

Données de la recherche : cadre juridique et licences



Travail de master réalisé par :
Anouk SANTOS

Sous la direction de :
René SCHNEIDER, professeur HES

Genève, le 17 août 2020

Master en Sciences de l'information
Haute école de gestion de Genève (HEG-GE)

Remerciements

Je tiens à adresser mes plus sincères remerciements à toutes les personnes qui m'ont aidée dans la réalisation de ce travail de master :

Pierre-Yves Burgi pour m'avoir confié ce mandat, la rapidité de ses réponses et sa grande disponibilité. Merci également à **Hugues Cazeaux**, qui lui aussi s'est rendu disponible et a pris le temps de m'aider dans la réalisation de ce mandat.

René Schneider, pour avoir accepté de me suivre dans ce travail, pour la grande confiance qu'il m'a témoignée dans l'exécution de ce mandat et pour ses conseils avisés.

Angélique Boschung, pour avoir accepté le rôle d'experte de ce travail de master.

Je souhaite également témoigner ma gratitude ici à toutes les personnes qui ont reçu avec bienveillance mes questions d'ordre juridique et qui ont pris le temps de leur apporter une réponse de manière claire par email, téléphone ou visioconférence. Dans l'ordre alphabétique, je remercie :

- **Natacha Hausmann**, directrice des Affaires juridiques à l'Université de Genève ;
- **Marc Jones**, avocat à CivicActions, Connecticut, USA ;
- **Suzanna Marazza** du Competence Center in Digital Law (CCDigitalLaw), que je remercie également ici d'avoir accepté de relire un extrait de ce travail ;
- **Grégoire Mercier**, juriste au service juridique de la HES-SO Genève ;
- **Odile Niederhauser**, responsable de la gestion documentaire et de l'archivage à la HES-SO Genève ;
- **Prof. Dr. Simon Schlauri**, partenaire chez Ronzani Schlauri Avocats, professeur à l'Université de Zurich et membre de l'équipe légale de Creative Commons Switzerland ;
- **Céline Vilmen**, bibliothécaire juridique en charge de la collection de droit et Data Protection Officer à l'Institut de hautes études internationales et du développement (IHEID) ;
- **Stéphane Werly**, préposé cantonal à la protection des données et à la transparence de l'Etat de Genève.

Je remercie aussi **Hester**, ma maman, pour sa précieuse relecture de ce travail ainsi que de tous les travaux d'étudiante qui l'ont précédé. Promis, c'est le dernier !

Pour finir, toute ma reconnaissance va à **Guy** qui ne cesse de me soutenir, de m'écouter et d'être fier de moi.

Notes préliminaires

Ce travail de master est essentiellement de nature juridique mais n'engage en rien la responsabilité de son auteur. Solliciter l'avis d'experts en droit en cas de question reste conseillé, cette étude ayant pour vocation de résumer et clarifier des enjeux légaux complexes et non pas d'y apporter des réponses sur le plan juridique.

Malheureusement, il est nécessaire de préciser également que ce mémoire n'a pas bénéficié d'un réflexe d'écriture inclusive, encore à acquérir. La forme masculine a ainsi été adoptée dans le texte pour parler indifféremment de l'homme ou de la femme.



Mon mémoire est mis à disposition sous licence [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/), à l'exception des parties du texte où il est fait mention d'une autre licence, notamment l'annexe 8.

Résumé

L'ouverture des données de la recherche, voulue par le mouvement de l'Open Science, implique d'en autoriser l'accès et de les partager, souvent dans un dépôt de données. Dans ce contexte, DLDM, qui a développé le dépôt Yareta pour le canton de Genève, souhaite connaître les enjeux légaux liés au partage des données, notamment dans les domaines du droit d'auteur, de la protection des données personnelles et des licences. Afin de répondre à ce mandat, une revue de la littérature spécialisée et une étude des pratiques encouragées ont été réalisées. Dans le but de garantir au mieux l'exactitude des propos avancés dans la partie légale de ce travail, des juristes, des avocats et d'autres professionnels du milieu académique ont ensuite été consultés par la suite.

Les principales conclusions de ce travail sont les suivantes :

- Les données scientifiques brutes ne sont pas protégées par le droit d'auteur mais d'autres peuvent l'être, et à plusieurs niveaux. Une analyse au cas par cas est nécessaire à chaque fois pour identifier ce qui est protégé (tant pour ses propres données que pour celles de tiers qui sont réutilisées) et qui détient les droits sur celles-ci. En effet, il est courant que les rapports de travail entre un chercheur et l'institution qui l'emploie prévoient une cession (partielle ou non) des droits d'auteur.
- Concernant la protection des données personnelles, celle-ci est réglementée à plusieurs niveaux (cantonal, national, européen), qui s'appliquent diversement selon le contexte de la recherche et édictent des dispositions différentes, rendant cet aspect très complexe. Dans tous les cas, deux éléments constituent des bonnes pratiques en la matière et sont encouragés : le recueillement d'un consentement éclairé à une granularité fine (comprenant l'archivage et la publication) et l'anonymisation des données personnelles et/ou sensibles lors de leur publication.
- En termes de licences, le domaine public est revendiqué pour la mise à disposition des données de la recherche, car c'est le seul qui ne pose pas de problèmes d'interopérabilité et qui lève toutes les restrictions légales qui empêchent la pleine réutilisation des données. En l'occurrence, la licence CC0 est celle qui fait consensus.

Ce mémoire porte sur des considérations larges qui ne concernent pas que le cas particulier de Yareta, que cela soit au niveau du cadre juridique ou des licences pour les données. Il peut ainsi apporter des réponses à d'autres institutions qui se poseraient des questions similaires. De plus, ce travail fournit en annexe un résumé des points essentiels sur le droit d'auteur et sur le choix d'une licence pour les données sous la forme de fiches informatives et d'un arbre de décision. Disponibles sous licence CC-BY (<https://doi.org/10.5281/zenodo.3967402>), ces documents peuvent être adaptés aux besoins de n'importe quelle équipe de soutien à la recherche en Suisse.

Mots-clefs : Open Science ; Open Research Data, données de recherche ; cadre juridique ; législation ; droit d'auteur ; protection des données ; licences ; Creative Commons ; dépôt de données ; Repository ; soutien à la recherche.

Table des matières

Remerciements	i
Notes préliminaires	ii
Résumé	iii
Table des matières	iv
Liste des tableaux	vii
Liste des figures	viii
Liste des abréviations utilisées	ix
1. Introduction	1
1.1 Présentation du mandat	1
1.2 Objectifs	1
1.3 Méthodologie	2
2. L'Open Science (la « science ouverte »)	4
2.1 L'Open Access (l'accès libre aux publications)	4
2.2 L'Open Research Data (l'accès libre aux données de la recherche)	4
2.3 Le dépôt de données Yareta de DLCM	7
2.3.1 Choix de licences offert dans la version 1.10 de mai 2020	8
3. L'ouverture des données de la recherche en Suisse	10
3.1 Exigences des principaux bailleurs de fonds	10
3.1.1 Swissuniversities	10
3.1.2 Le Fonds national suisse	10
3.1.3 La Commission européenne	12
3.2 Aperçu de la situation dans les hautes écoles suisses	13
4. Cadre juridique	15
4.1 Droit d'auteur	15
4.1.1 Loi sur le droit d'auteur – LDA (CH)	15
4.1.1.1 Notions d'œuvre, d'œuvre scientifique et d'auteur	15
4.1.1.2 Principe de territorialité	16
4.1.1.3 Champ d'application : quid des données ?	16
4.1.1.4 Droits conférés à l'auteur	17
4.1.1.5 Titularité des droits d'auteur	17
4.1.1.5.1 La propriété des données : un concept erroné	18
4.1.1.6 Utilisation d'œuvres protégées et régime d'exceptions	18
4.2 Protection des données personnelles	20
4.2.1 Loi sur la protection des données personnelles – LPD (CH)	21
4.2.1.1 Révision totale en cours	21
4.2.1.2 Champ d'application	21
4.2.1.3 Principes	22
4.2.1.4 Consentement	22
4.2.1.5 Données sensibles	23

4.2.1.6	Déclaration des fichiers de données personnelles.....	23
4.2.1.7	Devoir d'information	23
4.2.1.8	Droits de la personne concernée	23
4.2.1.9	Sous-traitance et communication de données	24
4.2.1.10	Règlementation spécifique sur le traitement de données à des fins de recherche	24
4.2.2	Loi sur l'information du public, l'accès aux documents et la protection des données personnelles – LIPAD (GE).....	25
4.2.2.1	Champ d'application	26
4.2.2.2	Le consentement.....	26
4.2.2.3	Déclaration des fichiers de données personnelles.....	26
4.2.2.4	Droits conférés à la personne concernée	27
4.2.2.5	Disposition spécifique sur le traitement de données à des fins de recherche	27
4.2.3	Règlement général sur la protection des données – RGPD (UE)	28
4.2.3.1	Champ d'application	28
4.2.3.2	Principes.....	29
4.2.3.3	Consentement	29
4.2.3.4	Données sensibles	30
4.2.3.5	Registre des traitements	30
4.2.3.6	Data Protection Officer (DPO)	30
4.2.3.7	Privacy Impact Assessment (PIA)	31
4.2.3.8	Droits conférés à la personne concernée	31
4.2.3.9	Sous-traitance et communication de données	32
4.2.3.10	Devoir d'information	32
4.2.3.11	Exceptions à des fins de recherche scientifique.....	32
4.2.4	Quelle loi de protection des données personnelles s'applique lorsqu'on fait de la recherche ?.....	34
4.2.5	Partager des jeux de données qui contiennent des données personnelles ou sensibles	35
4.3	Résumé du cadre juridique lié aux données de la recherche	35
5.	Licences	39
5.1	Enjeux	39
5.1.1	Pourquoi publier ses données sous licence ?.....	39
5.1.2	Des licences ouvertes pour des données ouvertes	40
5.1.2.1	Les licences non ouvertes.....	41
5.1.2.1.1	Les licences interdisant l'usage commercial	41
5.1.2.1.2	Les licences interdisant la modification	42
5.1.2.2	Les licences ouvertes mais néanmoins problématiques.....	43
5.1.2.2.1	Les licences demandant une attribution	43
5.1.2.2.2	Les licences de partage à l'identique (copyleft)	43
5.2	Les licences ouvertes pour les données de la recherche	44
5.2.1	Les licences Creative Commons	44
5.2.1.1	Licence CC-BY	47
5.2.1.2	Licence CC-BY-SA	48
5.2.1.3	Licence CC0 (Public Domain Dedication)	48
5.2.1.4	Label Public Domain Mark	49
5.2.2	Les licences Open Data Commons	50
5.2.2.1	ODC Attribution License (ODC-By)	52

5.2.2.2	ODC Open Database License (ODC-ODbL).....	53
5.2.2.3	Public Domain Dedication and License (PDDL).....	54
5.2.3	Les licences Community Data License Agreement.....	54
5.2.3.1	CDLA-Permissive-1.0.....	55
5.2.3.2	CDLA-Sharing-1.0.....	56
5.2.4	Le domaine public comme solution à l'incompatibilité des licences	56
6.	Recommandations pour Yareta.....	59
7.	Des fiches d'information et un arbre de décision à destination des chercheurs	61
8.	Conclusion	63
8.1	Difficultés rencontrées	64
8.2	Perspectives offertes par le travail	64
	Bibliographie	66
Annexe 1 :	Principes FAIR pour les données	78
Annexe 2 :	Aperçu de la situation Open Science dans les hautes écoles suisses.....	79
Annexe 3 :	Définitions liées à la LPD.....	81
Annexe 4 :	Définitions liées à la LIPAD	83
Annexe 5 :	Définitions liées au RGPD	84
Annexe 6 :	Qui peut utiliser du contenu sous licence Creative Commons de type NC et dans quel cas ?	86
Annexe 7 :	Compatibilité des licences Creative Commons entre elles	87
Annexe 8 :	Tableau comparatif des licences ouvertes pour les données.	88
Annexe 9 :	Avantages et inconvénients des licences ouvertes pour les données	92
Annexe 10 :	Fiches info sur le droit d'auteur et les licences pour les données	94
Annexe 11 :	Arbre de décision pour le choix d'une licence pour les données	99

Liste des tableaux

Tableau 1 : Bénéfices et craintes liées à l'ouverture des données de la recherche	6
Tableau 2 : Droit applicable aux données de la recherche	36
Tableau 3 : Les quatre éléments des licences CC.....	46
Tableau 4 : Les six licences CC	46
Tableau 5 : La licence CC0	49
Tableau 6 : Le label Public Domain Mark	49
Tableau 7 : Différence entre la licence CC0 et le label Public Domain Mark.....	50
Tableau 8 : Droits octroyés et obligations de la licence ODC-By	52
Tableau 9 : Droits octroyés et obligations de la licence ODC-ODbL	53
Tableau 10 : Droits octroyés et obligations de la licence CDLA-Permissive	55
Tableau 11 : Droits octroyés et obligations de la licence CDLA-Sharing.....	56

Liste des figures

Figure 1 : Licences activées dans Yareta (juin 2020)	8
Figure 2 : Bulle d'aide du champ licence dans Yareta (juin 2020).....	9
Figure 3 : Exigences Open Research Data du FNS et de H2020	13
Figure 4 : Raisons valables d'exception à l'obligation de publier les données du FNS et de H2020	13

Liste des abréviations utilisées

ARK	Archival Resource Key
CC	Creative Commons (licences)
CC0	CC Zero (Public Domain Dedication)
CC-BY	CC Attribution
CC-BY-SA	CC Attribution-ShareAlike
CC-BY-ND	CC Attribution-NoDerivatives
CC-BY-NC	CC Attribution-Noncommercial
CC-BY-NC-SA	CC Attribution- Noncommercial- ShareAlike
CC-BY-NC-ND	CC Attribution-Noncommercial-NoDerivatives
CNIL	Commission Nationale de l'Informatique et des Libertés (France)
DLCM	Data Life-Cycle Management (projet)
DMP	Data management plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
DRM	Digital Rights Management
EPF	Ecole polytechnique fédérale
EPFL	Ecole polytechnique fédérale de Lausanne
ETHZ	Eidgenössische Technische Hochschule Zürich
FAIR (principes)	Findable, accessible, interoperable, reusable (trouvable, accessible, interopérable, réutilisable)
FNS	Fonds national suisse
HES-SO	Haute école spécialisée de Suisse occidentale
H2020	Horizon 2020
IHEID	Institut des hautes études internationales et du développement
LDA	Loi fédérale sur le droit d'auteur
LIPAD	Loi sur l'information du public, l'accès aux documents et la protection des données personnelles (GE)
LHES-SO	Loi sur la Haute école spécialisée de Suisse occidentale - Genève
LPD	Loi fédérale sur la protection des données

LRH	Loi fédérale sur la recherche sur l'être humain
LU	Loi sur l'Université (GE)
NC	Non commercial (clause interdisant l'utilisation commerciale dans une licence)
ND	No derivatives (clause interdisant la modification dans une licence)
OA	Open Access (accès ouvert)
ODC	Open Data Commons (licences)
ODC-By	Open Data Commons Attribution License
ODC-ODbL	Open Data Commons Open Database License
OLPD	Ordonnance relative à la loi fédérale sur la protection des données
ORD	Open Research Data (données de recherche ouvertes)
OS	Open Science (science ouverte)
PIA	Privacy Impact Assessment (étude d'impact sur la vie privée)
PDDL	Public Domain Dedication and License (de Open Data Commons)
PPDT	Préposé à la protection des données et à la transparence (cantons)
PFPDT	Préposé fédéral à la protection des données et à la transparence
RDM	Research data management (gestion des données de la recherche)
RIPAD	Règlement d'application de la loi sur l'information du public, l'accès aux documents et la protection des données personnelles (GE)
RGPD	Règlement général sur la protection des données
SA	Share-alike (clause imposant de partager l'œuvre dérivée sous la même licence que l'œuvre originale)
SEFRI	Secrétariat d'État à la formation, à la recherche et à l'innovation
TDM	Text and data mining (fouilles de textes et de données)
UE	Union européenne
UNIGE	Université de Genève
UNIL	Université de Lausanne
URI	Uniform Resource Identifier
ZHAW	Zürcher Hochschule für Angewandte Wissenschaften

1. Introduction

La gestion des données de la recherche est une problématique actuelle dans le contexte de l'ouverture de la science, tant au niveau des instances qui financent cette recherche en Suisse que des hautes écoles qui l'effectuent. Des exigences nouvelles émergent, occasionnant l'émission de directives ou l'encouragement de bonnes pratiques de gestion des données, notamment quant à leur conservation et leur partage pour réutilisation. Dans ce contexte, de plus en plus de dépôts pour héberger les données de la recherche sont développés dans notre pays, que ce soit au niveau national, institutionnel ou spécialisé. Ce travail est un mandat de l'un d'entre eux.

1.1 Présentation du mandat

Cette étude a été réalisée sur mandat de Pierre-Yves Burgi, directeur du projet DLCM (Data Life-Cycle Management) et directeur adjoint à la division du Système d'Information de l'Université de Genève (UNIGE). DLCM est un consortium suisse créé en 2015 dont le but est d'accompagner les chercheurs dans la gestion de leurs données (Favre, Hahling 2020). Une de ses récentes réalisations est Yareta, une archive pour conserver et partager les données de la recherche développée par l'Université de Genève et destinée aux hautes écoles du canton.

Jusqu'à présent, le développement de tels dépôts de conservation de données était surtout entouré de beaucoup d'enjeux techniques. Mais l'introduction du Règlement général sur la protection des données européen (RGPD), la révision toujours en cours de notre Loi fédérale sur la protection des données (LPD), l'entrée en vigueur de la récente révision de la Loi fédérale sur le droit d'auteur (LDA) et l'obligation de réaliser un Data Management Plan (DMP) imposée par de plus en plus de bailleurs de fonds impliquent d'avoir une bonne compréhension également du contexte légal de la gestion des données, et pas seulement des aspects techniques, afin de réduire les risques judiciaires. Or, comme le cadre juridique des données de la recherche est loin d'être aisé à saisir, un besoin se ressent d'identifier les principales questions légales qui se posent et de les clarifier afin d'aider au mieux les chercheurs dans ce domaine (DMLawTool 2019). DLCM, dans le contexte de la mise à disposition de Yareta, souhaitait une étude dédiée à ces questions, ainsi qu'à celles des licences pour les données, qui prend donc la forme de ce travail de master.

1.2 Objectifs

Après une mise en contexte de l'ouverture des données de la recherche et de ses exigences en Suisse, le premier objectif de ce mémoire est de rendre compte de leur cadre juridique, complexe et souvent un obstacle à cette ouverture (Popkin 2019). Puisque les données de la recherche ne sont pas mentionnées explicitement dans les textes juridiques, les différents niveaux de protection qui peuvent les concerner sont loin d'être aisés à comprendre. Si une personne responsable d'un projet de recherche a un doute à leur propos, elle préférera probablement, par précaution, restreindre l'accès aux jeux de données liées à la recherche par crainte d'enfreindre la loi, limitant donc leur diffusion et réutilisation, ce qui est en opposition avec le mouvement pour l'ouverture des données et qu'il faudrait éviter. Au niveau légal, les données de la recherche sont principalement concernées par le droit d'auteur et la protection des données personnelles, ce sont donc les aspects légaux qui sont étudiés dans ce mémoire.

Son second objectif est de déterminer les licences les plus adaptées aux données de la recherche, ce qui concerne toujours le droit d'auteur. En effet, afin d'être réutilisées par des tiers, les données doivent être rendues disponibles par leur auteur sous une licence qui en explicite les conditions de réutilisation. Les licences détaillent donc, s'il y en a, les droits d'auteur qui restent protégés et ceux qui, au contraire, sont concédés. Il existe un nombre important de licences et toutes ne sont pas adaptées aux données, et encore moins aux données ouvertes, rendant la sélection difficile pour le chercheur. En réalité, les spécificités des différentes licences disponibles sont souvent complexes et mal connues, même pour celles largement utilisées comme les Creative Commons, occasionnant parfois des choix qui vont à l'encontre des principes encadrant l'ouverture des données.

Il est donc nécessaire d'expliquer à la fois le cadre juridique des données de la recherche et les licences les plus adaptées à leur partage, ce à quoi se consacre ce travail. Ses recommandations sont principalement adressées à Yareta, le dépôt de DLCM pour la conservation à long terme des données de la recherche, mais elles peuvent également être pertinentes pour d'autres dépôts du même type, ou simplement pour les personnes œuvrant au sein d'institutions universitaires au soutien à la gestion des données de la recherche.

1.3 Méthodologie

Les moyens mis en œuvre pour la réalisation de ce mandat, de nature plutôt théorique, sont une revue de la littérature et des pratiques, ainsi que la sollicitation d'experts sur les questions juridiques difficiles à résoudre ou comprendre.

L'importante revue de la littérature spécialisée, réalisée en premier, concernait les sujets du développement de l'Open Science en Suisse, du droit d'auteur, de la protection des données personnelles et des licences, étudiés sous l'angle des données de la recherche. Les pratiques encouragées au sein des hautes écoles et de la communauté des données ouvertes ont aussi été étudiées notamment pour la question des licences. Le besoin de se reposer sur de la littérature spécialisée en droit a été très fort car c'est un sujet d'étude complexe pour ceux qui y sont étrangers. Concrètement, les recherches ont été faites dans des bases de données spécialisées en droit (Swisslex) ou en sciences de l'information (LISA, LISTA), tout comme dans des archives institutionnelles (Archive Ouverte de l'UNIGE, ArODES) ou des moteurs de recherche d'articles académiques (Google Scholar). Bien entendu, le moteur de recherche Google a aussi été très utile pour trouver de la littérature pertinente, et pour identifier tous les sites web des hautes écoles en Suisse. Après l'identification d'articles pertinents, leurs bibliographies ont été attentivement consultées à la recherche d'autres écrits importants du domaine, permettant de constituer le corpus de référence qui compose la bibliographie de ce travail. Après la lecture de tous les textes et dans le but de garantir au mieux l'exactitude des propos avancés, la précieuse aide de juristes, d'avocats et d'autres professionnels du milieu académique a été sollicitée. Ils ont accepté de partager leur expertise du sujet et d'éclairer les nombreuses interrogations qui restaient sans réponse. Ces contacts ont été réalisés par email, téléphone ou visioconférence, les rencontres en personne étant impossibles avec la crise sanitaire que nous avons vécu en cette année 2020.

Concernant l'agencement des parties du mémoire, il a été fait le choix de procéder du général au particulier. Soit de commencer par contextualiser l'ouverture des données de la recherche et présenter sa mise en œuvre en Suisse. Ensuite seulement, il a semblé logique de définir le cadre juridique des données de la recherche en termes de droit d'auteur et de protection des données personnelles. La question des licences pour les données, elle, est abordée en dernier

car elle repose sur le droit d'auteur. En effet, lorsqu'il est fait mention de licences dans le milieu suisse de la recherche, c'est de licences de droit d'auteur qu'il s'agit. A cet égard, il a été choisi dans ce travail d'aborder et de présenter seulement les licences ouvertes pour les données car c'est celles qui sont recommandées dans la littérature et par les services de soutien à la gestion des données de recherche d'institutions académiques.

Le résultat de la démarche de recherche adoptée est un mémoire qui regroupe et présente les principaux enjeux légaux du domaine des données de la recherche en termes de droit d'auteur, de protection des données personnelles et de licences. C'est à ma connaissance le premier travail (en français) qui réunit étude de textes législatifs suisses et européen entièrement sous l'angle des données de la recherche et étude des licences pour les données, ce qui en fait son originalité.

En annexe du mémoire sont mises à disposition des fiches information sur le droit d'auteur et les licences, ainsi qu'un arbre décisionnel pour le choix de la licence lors de la publication. Leur conception a été la dernière étape de ce travail, après la rédaction. Ces documents, plus pratiques et concrets que le mémoire lui-même, résument ses principales conclusions et ont été conçus pour aider les chercheurs qui font face à de telles questions légales.

2. L'Open Science (la « science ouverte »)

L'Open Science est un terme qui regroupe de nombreux mouvements, visant tous à rendre la recherche scientifique disponible au plus grand nombre en abattant les obstacles à son partage (Swiss Academies of Arts and Science 2019) afin de la faire progresser. C'est une nouvelle approche de la conception du savoir, qui prône donc le partage des connaissances acquises, mais aussi la transparence sur la méthode de recherche utilisée, l'évaluation de sa rigueur ou encore la reproductibilité des résultats obtenus. Ouvrir la science permet donc d'améliorer sa qualité et de la rendre plus inclusive et intègre (Université de Lausanne 2019a). Les deux mouvements de l'Open Science les plus implantés à ce jour sont l'Open Access (le libre accès aux publications scientifiques) et l'Open (Research) Data (l'ouverture des données, dont celles de la recherche).

2.1 L'Open Access (l'accès libre aux publications)

Une considération sous-jacente importante de l'Open Access (OA) est le fait que la recherche scientifique est financée essentiellement par des fonds publics. Or, le système qui prévaut encore aujourd'hui veut que les Etats qui financent des projets de recherche payent une nouvelle fois des abonnements ou des licences d'accès pour consulter les résultats de ces recherches, traditionnellement publiés chez des éditeurs scientifiques commerciaux. L'Open Access pour les publications scientifiques a été la première réponse à cet état de fait et s'est depuis largement répandu dans la recherche scientifique internationale. Ce mouvement veut que les publications scientifiques soient librement accessibles. Deux moyens sont principalement utilisés pour atteindre cet objectif : la publication de l'article directement en accès libre auprès d'un éditeur (il n'y aura donc pas de restriction à sa lecture) ; ou bien le dépôt du manuscrit dans un répertoire de publications avant ou après sa révision auprès d'un comité de lecture par les pairs (*peer review*). En Suisse, le Fonds national suisse (FNS) oblige depuis le début de l'année 2020 tous les chercheurs des projets qu'il finance à publier leurs résultats en OA, tandis que de son côté Swissuniversities a fixé 2024 pour atteindre le même objectif. De plus, les principales hautes écoles de Suisse ont toutes développé un dépôt institutionnel pour archiver les publications en libre accès de leurs chercheurs.

2.2 L'Open Research Data (l'accès libre aux données de la recherche)

Dans la continuité de cette politique d'ouverture, le second aspect qui est en train d'être mis en œuvre dans le monde de la science est la disponibilité des données de la recherche avec le mouvement de l'Open Research Data (ORD).

Pour commencer, il faut expliciter ce qui est entendu par « données de la recherche ». Une des difficultés inhérentes à leur définition est leur hétérogénéité. Au vu de la multitude des domaines de recherche scientifique et de leurs différences, les données de la recherche se distinguent par la diversité de leurs objets, « la manière dont elles ont été obtenues, la forme sous laquelle elles ont été enregistrées et le traitement qu'elles ont subi pour être formatées, lisibles » (Gaillard 2014). L'Organisation de coopération et de développement économiques (OCDE) a proposé une définition en 2007 qui fait depuis consensus : les données de la recherche sont « des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche » (OCDE 2007). De manière plus large, il est possible de dire que ce sont « des

éléments factuels nécessaires à la validation des résultats de recherche, ces derniers étant habituellement délivrés sous la forme de publications » (Marie-Madeleine Géroudet cité par Gaillard 2014).

Il existe plusieurs types de données de la recherche. Dans un premier temps, il est possible de dissocier la donnée brute de la donnée traitée ou nettoyée (Gaillard 2014 ; Schneider 2018). Ensuite, traditionnellement, six types de données sont distingués : les données d'observation, les données expérimentales, les données computationnelles ou de simulation (Gaillard 2014), les données de digitalisation, les données de référence (par exemple des sources textuelles primaires) et encore les données dérivées (à partir d'autres données ou compilées) (Schneider 2018).

Un jeu de données (un *dataset*) est pour sa part « l'agrégation, sous une forme lisible, de données (...) présentant une certaine 'unité', rassemblée pour former un ensemble cohérent » (Gaillard 2014). Pour les données de la recherche, c'est donc « un enregistrement de données sous la forme d'un ou plusieurs fichiers électroniques, téléchargeables, citables et intelligibles – ce jeu étant accompagné de métadonnées descriptives suffisantes » (Gaillard 2014).

Quant au concept de données ouvertes, il n'existe pas de définition unanime à leur propos. Selon l'*Open Definition* (un projet de l'Open Knowledge Foundation), les données ouvertes peuvent être réutilisées librement, redistribuées, modifiées, séparées ou encore compilées à toutes fins et gratuitement. Les seules restrictions acceptées sont celles de l'obligation de l'attribution, du partage à l'identique ou encore l'interdiction des restrictions techniques lors de la distribution de l'œuvre (Open Definition 2015).

Ouvrir l'accès aux données de la recherche et encourager leur réutilisation remplit plusieurs buts et possède de nombreux avantages, mais est aussi l'objet de craintes justifiées de la part des chercheurs :

Tableau 1 : Bénéfices et craintes liées à l'ouverture des données de la recherche

Bénéfices de l'Open Research Data	Craintes des chercheurs
<p>Transparence et reproductibilité : l'ouverture des données améliore la qualité et l'intégrité des travaux scientifiques, notamment en permettant la validation et la reproductibilité des résultats de la recherche présentés dans la publication scientifique (Swiss Academies of Arts and Science 2019 ; Gaillard 2014, p.51 ; Bagnoud et al. 2019).</p> <p>Innovation et impact : la recherche scientifique est renforcée (Cassella 2013) et son impact social et économique est augmenté (Ball 2014). L'ouverture de la science encourage les nouvelles collaborations entre chercheurs (Popkin 2019), la diversité des analyses et des opinions, la recherche transversale et multidisciplinaire, le test de nouvelles hypothèses de recherche, l'exploration de domaines novateurs ou encore la création de nouveaux jeux de données par combinaison de ceux existants (Cassella 2013).</p> <p>Visibilité : celle des chercheurs et des institutions de recherche est augmentée avec les données et les publications ouvertes, puisqu'elles « sont davantage téléchargées, lues et partagées » (Bagnoud et al. 2019). Les données peuvent aussi être citées dans une liste de réalisations ou un CV (Popkin 2019). Mettre à disposition les données de recherche a aussi pour conséquence une augmentation de la citation des publications : une étude sur 531'899 articles publiés par les éditeurs OA PLOS et Biomed Central a montré que les scientifiques qui partagent leurs données dans des dépôts étaient associés à une augmentation des citations de leurs articles allant jusqu'à 25,36% (Colavizza et al. 2020).</p>	<p>Manque de temps et de ressources : avec la mise à disposition des jeux de données, il y a une peur que cela devienne très chronophage s'il faut répondre à des demandes d'information liées aux données publiées, par exemple fournir des explications supplémentaires, des outils analytiques, des métadonnées, etc. Aussi, le manque d'expérience en matière de gestion des données de la recherche est pointé du doigt, occasionnant de devoir consacrer plus de temps à cette tâche car les ressources n'augmentent pas en conséquence. (Griffiths 2009)</p> <p>Considérations légales et éthiques : il n'est pas aisé de savoir qui est titulaire des droits d'auteur sur les données, s'il y en a, et si celles-ci peuvent légalement être publiées. Dans les disciplines de recherche où des données personnelles sont collectées, il y a également des exigences en termes de confidentialité et de protection, assez complexes à identifier. Il y a donc une peur de violer la loi en publiant des données, ce qui pourrait avoir de graves conséquences pour le chercheur. (Griffiths 2009)</p> <p>Concurrence et peur de perte de contrôle sur les données : beaucoup de scientifiques désirent conserver l'usage exclusif des données qu'ils ont collectées, souvent après des années de recherche, afin d'en extraire toute la valeur de publication qu'ils peuvent. De plus, certains craignent également que leurs données soient déformées ou mal utilisées par la suite, menant à des conclusions erronées. (Griffiths 2009)</p> <p>Manque de reconnaissance : dans un système où les chercheurs sont évalués principalement en fonction du nombre de leurs publications dans des revues à fort impact, il n'existe pas encore vraiment de reconnaissance liée à la carrière pour la publication d'ensembles de données. (Griffiths 2009)</p>

Du côté des craintes des chercheurs, ce travail tentera donc d'éclaircir le cadre légal en Suisse pour montrer qu'il est possible de publier ses données sans risques lorsqu'on est bien renseigné sur les conditions légales qui s'appliquent.

Une des autres réticences formulées à l'encontre de l'ouverture des données concerne le manque d'expérience dans la gestion des données de la recherche, qui la rendrait chronophage. Il est vrai que cette notion de « bonne gestion » des données reste un concept assez vague et très récent. En 2016, afin de mieux le définir et d'édicter des lignes directrices précises, un consortium de scientifiques et d'organisations a rédigé un article intitulé « The FAIR Guiding Principles for scientific data management and stewardship » (Wilkinson et al. 2016). Depuis, cet article est devenu la référence en la matière et les principes qu'il énonce sont des standards de la bonne gestion des données. L'abréviation FAIR est celle de *Findable, Accessible, Interoperable and Reusable* (trouvable, accessible, interopérable et réutilisable). Ces quatre principes sont déclinés en plusieurs exigences (le détail est disponible en annexe 1). Une gestion des données FAIR ne demande pas aux chercheurs de partager toutes leurs données sans aucune restriction, mais défend simplement l'application de bonnes pratiques dans la gestion des données en vue de leur partage et leur réutilisation, afin que les humains et les machines puissent les trouver, les interpréter et les utiliser sous des conditions clairement définies (Milzow et al. 2020).

