycle). Cette visualisation permet la prise en compte de besoins différents des chercheurs à différentes étapes de leur processus. Un soutien spécifique doit donc être apporté en fonction de ceux-ci. La modélisation la plus fréquemment citée de ce cycle est celle de l'UK Data Archive (« Research data lifecycle », s. d.), reprise notamment par les hautes écoles suisses. A nouveau, ce cycle est présenté dans de multiples travaux et n'est pas le sujet de ce travail.

1.3. Exigences des organismes de financement

Les deux sources principales de financement de projets de recherche à l'IHEID sont le Fonds national suisse de la recherche scientifique (FNS) et le Conseil européen de la recherche (European Research Council, ERC). Ces deux institutions appliquent une politique favorisant l'open access. Les articles issus de recherches qu'elles financent doivent notamment être publiés sous ce modèle.

Le FNS exige en plus depuis le 1^{er} octobre 2017 (« Open Research Data : les requêtes devront inclure un plan de gestion des données », 2017) que toute demande de financement soit accompagnée d'un plan de gestion des données (voir plus bas) et prévoie la publication des données en open access sur un dépôt répondant à la norme FAIR (findable, accessible, interoperable, re-usable) (Wilkinson et al., 2016).

Du côté de l'ERC, les exigences similaires concernent pour l'instant des projets pilotes du programme Horizon 2020. Celles-ci s'étendront ensuite à l'ensemble des projets soumis. (« Open Access », 2017)

⁵ *"Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples). [...]"* (OMB, 1999)

Plan de gestion des données

La gestion des données de la recherche débute par une phase de planification. Celle-ci s'incarne notamment dans un document nommé *data management plan* (DMP). Il anticipe les questions et problèmes qui se poseront dans la démarche de traitement et de curation des données du projet, et permet de documenter les solutions envisagées pour au final rendre (ou pas) les données ouvertes et accessibles.

L'objectif habituel est de renseigner l'agence de financement sollicitée quant aux mesures prévues par les chercheurs pour garantir un bon traitement de leurs données. Il permet bien sûr au passage de sensibiliser les chercheurs aux bonnes pratiques de gestion des données de la recherche, notamment en matière de conservation, et d'assurer qu'elles seront appliquées aux projets financés.

Autres arguments

Au-delà de l'aspect économique et politique de l'open access, un objectif important de la bonne gestion des données de la recherche est de répondre à l'exigence scientifique de vérifiabilité et de reproductibilité des résultats. Le bon traitement des données et leur divulgation publique doit ainsi améliorer la transparence du processus de recherche. Enfin, l'ouverture des données doit permettre à d'autres chercheurs de répondre à des questions nouvelles, et d'accélérer l'avancée de la recherche. (Borgman, 2012)

2. Les chercheurs de l'IHEID et leurs données

Pour avoir un aperçu de la familiarité des chercheurs avec le sujet de la gestion des données, des entretiens ont été menés avec cinq d'entre eux (un par département académique). Les chercheurs m'ont été suggérés par les gestionnaires des différents départements, en fonction notamment de leur disponibilité et de leur intérêt.

Les entretiens ont été tenus sous la forme d'un entretien semi-structuré. Un questionnaire disponible en annexe a servi de point d'appui pour orienter la discussion par des relances, mais les questions n'ont pas été posées systématiquement.

2.1. Type de données

Les données récoltées varient énormément en fonction des interlocuteurs et des domaines. Dans plusieurs cas, la question de ce qu'étaient exactement les « données de la recherche » posait déjà problème. Une partie de nos échanges a permis de trouver des moyens de clarifier la question.

Lors du webinaire IASSIST « Qualitative Data Management: Lessons from QDR » suivi pendant la finalisation de ce travail, les intervenants ont insisté sur la nécessité d'adapter notre vocabulaire et d'éviter le jargon : parler de « sources primaires » peut ainsi être plus parlant pour des chercheurs en histoire.

Pour le chercheur en anthropologie (département ANSO), ces données sont avant tout des enregistrements d'entretiens (WAV, MP3), et des carnets de notes manuscrites prises sur le terrain. Les carnets n'ont pas été numérisés et sont conservés dans les dossiers du chercheur. Ces données seraient donc qualitatives, issues de l'observation. Logiciels utilisés : Atlas, Word, Excel, NVivo.

Pour la chercheuse en histoire internationales (département IH), les données récoltées sont avant tout des documents d'archives analogiques : photographies, correspondance, rapports, etc. La chercheuse les a photographiés et a conservé les copies numériques, établissant un registre de chaque document (et chaque image individuellement) dans Endnote selon une méthodologie développée par ses soins. Ses données sont donc avant tout qualitatives. Elle a également créé des tableaux Excel pour une forme d'analyse quantitative de la correspondance étudiée (qui écrit à qui, à quelle fréquence, etc.), mais ce n'est pas le cœur de son travail. Logiciels utilisés : Word, Excel, Endnote.

Le projet de la chercheuse en science politique (département PS) consiste en du text-mining sur des discours parlementaires. Des analyses sémantiques puis statistiques en sont tirées. Logiciels utilisés : Word, Excel, NVivo, Rstudio.

Les chercheurs en droit (département LAW) rencontrés sortent du profil habituel du domaine. Ils développent une recherche empirique s'appuyant sur des données qualitatives récoltées selon plusieurs méthodes, dans une approche « digital humanities » inhabituelle dans le domaine du droit international. Ils ne sont donc pas représentatifs de leurs collègues. Selon eux, les autres juristes ont une approche purement théorique, non-empirique, ne générant pas de « données ». On peut imaginer toutefois des enregistrements d'entretiens, par exemple. Cette recherche devra être approfondie au cours des prochains mois, tant la question de ce domaine semble absente des références trouvées sur la gestion des données de la recherche. Logiciels utilisés : Word, Excel, « base de données » ad hoc, sans informations supplémentaires.

La chercheuse en économie (département IE) utilise, sans surprise, des données quantitatives. Ce sont des relevés douaniers détaillés au niveau de chaque transaction, soit une quantité massive de données (plusieurs téraoctets). C'est la seule à utiliser une base de cette sorte, et les capacités techniques offertes par l'institut pour leur traitement sont relativement insuffisantes. Logiciels utilisés : notes LaTeX, Matlab, développement Python, Stata.

2.2. Pratiques de stockage

Les chercheurs ont paru comprendre les enjeux posés à ce niveau. Les risques de perte d'information étaient généralement pris en compte, et la question de la sécurité était également évoquée. La facilité d'usage et de collaboration reste toutefois la première préoccupation des chercheurs interrogés.

Dropbox est très utilisé par les chercheurs, dont plusieurs utilisent une option payante pour stocker leurs données et collaborer avec leurs partenaires. C'est un cas très répandu dans le monde académique, mais il reste relativement étonnant dans le cas de l'IHEID, car l'Institut fournit avec ses adresses e-mail un stockage illimité sur Google Drive, dont les fonctionnalités sont globalement similaires à celles de Dropbox.

Les disques réseau de l'Institut (Share Drive ou S:) sont mis en place pour chaque projet par le service informatique, avec une gestion des accès par le bureau de la recherche pour les données sensibles, mais ils sont très peu utilisés. Leur accès à distance n'est pas jugé pratique, même s'ils fonctionnent normalement avec un VPN. Ils seraient peut-être davantage utilisés pour des projets collaboratifs utilisant des données sensibles, mais je n'ai pas été confronté à cette situation : les chercheurs interrogés travaillaient majoritairement soit seuls, soit avec des données peu sensibles. Dans le cas du projet de recherche en science politique, il est exclusivement utilisé pour conserver les PDF originaux des discours étudiés.