Un autre outil de l'ouverture des données de la recherche est le Data Management Plan (DMP), ou plan de gestion des données, exigé par de plus en plus de bailleurs de fonds. Il est considéré comme « l'instrument naturel de la "FAIRification de la recherche" » (Université de Lausanne 2019b). Ce document est le fil rouge de la gestion des données lors d'un projet de recherche, aidant à planifier celle-ci pendant tout le cycle de vie de la donnée, soit de la création, à la collecte, la documentation, la description, au partage et à la conservation. Le DMP est un document évolutif, qui doit être rédigé au début du projet et mis à jour tout au long de la durée de celui-ci. Il n'en n'existe pas qu'un seul modèle, mais ils abordent toujours les mêmes thèmes. Parmi les sujets à détailler, le DMP demande d'explicitier les enjeux juridiques liés aux données en vue de leur utilisation ou réutilisation, par exemple de décrire la situation en termes de « restriction légale, propriété des données, propriété intellectuelle, obligations contractuelles [ou] données sensibles » (Université de Lausanne 2019b).

Afin de parvenir à l'ouverture des données, il faut également une infrastructure qui le permette. En l'occurrence, les dépôts jouent un rôle crucial dans la visibilité et la réutilisation des données de la recherche car ils associent un identifiant pérenne à chaque jeu de données, ce qui permet de le citer. Ce sont également des infrastructures qui garantissent la conservation des données à moyen ou long terme, bien plus adaptées que la publication sur le site web du projet de recherche ou en annexe d'un article scientifique. Les dépôts de données peuvent être spécialisés dans une discipline ou généralistes, et liés à une institution ou non (Schneider 2018).

2.3 Le dépôt de données Yareta de DLCM

La première phase du projet DLCM (Data Life-Cycle Management) a commencé en août 2015 avec un financement de swissuniversities. Les institutions partenaires étaient au nombre de huit : l'UNIGE, la HES-SO, l'EPFL, l'ETHZ, l'Université de Lausanne, l'Université de Bâle, l'Université de Zurich et SWITCH (DLCM 2015). Tant des spécialistes en information documentaire que des informaticiens et des chercheurs ont collaboré à ce projet (Burgi, Blumer 2018). La seconde phase, DLCM 2.0, a elle commencé le 1^{er} janvier 2019 et se terminera en décembre 2020 (swissuniversities 2019). Les partenaires sont cette fois l'UNIGE, la HES-SO et la Zürcher Hochschule für Angewandte Wissenschaften (ZHAW).

En parallèle du projet DLCM, dans le canton de Genève, il existe depuis le 1^{er} janvier 2018 un projet de loi (n°12146) qui porte sur les infrastructures et les services numériques pour la recherche (Burgi 2019). Son but est de répondre au besoin d'évolution « des infrastructures et services associés de calcul à haute performance et de stockage long terme afin d'optimiser et faciliter l'utilisation des données issues de la recherche dans les hautes écoles universitaires genevoises » (Burgi 2019). Un des quatre objectifs du projet de loi est la mise en place d'« une architecture de stockage sécurisée construite sur les standards internationaux permettant la conservation à moyen et long terme des données scientifiques » (Burgi 2019). Cet objectif a été concrétisé par la création de Yareta, une archive pour les données de la recherche, issue donc des projets DLCM et de loi cantonale (Université de Genève 2017). Sur volonté du Département de l'instruction publique, les infrastructures et les services issus du projet de loi, mutualisés et gérés par l'UNIGE, sont également mis à la disposition de la HES-SO Genève et de l'IHEID, dont les chercheurs peuvent ainsi profiter de Yareta (Burgi 2019). Un autre dépôt de données à préservation à long terme, OLOS, est actuellement en développement par l'équipe de DLCM afin d'être proposé au niveau national à toutes les hautes écoles.

Respectant les principes FAIR, construit sur la base de logiciels libres, basé sur une architecture OAIS (*Open archival information system*) et offrant des APIs (*application program interface*), Yareta a été mis à disposition de la communauté scientifique genevoise en juin 2019. C'est une solution suisse, non-commerciale, compatible avec tous les formats de données, conforme aux standards internationaux pour l'interopérabilité des données (avec les normes OAIS, DOI, OAI-PMH, PREMIS ou Datacite par exemple) et généraliste quant aux types de données acceptés (Burgi 2019).

2.3.1 Choix de licences offert dans la version 1.10 de mai 2020

Yareta permet de choisir parmi trois types d'accès lors du dépôt des données : public (ouvert), restreint (à un groupe de personnes autorisées) et fermé (il faut adresser une demande d'accès à l'auteur du jeu de données le cas échéant) (Burgi, Cazeaux, Mengelle 2019). Le champ de la licence est obligatoire quand un jeu de données est déposé en accès public. Dans le cas d'un accès restreint ou fermé, il est optionnel. Actuellement, les six licences Creative Commons et la licence de domaine public CC0 sont proposées lors d'un dépôt en accès ouvert, comme le montre la capture d'écran ci-dessous :

Figure 1 : Licences activées dans Yareta (juin 2020)

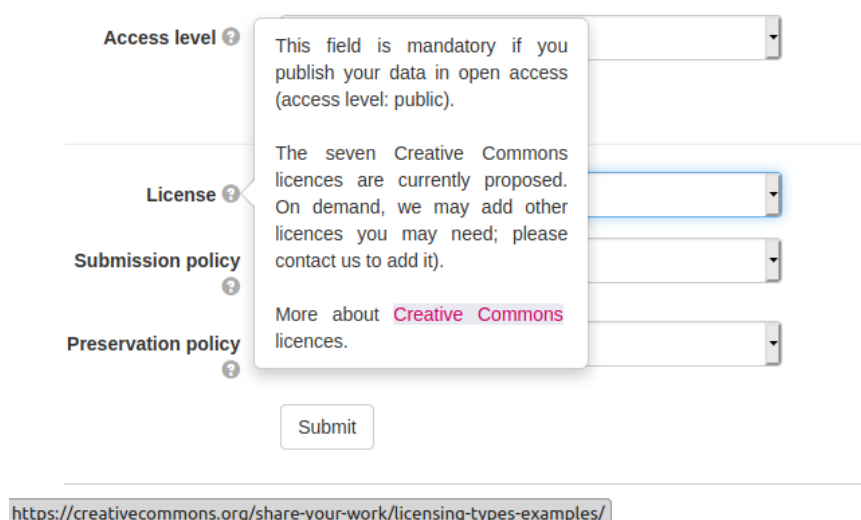
The screenshot shows the Yareta deposit form with the following fields and options:

- Language**: A dropdown menu.
- Collection begin**: A dropdown menu.
- Collection end**: A dropdown menu.
- Access level**: A dropdown menu.
- License**: A dropdown menu with the following options:
 - CC BY 4.0 (Creative Commons Attribution 4.0 International)
 - CC BY-NC 4.0 (Creative Commons Attribution-NonCommercial 4.0 International)
 - CC BY-NC-ND 4.0 (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International)
 - CC BY-NC-SA 4.0 (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International)
 - CC BY-ND 4.0 (Creative Commons Attribution-NoDerivatives 4.0 International)
 - CC BY-SA 4.0 (Creative Commons Attribution-ShareAlike 4.0 International)
 - CC0 1.0 Universal (Creative Commons Public Domain Dedication)
- Submission policy**: A dropdown menu with the option "Deposit without approval".
- Preservation policy**: A dropdown menu with the option "Keep It For 10 Years".
- Submit**: A button.

Capture d'écran transmise par email par Hugues Cazeaux le 18 juin 2020.

Afin de faciliter la description du jeu de données par le chercheur et de l'aider dans cette tâche, la licence CC-BY 4.0 est automatiquement attribuée lorsque l'accès libre est sélectionné. Ce choix a été fait car la licence CC-BY est une licence ouverte, mais qui n'induit pas une renonciation à tous les droits. La prudence a ainsi conduit à cette décision, la seule condition de l'attribution étant jugée comme acceptable. Bien entendu, si la personne qui enregistre ses données en accès ouvert désire une autre licence que la CC-BY, elle peut modifier ce champ pré-rempli. De plus, chaque responsable d'unité organisationnelle¹ peut décider d'attribuer une autre licence CC par défaut lors des dépôts en accès libre au sein de son unité. Finalement, dans le cas où la personne responsable de l'unité ou le chercheur qui fait le dépôt désire une autre licence que les Creative Commons, il peut en faire la demande au support de Yareta, comme le montre l'image ci-dessous :

Figure 2 : Bulle d'aide du champ licence dans Yareta (juin 2020)



Capture d'écran transmise par email par Hugues Cazeaux le 18 juin 2020.

¹ Une unité organisationnelle peut représenter plusieurs types d'entité, par exemple un laboratoire de recherche, un projet ou simplement un espace de travail partagé entre des chercheurs. Au sein de cette unité organisationnelle, des rôles sont attribués et tous les membres n'ont pas les mêmes droits (Favre, Hahling 2020, p. 11, 13).

3. L'ouverture des données de la recherche en Suisse

Cette partie fait un bref état des exigences actuelles des principaux bailleurs de fonds de la recherche en Suisse. Il faut garder à l'esprit que cet état des lieux sera probablement rapidement incomplet ou obsolète, car les pratiques et l'encadrement de l'Open Research Data se développent de plus en plus et de manière très rapide.

3.1 Exigences des principaux bailleurs de fonds

3.1.1 Swissuniversities

Dans le domaine des données de la recherche, swissuniversities, organisation faîtière des hautes écoles suisses, a été mandatée par le SEFRI (Secrétariat d'État à la formation, à la recherche et à l'innovation) en octobre 2019 pour étendre la stratégie nationale Open Access de 2017 « aux données de la recherche en accès libre » (swissuniversities 2020a). Un groupe de projet a été ainsi formé pour concevoir, à destination de la délégation Open Science de swissuniversities, les documents suivants :

- Un rapport de base, qui présente la situation actuelle au niveau national en ce qui concerne les projets, les infrastructures et les investissements et qui fournit une vue d'ensemble des infrastructures internationales, des solutions et des normes selon les domaines de spécialisation (SEFRI 2019a) ;
- Une stratégie Open Research Data ;
- Un plan d'actions Open Research Data. (swissuniversities 2020b)

Les travaux d'élaboration de ces documents sont en cours et la stratégie devait entrer en vigueur en 2021 (FNS 2020a). Cependant, le calendrier prévu initialement risque d'être modifié en raison de la crise sanitaire du Covid-19 de cette année 2020 (swissuniversities 2020b).

De plus, un nouveau programme *Open Science* succédera au programme *P-5 Information scientifique* en 2021. Il « créera les conditions permettant aux hautes écoles suisses de réutiliser, de propager et de reproduire des données et méthodes de recherche par voie électronique » et est fondé sur les principes FAIR. Ce nouveau programme aura deux phases (swissuniversities 2020c) :

- Phase A, 2021-2024 : Open Access (avec un budget de 11,8 millions de francs)
- Phase B, 2022-2024 : Open Research Data (avec un budget prévu de 33,2 millions de francs).

Un projet de stratégie OS globale a été rédigé dans le cadre de la préparation du programme *Open Science*. Ce document est disponible en téléchargement² afin de « soutenir les hautes écoles dans leurs propres démarches stratégiques relatives à Open Science » (swissuniversities 2020d).

3.1.2 Le Fonds national suisse

Depuis octobre 2017, le FNS exige un DMP lors de la soumission d'une demande de fonds pour un projet de recherche (FNS 2020a). Il s'attend à ce que (FNS 2017a ; 2020b) :

² SWISSUNIVERSITIES, 2019. Swiss National Open Science Strategy : Version 2. *Swissuniversities* [en ligne]. 9 septembre 2019. [Consulté le 20 juin 2020]. Disponible à l'adresse : <https://bit.ly/2zM4EqW>

- Les données sur lesquelles se basent les publications scientifiques soient rendues disponibles, en tout cas au minimum celles nécessaires pour reproduire les résultats présentés. Les données publiées peuvent être brutes comme traitées ;
- Qu'elles soient publiées aussi rapidement que possible, mais au plus tard lors de la parution de la publication scientifique auxquelles elles correspondent ;
- Les jeux de données soient correctement décrits à l'aide de métadonnées afin de permettre à d'autres scientifiques de saisir « comment et dans quelles conditions les données ont été collectées et comment elles peuvent être réutilisées » ;
- Si la réutilisation des données exige des outils spécifiques, que cela soit expliqué et, si possible, que ces outils soient mis à disposition également ;
- Les données et les métadonnées soient archivées dans une archive publique et dans des formats ouverts ;
- Si ces exigences ne peuvent pas être remplies, que l'explication de cet empêchement figure dans le DMP (clauses légales, éthiques, de droit d'auteur, de confidentialités ou autres).

Le DMP doit être rempli directement dans la plateforme *mySNF* et il est considéré comme une déclaration d'intention. Il est ensuite passé en revue par le secrétariat du FNS en vue de l'acceptation ou non des fonds demandés pour la recherche, mais il ne participe pas au processus d'évaluation de celle-ci (FNS 2017a). Le DMP doit donc être plausible et respecter les directives ORD du FNS. En cas de manquements, une demande de révision sera adressée. Le DMP doit être mis à jour tout au long du projet. A la fin de celui-ci, la version finale doit être soumise au FNS et les liens vers les jeux de données doivent être mentionnés pour être publiés dans la base de données publique de recherche P3 du FNS (Milzow et al. 2020). Globalement, les résultats de l'introduction de l'obligation de soumettre un DMP sont satisfaisants. Un *monitoring* des DMPs soumis en 2017 et 2018 pour des projets acceptés a en effet conclu que « many DMPs were very extensive and precise on the management of research data and a majority of them could be accepted without requesting a revision from the applicants. » (Milzow et al. 2020)

Pour encourager et soutenir la gestion des données de la recherche, le FNS accorde un subside de maximum 10'000 francs pour les frais qui « se rapportent spécifiquement au traitement des données de recherche en vue de l'archivage et à l'archivage proprement dit dans des bases de données », à condition que ces dépôts respectent les principes FAIR et n'aient pas de vocation commerciale (FNS 2019). Afin d'aider l'identification de tels dépôts, le FNS a défini des critères minimaux à remplir (FNS 2017a) :

- Attribution d'identifiants pérennes (par exemple DOI ou ARK) ;
- Possibilité de charger « les métadonnées intrinsèques (p. ex. nom de l'auteur, contenu du set de données, publications associées, etc.) et celles définies par la personne soumettant les données (p. ex. définition des variables, etc.) » ;
- Présence d'une licence clairement mentionnée ou possibilité pour l'utilisateur de choisir ou charger une licence. Le FNS ne formule pas de recommandations ou obligations quant à la licence à choisir ;
- Accès public aux informations de citation et aux métadonnées du jeu de données (même si l'accès est restreint) ;
- Formulaire de soumission qui demande « que les métadonnées intrinsèques respectent un format spécifique (afin d'assurer leur utilisation automatique/interopérabilité) » ;

- Plan de préservation à long terme des données archivées.

Pour savoir si le dépôt est non commercial, le FNS renvoie à la notice de celui-ci dans la base de donnée re3data.org (Registry of research data repositories) (FNS 2017a).

3.1.3 La Commission européenne

La Suisse est associée au programme-cadre de recherche de l'Union européenne *Horizon 2020* (SEFRI 2019b). Lors de sa révision en 2017, H2020 a étendu l'*Open Research Data Pilot*, qu'il menait déjà dans le cadre de certaines thématiques, à toutes celles du programme. L'accès libre aux données de la recherche s'applique donc par défaut dans H2020, mais la Commission européenne reconnaît qu'il y a parfois de bonnes raisons de ne pas publier certaines données (Commission européenne 2016). Un retrait (*opt-out*) partiel ou total de l'*Open Research Data Pilot* est donc possible, dans les cas suivants (Commission européenne 2017a) :

« participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited

participation is incompatible with the need for confidentiality in connection with security issues

participation is incompatible with rules on protecting personal data

participation would mean that the project's main aim might not be achieved

the project will not generate / collect any research data or there are other legitimate reasons »

La participation au pilote ne faisant pas partie de l'évaluation du projet, les propositions à la recherche de subventions qui ont choisi de se retirer partiellement ou totalement ne sont pas pénalisées par rapport à d'autres (Commission européenne 2017a).

Le DMP est obligatoire pour tout projet et doit être soumis dans les six premiers mois de la recherche. (Commission européenne 2017b). Les données et les métadonnées nécessaires à la validation des publications (au minimum) doivent être rendues accessibles, idéalement dans un dépôt de données de la recherche, à part si des raisons légales, éthiques, liées à la sécurité ou autres ne l'empêchent (Horizon 2020 France 2017 ; Commission européenne 2017b). Les bénéficiaires du financement de H2020 sont aussi encouragés à mettre en place des mesures pour permettre à des tiers d'accéder, d'exploiter, de reproduire et de diffuser (gratuitement et pour tous) leurs données de la recherche. Dans ce but, l'utilisation de la licence CC-BY ou CC0 lors de la publication est plébiscitée. Lors du dépôt, des informations doivent être fournies sur le projet ainsi que sur les outils disponibles et nécessaires à la validation des résultats (par exemple des logiciels, du code logiciel, des algorithmes ou des protocoles d'analyse). Dans la mesure du possible, ces instruments doivent être fournis avec les données également (Commission européenne 2017a). Certains coûts engendrés par le fait de rendre les données de la recherche accessibles peuvent être remboursés pendant la durée du projet (Commission européenne 2017b).

En 2021, le programme-cadre Horizon Europe (2021-2027) succédera à Horizon 2020. Dans ce nouveau programme, l'« Open Science deviendra la nouvelle approche standard », avec une politique d'accès libre allant plus loin que dans Horizon 2020 (SEFRI 2020).

Les graphiques ci-dessous, réalisés par la ZHAW, résument les exigences en termes d'ORD du FNS et de H2020 et les raisons acceptées pour justifier une impossibilité de publier les données :

Figure 3 : Exigences Open Research Data du FNS et de H2020

	Deposition of Data Underlying Publications	Deposition of all Project Data	Opt-Out	Data Management Plan (DMP) required	DMP Part of Evaluation	List of FAIR Repositories Provided	Funding for DMP provided	Min Duration of Deposition	Embargo Period before Deposition
H2020 Open Research Data Policy	YES	Recommended	YES	YES	NO	YES	YES	NO	YES
SNF Open Research Data Policy	YES	Optional	YES	YES	NO	NO	YES	YES	YES

(ZHAW Ressort Forschung & Entwicklung 2019)

Figure 4 : Raisons valables d'exception à l'obligation de publier les données du FNS et de H2020

SNF	EU
legal clauses (e.g. Schweizer Humanforschungsgesetz)	-
ethical clauses (e.g. personal data ⁷)	to protect personal data
copyright clauses (e.g. intention to patent)	to protect results that can reasonably be expected to be commercially or industrially exploited
confidentiality clauses (e.g. confidential company data)	confidentiality in connection with security issues
other reasons (e.g. no repositories available, not relevant for the wider community etc.)	other legitimate reasons
-	to protect the project's main aim

(ZHAW Ressort Forschung & Entwicklung 2019)

3.2 Aperçu de la situation dans les hautes écoles suisses

Dans les hautes écoles suisses la thématique de l'Open Research Data se développe également, mais à des rythmes différents selon les conclusions d'une étude récente : « on one hand, a number of institutions have established excellent RDM support services and tools for their researchers (e.g. ETH Zürich, EPF Lausanne, UNIGE). On the other hand, many (smaller) institutions so far lack the resources to set up dedicated RDM support units » (Brüwer 2019). Pour avoir une idée de l'avancement de la question dans les différentes hautes écoles suisses, il a fallu étudier ce qu'elles offrent en termes de documents institutionnels cadres (stratégie, directive ou politique), documentation, soutien et infrastructures dédiés aux données de la recherche. Les résultats de cette recherche d'information sont présentés brièvement dans un tableau en annexe 2 de ce travail. Les recherches se sont concentrées sur les EPFs, les universités et les hautes écoles spécialisées, répertoriées dans la base de données P3 du FNS. Elles n'ont pas été étendues aux hautes écoles pédagogiques au vu de leur nombre. Le tableau en annexe 2 n'est qu'un aperçu de la situation et ne se veut pas exhaustif, les informations étant basées que sur ce qu'il a été possible de trouver en ligne. De plus, comme

la situation ORD se développe très rapidement et que de nouvelles initiatives ne cessent de se mettre en place, il est donc probable que ces données deviennent rapidement obsolètes.

Ce qu'il est en tout cas possible d'affirmer à la lumière de cet aperçu de la situation OS dans les hautes écoles est que l'Open Access pour les publications est désormais bien encadré : quasiment toutes les écoles disposent d'une politique Open Access et d'une archive institutionnelle. Plus rares en revanche sont celles qui se sont dotées également d'une politique Open Research Data ou d'une stratégie cadre Open Science qui englobe également les données de la recherche. Du côté des dépôts pour les données de la recherche, certaines des archives institutionnelles se sont étendues à ce type de contenus ou des solutions ad hoc sont développées, à l'image de Yareta pour le canton de Genève, mais la plupart des hautes écoles n'offrent pas de dépôts de données institutionnels. Les chercheurs sont plutôt orientés vers des dépôts existants. Il est vrai que créer son propre *repository* n'est pas toujours nécessaire car il en existe déjà de nombreux qui satisfont aux critères FAIR, certains offrant même l'avantage d'être spécialisés dans une discipline et reconnus au sein de celle-ci. Au niveau suisse, il est possible de mentionner les dépôts disciplinaires FORSbase et DaSCH (Data and Service Center for the Humanities), dédiés aux sciences sociales, et aux humanités également pour le dernier. Parmi les projets de dépôts généralistes en cours au niveau national, il y a donc OLOS de DLCM, et également SwissUbase qui est mené conjointement par FORS, l'Université de Lausanne, l'Université de Zurich et SWITCH.

La majorité des hautes écoles propose de la documentation et a des équipes dédiées au soutien à la gestion des données de la recherche. Certaines exigent parfois un DMP pour des projets internes ou des demandes d'espace de stockage et ont réalisé leur propre modèle, comme l'Université de Lausanne ou la Zürcher Hochschule der Künste (ZHdK.) La HES-SO, quant à elle, a récemment dégagé une enveloppe budgétaire de 100'000 francs pour l'ouverture des données de la recherche de projets ayant reçu un financement de la HES-SO ou de tiers. Cet appel à projet *Open Data*, lancé en juin 2020, comporte deux conditions : les données doivent être archivées dans un dépôt FAIR et un rapport final décrivant le processus de gestion et de partage des données adopté dans le projet doit être déposé dans ArODES, l'archive institutionnelle de la HES-SO (Lucas 2020).

En somme, selon Michael Brüwer, responsable des services IT à l'Université de Bâle (2019) :

« Switzerland is currently dotted with a number of high-quality services and infrastructures providing excellent RDM support. Nevertheless, these services are currently not available to all researchers to the same extent. Institutions differ in profile, priorities and maturity. Existing services still partly lack visibility, governance and sustainability. The challenge and the vision of the P2/P5 programs that all researchers in Switzerland have access to the same high-level RDM support still remains valid and attractive. »

La transition vers l'Open Research Data est donc bien lancée dans notre pays et progresse rapidement. Le nouveau programme *Open Science* de swissuniversities qui commencera en 2021, continuera les efforts entrepris par les programmes précédents (P2/P5) afin d'offrir aux chercheurs suisses la gouvernance, le soutien et les infrastructures nécessaires à une bonne gestion des données de la recherche et à leur ouverture.

4. Cadre juridique

Le précédent chapitre a démontré que l'Open Research Data était actuellement le nouvel enjeu de l'Open Science et une volonté des bailleurs de fonds comme des institutions de recherche, qui encouragent et encadrent de plus en plus cette pratique. Seulement, partager des données en vue de leur réutilisation comporte de nombreuses implications au niveau légal. L'accompagnement juridique des chercheurs se montre complexe car il n'existe pas de définition légale des données de la recherche, ni de réglementation à ce sujet (Vilmen 2018a). Les conséquences en cas de violation de la loi peuvent s'avérer très lourdes pour les chercheurs. Il est donc essentiel d'être au clair sur le cadre juridique des données de la recherche, qui est principalement composé du droit d'auteur et de la protection des données personnelles.

4.1 Droit d'auteur

Le droit d'auteur est le premier composant de la propriété intellectuelle ; les brevets, les marques et les designs constituent le second. Dans le monde de l'enseignement et de la recherche, le droit d'auteur est primordial car les enseignants-chercheurs sont des utilisateurs de contenu, et souvent aussi des producteurs de ce dernier. Avec l'avènement des technologies et d'Internet, qui permettent la copie et la transmission immédiate de l'information, ceux-ci doivent veiller à ne pas violer les droits des autres lors de leur travail, ainsi qu'à bien protéger leurs propres œuvres tout en les diffusant de manière optimale (Favre, Germond 2020).

4.1.1 Loi sur le droit d'auteur – LDA (CH)

4.1.1.1 Notions d'œuvre, d'œuvre scientifique et d'auteur

Quelques définitions s'imposent avant d'aborder la LDA sous l'angle de la recherche, afin de bien comprendre la signification des termes utilisés dans la loi.

La LDA régit la protection de l'auteur et des œuvres (CCDigitalLaw 2017a). Une **œuvre** est considérée comme « toute création de l'esprit, littéraire ou artistique, qui a un caractère individuel » (art. 2, al.1). Les programmes d'ordinateurs (logiciels) sont aussi des œuvres (LDA art. 2, al. 3). Deux aspects sont donc essentiels à la définition d'une œuvre : la création de l'esprit et son caractère individuel. La création de l'esprit suppose que l'œuvre est le fruit d'une activité humaine, soit d'une manifestation de pensée. L'individualité, elle, résulte du travail de l'auteur, des décisions qu'il a prises et des combinaisons qu'il a effectuées. Cet état de fait veut dire que s'il y a un caractère individuel, il est impossible qu'un tiers réalisant la même tâche puisse créer une œuvre identique (Favre, Germond 2018, p. 5). Mais attention, depuis la récente révision de la loi, « les productions photographiques et celles obtenues par un procédé analogue à la photographie d'objets tridimensionnels », même dépourvues de caractère individuel, sont aussi considérées comme des œuvres (LDA art. 2, al. 3bis). En revanche, les productions de la nature ou de la technique créées sans intervention intellectuelle humaine (par exemple une photo prise par un radar) ne sont pas des œuvres au sens de la LDA. Une œuvre est protégée indépendamment de la forme de son expression (Favre, Germond 2018, p. 5), mais elle doit être exprimée d'une manière ou d'une autre pour bénéficier d'une protection (CCDigitalLaw 2020). Finalement, mentionnons qu'une œuvre qui est accessible librement n'en est pas moins protégée et que son utilisation reste soumise à l'accord de l'auteur (Métille 2017a, p. 70).

La LDA cite l'**œuvre scientifique** comme une création de l'esprit (art. 2, al. 2), qui est donc protégée. Avec ce type d'œuvre, le caractère individuel réside moins dans le contenu que dans la forme. En effet, la marge de manœuvre créative est moindre dans le cas des œuvres scientifiques en raison de la nature de l'objet. Dans le cas des articles scientifiques par exemple, ce qui distingue les créations de différents chercheurs œuvrant au sein de la même spécialité d'étude est « la forme de l'expression de l'idée, de l'opinion ou de la découverte », pour autant que cette forme soit « originale et individuelle ». C'est-à-dire que deux chercheurs avec la même spécialisation et le même sujet d'étude ne relateront pas leurs travaux scientifiques sous la même forme (Favre, Germond 2018, p. 7). En revanche, les **données scientifiques pures**, comme un calcul ou une formule mathématique par exemple, ne sont pas protégées.

L'**auteur**, lui, est défini comme « la personne physique qui a créé l'œuvre » (LDA art. 6). Si plusieurs personnes ont participé à la création de l'œuvre, ils sont titulaires du droit d'auteur en commun et donc **coauteurs** (LDA art. 7, al. 1). Généralement, les scientifiques reconnaissent que leurs résultats sont le fruit d'un travail d'équipe. Des conventions existent dans le milieu de la recherche pour savoir qui est nommé auteur d'un article et dans quel ordre les présenter pour signifier la reconnaissance de l'effort d'équipe auprès de la communauté scientifique (Carroll 2015). A ce titre, l'investissement à fournir pour obtenir le statut de coauteur est réglementée différemment selon les hautes écoles. Par exemple, dans les directives de l'EPFL, de l'ETHZ et de l'Académie suisse des sciences, il faut avoir contribué à la rédaction du manuscrit pour obtenir le statut de coauteur de la publication. En revanche, pour la plupart des universités suisses, notamment Genève, Lausanne et Berne, la contribution à la rédaction même du manuscrit n'est pas nécessaire, mais il est exigé une contribution essentielle au travail de recherche (Favre, Germond 2018, p. 8).

4.1.1.2 Principe de territorialité

La LDA s'applique selon le principe de la territorialité. Le droit suisse s'applique ainsi pour la recherche effectuée en Suisse, c'est le lieu d'utilisation de l'œuvre qui prime. Il y a donc un problème d'uniformité des régimes légaux applicables lors des recherches au caractère international, mais l'avantage est qu'un chercheur en Suisse ne devra pas s'inquiéter des lois des autres pays où le contenu a été créé, modifié ou publié avant qu'il ne l'utilise (Baggi [sans date], p. 1-2).

4.1.1.3 Champ d'application : quid des données ?

Est-ce que les contenus de type « données » sont protégés par la LDA ? Par précaution, la réponse est plutôt affirmative, tant la définition d'une œuvre est étendue. En réalité, plusieurs possibilités peuvent se présenter avec les données. Pour commencer, en droit suisse, il n'existe pas de protection *sui generis* pour les **bases de données** comme c'est le cas en Europe et dans quelques autres pays du monde. Ce droit protège les bases de données qui ne sont pas concernées par le droit d'auteur en raison d'un manque d'individualité. Il reconnaît l'investissement financier et l'effort lié à la création de la base de données et garanti à son créateur un droit exclusif dessus pendant 15 ans. Il interdit aussi de la réutiliser ou d'en extraire une partie substantielle (Aliprandi 2012). En Suisse, comme le droit *sui generis* n'existe pas, les bases de données sont considérées comme des recueils au sens de la LDA (Baggi [sans date], p. 2). Si la sélection et la mise en forme de leurs contenus est une création de l'esprit qui est individuelle, les bases de données sont protégées. A l'inverse, si les données ne sont organisées que sur une base logique, par exemple alphabétique ou chronologique, le

caractère individuel n'est pas présent et elles ne sont pas protégées (CCDigitalLaw 2017b). Viennent ensuite les **données** conservées dans ces bases de données. Celles-ci sont aussi protégées si elles satisfont singulièrement à l'exigence de la création de l'esprit et du caractère individuel. Dans le cas contraire, il sera possible de les utiliser librement par rapport au droit d'auteur (Baggi [sans date], p. 4). En général, la protection de quelque chose au niveau individuel va impliquer un choix expressif de l'auteur, comme dans un dessin ou un texte littéraire. Il est donc possible d'affirmer que les données brutes ne sont pas protégées, à l'image des données à caractère purement scientifique (Vilmen 2018a). Se présente ensuite le cas des **métadonnées**, qui est plus délicat. La majorité ne sont pas protégées car dépourvues d'individualité, par exemple le nom de l'auteur, la taille d'un fichier ou son format. En revanche, d'autres peuvent l'être, à l'image du titre, du résumé ou de la table des matières (CCDigitalLaw 2017c). Finalement, les **visualisations** de données, comme les schémas, les images, les tableaux ou les graphiques, peuvent également être protégées si ce sont des œuvres suffisamment originales (Carroll 2015).

Il est donc nécessaire de procéder au cas par cas dans l'analyse de ce qui compose le jeu de données, notamment dans le cas de la réutilisation de *datasets* d'auteurs tiers, pour savoir ce qui est protégé ou non par le droit d'auteur.

4.1.1.4 Droits conférés à l'auteur

Le droit d'auteur confère deux types de droits : les droits patrimoniaux et les droits moraux. Les droits patrimoniaux d'une œuvre peuvent être cédés, ce qui n'est pas le cas des droits moraux qui sont inaliénables et indissociables de la personne de l'auteur (Almansi et al. 2010).

Les droits patrimoniaux permettent à l'auteur de recevoir une rétribution sur son œuvre (Favre, Germond 2018, p. 10), en décidant si, quand et de quelle manière elle sera utilisée (CCDigitalLaw 2017d). La LDA fournit une liste exemplative et non exhaustive de ces droits à l'article 10, comme le droit de confection d'exemplaires, de représentation, de diffusion, de reproduction ou encore de retransmission.

Les droits moraux servent, quant à eux, à protéger la personnalité de l'auteur. Deux droits moraux sont d'intérêt dans le contexte de la recherche : le droit de décider « si, quand, de quelle manière et sous quel nom » se produira la première divulgation de son œuvre (LDA art. 9, al. 2) et le droit d'en revendiquer la paternité (LDA art. 9, al.1). Premièrement, le droit de décider des modalités de première divulgation de son œuvre est important car le régime des exceptions légales ne s'applique que si l'auteur a réalisé la première divulgation de l'œuvre. Ici, par divulgation, il est entendu que l'œuvre a été rendue accessible par l'auteur, ou avec son consentement, à un grand nombre de personnes ne constituant pas un cercle étroitement lié (Almansi et al. 2010, p. 29). Deuxièmement, le droit à la paternité implique l'éthique de citer l'auteur lorsque son œuvre est mentionnée. Il n'y a en réalité pas d'obligation de lier expressément l'auteur à son œuvre dans la LDA, à l'exception du cas de la citation (art. 25), mais les conventions font que l'auteur est généralement crédité, particulièrement dans le milieu de la recherche.

4.1.1.5 Titularité des droits d'auteur

L'article 10 LDA stipule que « l'auteur a le droit exclusif de décider si, quand et de quelle manière son œuvre sera utilisée ». Dans le cas d'un chercheur qui veut utiliser l'œuvre d'un tiers, il doit donc identifier l'auteur ou les ayants droits afin de leur demander leur consentement à cette utilisation. Si on se place cette fois du point de vue de la personne qui crée une œuvre,

et c'est celui qui nous intéresse le plus ici, il faut examiner la nature des différents rapports entre les personnes qui ont concouru à la création de l'œuvre pour savoir qui est titulaire des droits dessus (Baggi [sans date], p. 5). Dans le cas de coauteurs (LDA art. 7), conjointement titulaires des droits, le principe de la propriété en main commune a pour conséquence qu'ils ne peuvent disposer de leurs droits que d'un commun accord (Benhamou, Tran 2016, p. 582). Aucun des coauteurs ne peut refuser son accord pour des motifs qui vont à l'encontre de la bonne foi. Il est également permis à chacun des auteurs d'utiliser séparément sa contribution à l'œuvre, si les différents apports à celle-ci peuvent être différenciés, sous réserve que l'exploitation de la prestation commune ne soit pas entravée (Benhamou, Tran 2016, p. 582 ; Baggi [sans date], p. 5).

Les considérations évoquées ne valent que s'il n'y a aucun contrat, loi ou règlement valable entre les parties concernées qui n'établisse un état de fait différent. En effet, selon l'article 16 alinéa 1, un auteur peut transférer l'entière ou une partie de ses droits patrimoniaux à un tiers (CCDigitalLaw 2017d). La personne qui reçoit ces droits en devient la titulaire, mais ne gagne pas le statut d'auteur. Une fois qu'un transfert de droits a été réalisé, l'auteur « n'a plus le droit d'exercer les droits qu'il a transférés à un tiers » (CCDigitalLaw 2017d) sans lui en demander l'autorisation préalablement (CCDigitalLaw 2017d).

Dans le monde de la recherche, il est courant que les rapports de travail entre un chercheur et l'institution qui l'emploie soient régis par une loi, un règlement ou un contrat qui prévoit une cession (partielle ou non) des droits d'auteur (Baggi [sans date], p. 5). Pour les hautes écoles genevoises à qui sont destinées Yareta par exemple, la loi sur l'Université de Genève précise qu'« à l'exception des droits d'auteur sur les publications, l'université est titulaire des droits de propriété intellectuelle portant sur toutes les créations intellectuelles ainsi que les résultats de recherches, y compris les programmes informatiques » (LU art.15). La loi sur la HES-SO Genève formule exactement le même article sur la propriété intellectuelle, en ajoutant que les droits d'auteur sur les créations artistiques sont exclus en plus de ceux sur les publications (LHES-SO, art. 9). Du côté de l'IHEID, en revanche, les professeurs restent titulaires de leurs droits d'auteur sur leurs œuvres (Vilmen 2019). De manière générale, les chercheurs doivent donc regarder ce qui est prévu avec leur employeur pour savoir s'ils restent titulaires des droits d'auteur sur leurs œuvres ou non.

4.1.1.5.1 La propriété des données : un concept erroné

Dans le cas de données, il est juridiquement faux de parler de « propriété ». En effet, les données ne sont pas des choses, soit des biens matériels, sur lesquels la propriété porte en droit suisse (Benhamou, Tran 2016, p. 574-575). L'expression « propriétaire des données » est également imprécise, puisque qu'elle peut renvoyer à la personne qui est concernée par les données, à celle qui les a collectées ou encore à celle qui exploite une base de données. Il est cependant possible d'être propriétaire d'un support contenant des données ou de faire valoir des droits sur les données, comme les droits de propriété intellectuelle (Métille 2017a, p. 54) dont il est question ici. Sous l'angle de la LDA, il est donc correct de parler de titularité des droits et erroné de mentionner la propriété des données.

4.1.1.6 Utilisation d'œuvres protégées et régime d'exceptions

Si un chercheur veut utiliser des données protégées par la LDA, trois cas de figure se présentent à lui : renoncer à l'utilisation de ces données, utiliser une exception légale prévue par la LDA ou demander une autorisation à l'auteur ou aux titulaires des droits au préalable (Baggi [sans date], p. 4). A noter que si les données ont été transmises à la suite de l'adoption

d'un contrat, il convient d'en respecter les termes. De même, si les données sont mises à disposition par l'auteur sous une licence, il convient de respecter cette licence (Becard et al. 2017). Cette partie détaille les exceptions de la LDA qui peuvent encadrer la réutilisation d'une œuvre, qui peut prendre la forme de données. Pour rappel, la condition initiale pour bénéficier d'une exception au droit d'auteur est que l'œuvre ait été divulguée au sens de l'article 9 LDA. De plus, l'utilisation d'une exception ne veut pas dire qu'aucune rémunération à l'auteur ou au titulaire des droits n'est due (Baggi [sans date], p. 7).

Il n'existe qu'une seule exception légale pour l'utilisation d'œuvres à des fins de recherche scientifique. Elle concerne la reproduction de celles-ci et a été introduite au 1^{er} avril 2020 par une révision de la LDA. Cette dernière autorise « la reproduction d'une œuvre à des fins de recherche scientifique lorsqu'elle est nécessaire à l'application d'un procédé technique » (LDA art. 24d). Les logiciels sont exclus. Concrètement, cette restriction au droit d'auteur a été pensée pour la pratique du *text and data mining* (TDM), soit la fouille de textes et de données, mais la formulation est probablement assez large pour couvrir d'autres utilisations nécessitant un procédé technique et des technologies futures comparables (Conseil fédéral 2017). La confection automatique des copies nécessaires à la recherche est désormais autorisée sans l'accord de l'auteur, sous réserve d'un accès licite aux exemplaires à copier. Les reproductions ne sont pas soumises à rémunération et elles sont admises également à des fins commerciales. En outre, elles peuvent être conservées à des fins d'archivage et de sauvegarde après la fin de la recherche scientifique. La nouvelle disposition ne couvre donc que la reproduction d'œuvres, et à des conditions très précises. Elle n'englobe pas la mise à disposition ou la publication par la suite des œuvres utilisées (Conseil fédéral 2017, p. 37). L'article 24d complète ainsi d'autres exceptions déjà prévues par la LDA et que les scientifiques peuvent faire valoir dans leur travail (Conseil fédéral 2017, p. 37), notamment l'article 19 sur l'utilisation d'œuvres à des fins privées.

Cet article 19 comporte trois exceptions : pour l'usage privé, pour l'usage à des fins pédagogiques et pour la reproduction à des fins d'information interne ou de documentation. L'exception pour usage strictement personnel autorise toute utilisation « à des fins personnelles ou dans un cercle de personnes étroitement liées, tels des parents ou des amis » (art. 19, al. 1, let a), sans obligation de rémunération (Ling 2018). La loi ne mentionnant pas sous quelles conditions cette utilisation personnelle doit se réaliser, celle-ci pourrait être ludique, professionnelle ou de recherche :

« Sous les conditions strictes prévues par la loi, qui doivent être interprétées de façon à ne pas porter préjudice aux intérêts légitimes des titulaires des droits d'auteur, cette exception peut permettre l'utilisation d'une œuvre protégée aussi à des fins de recherche. Dans le cadre de cette exception et dans ces limites, un chercheur pourra donc effectuer toute utilisation que lui sera nécessaire à fin d'études comme par exemple la consultation, reproduction, stockage, archivage, indexation, etc. Par contre, tout l'aspect qui concerne le partage des résultats d'une recherche ainsi conduite avec des tiers n'est manifestement pas inclus dans cette exception et porterait atteinte aux droits exclusifs de l'auteur. »
(Baggi [sans date], p. 7)

L'exception ne peut manifestement pas être revendiquée non plus par une équipe de chercheurs, le fait d'être collègues ne constituant pas un cercle restreint de personnes étroitement liées (Baggi [sans date], p. 7).

Il n'est pas possible d'utiliser la deuxième exception pour usage didactique, soit « toute utilisation d'œuvres par un maître et ses élèves à des fins pédagogiques » (LDA art. 19, let. b)

dans le contexte de la recherche, à part si un chercheur souhaite présenter les résultats d'un de ses projets de recherche mené individuellement à sa classe (Baggi [sans date], p. 8). Cette exception est limitée à un extrait de l'œuvre et est soumise à la rémunération (Ling 2018).

Toujours dans l'optique d'un partage d'œuvres protégées et non plus dans l'utilisation de contenus protégés, un chercheur pourrait utiliser l'exception pour « la reproduction d'exemplaires d'une œuvre au sein des entreprises, administrations publiques, institutions, commissions et organismes analogues, à des fins d'information interne ou de documentation » (LDA art. 19, let. c). Ici, le cercle de personnes est bien plus large que celles « étroitement liées ». En revanche, l'exception ne se rapporte pas à « toute utilisation » comme précédemment, mais seulement à la reproduction (Baggi [sans date], p. 9). Elle est aussi limitée, comme celle pour l'usage didactique, à la reproduction d'une partie de l'œuvre protégée seulement et non pas de son entièreté (LDA art. 19, al.3, let. a) (Baggi [sans date], p. 9). De même, cette exception est soumise à une obligation de rémunération (Ling 2018).

En guise de conclusion à cette présentation du régime d'exceptions prévu par la LDA et de son application dans le contexte de la recherche, il est possible d'affirmer que les licences de droit d'auteur, qui seront détaillées par la suite, jouent un rôle très important pour faciliter les activités liées à la science. En effet, si chaque chercheur autorisait l'utilisation de son œuvre (si lui-même n'a pas utilisé d'œuvres de tiers protégées pour la concevoir) de manière plus ou moins libre en lui attribuant une licence, il n'y aurait plus besoin de demander une autorisation aux ayants droits (Baggi [sans date], p. 4), ni d'utiliser la petite marge de manœuvre permise par le régime d'exceptions légales.

4.1.1.7 Durée de la protection

La durée de la protection dépend du type d'œuvre et du nombre d'auteur. La protection prend fin :

- Pour les programmes d'ordinateur (logiciels) : 50 ans après le décès de l'auteur (LDA art. 29, al. 2, let. a) ou du dernier coauteur survivant (art. 30, al. 1, let. a) ;
- Pour les photographies : 50 ans après la confection (art. 29, al. 2, let. a^{bis}) ;
- Pour toutes les autres œuvres : 70 ans après le décès de l'auteur (art. 29, al. 2, let. b) ou le décès du dernier coauteur survivant (art. 30, al. 1, let. b).

A la fin de la durée de la protection, l'œuvre tombe dans le domaine public.

4.2 Protection des données personnelles

Les données à caractère personnel récoltées dans un cadre d'une activité de recherche sont aussi protégées. Des précautions s'imposent et le chercheur doit être vigilant pour ne pas enfreindre la loi. En Suisse, la protection de la sphère privée et la protection des données personnelles sont des droits fondamentaux garantis par la Constitution (art. 13). En 2014, un arrêt du Tribunal fédéral a interprété ces dispositions constitutionnelles comme un droit à l'autodétermination en matière d'information, soit la possibilité pour chacun de désigner « si et dans quel but des données qui la concernent peuvent être traitées et enregistrées par des tiers quels qu'ils soient, privés ou provenant du secteur public » (Cellier, Werly 2018, p. 1130).

La loi sur la protection des données personnelles vise à protéger les personnes qui font l'objet d'un traitement de données, et non pas les données elles-mêmes (Métille 2017b). Ce travail présente trois lois de protection des données personnelles qui s'appliquent à des niveaux différents. Il s'agit de la loi cantonale genevoise (puisque Yareta dépend du Canton de

Genève), de la loi fédérale et du récent règlement général européen sur la protection des données (RGPD). Concernant les rapports entre la loi fédérale et la loi cantonale genevoise, il faut retenir qu'elles ne s'appliquent pas aux mêmes secteurs. La LPD fédérale s'applique au secteur privé et au secteur public qui relève de la Confédération, tandis que la loi cantonale s'applique au secteur public communal et cantonal de Genève (PPDT GE 2015a). La LPD constitue la base de référence de la protection des données en Suisse et en pose les principes fondamentaux, qui sont ensuite repris dans les lois cantonales, si existantes. Si les cantons n'émettent pas leurs propres dispositions pour la protection des données, la LPD s'applique à titre supplétif.

Dans le domaine de la recherche médicale, et donc des données y afférentes, s'appliquent d'autres lois qui régissent la recherche sur l'humain ou le domaine de la santé, notamment la Loi fédérale relative à la recherche sur l'être humain (LRH) pour les recherches qui portent sur des maladies, sur la structure ou le fonctionnement du corps humain, ou encore qui ont pour objet des données de santé ou du matériel biologique. Ces types de recherches ne sont pas régies par la LPD car la LRH, par exemple, contient déjà des dispositions relatives à la protection des données. Ne faisant pas partie du champ de ce travail, ces textes légaux ne seront pas détaillés, mais il faut retenir que pour tout ce qui touche à l'être humain, d'autres lois que celles de la protection des données s'appliquent.

4.2.1 Loi sur la protection des données personnelles – LPD (CH)

L'analyse du cadre légal des données de la recherche en termes de protection des données commence par la LPD et non par la LIPAD simplement car cette dernière, en tant que loi cantonale, est fondée sur la LPD, dont elle reprend nombre de dispositions. Il est donc plus logique de commencer par la loi fédérale, pour ensuite présenter la loi cantonale. Pour les définitions des concepts de la LPD évoqués dans cette partie, il faut se référer à l'annexe 3 de ce travail.

4.2.1.1 Révision totale en cours

Une procédure de révision a été initiée en décembre 2011. Le Conseil fédéral a chargé le Département fédéral de justice et police d'évaluer la possibilité de renforcer la protection des données en Suisse, notamment en regard des évolutions au sein de l'Union européenne (UE) et du Conseil de l'Europe. Le projet de loi avec révision totale est sorti le 15 septembre 2017 (Office fédéral de la justice OFJ 2019). Selon le planning initial, il devait être traité rapidement afin que la loi révisée entre en vigueur avant la fin de l'année 2018 (Métille 2017c). Il est en effet important que notre loi s'aligne sur le RGPD européen, nouveau standard en la matière, afin que la Suisse continue à être jugée comme un pays de protection équivalente. Or, à ce jour, le projet de révision totale de la LPD n'a toujours pas abouti. Ce travail se base donc malheureusement sur la version de la LPD actuellement en vigueur, celle du 1^{er} mars 2019, qui devrait rapidement être remplacée.

4.2.1.2 Champ d'application

Le but de la LPD est de « protéger la personnalité et les droits fondamentaux des personnes qui font l'objet d'un traitement de données » (art. 1). Elle s'applique au traitement de données de personnes physiques et morales par des personnes privées ou par des organes fédéraux, lorsque le traitement a lieu en Suisse (Morscher, Seidl-Nussbaumer 2016). Il y a donc deux parties dans cette loi, l'une de droit privé et l'autre de droit public fédéral.

La LPD s'applique à tout traitement de données personnelles, pour autant qu'il ne relève pas d'une exception (Meier 2011, p. 187), mais elle ne s'applique pas aux données collectées anonymement ou anonymisées. Les données pseudonymisées, en revanche, sont concernées. Pour connaître la différence entre ces deux techniques de protection des données personnelles, se référer à l'annexe 3. Contrairement à d'autres pays, en Suisse ce n'est pas parce que les données personnelles sont publiquement accessibles qu'elles sont libres de droit. Ce qui autorise leur traitement est qu'elles soient rendues publiques par la personne concernée et, qu'après en avoir été informée, celle-ci ne se soit pas opposée à leur traitement (LPD, art. 12, al. 3) (Métille 2017a, p. 69).

4.2.1.3 Principes

La LPD édicte les grands principes à respecter quand des données personnelles sont traitées. Si ceux-ci sont violés ou que le traitement intervient contre la volonté de la personne concernée, il y alors atteinte à sa personnalité (Métille 2014). Celle-ci est illicite, « à moins d'être justifiée par le consentement de la victime, par un intérêt prépondérant privé ou public, ou par la loi » (LPD, art. 13, al. 2).

Dans la LPD figure le principe de la **licéité** du traitement (art. 4, al. 1), évoqué ci-dessus. C'est-à-dire que les données doivent être traitées légalement, soit en accord avec les principes de la loi ou, si ce n'est pas le cas, que cela soit justifié par une raison reconnue dans celle-ci (Morscher, Seidl-Nussbaumer 2016). Le principe de la **bonne foi** (LPD art. 4 al. 2), qui découle du principe de licéité (Meier 2011, p. 263), est aussi important. De la bonne foi résulte à son tour le principe de la **reconnaissabilité** (LPD art. 4 al. 4), qui indique que la collecte des données doit se faire en toute transparence et que les finalités du traitement doivent être reconnaissables pour la personne concernée. Le principe de **proportionnalité** est également évoqué (art. 4 al. 2). Il veut qu'une balance soit toujours effectuée entre le but du traitement et l'atteinte à la personne concernée : il doit y avoir une proportionnalité entre le but visé, les moyens mis en place pour l'atteindre et la préservation des droits de la personne concernée. De la proportionnalité découlent deux autres sous-principes, celui d'évitement (les données non nécessaires ne sont ni collectées, ni traitées) et de minimisation (il faut parvenir au but visé avec le moins de données possible) (Métille 2017b, p. 83-84). Le principe de la **finalité**, lui, veut que les données personnelles ne soient traitées que « dans le but qui est indiqué lors de leur collecte, qui est prévu par la loi ou qui ressort des circonstances » (LPD art.4, al. 3). Ce principe est caractéristique de la protection des données. Le but du traitement doit donc être déterminé préalablement et ne peut pas être modifié ou remplacé par un autre, sauf si une loi ou le consentement de la personne concernée l'autorise (Métille 2017b, p. 84). Le principe **d'exactitude**, quant à lui, implique que celui qui traite des données personnelles s'assure qu'elles soient correctes (LPD, art. 5, al.1), donc exactes et complètes. Pour terminer, le principe de **sécurité** des données veut qu'elles soient « protégées contre tout traitement non autorisé par des mesures organisationnelles et techniques appropriées » (LPD art. 7, al. 1).

4.2.1.4 Consentement

En l'absence d'une justification suffisante (par exemple un intérêt prépondérant), les données personnelles ne doivent pas être traitées sans le consentement de la personne concernée (Morscher, Seidl-Nussbaumer 2016). Celui-ci doit avoir lieu après information et être libre. Dans le cas de données sensibles, le consentement doit également être explicite (cette exigence n'est pas nécessaire dans le cas de données personnelles non sensibles). Il n'y a

pas d'obligation de donner le consentement par écrit, même s'il est bien évidemment conseillé d'en garder une trace (Morscher, Seidl-Nussbaumer 2016). Si le but du traitement de données est modifié dans le temps, de même que son étendue, une nouvelle information est nécessaire ainsi qu'un nouveau consentement en regard de ces modifications (Meier 2011, p. 331). La personne concernée a le droit de révoquer son consentement au traitement de ses données personnelles, ce qui aura bien entendu un impact sur un projet de recherche en cours par exemple. Si le traitement est terminé, cette révocation n'aura plus d'effet, à part sur des activités encore en cours, comme la conservation ou l'archivage de données (Meier 2011, p. 343).

4.2.1.5 Données sensibles

Les données sensibles sont : « les données personnelles sur : 1. les opinions ou activités religieuses, philosophiques, politiques ou syndicales, 2. la santé, la sphère intime ou l'appartenance à une race, 3. des mesures d'aide sociale, 4. des poursuites ou sanctions pénales et administratives (LPD art. 3, let. c) ». Les profils de la personnalité bénéficient de la même protection que les données sensibles. Les principes généraux de la LPD s'appliquent de la même manière aux données sensibles et aux profils de la personnalité qu'aux autres données personnelles. En revanche, moins de raisons peuvent légalement être invoquées pour justifier un irrespect de ces principes dans le cas de données sensibles et de profils de la personnalité (Morscher, Seidl-Nussbaumer 2016). De plus, sans motif justificatif, personne n'est autorisé à communiquer à des tiers des données personnelles sensibles ou des profils de la personnalité (Meier 2011, p. 218-220).

4.2.1.6 Déclaration des fichiers de données personnelles

Le préposé fédéral à la protection des données et à la transparence (PFPDT) tient un registre des fichiers, soit tout ensemble de données personnelles dont la structure permet de rechercher les données par personne concernée. Les organes fédéraux sont obligés de déclarer tous leurs fichiers pour enregistrement. Les privés doivent le faire dans deux cas : ils traitent régulièrement des données sensibles ou des profils de la personnalité, ou bien ils communiquent régulièrement des données personnelles à des tiers (LPD art. 11a).

4.2.1.7 Devoir d'information

Pour les privés, il suffit que la collecte de données et les finalités du traitement soit connues de la personne concernée, sauf dans le cas de données sensibles. Dans cette situation, même si les données sont librement accessibles et que la personne concernée les a elle-même publiées, les informations suivantes doivent être transmises au minimum : l'identité du maître du fichier, les finalités du traitement et les catégories de destinataires des données si la communication est envisagée (LPD art. 14). Les organes fédéraux, eux, ont un devoir général d'information et sont obligés d'aviser spontanément la personne concernée de toute collecte de données (Métille 2017b, p. 84-85). Doivent alors être transmises au minimum les informations suivantes : l'identité du maître du fichier, les finalités du traitement, les catégories de destinataires des données si la communication est envisagée, le droit d'accès de la personne concernée et les conséquences liées au refus de sa part de fournir les données personnelles demandées (LPD art. 18a).

4.2.1.8 Droits de la personne concernée

La personne concernée a plusieurs droits : le droit à la rectification de ses données, le droit de s'opposer au traitement ou à la communication de ses données à des tiers, le droit à la

destruction de celles-ci et d'y accéder (Morscher, Seidl-Nussbaumer 2016 ; Donzé, Zouhri 2019).

Le droit d'accès est détaillé à l'article 8 LPD : sur requête de la personne concernée, le maître du fichier doit lui transmettre toutes les données qu'il détient sur elle, la base juridique du traitement et si elles ont été transmises à des tiers (LPD art. 8, al. 2). Ce droit s'exerce par écrit et la communication des renseignements par le maître du fichier doit se faire dans les 30 jours qui suivent. En principe, ce service est gratuit.

4.2.1.9 Sous-traitance et communication de données

La sous-traitance est possible dans la LPD pour autant « qu'une convention ou la loi le prévoit », que seuls les traitements que le mandant peut légalement faire soient effectués, et qu'aucune obligation légale ou contractuelle de garder le secret ne l'empêche. La sécurité des données doit être garantie et le tiers « peut faire valoir les mêmes motifs justificatifs que le mandant » (LPD art. 10a). Lorsque le traitement est réalisé par une succursale, un autre service ou un collaborateur de la même entreprise ou institution en Suisse, ce n'est pas considéré comme de la sous-traitance. Il est entendu que les principes généraux de la protection des données doivent être respectés, « notamment la proportionnalité dans l'étendue des données transférées à l'interne en vue de traitement et le nombre de personnes auxquelles ces données sont communiquées » (Meier 2011, p. 423).

La communication transfrontalière de données personnelles (LPD art. 6) est impossible si la législation du pays destinataire ne prévoit pas une protection jugée égale (Morscher, Seidl-Nussbaumer 2016). L'article sur la communication transfrontalière s'applique dans tous les cas de délégation du traitement ou de communication à l'étranger, même dans le cas de transfert à une succursale, un autre service ou un collaborateur (Meier 2011, p. 423-424).

4.2.1.10 Règlementation spécifique sur le traitement de données à des fins de recherche

Les grands principes de la LPD détaillés jusqu'à maintenant s'appliquent donc dans le cas de la recherche scientifique. Mais une réglementation spécifique sur le traitement de données personnelles à des fins de recherche permet un régime juridique moins sévère, pour les privés comme pour les organes fédéraux, à deux conditions (LPD art. 13, al. 2, let. e et art. 22). Premièrement, le but du traitement des données personnelles dans le cadre d'une recherche ne doit pas se rapporter à des personnes, c'est-à-dire que « l'identité de la personne dont les données sont traitées ne joue aucun rôle pour le traitement et que des données anonymisées ou du moins pseudonymisées pourraient également être utilisées pour atteindre le même objectif » (PFPDT [sans date]). Deuxièmement, il faut que la publication des résultats de la recherche ne contienne aucun élément permettant d'identifier les personnes concernées. Ces deux articles 13 et 22 concernent donc des cas de figure où des données non anonymisées sont traitées à des fins qui ne se rapportent pas à des personnes (PFPDT [sans date]).

Pour les **organes fédéraux** (dont les deux EPFs), les deux conditions ci-dessus doivent être remplies mais également celles de rendre les données anonymes dès que le but du traitement le permet. Des allègements supplémentaires leur sont aussi octroyés :

- Les données personnelles peuvent être traitées dans un autre but que celui indiqué lors de leur collecte (art. 22, al. 2, let. a) ;
- Les données sensibles ou les profils de la personnalité peuvent être traités même en l'absence de loi ou d'exception qui le permet (art. 22, al. 2, let. b) ;

- La communication des données personnelles peut être réalisée sans base légale (art. 22, al. 2, let. c), mais le destinataire ne peut communiquer les données à des tiers qu'avec le consentement de l'organe fédéral qui les lui a transmises (art. 22, al. 1, let. b).

Les organes fédéraux peuvent aussi invoquer l'application du principe de la légalité pour le traitement de données personnelles à des fins de recherche (art. 17, al.1), ici la base légale pourrait être la Loi fédérale sur l'encouragement de la recherche et de l'innovation (LERI) (PFPDT [sans date]).

Concernant l'**anonymisation**, le chercheur doit privilégier l'utilisation de données anonymes dès le début de sa recherche. Si son sujet d'étude ne lui permet pas de travailler avec des données anonymisées (parce qu'il est nécessaire de pouvoir contacter souvent un participant par exemple), il doit les pseudonymiser et veiller à ce que le code et la liste de correspondance ne soit accessible qu'à un nombre restreint de personnes autorisées, et conservé de manière sécurisée et chiffrée. Les données doivent ensuite être anonymisées le plus rapidement possible, soit dès que le stade de la recherche le permet. Les résultats de la recherche doivent aussi être publiés sous forme anonyme (PFPDT [sans date]).

A propos de la **conservation** des données personnelles liées à la recherche, celle-ci est concernée par la prise de mesures organisationnelles et techniques appropriées, tout comme pour leur traitement (LPD art. 7). Une durée dans le temps doit être prévue pour la conservation des données. Celle-ci ne peut normalement pas dépasser l'atteinte du but fixé par le projet de recherche. Par la suite, les données personnelles doivent être détruites ou anonymisées (PFPDT [sans date]).

La **communication** de données personnelles **par un organe fédéral à un chercheur** qui en fait la demande est possible. Elles sont livrées à des fins de recherche moyennant un contrat de protection des données. L'organe fédéral reste responsable de la protection, dont il précise le niveau nécessaire à mettre en œuvre dans le contrat (PFPDT [sans date]). La communication **hors de la Suisse** de données personnelles non anonymisées dans un but de recherche reste en revanche soumise à l'article 6 LPD.

Pour finir, les **fichiers** de personnes ne doivent pas obligatoirement être déclarés au préposé fédéral dans le cadre d'une recherche scientifique (OLPD art. 4, al. 1, let. b et art. 18, al. 2, let. c).

En conclusion, un traitement de données personnelles non anonymisées à des fins de recherche peut donc se justifier même sans le consentement de la personne concernée, aux conditions énoncées ci-dessus. Mais dans le cas d'un projet de recherche, comme la participation se fait toujours sur une base volontaire, la personne concernée devra consentir au préalable au traitement de ses données dans le cadre du projet, et elle dispose toujours de son droit d'opposition. Il n'y a pas de dispense non plus au devoir d'informer la personne concernée objectivement sur le traitement de données envisagées (PFPDT [sans date]).

4.2.2 Loi sur l'information du public, l'accès aux documents et la protection des données personnelles – LIPAD (GE)

Cette loi régit la protection des données dans le Canton de Genève. Elle est l'objet d'une partie de ce mémoire car l'Université de Genève, qui mandate cette étude, y est soumise. La LIPAD comporte deux volets : un sur la transparence de l'information traitée par les institutions publiques genevoises et un sur la protection des données (Buche, Byrne-Sutton, Werly 2017,

p. 3). Les définitions utiles à la compréhension des termes utilisés dans cette partie sont disponibles en annexe 4 de ce travail.

4.2.2.1 Champ d'application

La loi s'applique pour les deux volets aux institutions publiques du canton. C'est donc le cas pour l'Université et la HES-SO Genève, deux hautes écoles genevoises à qui sont destinées Yareta. L'Institut de hautes études internationales et du développement (IHEID), lui, est soumis à la LPD uniquement car il est privé.

La plupart des principes fondamentaux déjà évoqués dans la LPD se retrouvent dans la LIPAD aux articles 35 à 38 (Cellier, Werly 2018, p. 1131). Parmi ceux-ci, citons celui de licéité (art. 35, al. 1 et 2) :

¹ Les institutions publiques ne peuvent traiter des données personnelles que si, et dans la mesure où, l'accomplissement de leurs tâches légales le rend nécessaire.

² Des données personnelles sensibles ou des profils de la personnalité ne peuvent être traités que si une loi définit clairement la tâche considérée et si le traitement en question est absolument indispensable à l'accomplissement de cette tâche ou s'il est nécessaire et intervient avec le consentement explicite, libre et éclairé de la personne concernée.

La LIPAD mentionne aussi les principes de finalité, de bonne foi, de reconnaissabilité de la collecte, de proportionnalité, d'exactitude, de sécurité et de destruction. Ce dernier (art. 40 et LIPAD art. 15) stipule que « les institutions publiques doivent détruire ou rendre anonymes les données personnelles dont elles n'ont plus besoin pour accomplir leurs tâches légales, sauf si ces données doivent être conservées en vertu d'une autre loi » (Cellier, Werly 2018, p. 1132).

4.2.2.2 Le consentement

Le consentement n'est mentionné que deux fois dans la LIPAD. Premièrement à l'article 35, al. 2, cité ci-dessus, donc dans le cas d'un traitement de données personnelles sensibles ou de profils de la personnalité. Le consentement apparaît une deuxième fois, à l'article 39, dans le cas de communication de données à une corporation ou un établissement de droit public étranger. Le consentement a donc un rôle limité en droit public (au contraire du droit privé). Il ne peut pas être un motif justificatif qui exclurait l'application des dispositions prévues dans la LIPAD³, par exemple celles de l'article 41 pour un traitement à des fins de recherche scientifique.

4.2.2.3 Déclaration des fichiers de données personnelles

Le responsable LIPAD de chaque institution publique doit être informé de toute création de fichier, requête de communication de données personnelles ou intention de destruction « à moins que ces opérations ne soient prévues explicitement par une loi, un règlement ou une décision du Conseil d'Etat » (LIPAD art. 51, al. 1). Les responsables LIPAD doivent communiquer la liste de leurs fichiers au préposé cantonal pour enregistrement dans le catalogue des fichiers (LIPAD art. 51, al. 2).

Le but de la déclaration des fichiers est de garantir la transparence sur les données personnelles en main des institutions publiques. Elle est obligatoire que le traitement soit automatisé (fichier informatique) ou manuel (fichier papier) et quel que soit son but (PPDT GE 2015a, p. 19), la recherche scientifique ne bénéficiant pas d'une exception. Les fichiers

³ Appel téléphonique avec Stéphane Werly, préposé à la protection des données et à la transparence de l'Etat de Genève, 4 juin 2020.

éphémères, dont la durée de vie ne dépasse pas une année et qui ne contiennent pas de données sensibles, ni de profil de la personnalité ne sont pas soumis à l'obligation de déclaration (PPDT GE 2015b).

4.2.2.4 Droits conférés à la personne concernée

La LIPAD confère à la personne concernée des droits similaires à ceux octroyés par la LPD. Le droit d'accès est régi à l'article 44. La demande doit être faite par écrit, sans justification, mais en prouvant son identité (PPDT GE 2015a, p. 20). Les données sont communicables quel que soit leur format ou leur type, soit en faisant venir la personne sur place, soit par l'envoi de copies. Normalement ce service est gratuit, mais s'il entraîne un travail disproportionné, il peut être soumis à des éventuels émoluments (PPDT GE 2015a, p. 20-21).