Les chercheurs conservent peu de données sur leurs disques personnels (ordinateur portable). C'est le cas de la chercheuse en histoire, mais elle effectue aussi des backups réguliers (toutes les deux semaines) sur deux disques durs externes dans deux lieux différents (travail et domicile) afin de limiter les risques de pertes de données.

La chercheuse en économie travaillant sur des données massives et nécessitant une protection importante, elle, utilise un coffre-fort numérique, disque externe sécurisé en deux exemplaires.

Enfin, le papier reste apprécié par le chercheur ANSO pour la conservation de son matériel d'observation, bien que son travail ultérieur soit numérique. Les documents numériques sont conservés sur son espace disque professionnel sur le réseau de l'Institut. J'ignore si des backups en sont réalisés.

La base de données créée pour l'un des projets de droit est hébergée dans le cloud chez Amazon. Je n'ai pas pu obtenir d'information concernant les éventuels backups, le projet étant majoritairement géré par un partenaire.

2.3. Pratiques de documentation

C'est à ce niveau que les problèmes et lacunes semblent être les plus nombreux.

La seule documentation détaillée a été réalisée pour le projet d'économie. La documentation du projet et des données de la recherche est réalisée en LaTeX par des étudiants en master ou PhD collaborant au projet.

Des guidelines et procédures de projet ont été rédigées à l'usage des doctorants intervenant aux côtés de la chercheuse en science politique. Les versions intermédiaires des transformations pour l'analyse sémantique sont conservées mais aucune documentation spécifique n'est réalisée, et les métadonnées ne sont pas gérées.

Le chercheur en anthropologie crée des inventaires et tables des matières de ses carnets sous Excel, tandis que la chercheuse en histoire référence, comme dit plus haut, chaque document sous Endnote, ce qui constitue une forme de métadonnées. Elle n'a pas davantage documenté son travail mais enseigne en revanche ses méthodes spécifiques à d'autres chercheurs et étudiants.

Aucun travail de documentation n'a été signalé par les chercheurs en droit, mais il est vraisemblable qu'il soit assumé par leur collaborateur externe.

2.4. Ouverture des données

L'idée de documenter et partager les données issues de leur recherche était généralement vue d'un bon œil, avec quelques bémols.

Dans deux cas (LAW & PS), les données recueillies et transformées avaient déjà vocation à devenir publiques à travers des sites web interactifs. Pour la chercheuse en PS, l'ajout de celles-ci dans un dépôt académique a été bien accueilli. La pertinence est moins certaine pour les chercheurs en droit (voir plus bas), mais le principe en était reconnu.

Le chercheur en anthropologie s'est en revanche interrogé longuement sur la pertinence des exigences du FNS pour son champ d'études. Nous verrons plus loin les obstacles perçus.

2.5. Obstacles à la publication des données

Un sondage de l'éditeur Wiley (Ferguson, 2014) a recensé en 2017 sept points fréquemment relevés comme obstacles à la publication des données, dont certains se sont retrouvés dans les témoignages recueillis auprès des chercheurs.

Copyright et propriété intellectuelle, exigence de confidentialité

Dans le cas de la chercheuse en économie, c'est la source des données qui empêche leur publication. Les données douanières détaillées sont confidentielles et empêchent toute divulgation, même sous forme agrégée et anonymisée. Quand un journal exige avant publication les données sur lesquelles l'article est basé, la pratique des économistes utilisant des données couvertes par ce genre de restriction est de proposer explicitement des données simulées permettant de tester les modèles développés et les hypothèses formulées.

Aucun obstacle n'a été relevé par la chercheuse en science politique : les données sont basées sur des informations dans le domaine public, ou nécessitant simplement l'ajout d'une mention légale sur l'origine des données. Lors de notre entretien, un inventaire était toutefois en cours pour s'assurer qu'aucun élément n'était couvert par d'autres droits.

Le premier projet des chercheurs en droit mine des données publiques de l'Organisation mondiale du commerce (OMC), dont la licence autorise l'exploitation. La redistribution de celles-ci ne pose selon les chercheurs pas de problème particulier. La base de données qui en est issue sera utilisable par le public et mise à jour en

permanence, et la pertinence d'un dépôt de données pour son contenu est donc discutable, mais le code pourrait entrer dans la catégorie des éléments à préserver et diffuser. L'aspect technique étant géré par un ancien assistant, aujourd'hui à Ottawa, les responsabilités, les droits et autres aspects n'étaient pas clairement définis.

Pas d'exigence de la part de l'institution, du financeur, ou du journal

C'est le cas de la plupart des chercheurs : si certains (PS et LAW) permettent l'accès à leurs données via des sites web, le faire dans des conditions FAIR ne se fera que parce que les institutions le demande. Ce n'est généralement pas par refus, mais parce que la question ne s'est jamais posée.

Crainte de se faire « doubler » par un autre chercheur

Cette crainte n'a pas été relevée dans le cadre de cette recherche. La publication des données est généralement envisagée en fin de projet, pas au cours de celui-ci.

Possibilité de mauvais usage ou interprétation par d'autres utilisateurs, notamment non-académiques

Cette crainte n'a pas été relevée dans le cadre de cette recherche.

Ethique et protection des individus concernés par les données

Dans le cas de la chercheuse en histoire, les sources primaires sont à priori sensibles : les personnes concernées sont encore vivantes et le sujet d'étude (planning familial et contraception) est encore sujet à polémique. Elle soupçonne par ailleurs avoir pu accéder à des documents qui n'auraient pas dû être disponibles, en raison du faible niveau de compétence du personnel des archives consultées (dans les Caraïbes). La légalité de son accès à certains éléments n'est pas certaine, et elle souhaite donc faire preuve de prudence dans le partage des données. Les données agrégées qu'elle a générées pourraient en revanche être partagées sans problème notable après anonymisation des personnes concernées.

Dans le deuxième projet des chercheurs en droit, des données concernant les profils, formation et carrières de juges d'arbitration internationaux ont été récoltées sur des sites web publics, et ils ne voyaient pas d'obstacle particulier à leur partage. J'ai exprimé mon interrogation relative au fait que ces données comportaient des éléments personnels, bien que publics. En France, ces éléments pourraient-ils nécessiter l'inscription auprès de la CNIL ?

Pour le chercheur en anthropologie, c'est le problème majeur : le consentement éclairé des sujets de recherche lui permet de traiter les éléments récoltés, mais les données brutes restent trop sensibles dans de nombreux cas et leur disponibilité pourrait mettre des participants, par exemple afghans, en danger. L'anonymisation des données devient rapidement difficile et fait perdre leur sens aux observations qualitatives, qui nécessitent une compréhension du contexte.

Crainte de ne pas être cité comme origine des données

Cette crainte n'a pas été relevée dans le cadre de cette recherche.

Ignorance des lieux de publication de données ouvertes

Ce défaut d'information a effectivement été constaté chez les chercheurs potentiellement intéressés par la publication de leurs données.

Mais aussi : protection du chercheur

Cette raison n'apparaissait pas dans les résultats du sondage de Wiley. Deux chercheurs de l'institut ont relevé un élément ne relevant pas de l'éthique concernant les sujets de la recherche, mais les chercheurs eux-mêmes.

Le chercheur en anthropologie a relevé l'aspect extrêmement cru et personnel de ses notes de terrain. Pour documenter son état d'esprit au cours de l'observation, il doit noter ses humeurs, problèmes physiologiques (diarrhée, maladie) et autres détails pouvant affecter sa perception (manque affectif, etc.) et donc les résultats de sa recherche. La documentation de ces éléments lui sert pour son analyse ultérieure, mais il ne tient pas à ce que d'autres chercheurs puissent les consulter.

Dans le même état d'esprit, le partage des notes prises lors de ses visites en archives poserait problème à la chercheuse en histoire internationale, en raison des multiples fautes linguistiques dans celles-ci. La honte est ainsi un facteur dissuasif.

On peut tirer de ces éléments que la répugnance à partager des données ne posant en apparence pas de problème peut avoir d'autres raisons que la crainte d'être démasqué après avoir falsifié des résultats.

2.6. Bilan des entretiens

Le point le plus évident ressorti de ces entretiens est la diversité des profils. D'une part, les sources primaires sont bien sûr extrêmement variées : documents d'archives, entretiens enregistrés, notes d'observation, statistiques structurées, textes numérisés, voire données collectées individuellement sur le web posent des problèmes largement différents.

Le niveau de littératie numérique est également extrêmement inégal. Sans surprise, un chercheur en économie n'aura probablement pas la même pratique des données ni les mêmes compétences informatiques qu'un chercheur en histoire ou en droit. Il peut toutefois être difficile de juger à partir de ces entretiens si ces connaissances sont purement individuelles ou principalement liées au domaine d'études : un professeur particulièrement connu pour son côté technophile militant est par exemple rattaché au département d'histoire.

Mon intuition est que malgré des variations individuelles notables, le niveau moyen de sensibilité aux données et les questions nécessitant des réponses varieront surtout en fonction du domaine et du département. Ceci implique que d'éventuels guides ou formations devront être plus ou moins pointus en fonction du public, à la fois relativement à des aspects techniques et à des problèmes de fond. La question de l'anonymisation des sujets de recherche ne sera par exemple pas abordée de la même manière pour les uns et les autres.

Une enquête plus détaillée va être menée avant la fin de l'année en collaboration avec le service de la recherche. Un questionnaire a déjà été envoyé à l'ensemble des chercheurs de l'Institut, et des entretiens personnels en plus grand nombre (2 par département) permettront d'affiner le diagnostic de la situation : un échantillon de 5 entretiens n'est évidemment pas représentatif des multiples profils présents à l'IHEID.

3. Services pour les données de la recherche

De multiples institutions ont développé des services pour aider les chercheurs dans la gestion de leurs données de la recherche. Yoon et Schultz ont choisi de répartir les types de soutien dans quatre catégories qui vont être utilisées ci-dessous (Yoon & Schultz, 2017). Elles seront classées par ordre d'implication demandée à la bibliothèque⁶.

Plusieurs travaux ont étudié les services RDM fournis par les bibliothèques à travers des sondages (Tenopir et al., 2017) ou leurs sites web (Si, Xing, Zhuang, Hua, & Zhou, 2015). J'ai choisi en complément de consulter les sites et les bibliothécaires d'institutions proches de l'IHEID pour avoir une idée plus précise des activités pouvant inspirer notre bibliothèque. Ces institutions sont l'European University Institute (EUI, Firenze), la London School of Economics (LSE), l'Université de Genève (Unige), et l'Ecole polytechnique fédérale de Lausanne (EPFL).

J'avais également l'intention d'étudier le cas de Sciences Po Paris, mais la page dédiée aux données de la recherche se résume à annoncer un projet pour 2018 et un groupe Zotero pour qui souhaite consulter des articles sur le sujet. (« Gestion des données de la recherche », 2017)

	EPFL	Unige	Sciences Po	EUI	LSE
Liens externes	Oui	Oui	Non	Oui	Oui
Veille	Non	Actualités	Groupe Zotero	Newsletter/Blog	Non
Guide	Oui	Oui	Non	Oui	Oui
Cours DMP	Oui	Oui	Non	Non	1 seul par an
Autres cours	Oui	Non	Non	Non	Intégrés
Référence	Online	Online	Non	Data librarian	Data librarian + Ethics officer
Consulting	Oui	Oui	Non	Oui	Oui
Hands-on	Non	Non	Non	Pour dépôt	Non
Dépôt	Non	Non	Non	Oui	Non
Note			Projet 2018	Data portal !	

Figure 1 : Services RDM par institution

3.1. Réseau

Selon Yoon et Schultz, ce niveau de service revient à renvoyer les chercheurs vers des ressources internes d'autres services, voire externes à l'institution. De nombreuses bibliothèques adoptent ainsi parmi d'autres stratégies la position de guide.

- Les ressources internes regroupent des liens vers le dépôt de données institutionnel, mais aussi vers le bureau de la recherche, ou encore vers le data/computing center de l'institution lorsqu'il existe.
- Les ressources externes peuvent être de tous types, mais il s'agit en général d'outils de gestion des données⁷, de dépôts de données recommandés, de liens vers les politiques gouvernementales (exigences d'agences de financement de

⁶ Suite à des choix de dénomination discutables dans l'article de Yoon et Schultz, j'ai repris le contenu de leurs catégories tout en modifiant leurs titres afin de réduire les malentendus possibles.

⁷ Voir à ce sujet la liste d'outils de gestion des données préparée à l'EPFL. (Krause, 2015)

la recherche), et dans certains cas de guides RDM plus complets mis en ligne par d'autres bibliothèques.

Les institutions étudiées proposent quasiment toutes cette forme de renseignements, en complément des autres informations proposées. Seule Sciences Po ne propose apparemment même pas cet élément, à moins qu'on considère le groupe Zotero comme tel (il recense notamment des sites et références).

Cette approche paraît indispensable pour l'IHEID, quel que soit l'investissement ultérieur de la bibliothèque dans le soutien à la gestion des données de la recherche. Elle est développable à court terme et permet de répondre à des besoins urgents. Elle restera d'ailleurs pertinente tant que l'ensemble des services possibles ne sera pas développé localement.

3.2. Information

Cette catégorie concerne la publication en ligne ou sous forme imprimée de guides et références pour la gestion des données de la recherche. Yoon et Schultz ont étudié quatre sujets qu'ils considèrent comme nécessaires.

1. Information générale (background information) : l'importance de la bonne gestion des données de la recherche, et la planification de cette gestion. Cet élément rejoint les préoccupations des chercheurs, pour qui le premier moteur poussant à s'intéresser à la gestion des données de la recherche est l'exigence des agences de financement.
2. Documentation des données : l'organisation des données et le développement de métadonnées, mais aussi des recommandations de logiciels de documentation ou de gestion du workflow. L'étude montre que les deux derniers points sont peu présents.
3. Administration des données : propriété, copyright, mais aussi confidentialité et sensibilité, qui sont peu traitées sur les sites surveillés par l'étude. Stockage et préservation entrent également dans cette catégorie, et sont plus largement évoqués.
4. Utilisation et réutilisation des données : questions de citation de données et sensibilisation à l'importance du partage de celles-ci.

Dans le cas de l'EUI, le guide est un élément central du soutien offert concernant les données de la recherche. L'objectif déclaré du bibliothécaire qui l'a développé est « d'aider les chercheurs à s'aider eux-mêmes » pour éviter des demandes trop nombreuses, difficiles à gérer pour une institution de cette taille.