Les autres droits sont mentionnés à l'article 47 LIPAD, notamment ceux de demander à :

- « *détruire les informations qui ne sont pas pertinentes ou nécessaires ;*
- *rectifier, compléter ou mettre à jour les informations inexactes, incomplètes ou dépassées ;*
- *ajouter une remarque sur les données dont ni l'exactitude ni l'inexactitude ne peuvent être prouvées ;*
- *refuser la communication de données qui ne sont ni nécessaires, ni pertinentes à l'accomplissement des tâches légales de l'institution concernée ou qui ne sont ni exactes, ni mises à jour. »* (Université de Genève 2012)

4.2.2.5 Disposition spécifique sur le traitement de données à des fins de recherche

Il existe dans la LIPAD une disposition pour le « traitement à des fins générales », mentionnée à l'article 41, soit :

¹ *Dans le cadre de l'accomplissement de leurs tâches légales, les institutions publiques sont en droit de traiter des données personnelles à des fins générales de statistique, de recherche scientifique, de planification ou d'évaluation de politiques publiques, pour leur propre compte ou celui d'une autre institution publique en ayant la mission légale, aux conditions cumulatives que :*

- a) le traitement de données personnelles soit nécessaire à ces fins ;*
- b) ces données soient détruites ou rendues anonymes dès que le but du traitement spécifique visé le permet ;*
- c) les données collectées à ces seules fins ne soient communiquées à aucune autre institution, entité ou personne ;*
- d) les résultats de ce traitement ne soient le cas échéant publiés que sous une forme excluant la possibilité d'identifier les personnes concernées ;*
- e) le préposé cantonal en soit préalablement informé avec les précisions utiles sur le traitement qu'il est prévu de faire des données personnelles et sa nécessité ;*
- f) le traitement portant sur des données personnelles sensibles ou impliquant l'établissement de profils de la personnalité fasse préalablement l'objet d'une autorisation du Conseil d'Etat, qui doit requérir le préavis du préposé cantonal et assortir au besoin sa décision de charges ou conditions.*

A la lecture de cet article, il est clair qu'il se révèle très limitant pour les institutions publiques genevoises qui réalisent des activités de recherche, à l'image de l'UNIGE et de la HES-SO Genève. L'article 41 n'est pas vraiment adapté à de telles activités, surtout lors d'études en collaboration nationale ou internationale. C'est essentiellement la lettre c) de cet article qui est problématique, car elle interdit toute communication des données personnelles collectées

à des fins de recherche à une autre institution, entité ou personne. La lettrine d), elle, pose moins de problèmes car elle ne fait que reprendre le principe de l'anonymisation des données personnelles dans les publications scientifiques, ce qui est déjà la règle (Hausmann 2020). Des chercheurs soumis à la LIPAD n'ont donc pas d'autre option que de publier des données personnelles anonymisées, le cas échéant.

La LIPAD est en voie d'être revue⁴, notamment pour donner suite à la révision de la LPD qui s'alignera sur le RGPD, la situation devrait donc s'améliorer à l'avenir.

4.2.3 Règlement général sur la protection des données – RGPD (UE)

Le RGPD est entré en application en mai 2018. L'UE a décidé de mettre à jour sa protection des données afin de créer un socle commun de protection et d'actualiser les standards, qui étaient assez hétérogènes. La dernière directive en la matière datait de 1995 et elle ne prenait pas en compte l'évolution des technologies et d'internet. Il a été cette fois décidé de promulguer un règlement et non une directive afin d'harmoniser la protection pour toute l'UE. En effet, un règlement s'impose tel quel à tous tandis qu'une directive a besoin d'être transposée dans les lois nationales des Etats de l'UE (Bayle 2018). Les définitions utiles à la compréhension des termes utilisés dans cette partie sont disponibles en annexe 5 de ce travail.

4.2.3.1 Champ d'application

Concernant le **champ d'application matériel**, il s'agit de tout traitement de données à caractère personnel par une institution, un organe, un établissement ou un organisme de l'UE à l'exception des dispositions prévues à l'article 2, par exemple une activité strictement personnelle ou domestique (Bayle 2018). Il ne s'applique pas aux données anonymes.

Si le RGPD fait partie du champ d'étude de ce travail, c'est que son champ d'application territorial peut concerner notre pays. En effet, pour être soumis au RGPD, le traitement doit (Bayle 2018) :

- Être effectué par un responsable de traitement ou un sous-traitant établi dans l'UE, c'est-à-dire que la personne qui traite est dans l'UE ;
- Être effectué dans l'UE (vente ou prestation de services), c'est-à-dire que le traitement est dans l'UE ;
- Être relatif à des personnes concernées résidant dans l'UE, c'est-à-dire que la personne concernée par le traitement est dans l'UE.

Le RGPD s'applique de manière **extraterritoriale** pour les cas prévus par les critères d'établissement et de ciblage. Pour l'**établissement**, sont concernés : un traitement dans le cadre des activités d'une institution ou d'une entreprise suisse qui a une succursale dans l'UE ; et un traitement qui a lieu dans une entreprise suisse en sous-traitance pour une entreprise de l'UE (art. 3, §1 et consid. 22). Concernant le **ciblage**, le RGPD s'applique dans le cas d'un traitement de données de résidents de l'UE par une entreprise suisse qui offre des biens et des services dans l'UE ou qui suit le comportement de résidents de l'UE (par exemple avec de la publicité ciblée) (art. 3, §2 et consid. 23 et 24) (PFPDT 2018). L'important n'est donc pas la nationalité européenne des personnes concernées, mais là où elles se trouvent quand des biens et des services leur sont proposés ou que leur comportement est suivi (Métille 2020).

⁴ Appel téléphonique avec Stéphane Werly, préposé à la protection des données et à la transparence de l'Etat de Genève, le 4 juin 2020.

Savoir si le RGPD s'applique pour des activités liées à la recherche n'est donc pas aisé. Les critères en relation avec le traitement sont assez limpides, mais ceux liés au ciblage le sont moins. La recherche scientifique suisse ne tombe probablement pas dans l'offre de biens et de services, mais la notion de « suivi de comportement » est assez vague. Elle n'est de plus pas définie plus précisément dans le règlement. Dans le doute, il vaudrait donc mieux appliquer le RGPD, qui, de plus, est un peu devenu le standard de protection par défaut (Junod, Gorbacheva 2018).

4.2.3.2 Principes

Le RGPD contient aussi les grands principes de protection des données que nous avons vus auparavant avec la loi suisse et la genevoise. Le traitement doit premièrement être **licite** (art. 5, al.1, let. a) en remplissant l'une des six conditions suivantes : la personne concernée a donné son consentement, le traitement est obligatoire à l'exécution d'un contrat, il découle d'une obligation légale, est obligatoire à la sauvegarde des intérêts vitaux de la personne concernée, est utilisé pour la réalisation d'une mission d'intérêt public ou encore dans l'intérêt légitime du responsable du traitement (Bayle 2018). Le traitement doit aussi respecter le principe de **loyauté** (art. 5, §1, let. a), dont il est possible d'inférer plusieurs obligations, par exemple celle d'information ou de garantir les droits des personnes concernées. Le principe de loyauté veut donc que le traitement soit réalisé en toute **transparence**, impératif auquel répond aussi par ailleurs le principe de licéité (Bamdé 2018). Le RGPD consacre également le principe de **limitation des finalités** (art. 5, §1, let. b), qui veut que les données personnelles soient collectées pour des finalités déterminées, explicites et légitimes, et ne soient pas traitées ultérieurement d'une manière incompatible avec celles-ci (Bayle 2018). Les principes de **proportionnalité**, **pertinence** et **minimisation** des données sont aussi présents (art. 5, §1, lettrine c). Le principe **d'exactitude** des données est aussi mis en avant (art. 5, §1, let. d), tout comme celui de la **conservation limitée** (art. 5, §1, let. e), qui veut que les données personnelles soient gardées « pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées ». Enfin, les données personnelles doivent être traitées selon les principes de **sécurisation** et de **confidentialité** (art. 5, §1, let. f).

Le RGPD instaure également un principe inconnu dans la législation suisse, celui de la **responsabilité du responsable de traitement** (ou d'**accountability**). Ce dernier doit être en mesure de démontrer qu'il respecte tous les principes cités ci-dessus (art. 5, §2). Il lui faudra donc être en possession d'une documentation qui prouve sa propre conformité avec le règlement (Bayle 2018).

Pour finir, deux autres principes sont essentiels au respect du RGPD et inconnus de la législation suisse, celui de **privacy by design** (art. 25, §1) et de **privacy by default** (art. 25, §2). Le premier veut que, dès les premières étapes de conception et de mise en place d'un projet qui impliquera des données à caractère personnel, leur traitement et son impact sur la vie privée des personnes concernées soit étudié, afin de minimiser les risques et de prévoir un respect rigoureux du RGPD. Le second principe veut que le plus haut niveau de protection soit garanti d'office aux personnes dont les données personnelles sont traitées, du début à la fin de ce traitement (Bayle 2018).

4.2.3.3 Consentement

Pour être valable, le consentement doit être (Université de Lausanne 2020) :

- Libre (obtenu sans contrainte et non conditionné à l'octroi d'un avantage) ;

- Spécifique (recueilli pour chaque finalité envisagée aux traitements et non pour un ensemble de finalités) ;
- Eclairé (il faut connaître l'identité du responsable du traitement, son but, sa base légale etc. dans le respect du principe de reconnaissabilité) ;
- Univoque (il doit être « donné par un acte positif clair », soit « une indication active de choix », que ne sont pas le silence ou la continuité de l'utilisation d'un service par exemple).

Il n'y a donc pas de possibilité de consentement par défaut dans le RGPD et celui-ci est retirable à tout moment, possibilité qui doit être précisée à la personne concernée. Le consentement doit être exprimé de manière compréhensible, daté et signé pour en avoir une preuve. S'il est donné oralement, il sera donc obligatoire de l'enregistrer (Bayle 2018).

Les participants à une étude ont le droit de se retirer de celle-ci ou de retirer leur consentement au traitement de leurs données. Les deux processus n'ont pas les mêmes implications pour la conservation des données (Krügel 2019) :

- Le retrait de la participation a pour impact que la collecte de données n'est pas poursuivie mais que les données collectées auparavant restent valables ;
- Le retrait du consentement au traitement implique l'effacement des données personnelles, les données collectées avant le retrait restant légales si elles ne peuvent plus être liées au participant.

4.2.3.4 Données sensibles

Le traitement des données sensibles est interdit par le RGPD. Celles-ci sont les « données à caractère personnel qui révèle[nt] l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique » (art. 9, §1). L'interdiction ne s'applique pas s'il y a un motif d'exception (art. 9, §.2) ou si la personne concernée y a consenti (Commission de la protection de la vie privée - CPVP 2018).

4.2.3.5 Registre des traitements

La tenue d'un registre des traitements sous forme écrite par chaque responsable de traitement ou bien son représentant est obligatoire (art. 30, §1 et §3). Les entreprises de moins de 250 employés n'y sont normalement pas soumises, à quelques exceptions près. Le contenu du registre en question est détaillé à l'article 30, §1 (voir annexe 5). Les sous-traitants doivent tenir un registre similaire des activités de traitement qu'ils effectuent pour le compte d'un responsable de traitement (art. 30, §2).

4.2.3.6 Data Protection Officer (DPO)

Le RGPD instaure l'obligation de nommer un DPO dans le cas d'un organisme public qui traite des données personnelles ou bien d'une société privée qui pratique le traitement à grande échelle de données personnelles, notamment de données sensibles. Le rôle du DPO est d'informer, de conseiller, de contrôler le respect du règlement et d'être l'intermédiaire entre l'organisme public ou l'entreprise privée et l'autorité de protection des données (Bayle 2018). En Suisse, des institutions ou des entreprises ont également nommé des DPOs, à l'image de l'IHEID. L'Université de Genève, elle, recourt pour le moment de manière ponctuelle à un DPO externe, en particulier lorsque des projets européens en créent le besoin (Hausmann 2020).

4.2.3.7 Privacy Impact Assesement (PIA)

Le PIA (art. 35) intervient lorsqu'un traitement occasionne un grand risque pour les droits et les libertés des personnes concernées (CNIL 2019). En français, on parle d'étude d'impact sur la vie privée (EIVP). Les traitements de données personnelles qui remplissent au moins deux des neuf critères suivants doivent faire l'objet d'un PIA (CNIL 2019) : 1. évaluation ; 2. scoring ou profilage ; 3. décision automatique avec effet légal ou qui y ressemble ; 4. surveillance systématique ; 5. collecte de données sensibles ; 6. collecte de données personnelles à large échelle ; 7. croisement de données ; personnes concernées vulnérables (patients, enfants, personnes âgées par exemple) ; 8. utilisation d'une nouvelle technologie de manière innovante ; 9. exclusion du bénéfice d'un droit ou d'un contrat.

La réalisation d'un PIA participe à l'application du principe de *privacy by design*. Il a trois composants : une description détaillée du traitement, une évaluation de nature juridique de la proportionnalité de ce dernier et une étude de nature plus technique des risques sur la sécurité des données et leur impact sur la protection de la vie privée. Le PIA permet ainsi de déterminer les mesures techniques et organisationnelles à prendre pour protéger les données (CNIL 2019).

4.2.3.8 Droits conférés à la personne concernée

Le RGPD confère de nombreux droits à la personne concernée (Bayle 2018) :

- Le **droit à la transparence** (art. 13 et 14) permet de demander à l'organe qui traite les données lesquelles il possède, pour quelle raison et combien de temps il compte les garder. Normalement, l'exercice de ce droit se réalise par écrit dans un souci de traçabilité.
- Le **droit d'accès** (art. 15) stipule que la personne concernée doit avoir accès à toutes les données qui la concernent et savoir où elles sont transmises. Les données doivent être envoyées à la personne qui fait valoir son droit d'accès sous forme papier ou électronique, dans un format accessible et sous une forme aisément compréhensible.
- Le **droit à la rectification** (art. 16) est également conféré. Celle-ci doit être effectuée dans les plus brefs délais.
- Le **droit à la limitation du traitement** (art. 18) permet, dans le cas où la personne concernée remarque qu'une entreprise traite ses données de manière illicite ou ne remplit pas ses devoirs, de demander à mettre le traitement en pause jusqu'à ce qu'il soit régularisé. Ici, c'est le principe de licéité qui est en jeu.
- Le **droit à l'opposition** (art. 21) confère à la personne concernée le droit de s'opposer à un traitement de ses données personnelles. Le responsable du traitement peut les garder dans sa base de données mais il ne peut plus les utiliser, sauf s'il a un motif impérieux de continuer.
- Le **droit d'introduire une réclamation auprès de l'autorité de contrôle et à une action de groupe** (art. 77) implique qu'il est possible de chiffrer un préjudice et de demander des dommages et intérêts. Attention, en Suisse, l'action de groupe n'existe pas.
- Le **droit à la portabilité de ses données** (art. 20) permet, par exemple, de demander à son ancien opérateur téléphonique de transmettre ses données au nouveau en cas de changement.
- Le **droit à l'oubli** (art. 17) demande un effacement des données, contrairement au déréférencement qui a seulement pour conséquence que la donnée ne soit plus

référéncée sur le web. Il y a beaucoup d'exceptions au droit à l'oubli et en pratique, il est difficilement applicable.

4.2.3.9 Sous-traitance et communication de données

En cas de **sous-traitance**, le RGPD instaure une logique de responsabilité partagée entre tous les acteurs impliqués dans le traitement des données personnelles. Ainsi, les sous-traitants ont aussi des obligations spécifiques à remplir et des garanties à présenter, détaillées à l'article 28. Le responsable de traitement doit vérifier l'implication de ces garanties, par exemple avec un audit, et établir le contrat avec le sous-traitant. Il a également la charge de la coopération avec l'autorité de contrôle (Bayle 2018). Le sous-traitant doit, lui, respecter ses engagements et collaborer avec le responsable de traitement dans sa conformité (CNIL 2017). Le sous-traitant peut contractualiser avec des sous-traitants ultérieurs s'il en a l'autorisation (Bayle 2018).

La **communication** de données en dehors de l'UE par les responsables de traitements et les sous-traitants est possible tant qu'un niveau de protection équivalent est garanti (CNIL 2018). Pour l'instant, la Suisse est reconnue comme offrant un niveau de protection équivalente, mais la révision de la LPD est attendue par l'UE (24 heures 2020).

4.2.3.10 Devoir d'information

Les responsables de traitement ou leurs sous-traitants ont un devoir d'information envers les personnes concernées par les traitements de données personnelles qu'ils effectuent (art. 13 et 14). Cette communication doit être proactive et est obligatoire. Les informations doivent être communiquées dès le début de la collecte pour la collecte directe et au plus tard un mois après l'obtention des données personnelles si cela s'est fait de manière indirecte (Commission de la protection de la vie privée - CPVP 2018). Doivent être communiquées les informations suivantes : l'identité et les coordonnées du responsable de traitement et de son représentant si échéant ; les coordonnées du DPO ; les finalités du traitement et leur base juridique ; les destinataires ; la présence d'un transfert en dehors de l'UE si échéant ; la durée de conservation ; l'existence et les modalités des droits des personnes concernées ; l'existence d'une prise de décision automatisée si échéant et les finalités analogues compatibles avec la finalité initiale, si une base légale permet une réutilisation des données à ce but (Bayle 2018). Toutes ces informations doivent être rédigées de manière facilement compréhensible et être accessibles aisément (Commission de la protection de la vie privée - CPVP 2018). C'est pour cette raison que de nombreux sites web se sont dotés d'une page de déclaration de protection des données.

4.2.3.11 Exceptions à des fins de recherche scientifique

Le RGPD définit la recherche scientifique de manière non exhaustive et assez extensive au considérant⁵ 159 : « le traitement de données à caractère personnel à des fins de recherche scientifique devrait être interprété au sens large et couvrir, par exemple, le développement et la démonstration de technologies, la recherche fondamentale, la recherche appliquée et la recherche financée par le secteur privé ». L'importance des activités de recherche est ainsi reconnue dans le règlement. Le traitement de données dans ce contexte devra quand même respecter les principes de la protection des données personnelles (consid. 159) et reste « soumis à des garanties appropriées pour les droits et libertés de la personne concernée »

⁵ Selon *Le Robert dico en ligne*, un considérant est une : « considération qui motive un décret, une loi, et qui en précède le texte. »

(consid. 156). Une balance des intérêts doit donc être réalisée entre les objectifs de la recherche et la protection des données personnelles des participants (Maurel 2018).

Les chercheurs peuvent être exemptés de plusieurs dispositions du RGPD. Certaines de ces exceptions sont directement prévues par le règlement, tandis que d'autres peuvent être introduites par les Etats membres (Junod, Gorbacheva 2018). La recherche scientifique est en effet un des rares cas de figure où ceux-ci ont une marge de manœuvre pour émettre leurs propres dispositions (Bayle 2018).

Les **exceptions prévues par le RGPD** sont pour la plupart mentionnées à l'article 89, paragraphe 1. Il y a deux conditions à celles-ci (Leonard et al. 2019) :

- L'anonymisation n'est pas possible pour atteindre les finalités de recherche ;
- Des garanties appropriées permettant de protéger les données personnelles sont mises en place via des mesures techniques et organisationnelles, en particulier pour assurer le principe de minimisation. Par exemple en utilisant la pseudonymisation si les finalités peuvent être atteintes de cette manière.

Sont détaillés ici les exceptions de l'article 89 qui concernent tous les types de recherche :

L'exception au principe de limitation des finalités est mentionnée au considérant 50 et stipule : « le traitement ultérieur à des fins (...) de recherche scientifique (...) devrait être considéré comme une opération de traitement licite compatible. » Celui-ci peut ainsi être mené sur des données qui n'ont pas initialement été collectées à ces fins (Maurel 2018 ; Junod, Gorbacheva 2018). Cependant, la licéité du traitement doit toujours être acquise (CNIL 2019).

L'exception au principe de limitation de la conservation permet aux chercheurs de conserver des données personnelles aussi longtemps qu'ils envisagent d'autres utilisations possibles pour la recherche (Junod, Gorbacheva 2018). Cependant, ils doivent pouvoir établir une durée de conservation proportionnée à la finalité poursuivie (CNIL 2019).

L'exception aux règles applicables aux données sensibles veut que celles-ci puissent être légalement traitées si c'est nécessaire pour l'intérêt public, ce qui inclut la recherche scientifique (Junod, Gorbacheva 2018) et ce sans le consentement des personnes concernées (Maurel 2018).

L'exception au droit d'information lorsque les données ont initialement été collectées auprès de tiers dispense les chercheurs du devoir d'information lorsque des données personnelles, qui n'ont pas été collectées directement auprès des personnes concernées, sont réutilisées à des fins de recherche. Cette exception est valable seulement si le devoir d'information s'avère impossible à réaliser ou nécessite des efforts disproportionnés (Junod, Gorbacheva 2018), ce qui doit par ailleurs être démontré (CNIL 2019).

L'exception au droit à l'effacement veut que, si des données à caractère personnel sont traitées à des fins de recherche, le droit à l'effacement soit retiré à la personne concernée. Deux conditions doivent cependant être remplies pour bénéficier de l'exception : des garanties de sécurité appropriées doivent être mises en œuvre et il doit être démontré que la suppression des données rendrait la recherche scientifique impossible, ou lui porterait gravement atteinte. (Junod, Gorbacheva 2018)

Les autres exceptions prévues par le RGPD en dehors de l'article 89 sont les suivantes :

L'**exception au consentement spécifique** (consid. 33) : il est admis ici qu'il n'est pas toujours possible de déterminer avant le début d'une recherche scientifique la finalité exacte du traitement de données qui sera effectué. Les chercheurs ont donc une marge de manœuvre pour formuler les finalités, car celles-ci deviennent souvent plus nombreuses ou se précisent durant le projet de recherche (Maurel 2018). Mais normalement, la personne concernée doit toujours pouvoir personnaliser son consentement et n'accepter l'utilisation de ses données que pour certains types de recherche (Junod, Gorbacheva 2018) ou certaines parties du projet.

Les **exceptions qui peuvent être introduites par les Etats** sont précisées à l'article 89, alinéa 2. Les droits concernés par ces exceptions supplémentaires sont : le droit d'accès, le droit de rectification, le droit à la limitation du traitement et le droit d'opposition (Maurel 2018). Les potentielles dispositions nationales restrictives qui peuvent être introduites par des pays sont en réalité un défi pour la recherche, car il y a un risque de créer des incompatibilités entre les Etats européens, pourtant tous soumis au RGPD (Thorogood 2019, p. 107).

Parmi les **obligations maintenues** dans le contexte de la recherche scientifique, il y a celle de nommer un DPO lorsqu'on traite des données sensibles ou que le traitement est effectué par une autorité publique, celle de réaliser un PIA avant de traiter des données sensibles à large échelle et celle de notifier de fuites de données aux autorités de contrôle nationales compétentes (Junod, Gorbacheva 2018).

4.2.4 Quelle loi de protection des données personnelles s'applique lorsqu'on fait de la recherche ?

Plusieurs lois peuvent s'appliquer selon la situation. Tout dépend de la nature de l'institution qui mène la recherche et de la nature de la recherche elle-même. Si celle-ci entre dans le champ de la Loi fédérale sur la recherche sur l'être humain (LRH), c'est celle-ci qui s'appliquera. Pour les activités de recherche qui n'entrent pas dans le périmètre de la LRH, les cas de figure suivants sont possibles (Universität Basel - Research Data Management-Network 2019) :

- A l'Université de Genève ou dans une autre institution publique cantonale ou communale, c'est la loi cantonale de protection des données qui s'applique ;
- Pour une recherche menée en collaboration par deux institutions cantonales, la loi de protection des données de chaque canton s'appliquera pour le traitement de données effectué dans le canton en question ;
- Pour la recherche privée ou la recherche menée par les EPFs, c'est la LPD qui s'applique ;
- Pour la recherche à caractère international, souvent, la protection des données du pays partenaire en question s'applique aussi ;
- A l'international toujours, si le partenaire est européen, le RGPD s'applique dans deux cas de figure :
 - Le responsable de traitement ou le sous-traitant est établi dans l'UE ;
 - Le responsable de traitement n'est pas établi en Europe mais offre des services ou des biens à des personnes qui se trouvent sur le territoire européen, ou bien effectue un suivi du comportement de ces personnes.

4.2.5 Partager des jeux de données qui contiennent des données personnelles ou sensibles

Comme les données personnelles et sensibles sont protégées, l'idéal serait de les enlever des *datasets* à partager (OpenAIRE 2019a). S'il est impossible de les ôter, le partage de celles-ci se réalisera alors à des conditions précises pour garantir leur protection. Pour le Centre de compétences suisse en sciences sociales FORS, il existe trois mesures de protection qui peuvent être mises en place, seules ou en combinaison, pour partager ce type de données tout en compensant les risques inhérents (FORS 2018). La première mesure est l'anonymisation, puisque les données anonymisées ne présentent aucun risque en regard de la protection des données, n'étant plus concernées par celle-ci. Mais bien conscient que l'anonymisation totale des données est parfois difficile à atteindre ou bien qu'elle impacte négativement leur valeur scientifique, FORS recommande également de toujours obtenir le consentement éclairé des participants au partage de leurs données personnelles. Ceux-ci doivent être mis au courant des données qui seront partagées, de quelle manière cela sera fait et aussi des risques potentiels que cela implique (par exemple leur réidentification par croisement de données). Après cette information, les participants consentiront ou non à ce traitement de leurs données personnelles. Il faut donc que le formulaire de consentement ait une granularité suffisante, qui aborde l'archivage des données et leur utilisation future. La troisième et dernière mesure de protection disponible est le contrat d'utilisation. Dans le cas de FORS, cette protection est mise en place lors de l'accès aux *datasets* du dépôt FORSbase : l'utilisateur doit signer un contrat qui stipule qu'il s'engage à ne pas essayer d'identifier les personnes qui ont participé au projet de recherche. Le contrôle d'accès peut aussi être renforcé par des conditions supplémentaires ou la soumission de cet accès à l'accord préalable de quelqu'un. Dans ce cas, le contrat d'utilisation devra être rédigé au cas par cas (FORS 2018).

4.3 Résumé du cadre juridique lié aux données de la recherche

Le cadre légal applicable aux données de la recherche est vaste. Il s'agit principalement du droit d'auteur et de la protection des données personnelles, détaillés dans ce travail. Mais les données de recherche peuvent également être concernées par le Code des obligations pour les contrats (Vilmen 2018a), la LRH pour la recherche sur l'être humain (Donzé, Zouhri 2019), ou encore d'autres législations ou règlements :

Tableau 2 : Droit applicable aux données de la recherche

	Europe	Suisse	Cantons
Droit d'auteur		Loi sur le droit d'auteur (LDA)	
Protection des données personnelles	<p>Règlement européen sur la protection des données personnelles (RGPD).</p> <p>Il s'applique en Suisse s'il y a un traitement :</p> <ul style="list-style-type: none"> -Dans le cadre des activités d'une institution ou entreprise suisse qui a une succursale dans l'UE ; -Qui a lieu dans une entreprise suisse en sous-traitance pour une entreprise de l'UE ; -De données de résidents de l'UE par une entreprise suisse qui offre des biens et des services dans l'UE ou qui suit le comportement de résidents de l'UE. 	Loi sur la protection des données personnelles (LPD) pour une collecte de données par le secteur privé et les organismes fédéraux.	<p>A Genève, la Loi sur l'information du public, l'accès aux documents et la protection des données personnelles (LIPAD), pour une collecte de données issues d'organismes cantonaux.</p> <p>Autres lois cantonales, si existantes, sinon la LPD s'applique à titre supplétif.</p>
Contrats		Code des obligations (CO)	
Recherche sur l'être humain		<p>Liste non exhaustive :</p> <p>Loi fédérale relative à la recherche sur l'être humain (LRH) ;</p> <p>Ordonnance fédérale relative à la recherche sur l'être humain (ORH) ;</p> <p>Ordonnance fédérale sur les essais cliniques (Oclin) ;</p> <p>Loi fédérale sur l'analyse génétique humaine (LAGH).</p>	<p>Genève : Loi sur la santé (LS) ;</p> <p>Autres lois cantonales, si existantes.</p>

La difficulté dans l'analyse du cadre juridique applicable aux données de la recherche réside également dans leur forme et dans l'évolution de celle-ci, qui peut occasionner un changement de régime légal (Benhamou, Tran 2016, p. 577) :

« À titre d'exemple, des données brutes, lorsqu'elles sont agrégées, doivent être traitées comme des données personnelles si l'agrégation permet de remonter à une personne identifiable (p. ex. désanonymisation de données brutes par agrégation), comme un bien protégé par le droit d'auteur si l'agrégation acquiert une certaine individualité (p. ex. algorithmes protégés en tant qu'œuvre ou compilation de données protégée en tant que recueil) ou encore protégé contre toute appropriation déloyale. »

En termes de propriété intellectuelle, les questions qui se posent généralement avec les données de la recherche sont (complétée de Carroll 2015) :

1. Qu'est-ce qui est protégé par le droit d'auteur ?
2. Quels droits sont liés aux données ?
3. Qui les possède ?
4. Comment celui qui les possède peut les utiliser pour partager ses données de manière à permettre ou encourager leur réutilisation ?

Pour réponse à ces questions avec la LDA, il est d'abord possible d'affirmer que les données de la recherche peuvent être protégées à plusieurs niveaux : celui des données individuelles, de la base de données, des métadonnées ou encore des visualisations de données. A chaque fois, pour savoir si c'est un contenu protégé ou non, il faut analyser s'il remplit les critères d'une œuvre, soit être une création de l'esprit qui a un caractère individuel. Généralement, dans le cas d'une œuvre scientifique, c'est plus l'individualité de la forme que du contenu qui est protégé. Dans tous les cas, les données scientifiques factuelles brutes, elles, ne sont pas protégées (par exemple formule mathématique, relevés de température, données géologiques etc.) (Labastida, Margoni 2020) :

« (...) The legal system has designed a mechanism whereby the basic bricks of knowledge such as ideas or factual information are freely available to all in order to learn and advance the knowledge in a field. But this is balanced by the possibility to protect some of the results obtained using those ideas and facts: original expressions of unprotected ideas and substantial amounts of obtained data (...) arranged. »

La LDA confère deux types de droit à l'auteur : les droits patrimoniaux (décider quand, si et comment son œuvre sera utilisée) et les droits moraux, irrévocables, qui protègent la personnalité de l'auteur. Quant à savoir qui est titulaire des droits sur les données de la recherche, cet élément est souvent régi par une loi ou un règlement émanant de l'institution ou l'entreprise qui emploie le chercheur, il faudra donc résoudre cette question au cas par cas. Dans l'éventualité où un chercheur veut réutiliser des données protégées par la LDA, actuellement il est face à trois options (Baggi [sans date], p. 4) :

1. Renoncer à l'utilisation de ces données ;
2. Utiliser une exception légale prévue ;
3. Demander une autorisation à l'auteur ou au titulaire des droits.

En l'occurrence, il n'existe qu'une exception spécifiquement pour la recherche, introduite récemment : la reproduction d'œuvres protégées est autorisée lorsqu'elle est nécessaire à l'application d'un procédé technique, ce qui permet la pratique du *text and data mining*. Les autres exceptions prévues de longue date dans la LDA, soit celles sur l'utilisation d'une œuvre à des fins privées (pour l'usage strictement personnel ou à des fins pédagogiques) et de reproduction à des fins d'information interne et de documentation, peuvent aussi être évoquées dans le contexte de la recherche. Cependant, dans les faits, comme elles sont assez limitées, elles ne permettent pas une utilisation d'œuvres protégées qui atteignent les objectifs visés par le mouvement de l'Open Science. Ainsi, les licences ont un rôle très important dans la facilitation de la science ouverte en autorisant l'utilisation des données de manière plus ou moins libre, selon les souhaits de leur auteur. Il n'y a alors plus besoin de demander l'autorisation préalablement à la réutilisation de données ou d'utiliser la petite marge de manœuvre offerte par les exceptions prévues dans la LDA. Les licences permettent aussi de résoudre les différences entre les droits d'auteur nationaux en simplifiant ce qui est permis ou

non selon le lieu d'utilisation (puisque ce droit s'applique de manière territoriale), tout comme la question du droit *sui generis* des bases de données, présent en Europe mais inconnu dans de nombreux autres pays, dont la Suisse.

Concernant la protection des données personnelles, il est difficile de résumer tous les aspects du sujet en lien avec les données de la recherche, tant il est complexe. De plus, le fait que la protection des données personnelles soit réglementée à plusieurs niveaux (cantonal, national, européen), qui s'appliquent différemment selon le contexte de la recherche, rend la tâche encore plus compliquée. D'une manière générale, il est possible de dire que la LPD et le RGPD offrent des exceptions pour un traitement de données personnelles et sensibles à des fins de recherche, tandis que la LIPAD genevoise est assez contraignante, notamment dans le cas de collaborations nationales ou internationales. Deux éléments sont importants à souligner à propos de la protection des données et de la recherche : le rôle du consentement éclairé et la pratique de l'anonymisation des données. Il est de manière générale recommandé d'offrir aux participants d'une recherche la possibilité de consentir avec une granularité fine et après information sur les usages potentiels et futurs de leurs données. Le consentement devra également porter sur l'archivage et le partage des données, notamment dans un dépôt. Puisque le consentement peut être retiré à tout moment, il conviendra également d'étudier l'impact d'une telle rétractation de la recherche, notamment sur les données déjà collectées (Donzé, Zouhri 2019) ou archivées. Normalement, en cas de retrait du consentement, il faut retirer du *dataset* les données de la personne qui en a fait la demande, ou du moins, les anonymiser. A ce propos, il est généralement conseillé de réfléchir si la collecte de données personnelles est absolument nécessaire à la recherche et de s'en abstenir si ce n'est pas le cas, ou bien de les anonymiser dès que la recherche le permet (Donzé, Zouhri 2019). De même, lors de la publication des jeux de données, la règle est de ne faire paraître que des données anonymes, qui ne sont alors plus concernées par la protection des données personnelles.