La LSE propose un site abordant les principaux sujets sur des pages dédiées : Anonymisation and data protection, Ethical review, File formats, Informed consent, Keeping data safe, Metadata and documentation, Organising research, et enfin Version control. Il existe en complément un cours Moodle que je n'ai pas pu consulter.

L'Unige a structuré son site RDM selon des étapes de la recherche : rédiger un plan, stocker vos données, partager vos données. Toutes les informations citées plus haut y sont présentées.

Le site RDM de l'EPFL suit à peu près la même structure avec trois étapes : plan and fund, work with data, et publish and preserve. Le même type d'informations y figure.

On notera au passage que pour l'aspect sensibilisation, l'EUI propose une newsletter concernant les données, tenue ainsi qu'un blog par le bibliothécaire de liaison du département d'économie. L'Unige a également prévu une page d'actualités disposant d'un flux RSS, mais son évolution est encore incertaine.

Au cours de mes recherches, j'ai pu constater que de nombreux guides dédiés à la gestion des données de la recherche, en particulier d'universités anglo-saxonnes, ont été développés sur une plateforme LibGuide, système déjà utilisé par la bibliothèque de l'IHEID pour d'autres sujets (« LibGuides IHEID », s. d.). C'est donc une piste de développement possible et relativement facile à mettre en place.

3.3. Formation

Cette catégorie regroupe l'organisation de cours et ateliers collectifs. Un peu plus d'un tiers des institutions étudiées par Yoon et Schultz (34.5%) en proposaient. Dans les cas que j'ai pu étudier, seul l'EUI (et bien sûr Sciences Po) ne propose apparemment pas de cours ou de formation.

Ateliers DMP

La Division de l'information scientifique (DIS) de l'Université de Genève et la bibliothèque de l'EPFL ont tous deux proposé au mois de septembre 2017 des ateliers pour les chercheurs en quête de financement du FNS. Ils avaient pour but de sensibiliser les chercheurs à leurs nouvelles obligations, de clarifier les questions incluses dans le plan de gestion des données du FNS, et de dissiper les craintes des participants face à ces demandes. Les notes prises durant les présentations et détaillant leur contenu sont disponibles en annexe.

La LSE organise également un atelier DMP par an depuis plusieurs années. D'après l'organisateur, le nombre de participants est trop faible pour justifier un plus grand nombre d'ateliers. L'EUI n'organise pas d'atelier DMP.

Dans les deux institutions suisses, les ateliers duraient 90 minutes, afin d'attirer plus facilement les chercheurs en nombre. Un format de 90 minutes n'est toutefois pas la seule option possible. Un webinaire IASSIST suivi le 28 septembre 2017 présentait par exemple les méthodes adoptées par des formateurs pour des ateliers DMP durant une journée.

Si une telle durée peut être difficile à justifier auprès d'enseignants ayant déjà un planning chargé, elle indique toutefois qu'une durée plus longue que les exemples précédents peut tout à fait être envisagée. Les ateliers seraient alors moins nombreux, mais permettraient d'aller plus dans le détail, et de développer des exercices pertinents en fonction du profil des chercheurs inscrits.

Autres formations

La bibliothèque de l'EPFL propose aux doctorants et jeunes chercheurs une série de cours intitulée « From research to publication » (« Doctorants », s. d.). Quatre demi-journées permettent de survoler l'ensemble du processus, et la quatrième partie du cours est spécifiquement consacrée à la gestion des données de la recherche. Le support de cours est disponible en ligne (« From research to publication, module 4 : Research Data Management », s. d.). Ce cours sensibilise les jeunes chercheurs à la question de la GDR, et plus particulièrement à son intégration dans la problématique de l'open science. La moitié de la session est ainsi dédiée à la publication et à la réutilisation des données.

Les autres institutions ne semblent pas proposer de cours spécifiques ou intégrés au cursus des étudiants en doctorat. A l'étranger, de grandes institutions proposent parfois des formations en ligne.

Evénements

A la LSE, des événements sur la gestion des données sont intégrés dans les activités de la PhD academy et du Centre for personal and professional development. L'année dernière, deux cours ont ainsi été organisés : l'un sur la gestion de données quantitatives, et l'autre sur l'éthique et la gestion de données de médias sociaux.

L'Unige et l'EPFL organisent également régulièrement des événements liés directement à l'open access et aux données de la recherche. Les midis de l'info scientifique, des séances de 45 minutes organisés par la DIS, pourraient également être une source d'inspiration (« Les Midis de l'info scientifique », 2017).

3.4. Participation

Cette catégorie de service est caractérisée par l'engagement actif de la bibliothèque avec des chercheurs. On peut encore la subdiviser selon les termes de Tenopir et al. entre les services de consultation (renseignements et expertise) et les services techniques (curation et dépôt de données) (Tenopir et al., 2017).

Service de renseignements RDM

C'est l'équivalent GDR du service de référence traditionnel « Ask a librarian », et il n'en est pas toujours séparé. Ils sont donc statistiquement très répandus dans les différentes enquêtes, sans qu'on soit certain de l'existence de conseils spécifiques sur le sujet.

L'EUI et la LSE proposent des heures de réception dédiées pour les questions liées aux données, et la seconde y adjoint même les services d'un spécialiste des questions éthiques. Les institutions suisses quant à elles offrent un contact en ligne et la proposition d'un rendez-vous.

Service d'expertise RDM

D'autre part, les services de consultation recouvrent l'existence d'une offre d'expertise plus avancée, basée sur le projet lui-même : analyse des mesures prises et conseils sur celles qui doivent être mises en place au cas par cas. Ce type de service requiert des connaissances spécifiques des méthodes de la discipline (voir plus loin).

La bibliothèque de l'EPFL propose ainsi des conseils personnalisés aux chercheurs sur demande. C'est le bureau de la recherche qui provoque en général la prise de contact. Jusqu'à l'été 2017, c'était l'outil privilégié pour le remplissage de DMP dans le cadre de projets pilotes Horizon 2020 (une vingtaine entre janvier et mai 2017). Cette option est devenue difficilement tenable avec les exigences du FNS en raison de l'explosion du nombre de demandeurs potentiels, d'où le choix d'organiser les ateliers DMP mentionnés plus haut.

Les autres institutions étudiées proposent toutes (sauf Sciences Po) ces services sur leurs pages dédiées, évidemment sur rendez-vous.

Services techniques

Tenopir et al. parlent de services techniques quand des *data librarians* interviennent directement sur le matériel (« hands-on »), notamment pour documenter les données/métadonnées ou pour préparer leur partage sur un dépôt local ou externe.

Ce type de service existe dans d'autres universités. Parmi les institutions étudiées pour ce travail, il n'a toutefois été mis en place qu'à l'EUI.

Dépôt de données

Les services techniques accompagnent souvent l'existence d'un dépôt de données institutionnel. C'est le cas à l'EUI, même si les services techniques mentionnés au point précédent ne s'appliquent pas uniquement à leur propre dépôt.

Là encore, les institutions suisses et même la LSE ne proposent pas (encore) de dépôt de données institutionnel.

4. La bibliothèque IHEID et ses ressources

4.1. Personnel

Comme évoqué plus haut, la bibliothèque compte 18 employés pour environ 16 ETP. Ce nombre comprend un poste d'archiviste et trois d'aide-bibliothécaires. La bibliothèque figurerait donc en queue de classement par ETP dans la statistique suisse des bibliothèques 2016, aux côtés des bibliothèques de l'Université de la Suisse italienne (Lugano et Mendrisio, voir fig. 2). La médiane des bibliothèques académiques et de recherche américaines se situe à 84 professionnels ETP, en excluant les professions de support (Wilder, 2018).