5. Licences

Les licences jouent un rôle très important pour les données de la recherche car elles explicitent ce que l'auteur accorde quant à leur utilisation, encourageant par là leur réutilisation. Les licences ne règlent que les questions de droit d'auteur et de droit des bases de données, et non de protection des données personnelles. Ces questions doivent être réglées à part. A nouveau, l'anonymisation des données personnelles est recommandée lors de la publication pour protéger au mieux les personnes concernées.

Dans cette partie j'ai fait le choix de me concentrer uniquement sur les licences ouvertes pour les données, en accord avec les objectifs que souhaite atteindre le mouvement de l'Open Science, dont fait partie l'ouverture des données de la recherche.

5.1 Enjeux

5.1.1 Pourquoi publier ses données sous licence ?

Il est important de publier son jeu de données sous une licence afin de garantir son interopérabilité, notamment légale (Thorogood 2019, p. 40). La *Research Data Alliance* définit l'interopérabilité légale de la sorte : « the ability to combine data from two or more sources without conflicts among restrictions imposed by data providers ... and without having to seek authorization from the data providers on a case-by-case basis. » (Doldirina et al. 2018 cités par Thorogood 2019, p.62). Ainsi, rendre disponible ses données sans mentionner clairement comment elles peuvent être utilisées est contre-productif car il n'y a pas de consensus légal clair à ce propos en raison des différentes juridictions nationales, qui ont chacune leurs critères pour juger ce qui est une œuvre protégée ou non par le droit d'auteur (créativité, originalité, compétences nécessaires, travail ou encore frais investis) (Ball 2014, p. 2). De plus, ces juridictions émettent des exceptions différentes au droit d'auteur et interprètent parfois diversement les mêmes concepts, par exemple ce qui est entendu par « extrait » ou « partie » d'une œuvre (Ball 2014, p. 3). En l'absence de notice légale, c'est donc le « tous droits réservés » qui prévaut par défaut (Labastida, Margoni 2020), ce qui est problématique dans le contexte de l'Open Science.

Les deux moyens les plus efficaces pour lever toutes ces incertitudes légales, pires ennemies de la réutilisation (Cassella 2013), sont les licences et la renonciation au droit d'auteur (Ball 2014, p. 3). Une licence de droit d'auteur est un contrat dans lequel un auteur précise les modalités d'utilisation qu'il permet de son œuvre. La licence offre à l'auteur la possibilité de ne pas transférer ses droits d'auteur, tout en concédant des droits d'utilisation non exclusifs. Les personnes qui prennent la licence n'accèdent donc pas au statut juridique d'auteur (CCDigitalLaw 2017e) mais peuvent utiliser l'œuvre aux conditions stipulées par la licence. Il n'y a pas besoin d'un accord entre les deux parties, le preneur de licence acceptant tacitement les termes du contrat. Une renonciation au droit d'auteur (un *waiver*) est un outil qui permet de mettre son œuvre dans le domaine public, ou bien de s'en approcher au plus près si la législation nationale n'autorise pas l'auteur à renoncer à tous ses droits. En Suisse, en l'occurrence, comme les droits moraux sont irrévocables, un auteur se trouve devant l'impossibilité de placer son œuvre volontairement dans le domaine public, une renonciation à ses droits n'ayant aucun effet. C'est pour cette raison que les renonciations aux droits d'auteur contiennent des clauses supplémentaires qui permettent de les faire fonctionner comme des licences dans les pays qui ne permettent pas de renoncer à ces droits, comme la Suisse.

Seule la personne qui détient les droits d'auteur peut placer une œuvre sous une licence ou dans le domaine public : il est donc impératif d'avoir établi quels droits de propriété intellectuelle sont liés aux données au préalable (Ball 2014, p. 3). Dans le cas d'une base de données, cela veut dire de contrôler au cas par cas tous les droits d'auteur qu'il pourrait y avoir (Vilmen 2018a). Les licences ne s'appliquent pas aux œuvres de tiers qui ont été cités ou dont le travail a été réutilisé : ceux-ci conservent leurs droits (OpenAIRE 2019a). Il conviendra donc soit d'enlever les données dont les droits de propriété intellectuelle appartiennent à un tiers (OpenAIRE 2019a), soit de respecter l'éventuelle licence sous laquelle les données ont été mises à disposition, soit de contacter ces tiers pour leur demander l'autorisation quant à la réutilisation envisagée de leur œuvre. L'auteur d'une œuvre qui la met sous licence est également tenu de respecter les conditions de la licence qu'il a choisi sur sa propre œuvre.

Il est généralement recommandé d'attribuer une licence standard à son jeu de données, c'est-à-dire une licence déjà existante, qui a un texte légal élaboré d'avance et adapté à plusieurs types de contenus. Les licences standards s'opposent aux licences individuelles (*custom license*), qui sont rédigées sur mesure. Cet exercice est une tâche complexe qui nécessite normalement l'intervention d'un service juridique. Les licences standards, elles, ont l'avantage d'offrir une explication en langage clair sur les utilisations qui sont permises ou non ainsi que les obligations à respecter. C'est utile en premier lieu pour l'auteur, qui, la majeure partie du temps, n'est pas un expert juridique et ne comprend donc pas ce qui est autorisé ou non. Cette compréhension aisée est aussi utile en second lieu pour le preneur de licence. Adopter une licence standard simplifie donc la transaction juridique entre l'auteur et l'utilisateur (Kreutzer 2014, p. 15), tout en favorisant aussi l'interopérabilité des données. Il existe trois grandes catégories de licences standards (Thorogood, 2019, p. 48) :

- Les licences permissives, qui autorisent la réutilisation, la copie, la modification et la distribution de l'œuvre sans restriction (à l'exception éventuellement d'une attribution à l'auteur) ;
- Les licences restrictives, qui imposent des conditions à la réutilisation, comme par exemple l'interdiction d'un usage commercial ou de la modification ;
- Les licences virales, appelées aussi copyleft, qui exigent que l'œuvre dérivée soit disponible sous la même licence. Celles-ci peuvent à la fois être qualifiées d'ouvertes (parce que l'œuvre et les éventuelles adaptations restent toujours ouvertes) et de restrictives (parce qu'elles restreignent la liberté de l'utilisateur à concéder une licence pour toute adaptation qu'il aurait faite, sous quelque forme que ce soit).

5.1.2 Des licences ouvertes pour des données ouvertes

Selon l'*Open Definition* (2015) de l'*Open Knowledge Foundation*, déjà citée au début de ce travail, des données sont ouvertes si elles peuvent être **librement utilisées, modifiées et partagées par tout le monde dans n'importe quel but**. Cette définition des données ouvertes, communément admise par d'autres organismes et reprise dans la littérature, a une implication directe sur les licences. Il n'y a en effet que deux types de restrictions qui sont alors acceptables : l'obligation d'attribution à l'auteur original et/ou une licence virale de type partage à l'identique (copyleft) (Dodds 2013). Une licence ouverte peut donc être soit de domaine public, soit permissive mais demandant l'attribution, soit copyleft, soit les deux derniers types à la fois.

Le but de rendre les données de la recherche ouvertes est de favoriser leur réutilisation et la découverte de connaissances nouvelles qui pourraient en émerger, dans l'esprit du

mouvement Open Science. Les licences permettent d'établir une transparence légale à la fois pour l'auteur et les utilisateurs (Kreutzer 2014). Réserver tous les droits sur ses données ou les mettre sous une licence non ouverte crée au contraire une sorte de chaos juridique. Une récente étude sur le potentiel de réutilisabilité de *datasets* dans le domaine biomédical résume bien la situation (Carbon et al. 2019) :

« A lack of licensing rigor and standardization forces data users to manually seek, often repeatedly and from multiple data providers, essential reuse and redistribution permissions. Issues include missing licenses, nonstandard licenses, and license provisions that are restrictive or incompatible. The legal interpretation of, and compliance with, database license and reuse agreements has become a significant burden and expense for many fields in the scientific community, where a complex and lengthy set of legal negotiations may be required for a data integration project to legally and freely redistribute all of its relevant data. »

Cette étude, nommée *(Re)usable Data Project*⁶, a évalué 64 bases de données biomédicales sur une échelle allant de 1 à 5 étoiles selon les critères suivants : la possibilité de trouver les données et le type de conditions de la licence, la portée et l'exhaustivité de la licence, la possibilité d'accéder aux données de manière raisonnable, les restrictions sur la manière dont les données peuvent être réutilisées et les restrictions sur les personnes qui peuvent réutiliser les données. Les résultats ont démontré des obstacles significatifs à la réutilisation et à l'interopérabilité des données actuellement. Le type de licence le plus important parmi celles utilisées dans les bases de données étudiées est par ailleurs la licence sur mesure (*custom license*) avec 34,4 %, dont la majorité est, de plus, de type restrictif⁷ (Carbon et al. 2019).

Actuellement, en Suisse, les bailleurs de fonds ne se prononcent pas en faveur d'un type de licence plutôt que d'un autre. Par exemple, le FNS exige que les jeux de données soient accompagnés d'une licence claire qui explicite sous quelles conditions ils peuvent être réutilisés ou non, et qui soit lisible par les machines comme les humains, ceci afin de respecter les principes FAIR (FNS 2017b). Les hautes écoles n'ont pas non plus émis des obligations quant aux licences dans leurs stratégies ou directives, pour celles qui en possèdent. En revanche, les services de soutien à la recherche de certaines de ces hautes écoles recommandent l'usage de certaines licences plus que d'autres, et en l'occurrence se sont des licences ouvertes. Certains dépôts ont aussi fait le choix d'imposer des licences ouvertes.

5.1.2.1 Les licences non ouvertes

Il existe deux types de licences non ouvertes couramment utilisées, les licences interdisant l'usage commercial et celles ne permettant pas la modification. Ces types de licences ne sont pas recommandées pour l'Open Research Data.

5.1.2.1.1 Les licences interdisant l'usage commercial

Une clause interdisant l'usage commercial signifie que l'auteur se réserve le droit d'exploiter commercialement son œuvre. Cette restriction n'est pas jugée ouverte car un contenu doit pouvoir être réutilisé dans n'importe quel but pour être ouvert. L'application de licences qui interdisent l'utilisation commerciale s'avère de plus très compliquée en raison de l'ambiguïté du terme « utilisation commerciale » (Ball 2014, p. 4), qu'il est difficile de définir légalement de façon précise (OpenAIRE 2019b). Selon l'interprétation qui en est faite, ces licences peuvent ou non empêcher que les données soient utilisées en contribution à des œuvres pour

⁶ Voir : <https://reusabledata.org/>

⁷ Résultats disponibles à l'adresse : <https://reusabledata.org/explore.html>

lesquelles un auteur est rémunéré (Ball 2014, p. 4). Par exemple, une telle clause pourrait interdire à un scientifique de réutiliser des données dans le cadre de ses recherches s'il a l'intention d'en publier les résultats par la suite. En effet, puisque la plupart des revues scientifiques sont détenues par des éditeurs commerciaux, même si l'auteur de l'article ne tire aucun revenu financier de sa publication, cette utilisation des données pourrait être qualifiée de commerciale. Cette situation, puisqu'elle affecte la diffusion, la reconnaissance et l'impact du jeu de données, fait qu'il n'est pas en accès ouvert (OpenAIRE 2019b). Un tel type de licence se montre aussi problématique dans le cas de partenariats public-privé (Kreutzer 2014, p. 47), qui sont courants dans le milieu de la recherche. Pour avoir une idée de ce qui est vraisemblablement permis ou non avec une licence interdisant l'usage commercial, voir le tableau en annexe 6, basé sur la licence de ce type chez Creative Commons.

Les licences non commerciales se justifient plutôt dans le cas de donneurs de licence ou de créateurs d'entreprise déjà bien connus et prospères. Dans ces situations, elles peuvent être utilisées pour soutenir une stratégie marketing élaborée. Elles sont donc plutôt recommandées s'il y a des perspectives réalistes de commercialisation. Un musicien, par exemple, pourrait mettre à disposition un extrait d'une de ses chansons (*abridged version / light version*) sous licence non commerciale pour attirer l'attention sur son œuvre et la commercialiser dans sa version complète par la suite si le succès est au rendez-vous. De plus, pour être cohérent, il faudrait que le donneur de licence soit prêt à intenter une action en justice en cas de violation de la licence pour faire valoir ses droits (Kreutzer 2014, p. 48), ce qui n'est probablement pas le cas de la majorité des personnes qui utilisent ce type de clause.

5.1.2.1.2 Les licences interdisant la modification

La restriction interdisant la modification dans certaines licences n'est pas jugée ouverte non plus, car elle restreint de manière significative la manière dont un jeu de données peut être réutilisé, notamment en combinaison avec d'autres. Pas de modifications ne veut pas dire qu'il est absolument impossible de modifier une œuvre, l'exception d'usage personnel pouvant être utilisée, mais qu'il est interdit de diffuser publiquement l'œuvre modifiée. Une modification opère lorsque l'œuvre est traduite, altérée, arrangée, transformée ou modifiée d'une manière qui demande une autorisation selon le droit d'auteur. Il est donc quand même permis de changer le format ou de numériser une œuvre disponible sous licence interdisant la modification, car l'œuvre elle-même reste inchangée. L'interprétation d'une telle clause se complique lorsqu'il s'agit de combiner l'œuvre avec une autre, et elle ne peut être résolue qu'au cas par cas en étudiant la loi en vigueur (Kreutzer 2014, p. 49-50).

Dans le cas de données de la recherche, le fait d'interdire les modifications est souvent utilisé par crainte que les données ne soient faussées ou déformées ultérieurement, ce qui indirectement pourrait entacher la réputation de l'auteur initial. En réalité, cette question est déjà réglée dans la plupart des licences ouvertes, qui interdisent de suggérer que les données ou les analyses dérivées ont un quelconque statut officiel, et qui précisent que l'auteur original des données ne donne pas son approbation aux actions de ses réutilisateurs. Dans les faits, une telle clause est problématique car elle empêche a priori de modifier des erreurs dans le jeu de données, de l'enrichir avec des données additionnelles ou encore de le combiner avec d'autres jeux de données (The Open Data Institute 2015). Puisque les données existent la plupart du temps pour être croisées entre elles, une telle clause est à éviter pour ce genre de contenu. De plus, la recherche scientifique est presque toujours basée sur des travaux antérieurs (Hirschmann 2020), d'où le besoin de pouvoir combiner des données. Ici aussi, il

faudrait également que le donneur de licence soit prêt à intenter une action en justice en cas de violation de celle-ci.

Un des rares cas où l'utilisation d'une licence de type ND se justifie pleinement est celui des certifications ou des normes, qui ne peuvent et ne doivent être modifiées que par l'institution de certification, sinon elles ne sont plus valables (Kreutzer 2014, p. 52-53). Une norme ISO modifiée n'aurait plus aucun intérêt par exemple.

5.1.2.2 Les licences ouvertes mais néanmoins problématiques

Pour satisfaire la définition de licence ouverte, une licence peut donc être : soit de domaine public, soit permissive mais demandant l'attribution, soit copyleft, soit les deux derniers types à la fois. Mais en réalité, les licences qui demandent l'attribution, celles qui sont copyleft, ou les deux à la fois posent également des problèmes.

5.1.2.2.1 Les licences demandant une attribution

L'obligation de l'attribution est une condition admise pour une licence ouverte. Mais des arguments s'élèvent quand même en sa défaveur dans le contexte des données de la recherche. Le problème principal est connu sous le nom d'empilement des attributions (*attribution stacking*). En effet, si un chercheur réutilise de nombreux jeux de données, qui ont chacun un grand nombre de contributeurs, il devra tous les créditer lorsqu'il mettra son jeu de données dérivé à disposition et la liste des crédits sera très conséquente et deviendra difficile à gérer (Ball 2014, p. 4). Ces attributions ne feront que de s'accumuler encore si quelqu'un d'autre réutilise ce jeu de données dérivé pour le combiner avec d'autres qui exigent aussi l'attribution, et ainsi de suite. De plus, si la méthode de compilation des données est automatique et que les exigences sur le format de la citation diffèrent entre les différents contributeurs, la liste des crédits devra être faite manuellement, et tout l'avantage de l'automatisation sera perdu (Ball 2014, p. 4).

Le cas des licences qui demandent l'attribution soulève aussi la question de savoir si les chercheurs doivent ou non articuler des normes communautaires, soit ici la citation par les pairs, sous une forme légale. Cette interrogation est un élément central du débat sur les licences pour les données de la recherche. La *Research Data Alliance* est elle-même assez sceptique des licences qui demandent une attribution, jugeant que les mettre dans le domaine public est préférable pour éviter que cette simple condition ne mène à une accumulation des attributions au fur et à mesure que les données sont combinées, modifiées et redistribuées. A ses yeux, l'attribution devrait être traitée comme une norme éthique plutôt qu'une condition légale (Thorogood 2019, p. 76). Il ne faut pas oublier que, la plupart du temps, les données de la recherche ne sont pas concernées par le droit d'auteur et que le droit des bases de données n'existe pas en Suisse. Il est donc inutile d'exiger une attribution qui est en réalité juridiquement inapplicable et qui peut être ignorée. De plus, l'éthique scientifique et les normes communautaires prônent déjà la citation, et pour des raisons qui n'ont rien à voir avec le droit d'auteur. En effet, des données qui proviennent apparemment de nulle part n'ont que peu de crédibilité scientifique. Si quelqu'un veut les utiliser en tant que preuves convaincantes, il devra indiquer à ses lecteurs ou examinateurs leur source (Fortney 2016), ne serait-ce que pour garantir l'intégrité de ses résultats scientifiques et la transparence sur sa méthode.

5.1.2.2.2 Les licences de partage à l'identique (copyleft)

Exigeant que le jeu de données dérivé soit mis à disposition sous la même licence que le jeu de données original, ces licences sont reconnues comme ouvertes et sont souvent utilisées

pour forcer le preneur de licence à diffuser son jeu de données dérivé également sous une licence ouverte (The Open Data Institute 2015).

Ce type de licence pose néanmoins un gros problème : celui de la compatibilité des licences copyleft entre elles, qui entrave l'interopérabilité des données. En effet, un nouveau jeu de données créé de la combinaison de deux jeux de données sous deux licences copyleft différentes ne pourra pas être publié en respectant les conditions de ces deux licences en même temps, puisque chacune exige de mettre à disposition l'œuvre dérivée seulement avec la même licence (Ball 2014, p. 4). Certaines licences copyleft montrent une certaine flexibilité en regard de ce problème en indiquant des licences compatibles (Ball 2014, p. 4), mais celles-ci sont généralement peu nombreuses.

Les clauses copyleft ne font en réalité qu'accentuer le problème d'incompatibilité des licences, déjà causé par leur prolifération. Le fait que de nombreuses familles de licences cohabitent n'aide en rien à réduire les incertitudes légales à leur propos. Avant d'appliquer une licence copyleft, il conviendra donc de déterminer ce qui est le plus important : assurer que l'œuvre reste ouverte ou encourager le plus d'intérêt possible dans son utilisation (Kreutzer 2014, p. 55), qui est retenue par une condition copyleft.

5.2 Les licences ouvertes pour les données de la recherche

Dans cette partie sont détaillées seulement les licences qui sont conformes à la définition des données ouvertes de l'*Open Knowledge Foundation* et adaptées aux données et/ou aux bases de données. Un tableau comparatif résumant les caractéristiques de toutes les licences qui sont passées en revue est disponible en annexe 8.

5.2.1 Les licences Creative Commons⁸

Ce sont des licences de droit d'auteur pour des œuvres (CCDigitalLaw 2017f). Le projet est né en 2001 aux USA sur l'initiative de juristes. Creative Commons (CC) est une organisation à but non lucratif. L'idée est de se positionner entre le domaine public (*no copyright* / pas de droits) et le droit d'auteur (*all rights reserved* / tous droits réservés) en réservant seulement certains droits (*some rights reserved*). Les licences CC réglementent donc l'utilisation par des tiers de certains droits sur l'œuvre (Cassella 2013). Leur principe est « tout ce qui n'est pas interdit est permis », alors que la logique du droit d'auteur est plutôt « tout ce qui n'est pas permis est interdit » (Vilmen 2018b). En tant que licences standards, ce sont des contrats simples et préformulés (CCDigitalLaw 2017f), qui permettent de céder certains droits patrimoniaux à titre gratuit (Favre, Germond 2018, p. 21).

Les licences CC sont conçues en trois couches (Favre, Germond 2018, p. 20) :

1. Le code juridique (*legal code*), qui définit les licences en termes juridiques ;
2. Le résumé explicatif (*common deed*), aussi appelé la version « lisible par les humains », qui résume le code juridique en termes simples et clairs à l'attention de non-juristes ;
3. Le code numérique (*digital code*), aussi appelé la version « lisible par les machines », soit « un résumé des autorisations et obligations de la licence écrit dans un langage que les logiciels, moteurs de recherches, et autres types de technologies peuvent comprendre » (Creative Commons 2017).

⁸ Toutes les icônes de cette partie proviennent de Creative Commons et sont sous licence CC-BY 4.0. Elles sont disponibles à l'adresse : <https://creativecommons.org/about/downloads/>




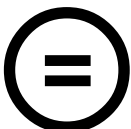
Au départ, ces licences ont été pensées pour des contenus comme la musique, les images ou les vidéos (Ball 2014, p. 6). Le droit des bases de données européen n'était pas abordé du tout, les Etats-Unis ne connaissant pas ce droit, ce qui était problématique dans le contexte de la recherche. CC a depuis corrigé ce manquement dans une nouvelle version de ces licences. Elles sont désormais largement utilisées pour tous types de contenus et adaptées aux données de la recherche. Les licences CC ne sont en revanche pas adaptées aux logiciels (Kreutzer 2014, p.12).

La dernière version des licences CC est la 4.0, sortie le 21 novembre 2013. Elle est internationale. Les anciennes versions, elles, étaient « portées » dans les juridictions nationales, c'est-à-dire non seulement traduites, mais également adaptées linguistiquement et légalement à la juridiction du pays en question (Kreutzer 2014, p. 32). Si cette approche servait le donneur de licence (qui avait un code légal dans sa langue, adapté au droit d'auteur de son pays et qui indiquait clairement aux preneurs de licences quelle était la juridiction applicable), elle a cependant été arrêtée car elle pouvait créer des incertitudes auprès des preneurs de licences étrangers à la langue et au droit dans lesquels la licence a été portée. Les licences CC 4.0, elles, sont valables internationalement, ce qui se montre bénéfique notamment pour les contenus en ligne ou les projets multinationaux ou collaboratifs (par exemple Wikipédia) (Kreutzer 2014, p. 33-34). Elles ont quand même été traduites dans plusieurs langues. Les versions des licences 3.0 portées, par exemple celles adaptées à la juridiction suisse, peuvent toujours être utilisées et se montrent pratiques dans certains cas, mais pas dans le contexte de la recherche scientifique effectuée au niveau international.

Les licences CC s'appliquent selon ce que prévoit le droit d'auteur. Cela veut dire que des chercheurs qui appliqueraient une licence CC sur des contenus non protégés par le droit d'auteur (comme des données factuelles scientifiques) outrepassent les droits qui sont les leurs sur ces contenus (Thorogood 2019, p. 72). Si la licence impose des restrictions ou des conditions sur un contenu qui n'est pas protégé par le droit d'auteur, il est en réalité possible d'utiliser ce contenu en ignorant totalement la licence, qui n'a pas lieu d'être. Il faudrait donc utiliser les licences CC que lorsqu'il y a de manière certaine des droits d'auteur sur les données (Thorogood 2019, p. 72).







Il existe six licences CC composées de quatre éléments différents. Ceux-ci règlent la portée des droits d'utilisation concédés et les conditions à remplir lorsqu'une reproduction ou un dérivé de l'œuvre sous licence CC est distribué, c'est-à-dire vendu, donné ou mis en circulation (CCDigitalLaw 2017f) :

Tableau 3 : Les quatre éléments des licences CC

	BY Attribution
	SA (<i>share-alike</i>), condition de type copyleft Partage à l'identique (sous la même licence)
	NC (<i>non commercial</i>) Pas d'utilisation commerciale
	ND (<i>no derivatives</i>) Pas de modification

Les six licences CC sont issues des combinaisons de ces quatre éléments différents. Toutes contiennent l'attribution :

Tableau 4 : Les six licences CC

Logo de la licence	Nom de la licence	Type de la licence
	CC-BY 4.0	Attribution
	CC-BY-SA 4.0	Attribution – partage dans les mêmes conditions
	CC-BY-NC 4.0	Attribution – pas d'utilisation commerciale
	CC-BY-NC-SA 4.0	Attribution – pas d'utilisation commerciale – partage dans les mêmes conditions
	CC-BY-ND 4.0	Attribution – pas de modification
	CC-BY-NC-ND 4.0	Attribution – pas d'utilisation commerciale – pas de modification

Les licences en rouge ne sont pas ouvertes car elles contiennent des clauses NC et/ou ND. N'étant pas recommandées pour les données de la recherche, celles-ci ne sont pas présentées plus en détails dans le cadre de ce travail.

Une licence ne peut pas contenir à la fois une clause SA et une clause ND, car la clause SA ne s'applique qu'aux œuvres dérivées, qu'empêche la clause ND (Meade 2020). La compatibilité des licences CC entre elles est disponible en annexe 7 de ce travail.

Les points communs à toutes ces licences sont les suivants (Kreutzer 2014, p. 35-43 ; CCDigitalLaw 2017g) :

- L'octroi d'une licence non exclusive, irrévocable⁹, libre de redevances et mondiale est accordé. L'auteur accorde au moins : le droit de reproduction, le droit de diffusion, le droit de mise en circulation et le droit de rendre l'œuvre perceptible et accessible ;
- Il n'est pas possible d'émettre des sous-licences ;
- Les conditions de la licence n'entrent en vigueur que lorsqu'une utilisation entre dans le champ d'application des droits exclusifs du droit d'auteur. Pour une utilisation qui n'est pas soumise au droit d'auteur, par exemple l'utilisation d'une exception à ce droit ou l'inexistence de droit *sui generis* des bases de données, l'utilisateur n'est pas tenu de respecter les obligations de la licence ;
- Les licences CC s'appliquent pour les droits d'auteur et le droit *sui generis* des bases de données ;
- Les brevets et le droit des marques ne peuvent pas être cédés sous licence CC ;
- Les droits moraux de l'auteur sont abandonnés au maximum de ce qui est permis par la juridiction applicable ;
- Toutes les licences mentionnent une clause d'exonération de garantie et de responsabilité. L'œuvre est donc partagée "telle quelle" et l'auteur n'est pas responsable des dommages ou des pertes qui pourraient résulter de l'utilisation de son œuvre ;
- Il est interdit de déployer des mesures techniques de protection (*Technological Protection Measures*) supplémentaires qui pourraient empêcher des tiers de procéder aux utilisations permises par la licence. Par exemple, il n'est pas possible de redistribuer une œuvre sur un CD protégé contre la copie, ou bien mettre en place des DRM (*Digital Rights Management*). Tout licencié est autorisé à contourner une telle mesure technologique de protection si besoin ;
- Les licences sont en vigueur jusqu'à ce que l'œuvre tombe dans le domaine public. Elle est révoquée automatiquement s'il y a une violation des conditions. La licence est réinstaurée automatiquement si le preneur de licence remédie à sa violation dans les 30 jours.

5.2.1.1 Licence CC-BY

Avec cette licence, toute exploitation de l'œuvre est possible. Cela comprend : partager, copier, reproduire, distribuer, communiquer, traduire, réutiliser et adapter l'œuvre, par tous les moyens, sous tous les formats possibles et à toutes fins. Il est possible de redistribuer ensuite son œuvre dérivée sous n'importe quelle licence (Favre, Germond 2018, p. 21).

Les obligations liées à l'attribution sont les suivantes (Kreutzer 2014, p. 38-40 ; Jones 2018) :

⁹ Il est possible d'arrêter de distribuer son œuvre sous licence CC, ou de changer de licence, mais les exemplaires déjà disponibles sous la licence CC originale restent valables (Vilmen 2018b ; Guibault, Wiebe 2013). Il est dans les faits impossible de révoquer une licence CC car « juridiquement parlant, tous les contrats de licence avec toute personne qui a reçu une copie devraient être résiliés avec son accord. » (CCDigitalLaw 2017g)

- Il est obligatoire d'attribuer l'œuvre à son auteur ou aux autres parties prenantes indiquées, ceci de manière raisonnable. Le titre de l'œuvre ne doit pas être obligatoirement rapporté ;
- Le donneur de licence peut demander que la citation de son nom soit supprimée (dans la mesure du possible) ;
- Si le donneur de licence fournit un avis de droit d'auteur (*copyright notice*) celui-ci doit être conservé ;
- Il est obligatoire de se référer à la licence et à l'exonération de garanties et de responsabilité (donc de fournir un lien vers la licence ou une copie de celle-ci) ;
- Dans la mesure du raisonnable, le preneur de licence doit conserver l'URI ou l'hyperlien vers l'œuvre qui est sous licence ;
- Il est obligatoire d'indiquer les modifications qui ont été faites. La mention des modifications déjà effectuées doit être conservée.

Comme l'élément BY est présent dans toutes les licences CC, ces obligations sont valables pour toutes les licences. Il est possible de le résumer en disant qu'il y a deux obligations liées à cet élément : attribuer l'œuvre (comme détaillé dans les points communs ci-dessus) et indiquer si des modifications ont été effectuées (Favre, Germond 2018, p. 22).

5.2.1.2 Licence CC-BY-SA

Cette licence, de type copyleft, permet : « toute exploitation de l'œuvre (partager, copier, reproduire, distribuer, communiquer, réutiliser, adapter, [traduire]) par tous moyens (...) », à toutes fins, « pour autant que la diffusion des nouvelles œuvres se fasse selon la même licence CC » (Favre, Germond 2018, p. 22). Il est obligatoire également de citer l'auteur de l'œuvre originale et d'indiquer les modifications effectuées, comme indiqué ci-dessus, puisque l'élément BY est présent dans la licence.

Il existe trois options pour partager un jeu de données dérivé d'un autre initialement sous licence CC-BY-SA (Kreutzer 2014, p. 53-54) :

1. Mettre à disposition le jeu de données dérivé sous licence CC-BY-SA ou une version postérieure de cette même licence ;
2. Distribuer le jeu de données dérivé à une licence CC avec les mêmes éléments que la licence originale. Ce cas s'applique en particulier aux licences portées dans des juridictions nationales. Un *dataset* initialement sous licence CC-BY-SA-3.0 CH peut donc être mis à disposition sous licence CC-BY-SA 4.0 et inversement ;
3. Mettre à disposition le jeu de données dérivé sous une licence compatible, mais la liste est plutôt restreinte¹⁰.

5.2.1.3 Licence CC0 (Public Domain Dedication)


En plus des six licences, CC offre deux outils de domaine public. Le premier est une renonciation aux droits d'auteur, indiquant qu'aucun droit n'est réservé sur l'œuvre. L'œuvre est ainsi placée par son auteur dans le domaine public. Or, en Suisse comme dans d'autres pays, il n'est pas possible de renoncer entièrement à ses droits d'auteur, et donc, de mettre son œuvre dans le domaine public. Une simple renonciation aux droits n'aura pas d'effets. C'est pourquoi la renonciation CC0 est dotée de deux options de repli. L'outil CC0 est donc une approche à trois niveaux (Labastida, Margoni 2020) :

¹⁰ Voir : <https://creativecommons.org/share-your-work/licensing-considerations/compatible-licenses/>

1. Une renonciation à tous ses droits, au maximum de ce que la loi nationale permet (impossible en Suisse) ;
2. Une licence qui ne place aucune restriction sur les droits auxquels il n'est pas possible de renoncer ;
3. S'il y a encore des droits auxquels il n'est pas possible de renoncer, ni permis de les licencier, le troisième niveau est une promesse de ne jamais faire valoir ces droits et de ne pas poursuivre en justice les personnes qui en font usage.

Cette partie se réfère à l'outil CC0 sous le terme de licence car la renonciation aux droits d'auteur n'est pas applicable en droit suisse :

Tableau 5 : La licence CC0

Logo de la licence	Nom de la licence	Type de la licence
	Public Domain Dedication (CC0 1.0)	Déclaration de domaine public

Il est à noter que certains dépôts de données ou éditeurs imposent la licence CC0 pour les *datasets*, par exemple Dryad, BioMed Central, Figshare ou encore le groupe Nature Publishing (Dedieu, Fily 2015 ; Creative Commons Wiki 2014). Le fait que des dépôts de données choisissent par défaut la licence CC0 implique que les dépositeurs aient réglé les cas de droits d'auteur de tiers et renoncé aux leurs, ce qui enlève l'incertitude légale et encourage une réutilisabilité productive des données (Carroll 2015).

5.2.1.4 Label Public Domain Mark

La licence CC0 est à ne pas confondre avec le label *Public Domain Mark*, qui n'est pas une licence mais une « étiquette » qui permet de marquer des contenus qui sont déjà dans le domaine public. Son but est de signaler de manière claire des œuvres qui sont tombées dans le domaine public car leur durée de protection s'est éteinte, et qui sont donc utilisables par n'importe qui. Ce label ne doit donc pas être utilisé pour indiquer que l'on renonce à ses droits, car ce n'est pas sa signification et utilisé de la sorte, il n'aura pas l'effet escompté (Ball 2014, p. 6)

Tableau 6 : Le label Public Domain Mark

Logo du label	Nom du label	Signification
	Public Domain Mark 1.0	Pas de droits d'auteurs

Tableau 7 : Différence entre la licence CC0 et le label Public Domain Mark

	Public Domain Mark	Public Domain Dedication (CC0)
What does it state?	The work has been identified as being free of known restrictions under copyright law, including all related and neighboring rights.	The copyright holder has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law.
Conditions for end user	The end user can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.	The end user can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.
When to apply?	Only use when either all copyright restrictions have expired or your data is not at all protected by copyright	Use when all or parts of your data are protected by copyright or you are not sure about it

(Hirschmann 2020)

Pour des données scientifiques factuelles, de toute façon non protégées par le droit d'auteur, il serait donc possible d'utiliser le label *Public Domain Mark*. Mais, si des doutes subsistent sur les droits liés aux données de la recherche, et il y en a souvent, il vaut mieux utiliser la licence CC0. De plus, la différence subtile entre les deux outils n'est pas connue de tous et peut prêter à confusion, ce qui est contre-productif lorsque le but visé est justement de clarifier les droits d'auteur liés à une œuvre.

5.2.2 Les licences Open Data Commons

Les licences Open Data Commons sont proposées par *The Open Knowledge Foundation*. Lancé en 2007, le projet ODC avait pour but de mettre en place des outils juridiques adaptés au partage des bases de données sur le web (Cassella 2013 ; Ball 2014), à une époque où les licences Creative Commons ne prenaient pas cet aspect en compte et où le mouvement des données ouvertes était un concept nouveau. Il n'y a pas eu de mise à jour depuis la sortie en 2009 de la version 1.0 de ces licences (Ball 2014, p. 8).

Les licences ODC reprennent certains principes des licences CC, tout en étant pensées spécifiquement pour les bases de données. Ce qui les rend unique est donc qu'elles se focalisent sur le droit *sui generis* des bases de données.¹¹ Ainsi, seuls les droits d'auteur sur la base de données (sur sa structure) et le droit *sui generis* des bases de données font l'objet de la licence. Le contenu de la base de données n'est pas compris (Labastida, Margoni 2020). ODC a choisi de distinguer la base de données de son contenu car ces deux éléments peuvent avoir des droits différents. Par exemple, pour une base de données contenant des photographies, cohabiteront le droit de la base de données (au sens européen du terme) et les droits d'auteur individuels sur les photographies (Open Data Commons 2009a). Ces licences sont conçues pour être appliquées à tout type de base de données, qu'elles soient commerciales ou issues du secteur public. Elles n'ont donc pas été pensées spécifiquement pour des résultats de recherche scientifique (Guibault, Wiebe 2013, p. 154).

¹¹ Entretien vidéo avec Marc Jones, avocat à CivicActions (Connecticut USA), le 8 juillet 2020.

Voici les caractéristiques des licences ODC :

- Ce sont des licences mondiales, libres de redevances et non exclusives, valables jusqu'à ce qu'il n'y ait plus de droit d'auteur ou de droit *sui generis* des bases de données (Guibault, Wiebe 2013, p. 156) ;
- ODC met à disposition un code juridique complet, une explication de ce dernier pour les non-juristes et un lien URL vers chaque licence (Ball 2014, p. 8). En revanche, ces licences ne sont pas disponibles dans un format lisible par les machines (Guibault, Wiebe 2013, p. 155) ;
- Elles autorisent au minimum, la copie, l'utilisation, la redistribution, la modification des données et la réalisation de travaux dérivés de la base de données (Dedieu, Fily 2015) ;
- Les licences ODC ne permettent pas d'attribuer une sous-licence (Jones 2018) ;
- Elles prévoient un système de révocation automatique de la licence en cas de violation (Jones 2018) ;
- Le contrat de licence présente des clauses d'exonération de garanties et de limitation de la responsabilité (Jones 2018) ;
- Les licences ne concernent pas les brevets ou le droit des marques (Jones 2018) ;
- Les licences choisissent comme juridiction applicable celle du lieu de l'exécution de la licence (Jones 2018).

Les licences ODC sont doublées d'un accord contractuel entre le donneur et le preneur de licence. Ainsi, si dans le pays en question il n'y a pas de propriété intellectuelle sur l'œuvre, que cela soit en termes de droit d'auteur ou de droit des bases de données, les conditions de la licence devraient s'appliquer par contrat (Open Data Commons 2009a). Dans les faits, en Suisse, il n'est pas sûr que cela soit possible d'imposer de telles obligations par contrat.

5.2.2.1 ODC Attribution License (ODC-By)

L'ODC *Attribution License* est une licence d'attribution, comme son nom l'indique. Voici le résumé des droits concédés et des obligations auxquelles il faut se conformer :

Tableau 8 : Droits octroyés et obligations de la licence ODC-By

Droits octroyés	Obligations
<p>Partager la base de données : la copier, la distribuer, la prêter, la rendre disponible et l'utiliser (partiellement ou en totalité).</p> <p>Créer des œuvres dérivées à partir de la base de données ou des bases de données combinées.</p> <p>Adapter la base de données : la modifier, la transformer ou construire autre chose à partir d'elle.</p> <p>Reproduire la base de données sous n'importe quelle forme et à toutes fins.</p>	<p>Attribuer toute réutilisation de la base de données, ou des œuvres produites à partir de celle-ci, de la manière indiquée dans la licence.</p> <p>Pour toute utilisation ou redistribution de la base de données, ou d'œuvre dérivée, il faut aussi :</p> <ul style="list-style-type: none">• Indiquer clairement que la base de données a été utilisée dans la création ;• Inclure une copie de la licence ou de son URI dans la base de données elle-même ;• Conserver intactes toutes les mentions de droit d'auteur, ainsi que celles en lien avec la licence, figurant sur la base de données originale.

(Open Data Commons 2010 ; Labastida, Margoni 2020 ; Cassella 2013 ; Jones 2018)

Il est nécessaire d'utiliser une autre licence pour le contenu de la base de données. ODC ne propose pas une licence déjà formulée pour le contenu dans le cas de la licence ODC-BY.

Une des spécificités de cette licence est que l'attribution n'est pas nécessaire dans le cas de contenus dérivés qui ne sont pas des bases de données, mais produits à partir de celles-ci, par exemple des graphiques, des schémas ou des cartes (Ball 2014, p. 8).

5.2.2.2 ODC Open Database License (ODC-ODbL)

La licence ODbL est une licence de type copyleft qui nécessite également l'attribution. Elle est donc semblable à la licence CC-BY-SA. Voici le résumé des droits concédés et des obligations auxquelles se conformer :

Tableau 9 : Droits octroyés et obligations de la licence ODC-ODbL

Droits octroyés	Obligations
<p>Partager la base de données : la copier, la distribuer, la prêter, la rendre disponible et l'utiliser (partiellement ou en totalité).</p> <p>Créer des œuvres dérivées à partir de la base de données ou des bases de données combinées.</p> <p>Adapter la base de données : la modifier, la transformer ou construire autre chose à partir d'elle.</p> <p>Reproduire la base de données sous n'importe quelle forme et à toutes fins.</p> <p>Lors de la redistribution de la base de données, ou d'une version adaptée de celle-ci, il est possible de mettre des mesures technologiques qui restreignent l'œuvre (comme les DRM), tant qu'une version sans ces mesures est également distribuée.</p>	<p>Attribuer toute réutilisation de la base de données, ou des œuvres produites à partir de celle-ci, de la manière indiquée dans la licence.</p> <p>Pour toute utilisation ou redistribution de la base de données, ou d'œuvre dérivée, il faut aussi :</p> <ul style="list-style-type: none">• Indiquer clairement que la base de données a été utilisée dans la création ;• Inclure une copie de la licence ou de son URI dans la base de données elle-même ;• Conserver intactes toutes les mentions de droit d'auteur, ainsi que celles en lien avec la licence, figurant sur la base de données originale. <p>Partager à l'identique la base de données elle-même ou la base de données dérivée (donc sous la licence ODbL).</p>

(Open Data Commons 2009b ; Labastida, Margoni 2020 ; Cassella 2013 ; Knight 2015 ; Jones 2018)

La condition copyleft s'applique aux nouvelles bases de données dérivées, mais pas aux collections de bases de données ou aux contenus dérivés qui ne sont pas des bases de données, mais produits à partir de celles-ci, comme par exemple des graphiques, des schémas ou des cartes. Par contre, au sein d'une collection, la base de données originale ou la base de données dérivée devra être sous licence ODbL (Ball 2014, p. 8 ; Guibault, Wiebe 2013, p. 159).

Il est nécessaire d'utiliser une autre licence pour le contenu de la base de données. Dans le cas de la licence ODbL, ODC propose une licence pour le contenu, la Database Contents License, qui mentionne que tous les droits sur les contenus individuels de la base de données sont levés (Open Data Commons, 2009a). Il est aussi possible d'utiliser une autre licence pour le contenu de la base de données.

Il y a trois options pour appliquer la clause copyleft sur l'œuvre dérivée (Open Data Commons 2009a) :

1. Mettre la base de données sous licence ODbL ;
2. Mettre la base de données sous une version future de la licence ODbL, qui est en accord avec l'esprit de la licence originale ;

3. Mettre la base de données à disposition sous une licence compatible. Il existe un problème avec cette option car ODC n'a jamais précisé quelles licences étaient compatibles.

5.2.2.3 Public Domain Dedication and License (PDDL)

Cette licence d'ODC ressemble fortement à la CC0, mais elle a été aussi formulée spécialement pour les bases de données. Elle peut s'appliquer à la base de données, à son contenu, ou les deux à la fois, mais n'est pas adaptée aux programmes d'ordinateurs (Jones 2018). Elle vise à se rapprocher au plus près du domaine public (Guibault, Wiebe 2013, p. 155). Si le fait d'abandonner tous ses droits n'est pas valable dans une juridiction nationale, la PDDL contient, comme la CC0, une licence de repli qui permet d'arriver au même résultat, ainsi qu'une promesse de ne pas faire valoir les droits auxquels il est impossible de renoncer (Labastida, Margoni 2020).

La licence PDDL permet de partager librement l'œuvre (copier, distribuer, utiliser), de l'adapter (modifier, transformer, construire à partir de celle-ci) et de créer des œuvres dérivées, à toutes les fins, sans aucune restriction, de manière permanente et irrévocable (Guibault, Wiebe 2013, p. 155).

5.2.3 Les licences Community Data License Agreement

Ces licences ont été développées très récemment, en 2017, par *The Linux Foundation*, qui affirme avoir voulu « to develop data license agreements that could enable sharing of data similar to what we have with open source software » (CDLA 2017a).

Les caractéristiques des licences CDLA sont les suivantes :

- Ce sont des licences mondiales, non exclusives et irrévocables qui donnent le droit d'utiliser et de publier les données (CDLA 2009a ; 2009b) ;
- Ces licences sont valables au niveau de la base de données et des données qu'elle contient (Thorogood 2019) ;
- Il n'y a pas de résumé des licences pour les non-juristes, seulement le code juridique et une foire aux questions ;
- Un système de résiliation automatique de la licence est prévu. Si le preneur de licence ne remédie pas à la situation « dans un délai raisonnable », il n'a plus de licence (CDLA 2017b) ;
- Le contrat de licence présente des clauses d'exonération de garanties et de limitation de la responsabilité (Jones 2018) ;
- Les licences ne concernent pas les brevets ou le droit des marques (Jones 2018) ;
- Elles ne sont pas lisibles par les machines.

Les licences CDLA, comme leur nom l'indique, sont à la fois des licences et des accords (contrats). C'est-à-dire que si une législation ne reconnaît pas de droits sur les données ou la base de données, ce qui est défini dans la licence devrait alors s'imposer par voie contractuelle (Thorogood 2019, p. 81). A nouveau, il n'est pas sûr que cette voie contractuelle soit valable en Suisse pour des contenus qui ne sont pas protégés par le droit d'auteur, ou qui sont protégés par le droit *sui generis* des bases de données qui n'existe pas dans notre pays.

Pensées pour des ensembles de données de grande envergure qui sont constamment modifiés, par exemple dans le domaine du *machine learning* ou encore de l'Internet des objets, ces licences ont l'avantage de n'imposer aucune obligation sur les résultats obtenus à partir

d'une utilisation computationnelle des données, pour autant que ces résultats ne contiennent que de petits extraits du jeu de données original (CDLA 2017b). L'autre avantage de ces licences est l'existence d'une clause dans le contrat de licence qui indique que le fournisseur de données déclare qu'un soin « raisonnable » a été apporté pour s'assurer que les données sont libres de droits de propriété intellectuelle de tiers et d'obligations en termes de protection de la vie privée ou de la confidentialité (Thorogood 2019, p. 85). Cette garantie est ainsi avantageuse pour ceux qui réutiliseront ces données.

5.2.3.1 CDLA-Permissive-1.0

La CDLA-Permissive est une licence qui demande l'attribution :

Tableau 10 : Droits octroyés et obligations de la licence CDLA-Permissive

Droits octroyés	Obligations
Utiliser, modifier, adapter ou partager les données, le jeu de données ou la base de données.	<p>Indiquer la licence originale (code juridique, nom et/ou hyperlien vers celle-ci) lors de la distribution des données ou des données dérivées.</p> <p>Faire en sorte que tout fichier de données contenant des données enrichies comporte des avis bien visibles indiquant que ces fichiers ont été modifiés.</p> <p>Préserver tout crédit ou attribution au fournisseur des données.</p> <p>(Ces obligations ne s'appliquent pas pour les résultats d'une utilisation computationnelle des données.)</p>

(Thorogood 2019, p. 48 ; CDLA 2017b ; data.world 2019 ; CDLA 2009a)

5.2.3.2 CDLA-Sharing-1.0

La licence CLDA-Sharing est de type copyleft et impose aussi l'attribution :

Tableau 11 : Droits octroyés et obligations de la licence CDLA-Sharing

Droits octroyés	Obligations
Utiliser, modifier, adapter ou partager les données, le jeu de données ou la base de données.	<p>Publier les données ou les données dérivées sous la même licence CLDA-Sharing. Indiquer la licence originale (code juridique, nom et/ou hyperlien vers celle-ci) lors de la distribution des données ou des données dérivées.</p> <p>Faire en sorte que tout fichier de données contenant des données enrichies comporte des avis bien visibles indiquant que ces fichiers ont été modifiés.</p> <p>Préserver tout crédit ou attribution au fournisseur des données.</p> <p>Ne pas restreindre l'habilité de quiconque qui reçoit les données de les publier d'une manière accessible au public, ni imposer de nouvelles restrictions aux droits octroyés par la licence et limiter l'utilisation des données.</p> <p>(Ces obligations ne s'appliquent pas pour les résultats d'une utilisation computationnelle des données.)</p>

(Thorogood 2019, p. 48 ; CDLA 2017b ; data.world 2019 ; CDLA 2009b)

En conclusion, les licences CDLA semblent prometteuses pour les recherches qui impliquent de gros volumes de données. Elles restent néanmoins complexes à comprendre par des non-juristes. Puisqu'elles sont très récentes, il faudra voir dans le temps si elles gagnent en utilisation et si elles s'imposent comme des licences de référence dans certains domaines.

5.2.4 Le domaine public comme solution à l'incompatibilité des licences

Choisir une licence standard ouverte est donc nécessaire pour le respect de l'*Open Definition* (2015) et pour garantir l'interopérabilité et la réutilisation des données. Seulement, même si tous les scientifiques publient leurs données sous des licences ouvertes demandant seulement l'attribution ou bien de type copyleft, celles-ci peuvent quand même empêcher des tiers de réutiliser les données. En effet, il existe des différences fondamentales à la fois entre les licences permissives et copyleft, et entre les licences copyleft entre elles (Lämmerhirt 2017). Cette incompatibilité est la raison pour laquelle, en pratique, le domaine public est plébiscité pour les données de la recherche. Pour le moment, seuls certains dépôts de données et surtout les équipes de soutien à la gestion des données de la recherche des institutions universitaires se positionnent sur la question. Les bailleurs de fonds, eux,

n'émettent pas encore de recommandations ou de directives concernant les licences. Par exemple, pour l'instant, le FNS demande que le choix se porte sur une licence standard, lisible par les humains et les machines et à laquelle on peut se référer par un URL, mais n'indique pas qu'il faudrait préférer une licence ouverte, ni de quel type (FNS 2017b).

Les raisons principalement avancées dans la littérature pour plébisciter le domaine public pour les données sont les suivantes :

- Résoudre l'ambiguïté autour de la protection ou non des données de la recherche (Ball 2014, p. 10). Avec une licence de domaine public, il n'y a plus vraiment d'importance à savoir si un jeu de données est protégeable ou non, puisque l'auteur renonce à tous les droits qu'il pourrait avoir dessus et indique très clairement qu'il ne compte pas en restreindre l'utilisation. Le domaine public évite donc le débat sur ce qui est protégé ou non (Fortney 2016) ;
- Mettre les données dans le domaine public est aussi très pratique pour résoudre le problème des données et/ou des bases de données protégées dans un pays, mais pas dans un autre (Fortney 2016) ;
- Le domaine public résout le problème de l'incompatibilité des licences entre elles et permet la meilleure interopérabilité possible entre les jeux de données, ce qui évite de créer des silos de données impossibles à combiner entre eux (Lämmerhirt 2017) ;
- L'Open Science est plus facile à réaliser si le moins de restriction possible n'entrave la réutilisation des données (Labastida, Margoni 2020) ;
- Mettre des données issues de recherches financées par des fonds publics dans le domaine public est cohérent. C'est aussi en logique avec l'éthique générale du partage et de la réutilisation qui prévaut généralement au sein de la communauté scientifique (Murray-Rust et al. 2010).

Quant à savoir quelle licence de domaine public est mise en avant pour les données de la recherche, sans trop de surprise, la licence CC0 l'emporte et ce pour une raison toute simple : elle est la plus connue et la plus facile à utiliser. La licence PDDL de ODC serait aussi compatible, mais elle est moins répandue et surtout moins connue, ce qui impliquerait de faire des recherches pour la comprendre. En réalité, en dehors des Creative Commons, les licences présentées dans ce travail se révèlent complexes pour des non-juristes.

Concernant l'attribution, si importante pour les chercheurs avec la méthode actuelle d'évaluation de la recherche basée entre autres sur leur nombre de citations, les partisans du domaine public pour les données expliquent que l'éthique scientifique s'en chargera. En réalité, rien que pour des raisons d'intégrité et de transparence, il semble assez improbable qu'un scientifique publie des résultats sans mentionner sur quels jeux de données il s'est basé dans sa démarche. La validation de ses résultats par la communauté ainsi que sa crédibilité sont en jeu, ce qui le poussera certainement à indiquer la provenance des données qu'il aura réutilisées.

Il est probablement vrai que le domaine public ne sera pas une option pour une minorité de jeux de données, en raison de caractéristiques très spécifiques et de besoins particuliers. Dans ce cas, un autre type de licence ouverte fera aussi l'affaire. Il faudra également faire attention aux données de la recherche liées à une potentielle invention qui pourrait être brevetée (Carroll 2015), ou qui pourraient mener à une réelle exploitation commerciale, qu'il ne faudra alors pas mettre à disposition de tous avant le dépôt du brevet ou sans en restreindre l'utilisation commerciale (Carroll 2015). Mais il est néanmoins recommandé aux chercheurs qui opteraient

pour une autre solution que le domaine public de soigneusement peser le pour et le contre de leur décision en regard des objectifs de l'Open Science afin de faire le meilleur choix (Labastida, Margoni 2020). Dans tous les cas, consulter le service juridique de son institution reste un bon réflexe et une pratique à encourager en cas de doute sur les droits existants, la protection des données personnelles ou encore les perspectives offertes par le jeu de données, notamment au niveau d'une potentielle commercialisation ou d'un dépôt de brevet. En revanche, il faudrait éviter à tout prix de rédiger une licence sur mesure afin de ne pas aggraver le problème de leur prolifération et de leur incompatibilité. Pour les mêmes raisons, la communauté des données ouvertes estime qu'il faudrait à tout prix éviter de créer encore des nouvelles licences standards pour les données, celles existantes étant déjà suffisantes (Lämmerhirt 2017).

6. Recommandations pour Yareta

Une analyse des avantages et des inconvénients des licences présentées dans ce travail, résumée sous la forme d'un tableau en annexe 9, a porté les recommandations qui sont faites ici sur les licences à proposer par défaut dans Yareta.

Les licences Creative Commons sont tout à fait adaptées aux données de la recherche, bien connues et généralement assez bien comprises. Le fait que Yareta les propose par défaut est ainsi tout à fait adéquat. Seulement, au vu des problèmes posés par les licences ouvertes de type attribution, il semblerait judicieux de proposer la licence CC0 par défaut lorsqu'un jeu de données est mis à disposition en accès libre, au lieu de la licence CC-BY comme c'est le cas actuellement. Pour Yareta, passer de la licence CC-BY, déjà permissive et très ouverte, à la licence CC0, ne serait qu'un pas de plus vers l'ouverture par défaut des jeux de données, qui garantirait toutes les possibilités de réutilisation en évitant le problème de l'empilement des attributions. L'éthique et les normes internationales de la recherche scientifique étant suffisantes pour concilier la mention, légitime, du producteur des données et l'abandon de ses droits sur les données (Gaillard 2014, p. 49).

Si certains dépôts imposent purement et simplement la licence CC0 sans aucun autre choix (la plus connue étant Dryad), cette optique ne semble pas judicieuse dans le contexte de Yareta, en tout cas pas pour le moment. Il est tout à fait vrai que les dépôts peuvent jouer un rôle vers l'ouverture des données en guidant les chercheurs vers un partage sans aucune restriction, voir en les obligeant à atteindre ce but, réduisant par-là les incertitudes légales et encourageant la réutilisation totale des données. Les dépôts peuvent aussi participer à déplacer la discussion sur l'utilisation éthique des données (y compris l'attribution), plutôt vers des normes communautaires que des outils juridiques mal adaptés (Thorogood 2019, p. 82-83). Car si les données ne sont pas protégées par le droit d'auteur (et c'est souvent le cas), droit qui est l'objet de ces licences, les éventuelles restrictions sur des données non protégées n'auront aucun effet légal et pourront être ignorées. Mais en réalité, l'obligation de mettre les données dans le domaine public est délicate dans le cas des dépôts qui reposent sur des soumissions volontaires, ce qui est le cas de Yareta. Il ne faudrait pas que celles-ci soient découragées par ces pratiques, surtout si les soumissions assurent le financement du dépôt. De plus, il pourrait y avoir des préoccupations quant à la responsabilité en cas de violation éventuelle de droits (Thorogood 2019, p. 82-83), d'autant plus quand ni les pouvoirs publics, ni les bailleurs de fonds, ni l'institution universitaire ne donnent d'indications claires ou émettent des obligations sur la licence à utiliser pour les données de la recherche.

Ainsi, l'approche d'ouverture par défaut, déjà adoptée par Yareta, est une très bonne alternative. La recommandation formulée ici est simplement qu'elle aille un peu plus loin en préférant la licence CC0 à la licence CC-BY. Les chercheurs resteraient autorisés à conserver des droits sur leurs données en changeant la licence par défaut afin d'en choisir une plus restrictive s'ils le veulent. Seulement, pour les licences à choix, il serait selon moi judicieux d'aller encore plus loin vers l'ouverture des données et de ne laisser à choix que les licences Creative Commons qui sont ouvertes, soit les licences CC0, CC-BY et CC-BY-SA. Les licences comprenant les éléments ND (*No derivatives*) et NC (*Non commercial*) pourraient alors être chargées sur demande, comme c'est le cas actuellement si la personne qui dépose ses données souhaite leur attribuer une autre licence que les CC. Il ne me semble pas nécessaire non plus de rajouter à la liste de licences à choix les Open Data Commons et les Community Data License Agreement. Bien que celles-ci soient ouvertes et adaptées aux

données, leur méconnaissance et leur complexité pour des non-juristes prêteront probablement à confusion lors du choix, d'autant plus que les licences CC sont déjà adaptées aux données. De plus, le FNS, en accord avec les principes FAIR, recommande d'utiliser une licence lisible par les machines (FNS 2017b), ce qui n'est pas le cas des licences ODC et CLDA. Le mieux serait donc d'offrir la possibilité de charger également une de ces licences sur demande. En parallèle, les chercheurs pourraient être informés de leur existence et de leurs avantages et spécificités respectives, par exemple dans le cas du *machine learning* pour les licences CDLA ou dans le cas d'une base de données d'œuvres artistiques pour les licences ODC. Elles seraient alors recommandées que dans des cas très précis. Il me semble par ailleurs que la pertinence de telles recommandations serait à étudier avec le service juridique de l'UNIGE et les facultés concernées, afin de s'assurer de leur réelle plus-value sur les licences CC et de leur applicabilité en Suisse.

De manière générale, un effort de sensibilisation sur la compréhension du droit d'auteur et la bonne utilisation des licences pour les données dans le contexte de la recherche est nécessaire. Pourquoi ne pas collaborer avec l'équipe de soutien à la recherche de la Bibliothèque de l'UNIGE pour introduire des formations spécifiques sur ces questions ? Il serait également possible de proposer un outil pour sélectionner la licence, à l'image du *license selector* d'EUDAT¹², une des plus grandes infrastructures de services de données intégrés et de ressources de soutien à la recherche en Europe, dont le code source est disponible.

Pour finir, je formulerai deux recommandations à propos de la protection des données personnelles. La première est, par précaution, de prévoir un moyen de modifier un jeu de donnée après son archivage, dans l'éventualité où une personne révoquerait son consentement et exigerait que ses données personnelles soient retirées. Actuellement, une telle modification est impossible par la personne qui soumet un jeu de données après l'archivage, mais une telle intervention pourrait se faire du côté du support de Yareta après évaluation du bienfondé de la demande. La seconde recommandation concerne l'anonymisation des données personnelles, dont la responsabilité en cas d'exigence légale est laissée aux chercheurs. Il m'a semblé avoir compris que pour les chercheurs soumis à la LIPAD, l'anonymisation des données personnelles en cas de publication des données de la recherche est obligatoire¹³. Tout en ne prenant toujours aucune responsabilité sur le sujet, il me semble pertinent que Yareta précise cette obligation légale dans ses futures conditions générales¹⁴, ou dans l'interface de dépôt elle-même. Pareillement, il pourrait être judicieux de mentionner quelque part que même pour les chercheurs non soumis à la LIPAD, l'anonymisation reste fortement conseillée dans tous les cas où elle est possible et notamment lors de la publication des données.

¹² <https://ufal.github.io/public-license-selector/>

¹³ Affirmation qui n'engage pas la responsabilité de l'auteur et qui est à confirmer avec le service juridique de l'UNIGE.

¹⁴ Les conditions générales de Yareta étaient en cours de rédaction au moment de l'écriture de ce mémoire.

7. Des fiches d'information et un arbre de décision à destination des chercheurs

Ce travail a surtout une vocation d'information. Il répond au mandat de clarifier et de synthétiser les enjeux légaux des données de la recherche et de rendre compte des bonnes pratiques liées aux licences. Ce travail n'entend pas apporter toutes les réponses aux questions juridiques qui se posent et résoudre des cas particuliers, car ce n'est pas le rôle d'un spécialiste en information documentaire mais de professionnels formés en droit. En revanche, puisque le soutien aux données de la recherche est un des domaines où le spécialiste en information documentaire peut œuvrer, ces questions lui sont souvent adressées et il se doit d'en avoir une bonne compréhension pour renseigner les chercheurs, même si souvent la consultation d'un service juridique sera également nécessaire et conseillée.

Dans ce contexte et pour donner une plus grande portée à ce travail, il a été décidé en concertation avec René Schneider, directeur de ce travail, d'ajouter au mandat initial la réalisation d'un guide qui regrouperait et résumerait les conclusions du mémoire à destination des chercheurs. En effet, bien que ce travail ait été mandaté par DLCM pour Yareta, ses recommandations et ses résultats peuvent être utiles à toute la communauté des données de la recherche en Suisse. Ce nouvel objectif a finalement pris la forme de cinq fiches sur le droit d'auteur et les licences pour les données, disponibles en annexe 10, ainsi que d'un arbre de décision pour le choix d'une licence lors de la publication des données, disponible en annexe 11. La conception des fiches information a été inspirée par les guides pour le *Research Data Management* que l'EPFL¹⁵ et l'Université de Bâle¹⁶ ont réalisés. L'arbre de décision pour les licences est basé sur le *Data publication decision tree* de l'EPFL également¹⁷, ainsi que le *License selector decision flow diagram* d'EUDAT¹⁸. A noter que pour la question des droits d'auteur seulement, il existe aussi un arbre de décision créé en 2011 par le projet DICE (devenu maintenant le CCDigitalLaw)¹⁹.

Idéalement, dans mes fiches information, il aurait fallu laisser plus de place pour des éléments graphiques, utiliser plus de couleurs et proposer une mise en page plus originale, un peu à l'image de ce qui se fait avec les posters scientifiques. Malheureusement, le manque de temps ne m'a pas permis de concevoir un guide au graphisme très soigné. La version PDF initiale, de meilleure qualité que les images en annexe de ce travail, est disponible sur Zenodo²⁰. Les fichiers de travail y sont aussi mis à disposition, permettant de les modifier pour adapter ce guide.

¹⁵ BLUMER, Eliane Ninfa, CHAPITINEL, Jérôme Julien, MASSON, Antoine, REICHLER, Fantin, SAMATH, Sitthida, VARRATO, Francesco et MILFORT, Frank, 2019. EPFL Library Research Data Management Fastguides. *Zenodo* [en ligne]. Fév 2019. [Consulté le 31 juillet 2020]. Disponible à l'adresse : <http://doi.org/10.5281/zenodo.3327830>

¹⁶ UNIVERSITÄT BASEL - RESEARCH DATA MANAGEMENT-NETWORK, 2019. RDM Cheat Sheets. *researchdata.unibas.ch* [en ligne]. [Consulté le 21 juin 2020]. Disponible à l'adresse : https://researchdata.unibas.ch/fileadmin/user_upload/researchdata/FastGuide_v1_print_1_.pdf

¹⁷ HENRY, Luc, KRAUSE, Jan et GROLIMUND Raphaël, 2017. Data publication decision tree. Lausanne : EPFL. [Consulté le 21 juin 2020]. Disponible à l'adresse : <https://infoscience.epfl.ch/record/230281>.

¹⁸ EUDAT, 2017. License selector decision flow diagram. <https://eudat.eu/services/userdoc/license-selector>

¹⁹ DICE, 2011. DICE decision tree. [Consulté le 4 septembre 2020]. Disponible à l'adresse : http://www.diceproject.ch/wp-content/uploads/2011/06/DICE_Albero_final_1.pdf

²⁰ A l'adresse : <https://doi.org/10.5281/zenodo.3967402>

Les fiches informations n'abordent pas du tout la question de la protection des données, qui représente pourtant une partie importante de ce mémoire. Ce choix a été fait délibérément pour trois raisons. La première est que j'ai rencontré des difficultés avec ce sujet très complexe. Malgré toutes mes lectures et mes questions adressées à des avocats ou juristes, je n'ai pas suffisamment confiance en ma compréhension des enjeux en termes de protection des données pour réussir à identifier les éléments saillants et les résumer sous la forme de fiches. La seconde raison est que les deux guides de l'EPFL et de l'Université de Bâle abordent déjà ces questions de protection des données, tandis que les questions du droit d'auteur et des licences ne sont pas mises en exergue. En l'occurrence, ma compréhension de ces enjeux était plus assurée, me poussant à me concentrer dessus. La troisième et dernière raison pour écarter la protection des données des fiches information se rapporte aux objectifs du mandat qui m'a été confié. Dès le début, la question des licences était centrale dans le travail à produire. J'ai donc choisi de me concentrer sur cet aspect, ainsi que sur celui du droit d'auteur, puisqu'ils sont liés.

8. Conclusion

En Suisse, le mouvement vers l'Open Science, déjà bien entamé avec l'Open Access pour les publications, se poursuit à bon train avec l'ouverture des données de la recherche. Le cadre juridique de celles-ci est complexe et il est nécessaire de l'expliquer pour qu'il n'entrave pas leur partage. Principalement concernées par le droit d'auteur et la protection des données personnelles, les données de recherche peuvent cependant aussi être encadrées par d'autres lois, notamment dans le domaine de la santé ou de la recherche sur l'être humain. En termes de droit d'auteur, les données ne sont pas explicitement mentionnées dans la LDA, mais elles sont protégées à plusieurs niveaux (celui des données individuelles, des bases de données, des métadonnées et des visualisations de données), pour autant qu'elles soient des créations de l'esprit qui ont un caractère individuel. De plus, il faut examiner au cas par cas qui détient les droits sur les données, certains pouvant être cédés, notamment dans le cadre d'une relation de travail entre un employé et un employeur. Pour la protection des données personnelles, c'est cette fois plusieurs niveaux de lois qui peuvent être concernés : les lois cantonales, la loi fédérale et le récent RGPD européen. Même si les grands principes de protection des données restent les mêmes, ces textes édictent des dispositions différentes, qui sont parfois en contradiction les unes avec les autres. Leurs relations entre eux ne sont pas aisées à comprendre, ni quelle attitude adopter si plusieurs textes législatifs différents sont concernés par une recherche, notamment dans le cadre de collaborations nationales ou internationales.