Bibliothèques universitaires, données 2016	Total employés (total personnes)	Equivalents plein temps (total postes)
Total	1,904	1,289
Bibliothek der Eidgenössische Technische Hochschule Zürich (ETHZ)	298	199.4
Zentralbibliothek Zürich	241	189.6
Universitätsbibliothek Bern	282	180.3
Universitätsbibliothek Basel	185	150.0
Bibliothèque cantonale et universitaire de Lausanne	261	130.4
Bibliothèque de l'Université de Genève	188	114.4
Bibliothèque cantonale et universitaire de Fribourg	132	93.6
Zentral- und Hochschulbibliothek Luzern	92	69.3
Hauptbibliothek Universität Zürich	54	41.4
Bibliothèque centrale de l'Ecole Polytechnique Fédérale de Lausanne (EPFL)	46	38.8
Bibliothèque publique et Universitaire de Neuchâtel	46	31.1
Bibliothek der Universität St. Gallen	45	29.9
Biblioteca universitaria di Lugano	22	11.8
Biblioteca dell'Accademia di architettura, Mendrisio	12	9.5

*Figure 2 : Employés des bibliothèques universitaires suisses
(« Bibliothèques universitaires », 2017)*

Pour poursuivre la comparaison on peut constater dans la figure 3 que le nombre d'employés de la bibliothèque de l'IHEID est très élevé par rapport au nombre d'étudiants, en raison de l'absence d'étudiants en bachelor à l'Institut. Ceci signifie que les tâches traditionnelles (développement de collection) sont assurées sans difficulté. Pour le soutien à la recherche en revanche, les ratios employés/doctorants et employés/professeurs sont moins impressionnants et plus proches de ceux de l'Université de Genève.

	ETP biblio	Etudiants	ETP/Etu	Doctorants	ETP/Doc	Professeurs	ETP/Prof
EPFL	39	10536	0.37%	2124	1.84%	346	11.27%
Unige	114	16530	0.69%	2231	5.11%	753	15.14%
IHEID	16	897	1.78%	344	4.65%	68	23.53%

*Figure 3 : Comparaison du nombre d'étudiants et professeurs
(Peila, 2017) (« Chiffres détaillés », 2017)*

Il faut toutefois noter que le nombre de bibliothécaires reste faible en valeur absolue, ce qui peut rendre le développement de nouveaux services plus complexe : le travail

nécessaire au développement d'une formation ou d'un site web est quasiment le même quelle que soit la taille de l'institution.

Pour la même raison, il est plus difficile d'envisager l'engagement de nouveaux collaborateurs pour ces missions : l'impact sur le budget serait comparativement plus fort. La tendance à la bibliothèque de l'IHEID est d'ailleurs plutôt à la réduction des postes : deux ont été supprimés en 2015 après des départs. Il est donc raisonnable d'imaginer que les nouveaux services devront être assurés par le personnel déjà en place, sans engagement d'un *data librarian*.

D'après Tenopir et al., 27,7% des bibliothèques européennes se préoccupant de services RDM répondaient en 2016 avoir engagé de nouveaux employés pour ces tâches, et 14,3% comptaient le faire. Selon la même étude, 45,4% avaient réassigné des employés existants à ces nouvelles tâches, et 21,8% comptaient le faire.

Une reconversion d'employés déjà présents à ce type de fonctions est donc possible, mais nécessite de proposer des possibilités de formation et de les encourager fortement. Le développement du soutien aux chercheurs peut se faire en prenant en compte les limitations de la bibliothèque et en anticipant les problèmes par le développement des compétences des bibliothécaires actuels.

4.2. Identifier les besoins en compétences

Les connaissances impliquées dans un processus de GDR sont nombreuses. Les besoins pour le développement de services de données de la recherche sont de plusieurs types, en fonction des rôles et des objectifs de la bibliothèque. Cox et al. en ont défini une série (Cox, Verbaan, & Sen, 2012), que je me suis permis de compléter ci-dessous :

Compétences générales

Ces compétences ne sont qu'une évolution des rôles existants des bibliothèques académiques, et sont présentes chez différents bibliothécaires. Des formations complémentaires pourraient toutefois être nécessaires.

- Conviction et argumentation (advocacy)
- Compétences de rédaction techniques (guides)
- Compétences pédagogiques (cours et ateliers)
- Entretien de référence (service de référence)
- Savoir identifier et localiser les sources de données : car la gestion commence par l'acquisition – les services de renseignement traditionnels trouvent ici une extension naturelle
- Compréhension stratégique de la gestion des données de la recherche : au-delà de comprendre le « comment » et les bonnes pratiques, il faut que les bibliothécaires maîtrisent le « pourquoi »

L'offre de formation est à la base des services de GDR, et signifie que les besoins en compétences pédagogiques des bibliothécaires vont croissant. En fonction des options prises pour le projet Information literacy, il est vraisemblable que d'autres bibliothécaires puissent avoir besoin d'une formation dans le domaine.

Compétences techniques

Ces compétences spécifiques sont au cœur de la GDR et doivent être développées :

- Copyright et licences : à la fois pour l'utilisation de données et pour leur diffusion

- Ethique et anonymisation des données
- Métadonnées et documentation
- Nommage de fichiers et classement
- Logiciels et formats de fichiers pour la conservation

Si les connaissances théoriques existent déjà à la bibliothèque, une expérience pratique reste indispensable.

Mais surtout compétences appliquées aux disciplines spécifiques

Brown et al. présentent un cas pratique d'adaptation d'une bibliothécaire spécialisée à ses nouvelles tâches. L'élément le plus important qui ressort de cet article est qu'une formation théorique est nécessaire mais largement insuffisante.

Pour pouvoir assister à la création d'un DMP de bonne qualité dans une discipline, il doit avoir étudié des exemples de DMP existants de celle-ci. Pour pouvoir fournir des recommandations pertinentes sur l'organisation de données, un bibliothécaire doit avoir été confronté à des exemples réels de données du champ de recherche concerné (Brown, Wolski, & Richardson, 2015).

4.3. Qui, combien, comment ?

Données + bibliothécaire ≠ *data librarian*

Les connaissances théoriques de base concernant données et formats de fichiers ne font pas d'un bibliothécaire un *data librarian*. Rice et Southall appuient sur la nécessité d'acquérir en plus des compétences de manipulation de celles-ci. Un tel bibliothécaire doit être à l'aise avec les logiciels utilisés par les chercheurs, et apprendre des compétences tel que le développement informatique (Rice & Southall, 2016).

Actuellement, les connaissances sont maigres à la bibliothèque dans ce domaine, au-delà de bases de Python ou de PHP/MySQL. Encourager la formation dans ce domaine pourrait être nécessaire pour le développement de services de type participatif. Ce manque est un obstacle évident pour le développement de services avancés.

Elargir l'équipe et développer les liens

Sans même parler d'un véritable *data librarian*, dont l'engagement ou la formation pourrait être difficile, la bibliothèque ne compte actuellement qu'une personne ayant suivi des formations relatives à la GDR. L'absence d'un collaborateur ne doit pas empêcher la bibliothèque de répondre aux questions de base des chercheurs, et il est important que les données de la recherche à la bibliothèque ne se résument pas à une personne.

D'autres collaborateurs disposant au moins de connaissances superficielles du sujet constitueraient un avantage notable lorsque des formations devront être organisées pour les chercheurs : deux intervenants donneront généralement une présentation plus complète et dynamique qu'un seul. La solution pourrait venir des...

Bibliothécaires de liaison

Comme nous l'avons vu précédemment, pour des services au-delà de la sensibilisation et de l'information, la pratique de la gestion des données de la recherche ne prend tout son sens qu'au niveau disciplinaire, voire du projet. Il serait donc nécessaire de se rapprocher des chercheurs pour leur offrir un service de meilleure qualité.

Les bibliothécaires de liaison de l'IHEID ont pour certains étudié les domaines de recherche de l'Institut, et cherchent à nouer de tels liens avec les chercheurs depuis la définition de la stratégie 2016-2019. Une plus grande connaissance de chaque domaine, par exemple en assistant à des cours de méthodologie ou en observant des processus de recherche réels, seraient à même d'apporter des compétences qui leur seraient utiles dans leur travail habituel, et qui complèteraient celles du bibliothécaire qui se spécialisera dans les données de la recherche.

Pour ce faire, l'ensemble des bibliothécaires de liaison doit comprendre les bases de la GDR, afin de pouvoir échanger avec les chercheurs de leur département académique sur leurs pratiques spécifiques. Ils doivent être formés à un niveau de base, soit en utilisant des ressources déjà disponibles (formations en ligne), soit à travers un ou des ateliers spécifiques, qui pourraient au passage servir d'entraînement pédagogique au bibliothécaire spécialisé cité précédemment.

4.4. Partenaires potentiels (internes et externes)

Pour la partie politique et guidelines, le partenaire privilégié à l'IHEID est évidemment le bureau de la recherche. Les contacts pris avec ce service au printemps ont déjà deux résultats : d'une part l'information des chercheurs concernant les exigences des agences de financement avant la récente soumission de demande de fonds, et d'autre part la rédaction d'un projet de politique de l'Institut sur les données de la recherche.

Des liens doivent également être noués avec les chercheurs eux-mêmes et avec les départements académiques. Cela permettrait de mieux cerner leurs besoins, et de pouvoir éventuellement développer des capacités « hands-on » ou au moins une connaissance pratique de leurs activités pour nourrir les projets de formation de la bibliothèque.

La bibliothèque peut collaborer avec le service informatique pour mieux renseigner les chercheurs, et obtenir des informations qui seront certainement utiles pour toute formation sur la GDR. L'aspect du stockage par exemple ne doit pas être négligé. Au cours des entretiens, les principales demandes des chercheurs avaient trait à des besoins non satisfaits dans ce domaine, et à un manque d'information.

Les autres universités romandes sont évidemment des partenaires particulièrement intéressants. Les échanges avec les bibliothécaires permettent d'établir de bonnes pratiques et d'écarter par des retours d'expérience des projets trop ambitieux.

Des universités suisses ont par ailleurs lancé en commun un projet nommé DLCM (Data Life-Cycle Management) en vue de fournir des services divers pour la gestion des données de la recherche au niveau national. Une partie des solutions développées (livrables en 2018) pourraient être utilisables par l'IHEID. (« Projet DLCM », s. d.)

5. Propositions de services et priorités

Maintenant que les moyens nécessaires ou envisageables ont été détaillés, nous pouvons nous avancer à imaginer ce que seront les services offerts par la bibliothèque une fois prises toutes les mesures nécessaires.

5.1. Sensibilisation

La première tâche à remplir par la bibliothèque est le développement de la conscience des enjeux de la gestion des données de la recherche. C'est un travail de longue haleine, qui doit se faire notamment en collaboration avec le service de la recherche, et peut-être celui de la communication. Parmi les premiers résultats, on peut déjà citer le développement en cours de guidelines par le service de la recherche, ainsi que la grande enquête envoyée aux professeurs et chercheurs, qui servira à préciser leurs besoins.

Il ne faut pas s'arrêter aux premiers résultats obtenus. Des informations doivent être publiées régulièrement pour sensibiliser les chercheurs à la question, notamment à travers les newsletters de l'institut, de nouveaux supports de communication, ou en organisant des événements publics sur le sujet comme le font de nombreuses institutions. La bibliothèque peut être un soutien pour l'activité d'autres services.

Par ailleurs, des événements ponctuels lors desquels des chercheurs internes ou externes viendraient partager leurs expériences sur la gestion des données de la recherche pourraient être organisés en collaboration avec un département ou le bureau de la recherche.

Plus directement, les projets de formation évoqués plus bas doivent aussi mettre en valeur les avantages que les chercheurs peuvent tirer d'une bonne GDR, et pas seulement leurs obligations. C'est l'engagement des chercheurs eux-mêmes qui fera le succès ou l'échec de la stratégie de la bibliothèque.

5.2. Documentation de référence

Le développement d'un guide de bonnes pratiques de gestion des données de la recherche nécessite un fort investissement préalable en temps, mais permet ensuite de répondre à de nombreuses questions sans ressources supplémentaires.

Comme nous l'avons vu plus haut, ce type de référence peut s'appuyer sur des ressources tierces développées par d'autres universités, par exemple. Afin de ne pas handicaper le développement d'autres activités durant la phase de rédaction, cette option devra probablement être privilégiée.

Si l'objectif reste une documentation propre à l'Institut, celle-ci pourrait être créée progressivement pour publication dans le cadre des démarches de sensibilisation, par exemple sur un blog ou dans une newsletter. Ainsi, les publications successives viendront enrichir un guide qui deviendra peu à peu complet.

Notons que le service de la recherche a pris la décision de développer des guidelines sommaires plutôt qu'une simple politique institutionnelle. Ces guidelines pourraient permettre aux chercheurs de se poser les questions pertinentes, et il reviendra à la bibliothèque de suggérer des pistes de réponses.

5.3. Ateliers DMP

Comme j'ai pu le constater à l'EPFL et à l'Université de Genève, les ateliers d'information sur ce sujet sont un moyen efficace d'atteindre les chercheurs, et doivent donc être développés en priorité. La durée de ces ateliers, leur contenu et d'éventuels partenariats doivent être étudiés afin de garantir leur tenue avant la prochaine date de soumission de demandes de financement.

Nous avons conclu précédemment que les besoins sont très variés selon le domaine et le parcours des chercheurs. Des ateliers de 90 minutes ne semblent pas permettre de réelle différenciation en fonction des participants. La solution pourrait être l'organisation d'ateliers plus longs, avec notamment des exercices basés sur des cas réels du domaine de recherche. Notons toutefois la faible demande notée par la LSE, qui pourrait ne pas justifier une préparation de séance trop importante.

5.4. Programme de formation

De petits ateliers dédiés à un point spécifique de la bonne gestion des données à destination des doctorants pourraient être organisés. Il faut avant cela les consulter, par exemple au cours de focus groups organisés par la bibliothèque, sur leurs besoins et leur intérêt à ce sujet. Une intégration au sein des cours de méthodologie des différents départements pourrait être pertinente, et permettrait de moduler le niveau de complexité de l'atelier en fonction des destinataires.

Le développement du projet de cours en ligne d'information literacy à destination des nouveaux étudiants pourrait être l'occasion de développer les capacités des bibliothécaires dans ce domaine, et d'envisager une offre de ce type. On notera toutefois que d'autres formations en ligne de qualité existent également, et que ce type d'offre moins interactive apporterait probablement moins de valeur ajoutée à notre activité.

Des équivalents de Midis de l'info scientifique mentionnés plus haut pourraient également être organisés sur des sujets précis. Entre sensibilisation et formation, ceux-ci auraient l'avantage d'attirer les chercheurs curieux sans leur paraître inutilement lourds.