En l'occurrence, il faut concilier toutes ces contraintes légales avec le partage et l'ouverture des données pour satisfaire les exigences du mouvement de l'Open Science, promu par les bailleurs de fonds et les institutions de recherche. Or, le fait que ce cadre juridique soit si compliqué encourage parfois les chercheurs à limiter l'ouverture ou le partage de leurs données, car ils craignent de violer une loi en procédant autrement. De même, si des doutes subsistent sur les droits ou les obligations liés aux données, la tendance sera de limiter leur diffusion et leurs possibilités de réutilisation. En réalité, contraintes juridiques et ouverture des données ne sont pas irréconciliables, mais un effort est à faire pour identifier ce que les lois permettent, les obligations qu'elles fixent et les droits qu'elles octroient ; tout comme pour en résumer les principaux enjeux de manière compréhensible à destination des chercheurs.

Pour concilier ouverture des données de la recherche et protection des données personnelles ou sensibles, la solution la plus évidente et celle qui est communément recommandée est d'anonymiser les données concernées, car dès lors elles ne sont plus protégées. Cet acte n'est pas anodin et s'il est impossible à réaliser sans dénaturer la valeur des données de la recherche, il faut alors consulter un service juridique pour trouver d'autres solutions. De manière générale, il est également conseillé de recueillir le consentement éclairé des participants à une étude à un niveau très fin, qui comprend l'ouverture, le partage et l'archivage des données. Du côté du droit d'auteur, le meilleur moyen de réduire les problèmes de propriété intellectuelle liés aux données est de les mettre à disposition sous une licence ouverte. En effet, il y a des problèmes d'interopérabilité légale ainsi que de l'insécurité lors de la réutilisation de données de tiers, ou de leur combinaison avec d'autres *dataset*, si ces données ne sont pas mises à disposition sous licence, ou sous une licence restrictive. C'est pourquoi il est conseillé de rendre les données ouvertes par défaut :

« If researchers commit (to funders and participants) to make the products of their research publicly available, they should also explicitly commit to do so under a standard, open-

conformant license. Ideally, this commitment should be made at the time the research project is proposed, so it is clear to funders and participants exactly how open and interoperable the outputs will be. Otherwise, there is a risk that researchers make data technically, but not legally, available. » (Thorogood 2019, p. 81)

Pour éviter le problème de la non-interopérabilité des licences entre elles, et les autres problèmes liés aux licences non ouvertes, ou aux licences ouvertes mais problématiques, les licences plébiscitées pour les données de la recherche sont celles qui se rapprochent au plus près du domaine public, notamment la licence CC0 de Creative Commons. Actuellement, cette dernière est recommandée par les personnes œuvrant dans le milieu des données de la recherche, que cela soit en Suisse ou à l'étranger. Un pas encore plus en avant vers l'ouverture des données serait une approche « top-down », où les agences de financement, les institutions de type universitaire ou même les législateurs, imposeraient une licence standard ouverte, ou un choix limité de licences ouvertes approuvées, pour le partage des données de la recherche financées par des fonds publics (Thorogood 2019, p. 83). Si des exigences ont été formulées par le FNS et swissuniversities en termes d'Open Access des publications scientifiques, ne pourrait-on pas en imaginer également en termes de licences pour les données de la recherche ?

8.1 Difficultés rencontrées

La partie sur la protection des données personnelles s'est révélée particulièrement difficile. J'ai senti que ma compréhension arrivait à ses limites et qu'il aurait fallu que je sois formée en droit pour réussir à identifier avec certitude les éléments importants pour les données de la recherche et les analyser. En réalité, la partie que je propose sur la protection des données est plutôt un inventaire, pour chaque texte juridique, des grands principes de protection, des obligations auxquelles il faut se plier et des exceptions qui existent pour la recherche. L'analyse a été compliquée en raison d'un manque de compréhension sur certains points, et surtout d'un manque d'assurance et de légitimité dans cette tâche. De plus, très souvent, lorsqu'il y a des données personnelles ou sensibles, notamment lors de recherches en médecine et en sciences sociales, ce ne sont pas les lois cantonales ou la LPD qui sont concernées, mais la LRH et d'autres lois régissant la recherche sur l'humain ou dans le domaine de la santé. Celles-ci ne faisaient pas partie du champ de ce mandat et il a été impossible par manque de temps de les inclure après coup, lorsque leur importance a été réalisée. Ainsi, pour rendre compte de tous les enjeux des données de la recherche en termes de protection des données, il serait peut-être judicieux de mandater une nouvelle étude qui se concentrerait seulement sur cet aspect, qui prendrait en compte toutes les lois concernées, et qui serait mené par quelqu'un de spécialisé dans ce domaine.

8.2 Perspectives offertes par le travail

Malgré ces difficultés, le cadre juridique des données de la recherche tel que présenté dans ce travail offre un panorama assez large des exigences légales, qui peut se révéler instructif pour toutes les institutions qui font de la recherche à Genève, voir en Suisse s'il est complété par une analyse d'autres lois cantonales que la LIPAD.

A l'heure où de plus en plus de projets développent des solutions d'archivage des données de la recherche, que cela soit au niveau institutionnel, cantonal ou national, les recommandations formulées pour Yareta à l'égard des licences ouvertes, et notamment le plébiscite envers la licence CC0, ont une portée plus large et sont en réalité applicables à tout dépôt de données

généraliste. Par ailleurs, les fiches que j'ai créées, qui expliquent pourquoi les licences ouvertes sont si importantes et qui détaillent les plus appropriées pour les données, ont été pensées pour aider les chercheurs dans le choix de la bonne licence, et pour les convaincre du bien-fondé des licences ouvertes pour la science ouverte. Ces contenus, mis à disposition sous licence CC-BY, pourraient aussi être utiles à des services de soutien à la gestion des données de la recherche, qui ont la possibilité de se les approprier, de les traduire ou de les modifier pour qu'ils correspondent au mieux à leurs besoins et aux caractéristiques de leurs institutions.

Bibliographie

Textes juridiques :

- LDA Loi fédérale du 9 octobre 1992 sur le droit d'auteur et les droits voisins (LDA ; RS 231.1). *Confédération suisse* [en ligne]. 9 octobre 1992. Mise à jour le 1^{er} avril 2020. [Consulté le 14 avril 2020]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/19920251/>
- LIPAD Loi sur l'information du public, l'accès aux documents et la protection des données personnelles du 5 octobre 2001 (LIPAD ; RSGe A 2 08). *Site officiel de l'Etat de Genève* [en ligne]. 5 octobre 2001. Mise à jour le 30 juin 2018. [Consulté le 15 avril 2020]. Disponible à l'adresse : https://www.ge.ch/legislation/rsg/f/s/rsg_a2_08.html
- LHES-SO Loi sur la Haute école spécialisée de Suisse occidentale – Genève du 29 août 2013 (LHES-SO-GE; RSGe C 1 26). *Site officiel de l'Etat de Genève* [en ligne]. 29 août 2013. Mise à jour le 8 juin 2019. [Consulté le 15 avril 2020]. Disponible à l'adresse : https://www.ge.ch/legislation/rsg/f/s/rsg_c1_26.html
- LPD Loi fédérale du 19 juin 1992 sur la protection des données (LPD ; RS 235.1). *Confédération suisse* [en ligne]. 19 juin 1992. Mise à jour le 1^{er} mars 2019. [Consulté le 14 avril 2020]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/19920153/index.html>
- LRH Loi fédérale du 30 septembre 2011 relative à la recherche sur l'être humain (LRH ; RS 810.30). *Confédération suisse* [en ligne]. 30 septembre 2011. Mise à jour le 1^{er} janvier 2020. [Consulté le 15 avril 2020]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/20061313/index.html>
- LU Loi sur l'université du 13 juin 2008 (LU ; RSGe C 1 30). *Site officiel de l'Etat de Genève* [en ligne]. 13 juin 2008. Mise à jour le 8 juin 2019. [Consulté le 15 avril 2020]. Disponible à l'adresse : https://www.ge.ch/legislation/rsg/f/s/rsg_c1_30.html
- OLPD Ordonnance du 14 juin 1993 relative à la loi fédérale sur la protection des données (OLPD ; RS 235.11). *Confédération suisse* [en ligne]. 14 juin 1993. Mise à jour le 16 octobre 2012. [Consulté le 22 avril 2020]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/19930159/index.html>
- RIPAD Règlement d'application de la loi sur l'information du public, l'accès aux documents et la protection des données personnelles du 21 décembre 2011 (RIPAD; RSGe A 2 08.01). *Site officiel de l'Etat de Genève* [en ligne]. 21 décembre 2011. Mise à jour le 3 juillet 2019. [Consulté le 15 avril 2020]. Disponible à l'adresse : https://www.ge.ch/legislation/rsg/f/s/rsg_A2_08p01.html

RGPD Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE). *EUR-Lex* [en ligne]. 27 avril 2016. Mise à jour le 4 mai 2016. [Consulté le 24 avril 2020]. Accessible à l'adresse : <http://data.europa.eu/eli/reg/2016/679/oj/fra>

Littérature secondaire :

24 HEURES, 2020. Protection des données : Bruxelles temporise. *24 heures* [en ligne]. 24 juin 2020. [Consulté le 26 juin 2020]. Disponible à l'adresse : <https://www.24heures.ch/protection-des-donnees-bruxelles-temporise-972230548499>

ALIPRANDI, Simone, 2012. Open licensing and databases. *Journal of Open Law, Technology & Society* [en ligne]. 9 mai 2012. Vol. 4, n°1, pp. 5-18. [Consulté le 26 juin 2020]. Disponible à l'adresse : <https://www.jolts.world/index.php/jolts/article/view/62>

ALMANZI, Claude, BAGGI, Marcello, CONTEL, Raphaël, COTTIER, Bertil et DE WERRA, Jacques, 2010. *Le droit d'auteur dans le contexte de l'enseignement* [en ligne]. Lugano & Genève : DICE: Digital Copyrights for E-Learning. [Consulté le 14 avril 2020]. Disponible à l'adresse : http://www.diceproject.ch/wp-content/uploads/2011/05/Handbook_web_fr.pdf

BAGGI, Marcello, [sans date]. Droit d'auteur dans le contexte de la recherche. *Parasolcorpus.org* [en ligne]. [Consulté le 16 avril 2020]. Disponible à l'adresse : http://parasolcorpus.org/swisscorpora/templates/bee5sw/document/Guideline_fr.pdf

BALL, Alex, 2014. *How to License Research Data* [en ligne]. Edinburgh : Digital Curation Centre. [Consulté le 3 février 2020]. DCC How-to Guides. Disponible à l'adresse : <http://www.dcc.ac.uk/resources/how-guides/license-research-data>

BAMDÉ, Aurélien, 2018. RGPD : les principes de loyauté et de licéité du traitement des données. A. Bamdé & J. Bourdoiseau : *le Droit dans tous ses états* [en ligne]. 9 décembre 2018. [Consulté le 22 mai 2020]. Disponible à l'adresse : <https://aurelienbamde.com/2018/12/09/rgpd-les-principes-de-loyaute-et-de-liceite-du-traitement-des-donnees/>

BAYLE, Aurélie, 2018. *RGPD et protection des données : panorama d'une réglementation et ses impacts* [document PPT]. Support de cours : « Open Science ». Haute école de gestion de Genève. Master en Science de l'information, année académique 2018-2019. 8 novembre 2018.

BECARD, Nicolas, CASTETS-RENARD, Céline, CHASSANG, Gauthier, COURTOIS, Mary-Audrey, DANTANT, Martin, GANDON, Nathalie, MARTIN, Caroline, MARTELLETTI, Andrea, MENDOZA-CAMINADE, Alexandra, MORCLETTE, Nathalie et NEIRAC, Claire, 2017. Ouverture des données de recherche. Guide d'analyse du cadre juridique en France - Version 2. [en ligne]. Décembre 2017. [Consulté le 15 avril 2020]. Disponible à l'adresse : <http://prodinra.inra.fr/record/382263>

BENHAMOU, Yaniv et TRAN, Laurent, 2016. Circulation des biens numériques : de la commercialisation à la portabilité. *Sic !* N°11, pp. 571-591.

BRÜWER, Michael, 2019. White Paper - A View of Swissuniversities' Network of ICT Experts on the Research Data Management Landscape in Switzerland. *swissuniversities* [en ligne]. 6 mai 2019. [Consulté le 21 juin 2020]. Disponible à l'adresse : https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Organisation/SUK-P/SUK_P-2/RDM_Landscape_CH_WhitePaper_v1.1_final.pdf

BUCHE, Eric, BYRNE-SUTTON, Pascale et WERLY, Stéphane, 2017. *La LIPAD*. Genève : République et Canton de Genève.

BURGI, Pierre-Yves et BLUMER, Eliane, 2018. Le projet DLCM : gestion du cycle de vie des données de recherche en Suisse. In : *Bibliotheken der Schweiz: Innovation durch Kooperation. Festschrift für Susanna Bliggenstorfer anlässlich ihres Rücktrittes als Direktorin der Zentralbibliothek Zürich* [en ligne]. Berlin : De Gruyter. pp. 235-249.

[Consulté le 24 février 2020]. ISBN 978-3-11-055379-6. Disponible à l'adresse :

<https://archive-ouverte.unige.ch/unige:105931>

BURGI, Pierre-Yves, CAZEAUX, Hugues et MENGELLE, Thierry, 2019. Yareta : solution numérique pour archiver et partager vos données de recherche. *YouTube* [en ligne]. 9 octobre 2019. [Consulté le 8 avril 2020]. Disponible à l'adresse :

https://www.youtube.com/watch?v=IMSy_K-K0Sc

BURGI, Pierre-Yves, 2019. Le Projet de Loi 12146 : Infrastructures et services numériques pour la recherche. *Revue électronique suisse de science de l'information* [en ligne]. Vol. 20.

[Consulté le 24 février 2020]. Disponible à l'adresse : [https://archive-](https://archive-ouverte.unige.ch/unige:128845)

[ouverte.unige.ch/unige:128845](https://archive-ouverte.unige.ch/unige:128845)

CARBON, Seth, CHAMPIEUX, Robin, MCMURRY, Julie A., WINFREE, Lilly, WYATT, Letisha R. et HAENDEL, Melissa A., 2019. An analysis and metric of reusable data licensing practices for biomedical resources. *PLoS ONE* [en ligne]. 27 mars 2019. Vol. 14, n°3.

[Consulté le 20 mars 2020]. Disponible à l'adresse :

<https://doi.org/10.1371/journal.pone.0213090>

CARROLL, Michael W., 2015. Sharing Research Data and Intellectual Property Law: A Primer. *PLOS Biology* [en ligne]. 27 août 2015. Vol. 13, n°8. [Consulté le 5 mars 2020].

Disponible à l'adresse : <https://doi.org/10.1371/journal.pbio.1002235>

CASSELLA, Maria, 2013. Dati aperti e ricerca scientifica: aspetti gestionali e normativi nel contesto dell'e-science. *AIB Studi* [en ligne]. Vol. 53, n°3, pp. 223-238. [Consulté le 11 mars 2020].

Disponible à l'adresse : <https://dx.doi.org/10.2426/aibstudi-9131>

CCDIGITALLAW, 2017a. 2. QUOI... Qu'est-ce qu'une oeuvre protégée ? *CCdigitallaw.ch* [en ligne]. [Consulté le 13 mai 2020]. Disponible à l'adresse :

<https://ccdigitallaw.ch/index.php/french/copyright/2-quoi-quest-ce-quune-oeuvre>

CCDIGITALLAW, 2017b. 2.5 Recueils. *CCdigitallaw.ch* [en ligne]. [Consulté le 13 mai 2020].

Disponible à l'adresse : <https://ccdigitallaw.ch/index.php/french/copyright/2-quoi-quest-ce-quune-oeuvre/25-recueils>

CCDIGITALLAW, 2017c. 2.1.4 Titres et parties d'œuvres, projets. *CCdigitallaw.ch* [en ligne]. [Consulté le 13 mai 2020]. Disponible à l'adresse :

<https://ccdigitallaw.ch/index.php/french/copyright/2-quoi-quest-ce-quune-oeuvre/21-livre-protegee-par-le-droit-dauteur/214-titres-et-parties-duvres-projets>

CCDIGITALLAW, 2017d. 3.4 Autres titulaires de droits. *CCdigitallaw.ch* [en ligne].

[Consulté le 13 mai 2020]. Disponible à l'adresse :

<https://ccdigitallaw.ch/index.php/french/copyright/3/34-autres-titulaires-de-droits>

CCDIGITALLAW, 2017e. 5b. COMMENT... des tiers sont-ils autorisés à utiliser une œuvre ? - Licences. *CCdigitallaw.ch* [en ligne]. [Consulté le 26 mai 2020]. Disponible à l'adresse : <https://ccdigitallaw.ch/index.php/french/copyright/5b-comment-des-tiers-sont-ils-autorises-utiliser-une-oeuvre-licences>

CCDIGITALLAW, 2017f. 5b.1 Licences Creative Commons. *CCdigitallaw.ch* [en ligne]. [Consulté le 26 mai 2020]. Disponible à l'adresse : <https://ccdigitallaw.ch/index.php/french/copyright/5b-comment-des-tiers-sont-ils-autorises-utiliser-une-oeuvre-licences/5b1-licences-creative-commons>

CCDIGITALLAW, 2017g. 5b.1.1 Quel contenu ces six licences Creative Commons réglementent-elles ? *CCdigitallaw.ch* [en ligne]. [Consulté le 11 juin 2020]. Disponible à l'adresse : <https://ccdigitallaw.ch/index.php/french/copyright/5b-comment-des-tiers-sont-ils-autorises-utiliser-une-oeuvre-licences/5b1-licences-creative-commons/5b11-quel-contenu-ces-six-licences-creative-commons-reglementent-elles>

CCDIGITALLAW, 2020. 2.1 L'œuvre protégée par le droit d'auteur. *CCdigitallaw.ch* [en ligne]. 1 avril 2020. [Consulté le 13 mai 2020]. Disponible à l'adresse : <https://ccdigitallaw.ch/index.php/french/copyright/2-quoi-quest-ce-quune-oeuvre/21-livre-protgee-par-le-droit-dauteur>

CDLA, 2009a. Community Data License Agreement -Permissive-Version 1.0. *Community Data License Agreement* [en ligne]. [Consulté le 22 juillet 2020]. Disponible à l'adresse : <https://cdla.io/wp-content/uploads/sites/52/2017/10/CDLA-Permissive-v1.0.pdf>

CDLA, 2009b. Community Data License Agreement -Sharing-Version 1.0. *Community Data License Agreement* [en ligne]. [Consulté le 22 juillet 2020]. Disponible à l'adresse : <https://cdla.io/wp-content/uploads/sites/52/2017/10/CDLA-Sharing-v1.0.pdf>

CDLA, 2017a. Home. *Community Data License Agreement* [en ligne]. [Consulté le 11 juin 2020]. Disponible à l'adresse : <https://cdla.io/>

CDLA, 2017b. FAQ. *Community Data License Agreement* [en ligne]. [Consulté le 11 juin 2020]. Disponible à l'adresse : <https://cdla.io/faq/>

CELLIER, Léonore et WERLY, Stéphane, 2018. La protection des données personnelles relatives à la santé dans le canton de Genève. *Pratique juridique actuelle*. Dike Verlag AG. 2018. pp. 1128-1140.

CNIL, 2017. Règlement européen sur la protection des données personnelles : guide du sous-traitant. *Commission Nationale de l'Informatique et des Libertés* [en ligne]. Septembre 2017. [Consulté le 22 mai 2020]. Disponible à l'adresse : https://www.cnil.fr/sites/default/files/atoms/files/rqpd-guide_sous-traitant-cnil.pdf

CNIL, 2018. Transferts de données hors UE : ce qui change avec le règlement général sur la protection des données (RGPD). *Commission Nationale de l'Informatique et des Libertés* [en ligne]. 24 mai 2018. [Consulté le 22 mai 2020]. Disponible à l'adresse : <https://www.cnil.fr/fr/transferts-de-donnees-hors-ue-ce-qui-change-avec-le-reglement-general-sur-la-protection-des-donnees>

CNIL, 2019. Ce qu'il faut savoir sur l'analyse d'impact relative à la protection des données (AIPD). *Commission Nationale de l'Informatique et des Libertés* [en ligne]. 22 octobre 2019. [Consulté le 22 mai 2020]. Disponible à l'adresse : <https://www.cnil.fr/fr/ce-quil-faut-savoir-sur-lanalyse-dimpact-relative-la-protection-des-donnees-aipd>

COMMISSION DE LA PROTECTION DE LA VIE PRIVÉE - CPVP, 2018. RGPD : vademecum pour les PME. *autoriteprotectiondonnees.be* [en ligne]. Janvier 2018. [Consulté le 22 mai 2020]. Disponible à l'adresse : https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/PME_FR_0.pdf

COMMISSION EUROPÉENNE, 2016. H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. *Commission européenne* [en ligne]. 26 juillet 2016. [Consulté le 22 mars 2020]. Disponible à l'adresse : https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

COMMISSION EUROPÉENNE, 2017a. H2020 Programme: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. *Commission européenne* [en ligne]. 12 mars 2017. [Consulté le 20 juin 2020]. Disponible à l'adresse : https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf#page=8

COMMISSION EUROPÉENNE, 2017b. H2020 Online Manual: Data management. *Commission européenne* [en ligne]. [Consulté le 20 juin 2020]. Disponible à l'adresse : https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

CONSEIL FÉDÉRAL, 2017. *Message relatif à la modification de la loi sur le droit d'auteur, à l'approbation de deux traités de l'Organisation Mondiale de la Propriété Intellectuelle et à leur mise en œuvre* [en ligne]. 22 novembre 2017. [Consulté le 9 janvier 2018]. Disponible à l'adresse : <https://www.ejpd.admin.ch/dam/data/ejpd/aktuell/news/2017/2017-11-22/bot-f.pdf>

CREATIVE COMMONS, 2017. A propos des licences. *Creative Commons* [en ligne]. [Consulté le 5 juillet 2020]. Disponible à l'adresse : <https://creativecommons.org/licenses/?lang=fr>

CREATIVE COMMONS WIKI, 2014. CC0 use for data. *CC Wiki* [en ligne]. 3 décembre 2014. [Consulté le 6 juin 2020]. Disponible à l'adresse : https://wiki.creativecommons.org/index.php?title=CC0_use_for_data&oldid=101674

DATA.WORLD, 2019. Common license types for datasets. *data.world help center* [en ligne]. [Consulté le 20 mars 2020]. Disponible à l'adresse : <http://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets>

DEDIEU, Laurence et FILY, Marie-Françoise, 2015. *Rendre publics ses jeux de données scientifiques en 6 points* [en ligne]. Montpellier : Cirad. [Consulté le 10 avril 2020]. Disponible à l'adresse : <https://collaboratif.cirad.fr/alfresco/s/d/workspace/SpacesStore/4aad9e30-6e00-4c97-8126-4c6b4b2d4bc6/rendre-public-ses-jeux-de-donnees-avril-2015.pdf>

DLCM, 2015. Partners :: DLCM. *DLCM* [en ligne]. [Consulté le 15 juin 2020]. Disponible à l'adresse : <https://www.dlcm.ch/about-us/partners>

DMLAWTOOL, 2019. DMLawTool (192-009). *swissuniversities* [en ligne]. 26 août 2019. [Consulté le 26 juin 2020]. Disponible à l'adresse : https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Organisation/SUK-P/SUK_P-2/192-009_DMLawTool_Abstract_26.08.2019.pdf

DODDS, Leigh, 2013. Publisher's Guide to Open Data Licensing. *The Open Data Institute* [en ligne]. 15 décembre 2013. [Consulté le 11 mars 2020]. Disponible à l'adresse : <https://theodi.org/article/publishers-guide-to-open-data-licensing/>

DONZÉ, Dimitri et ZOUHRI, Talal, 2019. Gestion des données personnelles et sensibles : obligations et bonnes pratiques. *Bibliothèque de l'Université de Genève* [en ligne]. Novembre 2019. [Consulté le 22 avril 2020]. Disponible à l'adresse : https://www.unige.ch/biblio/files/7615/7556/2200/2019_midi_donnees_sensibles.pdf

DOLDIRINA, Catherine, EISENSTADT, Anita, ONSRUD, Harlan et UHLIR, Paul, 2018. Legal Approaches for Open Access to Research Data. *LawArXiv Papers* [en ligne]. 1 avril 2018. [Consulté le 6 juin 2020]. Disponible à l'adresse : <https://doi.org/10.31228/osf.io/n7qfa>

FAVRE, Janelise et GERMOND, Tania, 2018. Augmenter la visibilité et l'impact d'une publication scientifique en maîtrisant le droit d'auteur. *oa100.snf.ch* [en ligne]. [Consulté le 26 juin 2020]. Disponible à l'adresse : <https://oa100.snf.ch/wp-content/uploads/2019/02/Augmenter-la-visibilit%C3%A9-et-l'impact-d'une-publication-scientifique-en-ma....pdf>

FAVRE, Janelise et GERMOND, Tania, 2020. Tania Germond & Janelise Favre : Le B.A-BA du droit d'auteur pour l'enseignant et le chercheur. *YouTube* [en ligne]. 25 mars 2020. [Consulté le 13 mai 2020]. Disponible à l'adresse : https://www.youtube.com/watch?v=10uaGgHo6LY&list=PLMn8MvVsX6LNRX_woTuiqLOuXPEHYdrSP&index=8

FAVRE, Marion et HÄHLING, Héloïse, 2020. *Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données* [en ligne]. Genève : Haute école de gestion de Genève. Mémoire de recherche. [Consulté le 26 juin 2020]. Disponible à l'adresse : <https://doc.rero.ch/record/328470>

FNS, 2017a. Data Management Plan (DMP) - Directives pour les chercheuses et chercheurs. *Fonds national suisse de la recherche scientifique* [en ligne]. [Consulté le 8 avril 2020]. Disponible à l'adresse : http://www.snf.ch/fr/leFNS/points-de-vue-politique-de-recherche/open_research_data/Pages/data-management-plan-dmp-directives-pour-les-chercheuses-et-chercheurs.aspx

FNS, 2017b. Explanation of the FAIR data principles. *Fonds national suisse de la recherche scientifique* [en ligne]. [ca. 2017]. [Consulté le 24 février 2020]. Disponible à l'adresse : http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

FNS, 2019. Règlement d'exécution général relatif au règlement des subsides. *Fonds national suisse de la recherche scientifique* [en ligne]. 1 décembre 2019. [Consulté le 20 mars 2020]. Disponible à l'adresse : <http://www.snf.ch/fr/encouragement/documents-telechargements/Pages/reglement-execution-reglement-des-subsides.aspx>

FNS, 2020a. Elaboration d'une stratégie nationale sur les données de recherche ouvertes. *Fonds national suisse de la recherche scientifique* [en ligne]. 9 avril 2020. [Consulté le 20 juin 2020]. Disponible à l'adresse : <http://www.snf.ch/fr/pointrecherche/newsroom/Pages/news-200409-strategie-nationale-des-donnees-ouvertes.aspx>

FNS, 2020b. Open Research Data. *Fonds national suisse de la recherche scientifique* [en ligne]. [ca. 2020]. Disponible à l'adresse : http://www.snf.ch/fr/leFNS/points-de-vue-politique-de-recherche/open_research_data/Pages/default.aspx

FORS, 2018. Data management - Comment assurer le partage de mes données ? *forscenter.ch* [en ligne]. 19 juin 2018. [Consulté le 3 mars 2020]. Disponible à l'adresse : <https://forscenter.ch/data-services/data-management/?lang=fr>

FORTNEY, Katie, 2016. CC BY and data: Not always a good fit. *Office of Scholarly Communication (University of California)* [en ligne]. 15 septembre 2016. [Consulté le 12 juin 2020]. Disponible à l'adresse : <https://osc.universityofcalifornia.edu/2016/09/cc-by-and-data-not-always-a-good-fit/>

GAILLARD, Rémi, 2014. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?* [en ligne]. Lyon : Ecole nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB). Mémoire d'étude. [Consulté le 16 avril 2020]. Disponible à l'adresse : <https://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche>

GRIFFITHS, Aaron, 2009. The Publication of Research Data: Researcher Attitudes and Behaviour. *The International Journal of Digital Curation* [en ligne]. 2009. Vol. 4, n°1, pp. 46-56. [Consulté le 20 juin 2020]. Disponible à l'adresse : <https://doi.org/10.2218/ijdc.v4i1.77>

GUIBAULT, Lucie et WIEBE, Andreas, 2013. *Safe to be open - study on the protection of research data and recommendations for access and usage* [en ligne]. Göttingen : Universitätsverlag Göttingen. [Consulté le 3 mars 2020]. ISBN 978-3-86395-147-4. Disponible à l'adresse : <https://www.fosteropenscience.eu/content/safe-be-open-study-protection-research-data-and-recommendation-access-and-usage>

HAUSMANN, Natacha, 2020. *Relance : LIPAD, RGPD et données de la recherche à l'UNIGE* [message électronique]. 29 juin 2020.

HIRSCHMANN, Barbara, 2020. Creative Commons Licenses. *Manual Research Collection (ETHZ Library)* [en ligne]. 20 janvier 2020. [Consulté le 22 mars 2020]. Disponible à l'adresse : <https://documentation.library.ethz.ch/display/RC/Creative+Commons+Licenses>

HORIZON 2020 FRANCE, 2017. Le libre accès aux publications et aux données de recherche. *Horizon 2020* [en ligne]. [Consulté le 12 mai 2020]. Disponible à l'adresse : <https://www.horizon2020.gouv.fr/cid82025/le-libre-acces-aux-publications-aux-donnees-recherche.html?menu=2>

JONES, Marc, 2018. What's the difference between all those open data licenses? *FOSDEM 2018, Bruxelles, Belgique, 3-4 février 2018* [en ligne]. [Consulté le 12 juin 2020]. Disponible à l'adresse : https://archive.fosdem.org/2018/schedule/event/open_data_licenses/attachments/slides/2535/export/events/attachments/open_data_licenses/slides/2535/Open_Data_Licenses.pdf

JUNOD, Valérie et GORBACHEVA, Daria, 2018. Medical research on pre-existing personal health data: the GDPR exemptions in the European Union. *Life Science Recht*. 2018. Vol. 4, pp. 223-236.

KNIGHT, Gareth, 2015. Introduction to Data Licences. *London School of Hygiene & Tropical Medicine* [en ligne]. 14 juillet 2015. [Consulté le 10 avril 2020]. Disponible à l'adresse : <https://vimeo.com/133429240>

KREUTZER, Till, 2014. Open Content: A Practical Guide to Using Creative Commons Licences. *Wikimedia* [en ligne]. 13 novembre 2014. [Consulté le 30 juin 2020]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:Open_Content_-_A_Practical_Guide_to_Using_Creative_Commons_Licences.pdf

KRÜGEL, Sybil, 2019. *The informed consent as legal and ethical basis of research data production* [en ligne]. Lausanne : Swiss Centre of Expertise in the Social Sciences FORS. [Consulté le 3 février 2020]. FORS Guide, 5. Disponible à l'adresse : <http://doi.org/10.24449/FG-2019-00005>

LABASTIDA, Ignasi et MARGONI, Thomas, 2020. Licensing FAIR Data for Reuse. *Data Intelligence* [en ligne]. Janvier 2020. Vol. 2, n°1-2, pp. 199-207. [Consulté le 5 mars 2020] Disponible à l'adresse : https://doi.org/10.1162/dint_a_00042

LÄMMERHIRT, Danny, 2017. Avoiding data use silos: How governments can simplify the open licensing landscape. *research.okfn.org* [en ligne]. Décembre 2017. [Consulté le 27 juin 2020]. Disponible à l'adresse : <https://research.okfn.org/avoiding-data-use-silos/>

LING, Peter, 2018. Cours I - Le cadre légal des bibliothèques [document PPT]. Support de cours : Cours « Open Science ». Haute école de gestion de Genève. Master en Science de l'information, année académique 2018-2019. 6 décembre 2018.

LUCAS, Isabelle, 2020. Appel à projets Open Data HES-SO. *HES-SO* [en ligne]. 2 juin 2020. [Consulté le 4 juin 2020]. Disponible à l'adresse : <https://www.hes-so.ch/fr/appel-projets-open-data-hes-so-16822.html>

MAUREL, Lionel, 2018. Données personnelles et recherche scientifique : quelle articulation dans le RGPD ? - *S.I.Lex* - [en ligne]. 18 juillet 2018. [Consulté le 24 février 2020]. Disponible à l'adresse : <https://scinfolex.com/2018/07/18/donnees-personnelles-et-recherche-scientifique-quelle-articulation-dans-le-rgpd/>

MEADE, Clair, 2020. Library Guides: Research Data Management Toolkit: Copyright and Licensing. *James Cook University* [en ligne]. 2 juin 2020. [Consulté le 6 juin 2020]. Disponible à l'adresse : <https://libguides.jcu.edu.au/research-data-management-toolkit/copyright-licensing>

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

MÉTILLE, Sylvain, 2014. Protection des données : tout savoir sur le consentement. *Sylvain Métille* [en ligne]. 20 novembre 2014. [Consulté le 18 juin 2020]. Disponible à l'adresse : <https://smetille.ch/2014/11/20/protection-des-donnees-tout-savoir-sur-le-consentement/>

MÉTILLE, Sylvain, 2017a. IV. L'individu et ses données. In : *Internet et droit : protection de la personnalité et questions pratiques*. Genève : Schulthess Editions romandes. pp. 53-72. *Quid iuris ?*, 20. ISBN 978-3-7255-8589-2.

MÉTILLE, Sylvain, 2017b. V. La protection des données personnelles. In : *Internet et droit : protection de la personnalité et questions pratiques*. Genève : Schulthess Editions romandes. pp. 72-83. *Quid iuris ?*, 20. ISBN 978-3-7255-8589-2.

MÉTILLE, Sylvain, 2017c. Révision de la LPD : le projet du Conseil fédéral. *Sylvain Métille* [en ligne]. 19 septembre 2017. [Consulté le 23 avril 2020]. Disponible à l'adresse : <https://smetille.ch/2017/09/19/revision-de-lpd-projet-conseil-federal/>

MÉTILLE, Sylvain, 2020. Protection des données - Partie 1/5. *YouTube* [en ligne]. 25 mars 2020. [Consulté le 1 mai 2020]. Disponible à l'adresse : <https://www.youtube.com/watch?v=mMpw2lgIMAg>

MILZOW, Katrin, VON ARX, Martin, SOMMER, Cornélia, CAHENZLI, Julia et PERINI, Lionel, 2020. *Open Research Data: SNSF monitoring report 2017-2018* [en ligne]. [Consulté le 24 février 2020]. Disponible à l'adresse : <https://zenodo.org/record/3618123#.XIOOc0FCeUn>

MORSCHER, Lukas et SEIDL-NUSSBAUMER, Kaj, 2016. SWITZERLAND. In : *Data protection & privacy: International series (Kuschewsky Monika ed.)* [en ligne]. London : Thomson Reuters. [Consulté le 30 mars 2020]. ISBN 978-0-414-05733-3. Disponible à l'adresse : https://www.lenzstaehelin.com/uploads/tx_netvlsldb/SwissDataProtectionPrivacyChapter_01.pdf

MURRAY-RUST, Peter, NEYLON, Cameron, POLLOCK, Rufus et WILBANKS, John, 2010. *Panton Principles: Principles for Open Data in Science*. *Panton Principles* [en ligne]. 19 février 2010. [Consulté le 12 juin 2020]. Disponible à l'adresse : <https://pantonprinciples.org/>

OCDE, 2007. Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. *oecd.org* [en ligne]. 2007. [Consulté le 24 février 2020]. Disponible à l'adresse : <http://www.oecd.org/fr/science/inno/38500823.pdf>

OFFICE FÉDÉRAL DE LA JUSTICE OFJ, 2019. Renforcement de la protection des données. *Confédération suisse* [en ligne]. 30 janvier 2019. [Consulté le 23 avril 2020]. Disponible à l'adresse : <https://www.bj.admin.ch/bj/fr/home/staat/gesetzgebung/datenschutzstaerkung.html>

OPEN DATA COMMONS, 2009a. Licenses FAQ. *Open Data Commons* [en ligne]. 27 février 2009. [Consulté le 24 février 2020]. Disponible à l'adresse : <https://opendatacommons.org/faq/licenses/>

OPEN DATA COMMONS, 2009b. ODC Open Database License (ODbL) Summary. *Open Data Commons* [en ligne]. 12 septembre 2009. [Consulté le 12 juin 2020]. Disponible à l'adresse : <https://opendatacommons.org/licenses/odbl/summary/>

OPEN DATA COMMONS, 2010. ODC Attribution Summary. *Open Data Commons* [en ligne]. 24 juin 2010. [Consulté le 12 juin 2020]. Disponible à l'adresse : <https://opendatacommons.org/licenses/by/summary/>

OPEN DEFINITION, 2015. Open Definition 2.1. *Open Definition* [en ligne]. [Consulté le 11 juin 2020]. Disponible à l'adresse : <http://opendefinition.org/od/2.1/en/>

OPENAIRE, 2019a. Specifications of licensing Research Data. *OpenAIRE* [en ligne]. [Consulté le 24 février 2020]. Disponible à l'adresse : <https://www.openaire.eu/how-do-i-license-my-research-data>

OPENAIRE, 2019b. How to apply licenses for Research Data. *OpenAIRE* [en ligne]. [Consulté le 24 février 2020]. Disponible à l'adresse : <https://www.openaire.eu/how-do-i-license-my-research-data>

PFPDT, [sans date]. Protection des données et recherche en général. *Confédération suisse* [en ligne]. [Consulté le 20 mars 2020]. Disponible à l'adresse : <https://www.edoeb.admin.ch/edoeb/fr/home/datenschutz/statistik--register-und-forschung/forschung/datenschutz-und-forschung-im-allgemeinen.html>

PFPDT, 2018. Le RGPD et ses conséquences sur la Suisse. *Confédération suisse* [en ligne]. mars 2018. [Consulté le 7 mai 2020]. Disponible à l'adresse : https://www.edoeb.admin.ch/dam/edoeb/fr/dokumente/2018/Le%20RGPD%20et%20ses%20cons%C3%A9quences%20sur%20la%20Suisse_FR%20Jan%2018.pdf.download.pdf/Le%20RGPD%20et%20ses%20cons%C3%A9quences%20sur%20la%20Suisse_FR%20Jan%2018.pdfBe9NK

POPKIN, Gabriel, 2019. Data sharing and how it can benefit your scientific career. *Nature* [en ligne]. 13 mai 2019. Vol. 569, n°7756, pp. 445-447. [Consulté le 30 mars 2020]. Disponible à l'adresse : <https://doi.org/10.1038/d41586-019-01506-x>

PPDT GE, 2015a. *Secret médical et protection des données : fondements juridiques et aspects pratiques*. [en ligne]. Genève : République et canton de Genève. [Consulté le 21 avril 2020]. Fiches info du PPDT. Disponible à l'adresse : <https://www.ge.ch/ppdt/fiches-info/doc/secret-medical.pdf>

PPDT GE, 2015b. Accueil. *PPDT Catalogue des fichiers* [en ligne]. [Consulté le 25 mai 2020]. Disponible à l'adresse : <http://outil.ge.ch/chacatfich/#/home>

SCHNEIDER, René, 2018. Introduction à la gestion des données de la recherche [document PDF]. Support de cours : Cours « Gestion des données de la recherche ». Haute école de gestion de Genève. Master en Science de l'information, année académique 2018-2019. 15 novembre 2018.

SEFRI, 2019a. Vereinbarung betreffend die Erarbeitung einer «Nationalen Strategie Open Research Data» und eines zugehörigen «Umsetzungsplans». *swissuniversities* [en ligne]. [Consulté le 20 juin 2020]. Disponible à l'adresse : https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Organisation/SUK-P/SUK_P-2/Vereinbarung_Open_Research_Data-sign.pdf

SEFRI, 2019b. Statut de la Suisse dans Horizon 2020. *Secrétariat d'État à la formation, à la recherche et à l'innovation* [en ligne]. [Consulté le 20 juin 2020]. Disponible à l'adresse : <https://www.sbfi.admin.ch/sbfi/fr/home/themen/internationale-forschungs-und-innovationszusammenarbeit/forschungsrahmenprogramme-der-europaeischen-union/horizon-2020/h2020.html>

SEFRI, 2020. Horizon Europe : 9e programme-cadre de recherche et d'innovation de l'Union européenne (2021-2027). *Secrétariat d'État à la formation, à la recherche et à l'innovation* [en ligne]. [Consulté le 20 juin 2020]. Disponible à l'adresse : <https://www.sbfi.admin.ch/sbfi/fr/home/themen/internationale-forschungs-und-innovationszusammenarbeit/forschungsrahmenprogramme-der-europaeischen-union/horizon-europe.html>

SWISS ACADEMIES OF ARTS AND SCIENCE, 2019. Open Science in Switzerland: Opportunities and Challenges. *Swiss Academies Factsheets* [en ligne]. Vol. 14, n°2. [Consulté le 19 mars 2020]. Disponible à l'adresse : http://www.swiss-academies.ch/index/Publikationen/Swiss-Academies-Factsheets/mainColumnParagraphs/0111/download_website_en.pdf

SWISSUNIVERSITIES, 2019. DLCM Phase 2. *swissuniversities* [en ligne]. [Consulté le 15 juin 2020]. Disponible à l'adresse : <https://www.swissuniversities.ch/fr/themes/digitalisation/p-5-information-scientifique/projets/dlcm-phase-2>

SWISSUNIVERSITIES, 2020a. Stratégie nationale et plan d'action. *swissuniversities* [en ligne]. [Consulté le 19 mars 2020]. Disponible à l'adresse : <https://www.swissuniversities.ch/fr/themes/digitalisation/open-access/translate-to-franzoesisch-nationale-strategie-open-access>

SWISSUNIVERSITIES, 2020b. Open Science - le programme pour des sciences ouvertes. *swissuniversities* [en ligne]. [Consulté le 19 mars 2020]. Disponible à l'adresse : <https://www.swissuniversities.ch/fr/themes/digitalisation/open-science>

SWISSUNIVERSITIES, 2020c. Programme. *swissuniversities* [en ligne].

[Consulté le 19 mars 2020]. Disponible à l'adresse :

<https://www.swissuniversities.ch/fr/themes/digitalisation/open-science/programme>

SWISSUNIVERSITIES, 2020d. Préparations stratégiques. *swissuniversities* [en ligne].

[Consulté le 19 mars 2020]. Disponible à l'adresse :

<https://www.swissuniversities.ch/fr/themes/digitalisation/programme-open-science/preparations-strategiques>

THE OPEN DATA INSTITUTE, 2015. What are the impacts of non-open licences? *The Open Data Institute* [en ligne]. 22 septembre 2015. [Consulté le 27 juin 2020]. Disponible à

l'adresse : <https://theodi.org/article/what-are-the-impacts-of-non-open-licences/>

THOROGOOD, Adrian, 2019. *Towards Legal Interoperability in International Health Research* [en ligne]. Toronto : University of Toronto. Master of Laws' thesis.

[Consulté le 11 mars 2020]. Disponible à l'adresse : <http://hdl.handle.net/1807/98411>

UNIVERSITÄT BASEL - RESEARCH DATA MANAGEMENT-NETWORK, 2019. RDM Cheat Sheets. *researchdata.unibas.ch* [en ligne]. [Consulté le 21 juin 2020]. Disponible à l'adresse :

https://researchdata.unibas.ch/fileadmin/user_upload/researchdata/FastGuide_v1_print_1.pdf

UNIVERSITÉ DE GENÈVE, 2012. Loi sur l'information du public, l'accès aux documents et la protection des données personnelles (LIPAD) : application à l'Université. *Memento de l'Université de Genève* [en ligne]. 26 novembre 2012. [Consulté le 15 avril 2020]. Disponible

à l'adresse : <https://memento.unige.ch/doc/0160/#e-protection-des-donn>

UNIVERSITÉ DE GENÈVE, 2017. Préservation à l'UNIGE. *Données de recherche* [en ligne]. 31 janvier 2017. [Consulté le 18 juin 2020]. Disponible à l'adresse :

<https://www.unige.ch/researchdata/fr/preserve/all/storage/>

UNIVERSITÉ DE LAUSANNE, 2019a. Open Science. *L'Open Science à l'UNIL* [en ligne]. 19 décembre 2019. [Consulté le 8 avril 2020]. Disponible à l'adresse :

<https://www.unil.ch/openscience/fr/home/menuinst/strategieplan-dactions/open-science.html>

UNIVERSITÉ DE LAUSANNE, 2019b. Data Management Plan (DMP). *L'Open Science à l'UNIL* [en ligne]. [Consulté le 13 juin 2020]. Disponible à l'adresse :

<https://www.unil.ch/openscience/fr/home/menuinst/open-research-data/gerer-ses-donnees-de-recherche/data-management-plan-dmp.html>

UNIVERSITÉ DE LAUSANNE, 2020. Données personnelles & sensibles. *L'Open Science à l'UNIL* [en ligne]. Avril 2020. [Consulté le 3 février 2020]. Disponible à l'adresse :

<https://www.unil.ch/openscience/home/menuinst/open-research-data/conformite-exigences/donnees-personnelles-sensibles.html>

VILMEN, Céline, 2018a. Droit d'auteur : Données de la recherche. *LibGuides* [en ligne]. 7 mars 2018. [Consulté le 3 mars 2020]. Disponible à l'adresse :

<https://libguides.graduateinstitute.ch/droit-dauteur/donnees-recherche>

VILMEN, Céline, 2018b. Droit d'auteur : J'utilise les Creative Commons. *LibGuides* [en ligne]. 7 mars 2018. [Consulté le 15 avril 2020]. Disponible à l'adresse :

<https://libguides.graduateinstitute.ch/droit-dauteur/creative-commons>

VILMEN, Céline, 2019. Droit d'auteur : Je suis professeur. *LibGuides* [en ligne]. 2019. [Consulté le 15 avril 2020]. Disponible à l'adresse :

<https://libguides.graduateinstitute.ch/droit-dauteur/professeur>

WILKINSON, Mark D., DUMONTIER, Michel, AALBERSBERG, IJsbrand Jan, APPLETON, Gabrielle, AXTON, Myles, BAAK, Arie, BLOMBERG, Niklas, BOITEN, Jan-Willem, DA SILVA SANTOS, Luiz Bonino, BOURNE, Philip E., BOUWMAN, Jildau, BROOKES, Anthony J., CLARK, Tim, CROSAS, Mercè, DILLO, Ingrid, DUMON, Olivier, EDMUNDS, Scott, EVELO, Chris T., FINKERS, Richard, GONZALEZ-BELTRAN, Alejandra, GRAY, Alasdair J.G., GROTH, Paul, GOBLE, Carole, GRETHE, Jeffrey S., HERINGA, Jaap, 'T HOEN, Peter A.C, HOOFT, Rob, KUHN, Tobias, KOK, Ruben, KOK, Joost, LUSHER, Scott J., MARTONE, Maryann E., MONS, Albert, PACKER, Abel L., PERSSON, Bengt, ROCCA-SERRA, Philippe, ROOS, Marco, VAN SCHAIK, Rene, SANSONE, Susanna-Assunta, SCHULTES, Erik, SENGSTAG, Thierry, SLATER, Ted, STRAWN, George, SWERTZ, Morris A., THOMPSON, Mark, VAN DER LEI, Johan, VAN MULLIGEN, Erik, VELTEROP, Jan, WAAGMEESTER, Andra, WITTENBURG, Peter, WOLSTENCROFT, Katherine, ZHAO, Jun et MONS, Barend, 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* [en ligne]. Décembre 2016. Vol. 3, n° 1. [Consulté le 24 février 2020]. Disponible à l'adresse : <https://doi.org/10.1038/sdata.2016.18>

ZHAW RESSORT FORSCHUNG & ENTWICKLUNG, 2019. Arbeitsanleitung zu Open Research Data und Datenmanagement-Plänen. *zhaw* [en ligne]. 19 décembre 2019. [Consulté le 21 juin 2020]. Disponible à l'adresse : https://gpmpublic.zhaw.ch/GPMDocProdDPublic/Vorgabedokumente_ZHAW/Z_AA_Arbeitsanleitung_ORD.pdf

Annexe 1 : Principes FAIR pour les données

The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

(Wilkinson et al. 2016)

Annexe 2 : Aperçu de la situation Open Science dans les hautes écoles suisses

Les informations réunies ici ne sont pas exhaustives, des initiatives en cours ou terminées ayant probablement échappé à mon attention. Ce tableau n'est ainsi qu'un aperçu de la situation Open Science dans les hautes écoles suisses, peut-être déjà en partie obsolète car la collecte d'informations a majoritairement eu lieu en mars 2020 avec une rapide vérification en juin 2020. Celle-ci s'est concentrée sur trois types de sources : les sites web institutionnels des hautes écoles, le *Registry of Open Access Repository Mandates and Policies* (ROARMAP) et le *Registry of Research Data Repositories* (re3data.org). Manquent dans ce tableau la Fachhochschule Nordwestschweiz (FHNW), la Fachhochschule Ostschweiz (FHO) et la Fachhochschule Graubünden (FHGR) car aucune information sur l'Open Science ne figure sur leur site web.

Nom de l'institution	Stratégie cadre Open Science	Politique ou directive Open Access pour les publications	Politique ou directive pour l'Open Research Data	Pages web dédiées à l'Open Research Data et soutien offert aux chercheurs	Dépôt pour les publications	Dépôt pour les données
École polytechnique fédérale de Lausanne (EPFL)		X		X	Infoscience	
Eidgenössische Technische Hochschule Zürich (ETHZ)		X		X	Research collection	Research collection
Université de Genève (UNIGE)		X	X	X	Archive ouverte	Yareta
Université de Lausanne (UNIL)	X	X	X	X	SERVAL	
Université de Fribourg (UNIFR)		X		X	RERO DOC	
Université de Neuchâtel (UNINE)		X			Libra	
Universität Basel (UNIBAS)		X		X	edoc	Projet en cours
Universität Bern (UNIBE)		X		X	BORIS BORIS Theses	BORIS

Universität Luzern (UNILU)		X		X via la Zentral- und Hochschulbibliothek Luzern	LORY	LORY
Universität St.Gallen (HSG)		X		X	Alexandria Research Platform et EDIS (thèses)	
Universität Zürich (UZH)		X		X	ZORA	
Università della Svizzera italiana (USI)				X	RERO DOC	
Haute école spécialisée de Suisse occidentale (HES-SO)	X			X	ArODES	
Zürcher Hochschule für Angewandte Wissenschaften (ZHAW)		X		X	ZHAW digitalcollection	
Zürcher Hochschule der Künste (ZHdK)		X	X	X	ZOPAR	
Hochschule Luzern (HSLU)		X		X via la Zentral- und Hochschulbibliothek Luzern	LORY	LORY
Berner Fachhochschule (BFH)		X			ARBOR	
Scuola universitaria professionale della Svizzera italiana (SUPSI)	X			X	SUPSI Instory	

Annexe 3 : Définitions liées à la LPD

Données personnelles : « toutes les informations qui se rapportent à une personne identifiée ou identifiable » (LPD, art. 3 let. a). La possibilité d'identification peut venir d'un seul élément (ex. le numéro AVS) ou d'une combinaison d'éléments (nom, prénom, adresse) (Meier 2011, p. 202-203).

Données sensibles : « les données personnelles sur : 1. les opinions ou activités religieuses, philosophiques, politiques ou syndicales, 2. la santé, la sphère intime ou l'appartenance à une race, 3. des mesures d'aide sociale, 4. des poursuites ou sanctions pénales et administratives. » (LPD, art. 3, let. c)

Profil de la personnalité : « un assemblage de données qui permet d'apprécier les caractéristiques essentielles de la personnalité d'une personne physique. ». (LPD, art.3, let. d)

Personne concernée : « la personne physique ou morale au sujet de laquelle des données sont traitées » (LPD art. 3, let. b). La personnalité physique prenant fin avec la mort, les données personnelles de personnes décédées ne sont plus protégées par la LPD, mais leur accès n'en devient pas libre pour autant (par exemple en raison du secret médical) et « leurs ayants droits peuvent invoquer la protection de leur propre personnalité pour les protéger. » (Meier 2011, p. 213-214).

Organe fédéral : « l'autorité ou le service fédéral ainsi que la personne en tant qu'elle est chargée d'une tâche de la Confédération. » (LPD art. 3, let. h)

Traitement : « toute opération relative à des données personnelles – quels que soient les moyens et procédés utilisés – notamment la collecte, la conservation, l'exploitation, la modification, la communication, l'archivage ou la destruction de données. » (LPD art. 3, let. e)

Consentement : il n'y a pas de réelle définition du consentement dans la LPD. Il est mentionné que « la personne concernée ne consent valablement que si elle exprime sa volonté librement et après avoir été dûment informée. Lorsqu'il s'agit de données sensibles et de profils de la personnalité, son consentement doit être au surplus explicite. » (LPD art. 4, al. 5)

Consentement explicite : ce qui doit être explicite n'est pas « la forme de la manifestation de la volonté mais l'objet matériel du consentement ». Il est donc nécessaire dans ce cas que le traitement soit formulé correctement (qu'il ne soit donc pas implicite) (Meier 2011, p. 343).

Fichier : « tout ensemble de données personnelles dont la structure permet de rechercher les données par personne concernée. » (LPD art. 3, let. g)

Registre des fichiers : il doit être tenu à jour et déclaré au préposé fédéral à la protection des données dans tous les cas pour les organes fédéraux et si elles traitent des données sensibles ou communiquent régulièrement des données personnelles à des tiers pour les privés. (LPD art. 11a, al. 1 à 3)

Maître du fichier : « la personne privée ou l'organe fédéral qui décide du but et du contenu du fichier » (LPD art. 3 let. i). C'est donc le responsable du traitement et celui qui le contrôle sur le fichier, pouvant décider « de sa création, de ses buts et des moyens de traitement utilisés. » (Métille 2017b, p. 82)

Communication : « le fait de rendre des données personnelles accessibles, par exemple en autorisant leur consultation, en les transmettant ou en les diffusant. » (LPD art. 3, let. f)

Anonymisation : les données personnelles sont anonymisées quand elles ne peuvent plus être attribuées à une personne identifiée et identifiable. Rendre anonyme, c'est réaliser toutes les démarches pour empêcher l'identification des personnes concernées, ou si l'identification est encore possible, faire qu'elle le soit seulement « au prix d'efforts démesurés ». L'anonymisation suppose une suppression définitive de toutes les données identifiantes et de ne laisser aucun moyen de retrouver les données originales, elle est ainsi irréversible (PFPDT [sans date]). Si les données personnelles sont anonymes, elles ne sont plus concernées par la LPD.

Pseudonymisation : les données personnelles sont pseudonymisées quand l'ensemble des données identifiantes a été remplacé par un identifiant neutre (un pseudonyme). La pseudonymisation est donc réversible, une table de correspondance permettant normalement de lier le pseudonyme aux données identifiantes de la personne) (PFPDT [sans date]). La pseudonymisation est une mesure de sécurité et permet de limiter l'atteinte à la personne concernée, les personnes ayant accès aux données codées n'ayant pas toutes accès à la table de correspondance également (Métille 2017b). Les données personnelles pseudonymisées sont toujours concernées et protégées par la LPD.

Annexe 4 : Définitions liées à la LIPAD

Données personnelles : « toutes les informations se rapportant à une personne physique ou morale de droit privé, identifiée ou identifiable. » (LIPAD art. 4, let. a)

Données personnelles sensibles : « données personnelles sur : 1° les opinions ou activités religieuses, philosophiques, politiques, syndicales ou culturelles, 2° la santé, la sphère intime ou l'appartenance ethnique, 3° des mesures d'aide sociale, 4° des poursuites ou sanctions pénales ou administratives. » (LIPAD art. 4, let. b)

Profil de la personnalité : « un assemblage de données qui permet d'apprécier les caractéristiques essentielles de la personnalité d'une personne physique. » (LIPAD art. 4, let. c)

Personne concernée : « la personne physique ou morale au sujet de laquelle des données sont traitées. » (LIPAD art. 4, let. g)

Organe : « tout membre ou tout mandataire d'une institution visée à l'article 3 [champ d'application] et assumant, pour le compte de celle-ci, la diffusion active des informations prévue à l'article 18, le traitement des demandes d'accès aux documents régies par la présente loi, ou celui de données personnelles. » (LIPAD art. 4, let. h)

Traitement : « toute opération relative à des données personnelles – quels que soient les moyens et procédés utilisés – notamment la collecte, la conservation, l'exploitation, la modification, la communication, l'archivage ou la destruction de données. » (LIPAD art. 4, let. e)

Consentement : le consentement n'est pas défini dans la LIPAD. Concernant le traitement de données personnelles sensibles, il est en revanche mentionné : « des données personnelles sensibles ou des profils de la personnalité ne peuvent être traités que si une loi définit clairement la tâche considérée et si le traitement en question est absolument indispensable à l'accomplissement de cette tâche ou s'il est nécessaire et intervient avec le consentement explicite, libre et éclairé de la personne concernée. » (LIPAD art. 35, al. 2)

Pour la définition du **consentement éclairé**, voir celle proposée en annexe 3.

Fichier : « tout système destiné à réunir, sur quelque support que ce soit, des données personnelles d'un segment de population déterminé, et structuré de manière à permettre de relier les informations recensées aux personnes qu'elles concernent. » (LIPAD art. 4, let. d)

Communication : « le fait de rendre accessibles des données personnelles ou un document, par exemple en autorisant leur consultation, en les transmettant ou en les diffusant. » (LIPAD art. 4, let. f)

Anonymisation : voir définition proposée en annexe 3.

Pseudonymisation : voir définition proposée en annexe 3.

Annexe 5 : Définitions liées au RGPD

Données à caractère personnel : « toute information relative à une personne physique identifiée ou identifiable directement ou indirectement (nom, prénom, date de naissance, sexe, email privé ou professionnel, pseudo, adresse IP, situation familiale, données bancaires, géolocalisation, données de connexion ou de navigation, etc.) » (RGPD §4, al. 1 ; Vilmen 2018c)

Données sensibles : « données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique ». (RGPD §9, al. 1)

Profilage : « toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique. » (RGPD §4, al. 4)

Données concernant la santé : « les données à caractère personnel relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne. » (RGPD §4, al. 15)

Traitement : « toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel, telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction » (RGPD §4, al. 2). Il n'y a donc pas besoin de manipuler des données pour être concerné, même si elles ne font que transiter par une entreprise, celle-ci est concernée par le RGPD (Bayle 2018).

Consentement : « toute manifestation de volonté, libre, spécifique, éclairée et univoque par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement. » (RGPD §4, al. 11)

Fichier : « tout ensemble structuré de données à caractère personnel accessibles selon des critères déterminés, que cet ensemble soit centralisé, décentralisé ou réparti de manière fonctionnelle ou géographique. » (RGPD §4, al. 6)

Responsable du traitement : « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement. » (RGPD §4, al. 7)

Registre des traitements : son contenu est détaillé à l'article 30, paragraphe 1 : le nom et les coordonnées du responsable de traitement ; les finalités du traitement ; une description des

catégories de personnes concernées et des catégories de données à caractère personnel ; les catégories de destinataires auxquels les données sont communiquées ou pourraient l'être ; les transferts de données à caractère personnel vers un pays tiers ou vers une organisation internationale (si applicable) ; les délais prévus pour l'effacement des différentes catégories de données et une description générale des mesures de sécurité techniques et organisationnelles qui sont mises en place.

Pseudonymisation : « le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable. » (RGPD §4, al. 5)

Traitement transfrontalier :

- a) « un traitement de données à caractère personnel qui a lieu dans l'Union dans le cadre des activités d'établissements dans plusieurs États membres d'un responsable du traitement ou d'un sous-traitant lorsque le responsable du traitement ou le sous-traitant est établi dans plusieurs États membres; ou
- b) un traitement de données à caractère personnel qui a lieu dans l'Union dans le cadre des activités d'un établissement unique d'un responsable du traitement ou d'un sous-traitant, mais qui affecte sensiblement ou est susceptible d'affecter sensiblement des personnes concernées dans plusieurs États membres. » (RGPD §4, al. 23)

Violation de données à caractère personnel : « une violation de la sécurité entraînant, de manière accidentelle ou illicite, la destruction, la perte, l'altération, la divulgation non autorisée de données à caractère personnel transmises, conservées ou traitées d'une autre manière, ou l'accès non autorisé à de telles données. » (RGPD §4, al. 12)

Annexe 6 : Qui peut utiliser du contenu sous licence Creative Commons de type NC et dans quel cas ?

CHART 1: WHO CAN USE NC CONTENT IN WHAT USE CASE?

















ABBREVIATIONS:
Yes = Use of NC content is permitted
No = Use of NC content is not permitted
n.a. = Not applicable, i.e. such constellation is inconceivable in the logic of the chart as explained above

User type	Company	Public institution	Non-profit NGO	Free-lancer	Private person
Sell hardcopies	No	No	No	No	No
Licence content against payment	No	No	No	No	No
Use for advertising	No	No	No	No	No
Use to make money	No	No	No	No	No
Use for the job	n.a.	n.a.	n.a.	No	n.a.
Use on a website that displays ads to recover hosting costs	No	Yes	Yes	No	Yes
Use on a platform, where the platform provider (not the content provider) displays ads	No	Yes	Yes	No	Yes
Use for inhouse education and information	No	Yes	Yes	No	n.a.
Use for private entertainment and to entertain friends/family of the user	n.a.	n.a.	n.a.	n.a.	Yes
Use to inform/entertain customers/clients/audience	No	Yes	Yes	No	Yes
Use in tuition-free courses for educational purposes	No	Yes	Yes	No	Yes
Use in tuition-based courses for educational purposes	No	No	No	No	No
Use for corporate-funded research	No	No	No	No	n.a.
Use for tax-funded research	No	Yes	Yes	No	n.a.
Use for inhouse corporate research	No	n.a.	n.a.	No	n.a.

KREUTZER, Till, 2014. *Open Content: a practical guide to using Creative Commons Licences* [en ligne]. Bonn : German Comm. for UNESCO. [Consulté le 30 juin 2020]. ISBN 978-3-940785-57-2. Disponible à l'adresse : https://meta.wikimedia.org/wiki/File:Open_Content_-_A_Practical_Guide_to_Using_Creative_Commons_Licences.pdf. Page 46.

Tableau disponible sous licence [CC-BY 4.0](#)

Annexe 7 : Compatibilité des licences Creative Commons entre elles

	 PUBLIC DOMAIN	 PUBLIC DOMAIN	 CC BY	 CC BY SA	 CC BY NC	 CC BY ND	 CC BY NC SA	 CC BY NC ND
 PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
 PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
 CC BY	✓	✓	✓	✓	✓	✗	✓	✗
 CC BY SA	✓	✓	✓	✓	✗	✗	✗	✗
 CC BY NC	✓	✓	✓	✗	✓	✗	✓	✗
 CC BY ND	✗	✗	✗	✗	✗	✗	✗	✗
 CC BY NC SA	✓	✓	✓	✗	✓	✗	✓	✗
 CC BY NC ND	✗	✗	✗	✗	✗	✗	✗	✗

KENNISLAND, 2013. CC License Compatibility Chart. *Creative Commons Wiki*. [en ligne]. 31 mai 2013. [Consulté le 1 juillet 2020]. Disponible à l'adresse : https://wiki.creativecommons.org/wiki/File:CC_License_Compatibility_Chart.png. Disponible sous licence [CC0](#).

Annexe 8 : Tableau comparatif des licences ouvertes pour les données

Ce tableau est une adaptation de la présentation de Marc Jones : *What's the difference between all those open data licenses?* présentée lors de l'évènement FOSDEM 2018 à Bruxelles, en Belgique, les 3 et 4 février 2018.

L'œuvre originale est disponible sous licence [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) à l'adresse :

https://archive.fosdem.org/2018/schedule/event/open_data_licenses/attachments/slides/2535/export/events/attachments/open_data_licenses/slides/2535/Open_Data_Licenses.pdf

La présentation a ici été réorganisée en présentant le contenu des slides sous la forme d'un tableau. Le contenu a volontairement été laissé en anglais pour éviter des erreurs de traduction. Deux ajouts ont été fait quant aux licences CDLA : ils figurent en bleu.

Ce tableau est également disponible sous la licence [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).

Name	Type	Scope	Representations	Conditions
Creative Commons CCZero (CC0)	Public Domain Dedication <ul style="list-style-type: none"> •Applies to Copyrights •Applies to Database rights •Falls back to a world wide license •Waiver of Moral Rights 	<ul style="list-style-type: none"> •Content of the database protected by copyright •Structure of the database protected by sui generis right 	<ul style="list-style-type: none"> •No warranties •No license for Patent and trademark rights •No clearance of privacy or other third party rights 	
Open Data Commons Public Domain Dedication and License (PDDL)	Public Domain Dedication <ul style="list-style-type: none"> •Applies to Copyrights •Applies to Database rights •Falls back to a world wide license •Waiver of Moral Rights •Suggests using "Community Norms" 	<ul style="list-style-type: none"> •Content of the database protected by copyright •Structure of the database protected by sui generis right •Does not apply to computer programs 	<ul style="list-style-type: none"> •No warranties •No license for Patent and trademark rights •Does not address other rights •Liability Limitation 	

Name	Type	Scope	Representations	Conditions
Creative Commons Attribution 4.0 (CC-BY)	License <ul style="list-style-type: none"> ● Applies to Copyrights ● Applies to Database rights ● Waiver of Moral Rights ● Contemplates being a contract 	<ul style="list-style-type: none"> ● Content of the database protected by copyright ● Structure of the database protected by sui generis right ● Prohibits “Effective Technology Measures ● May not limit rights granted on work ● Automatic termination & reinstatement 	<ul style="list-style-type: none"> ● No warranties ● No license for Patent and trademark ● No license for publicity/personality ● Does not address other rights ● Liability Limitation 	<ul style="list-style-type: none"> ● Retain name, notices, and URI ● Indicate modification ● Indicate License on original work ● Conditions done in reasonable manner
Creative Commons Attribution-ShareAlike 4.0 (CC-BY-SA)	License <ul style="list-style-type: none"> ● Applies to Copyrights ● Applies to Database rights ● Waiver of Moral Rights ● Contemplates being a contract 	<ul style="list-style-type: none"> ● Content of the database protected by copyright ● Structure of the database protected by sui generis right ● Prohibits “Effective Technology Measures” ● May not limit rights granted