5.5. Expertise et consultation

Idéalement, c'est un type de service que la bibliothèque devrait mettre en place à plus ou moins long terme. Il répond aux recommandations de l'audit de la bibliothèque et la profile comme partenaire de la recherche. Il nécessite toutefois un développement des compétences important, comme évoqué dans le chapitre 4.

Il est probable que ce développement de compétences devra se faire sur le tas, par la pratique et l'observation. Il serait donc bon de démarrer cette phase d'observation relativement tôt auprès de chercheurs bienveillants à l'égard de la bibliothèque.

5.6. Le cas des services techniques

La mise en place d'un dépôt de données paraît exclue car non pertinente : l'IHEID reste une petite institution académique à l'échelle mondiale, et ne dispose pas des ressources humaines et financières pour gérer un tel projet. Des universités suisses plus importantes y ont également renoncé. Les bibliothécaires doivent toutefois pouvoir orienter les chercheurs vers les dépôts les plus pertinents en fonction de leurs souhaits.

Pour ce qui est des services « hands-on », comme nous l'avons vu plus haut, un véritable *data librarian* devrait acquérir des compétences informatiques spécialisées dont la bibliothèque ne dispose pas encore. De plus, ce type de service est coûteux en temps, et nous avons vu plus haut que les ressources de la bibliothèque ne sont pas illimitées. Ce type de services n'est donc pas envisageable en l'état.

6. Conclusion

Les chercheurs de l'IHEID ont des besoins variés en matière de services pour la gestion des données de la recherche. Les besoins divers d'information, de formation et de soutien doivent être remplis de manière différenciée pour obtenir des résultats probants. Un modèle unique ne conviendra pas à l'ensemble des situations.

La mise en place efficace de ces services nécessite la mise à niveau des compétences des bibliothécaires et leur collaboration avec le bureau de la recherche et les départements académiques. Il n'est pas trop tard pour bien faire, mais le problème doit être pris au sérieux si l'on veut saisir cette opportunité de profiler la bibliothèque comme partenaire de la recherche.

En guise de conclusion personnelle, j'ajouterai que le sujet est vaste et fascinant : je compte rédiger à l'avenir des billets de blogs pour aborder de nombreux sujets qui n'avaient pas leur place dans ce travail. J'espère aussi pouvoir transmettre à mes collègues l'intérêt que j'ai développé pour cette question, et pouvoir mettre en application rapidement les intentions que le travail exprime.

7. Bibliographie

- Bibliothèques universitaires. (2017, juillet 25). Consulté 12 octobre 2017, à l'adresse <https://www.bfs.admin.ch/bfs/fr/home/statistiques/culture-medias-societe-information-sport.assetdetail.3104685.html>
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078. <https://doi.org/10.1002/asi.22634>
- Brown, R. A., Wolski, M., & Richardson, J. (2015). Developing new skills for research support librarians. *The Australian Library Journal*, 64(3), 224-234. <https://doi.org/10.1080/00049670.2015.1041215>
- Chiffres détaillés. (2017). Consulté 13 octobre 2017, à l'adresse <https://information.epfl.ch/cms/site/accueil/lang/fr/chiffres-detailles>
- Concordat on Open Research Data. (2016, juillet 28). Consulté à l'adresse <http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/>
- Cox, A., Verbaan, E., & Sen, B. (2012). Upskilling Liaison Librarians for Research Data Management. *Ariadne*, (70). Consulté à l'adresse <http://www.ariadne.ac.uk/issue70/cox-et-al>
- Doctorants. (s. d.). Consulté 2 octobre 2017, à l'adresse <https://library.epfl.ch/training-phd/>
- Ferguson, L. (2014, novembre 3). How and why researchers share data (and why they don't). Consulté 2 octobre 2017, à l'adresse <https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont>
- From research to publication, module 4 : Research Data Management. (s. d.). Consulté 18 septembre 2017, à l'adresse <https://library.epfl.ch/files/content/sites/library3/files/Formations/4.Research-Data-Management-Course-Notes.pdf>
- Gestion des données de la recherche. (2017). Consulté 2 octobre 2017, à l'adresse <http://www.sciencespo.fr/bibliotheque/fr/enseignants-chercheurs/gestion-donnees-recherche>
- Gurria, A. (Éd.). (2007). *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics*. Paris: Editions de l'OCDE. Consulté à l'adresse <https://www.oecd.org/fr/sti/sci-tech/38500823.pdf>
- Krause, J. (2015). A Selection of Research Data Management Tools Throughout the Data Lifecycle. Consulté à l'adresse <https://infoscience.epfl.ch/record/211157>
- Les Midis de l'info scientifique. (2017, mai 5). Consulté 12 octobre 2017, à l'adresse <https://www.unige.ch/biblio/fr/formation/midis-info-scientifique/>
- LibGuides IHEID. (s. d.). Consulté 2 octobre 2017, à l'adresse <http://libguides.graduateinstitute.ch/>

- L'Institut en 2016-2017. (2017). Institut de hautes études internationales et du développement. Consulté à l'adresse http://graduateinstitute.ch/files/live/sites/iheid/files/sites/about-us/presentation/factsheet_Graduate_Institute_Geneva_fr.pdf
- OMB. (1999, septembre 30). Circular A-110 Revised 11/19/93 As Further Amended 9/30/99. Consulté 2 octobre 2017, à l'adresse https://www.whitehouse.gov/omb/circulars_a110
- Open Access. (2017, février 22). Consulté 2 octobre 2017, à l'adresse <https://erc.europa.eu/funding-and-grants/managing-project/open-access>
- Open Research Data : les requêtes devront inclure un plan de gestion des données. (2017, mars 6). Consulté 2 octobre 2017, à l'adresse <http://www.snf.ch/fr/pointrecherche/newsroom/Pages/news-170306-open-research-data-bientot-une-realite.aspx>
- Peila, O. (2017). *Statistique universitaire : Étudiantes et étudiants, diplômés et personnel 2016* (p. 56). Genève: Université de Genève. Consulté à l'adresse https://www.unige.ch/stat/files/8915/0219/4147/Statistiques_univ._2016_-_en_ligne_corr2.pdf
- Projet DLCM. (s. d.). Consulté 2 octobre 2017, à l'adresse <https://dlcm.ch/>
- Research data lifecycle. (s. d.). Consulté 2 octobre 2017, à l'adresse <https://www.ukdataservice.ac.uk/manage-data/lifecycle>
- Rice, R., & Southall, J. (2016). *The data librarian's handbook*. London: Facet Publishing.
- Si, L., Xing, W., Zhuang, X., Hua, X., & Zhou, L. (2015). Investigation and analysis of research data services in university libraries. *The Electronic Library*, 33(3), 417-449. <https://doi.org/10.1108/EL-07-2013-0130>
- Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., ... Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, 27(1), 23-44. <https://doi.org/10.18352/lq.10180>
- Wilder, S. (2018). Hiring and staffing trends in ARL libraries. *Research Library Issues*, à paraître. Consulté à l'adresse <http://www.arl.org/storage/documents/publications/rli-2017-stanley-wilder-article2.pdf>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yoon, A., & Schultz, T. (2017). Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites. *College & Research Libraries*, à paraître. Consulté à l'adresse <http://crl.acrl.org/index.php/crl/article/view/16620>

Annexes

Annexe 1 : Atelier DMP de l'Université de Genève

Une douzaine d'ateliers ont été organisés par la Division de l'information scientifique de l'Université (DIS). Ils ont eu lieu sur les différents sites de l'Université afin de favoriser la participation des chercheurs, et le nombre d'inscrits s'est élevé à plusieurs dizaines issus de tous les domaines de recherche.

Après quelques définitions (les données de la recherche selon l'OCDE, data lifecycle, principes FAIR), la présentation s'est appliquée à expliquer pourquoi les données doivent être gérées, puis au détail du DMP du Fonds national suisse. Les intervenants ont alors parcouru un formulaire de DMP champ par champ afin de détailler les renseignements attendus.

Un exercice a ensuite été distribué. Il constituait à identifier, dans un DMP anglo-saxon écrit sous forme libre, les éléments de réponse aux questions du modèle de DMP du FNS. Enfin, une présentation du site research data de l'Université⁸ concluait la présentation.

Les questions des participants ont souligné plusieurs points importants, souvent dus à des incompréhensions ou un manque d'information :

- Craintes face aux risques d'un DMP mal rempli. En réalité, le DMP ne fait pas partie des critères d'évaluation de la demande de fonds.
- Oppositions au partage des données : certains chercheurs craignaient que les données doivent être ouvertes pendant le processus de recherche. D'autres ont évoqué la sensibilité des données.
- Certaines difficultés (définition des données, lacunes techniques) parfois dues à leur domaine d'étude.
- Interrogations sur la propriété des données de la recherche (Unige vs chercheur) – apparemment, les données brutes appartiendraient à l'Université.
- La question du prix des dépôts a été posée. Zenodo est gratuit mais limité à 50 Go. Dryad est gratuit jusqu'à 20 Go. Le FNS prévoit une enveloppe supplémentaire de 10'000 chf pour couvrir les frais de curation et conservation des données qui peuvent notamment couvrir le coût du dépôt de données massives.

⁸ <https://www.unige.ch/researchdata/fr/>

Annexe 2 : Atelier DMP de l'EPFL

Contrairement aux ateliers de l'Université de Genève, les ateliers organisés à l'EPFL étaient à chaque fois destinés à une ou deux facultés spécifiques, et étaient localisés et nommés en fonction de celles-ci.

En pratique, les participants restaient toutefois libres de s'inscrire à n'importe quelle présentation en fonction de leurs disponibilités, et la présentation restait agnostique. Toutes les présentations ont été effectuées en anglais.

Les supports de cours distribués incluaient le modèle de DMP développé en collaboration avec l'ETHZ, mais aussi les recommandations de formats de fichiers pour la conservation. Celles-ci sont aussi disponibles sur le site web de l'équipe RDM⁹. La présentation est également téléchargeable.¹⁰

Après une introduction (chaque participant à la séance s'est présenté) et la diffusion d'une vidéo « The what, why and how of data management planning » sur Youtube¹¹, la séance a consisté en une présentation et démonstration de quelques outils.

Les chercheurs ont pu voir comment trouver un dépôt disciplinaire sur RE3data, découvrir brièvement l'interface de Zenodo, et le système de remplissage de DMP sur MySNF. L'atelier se termine ensuite par un exercice consistant à identifier à quels points d'un questionnaire DMP répondaient des extraits divers. Lors de la séance à laquelle j'ai assisté, le temps était insuffisant pour compléter cet exercice.

Les intervenants ont insisté sur l'utilité du DMP pour les chercheurs eux-mêmes (réflexion et gain de temps pour l'avenir), et sur le fait que c'est un document vivant et évolutif jusqu'à la fin du projet.

Encore une fois, les questions étaient larges et dépendaient parfois de domaines spécifiques de recherche :

- la question « What is data ? » est revenue, notamment pour parler de lab books
- question sur le coût de la curation/conservation, et réponse sur les 10000 chf du FNS
- « quid des découvertes brevetables ? », un enjeu important dans une école d'ingénieurs. Le DMP doit être mis à jour pour indiquer cet aspect.
- est-ce que Zenodo supporte la gestion de versions ? Oui. Le DOI dirige vers la version à jour mais les autres sont accessibles.
- la bibliothèque fournit des feedbacks sur les projets de DMP

⁹ <https://library.epfl.ch/research-data/>

¹⁰ <http://go.epfl.ch/aCs>

¹¹ <https://youtu.be/gYDb-GP1CA4>

Annexe 3 : Questionnaire d'entretiens IHEID

1. What percentage of your professional time do you dedicate to research?
2. Identify the main discipline of your work
3. Are you aware of the references to the proper management of research data set by SNSF and other funders?
4. Are you aware of your responsibilities and the Institute's responsibilities for proper management of research data?
5. The SNSF states they expect "that data generated by funded projects will be publicly accessible in digital databases provided there are no legal, ethical, copyright or other issues." To what extent do you agree with this statement?
6. When starting a research project, do you document a data management plan outlining how the data will be collected, managed and stored?
7. What research projects are associated with your research data?
8. Additional researchers (that worked on the project)
9. Do you keep a registry or list of your research data collections?
10. What are the main types or formats of digital and non-digital data you generate or collect?
11. Can you provide an indication of how much digital research data you currently have and how much you expect to generate over the next 12 months?
12. What advanced IT practices or technologies do you employ in your collection and management of data?
13. Would you use someone else's data in your research?
14. Where would you look for existing research data?
15. Do you currently use a research data repository for your data collections?
16. Does the data have spatial / temporal elements?
17. How long do you keep your data? Do you revisit it?
18. Does your data have links to other data and/or related information (eg publications)? What are meaningful sets of data (and what level(s) of granularity are required?
19. What metadata do you capture?
20. Where does your metadata come from?
21. Other sources? (Proposals database/Grants management system/Project management system/Image headers/Entered manually)
22. What formats does your data / metadata come in? Are these formats proprietary or open? If they are proprietary, how do you convert them to other formats? What formats do you need your data / metadata in?
23. How do you currently store your data / metadata in the short and long term?
24. How adequate is the data / metadata you capture for :
 - a. your own work (eg reuse)?
 - b. other people using / reusing your data?
25. Are there current access restrictions applied to your research data? If there are access restrictions, please specify:
26. Identify the appropriate access conditions you wish to apply to your research data:
 - a. Open to all / no restrictions
 - b. Metadata only open to all (not the data and objects)
 - c. No access to anyone.
 - d. (Why?)
27. When in the project do you want to share data / metadata (during the lifetime of the project, at selected stages of the process, after publication)?

28. How important do you think sharing your research data is for boosting your research profile?
29. What data / metadata does your discipline require for publication?
30. Are there any legal or ethical considerations associated with your research data that require attention?
31. What is your biggest research data management problem?
32. Do you want more information and guidelines on research data management? Which topics in particular?
33. What would you like to be able to do? What would you need?

Ce questionnaire s'est basé sur celui de l'Australian National Data Service, avec quelques modifications liées au contexte de l'IHEID : <http://www.ands.org.au/working-with-data/data-management/institutional-dm-frameworks/research-data-interviews>

DÉCLARATION SUR L'HONNEUR

Je soussigné(e) déclare sur l'honneur avoir rédigé personnellement ce travail écrit. Celui-ci n'a pas d'autres sources que celles que j'ai indiquées systématiquement dans le texte (avec les citations entre guillemets) et dans les références bibliographiques. J'ai pris connaissance que la fraude et le plagiat seront sanctionnés par un échec et seront également communiqués au Rectorat qui est compétent pour prendre des sanctions disciplinaires.

Lieu:

Date:

Signature: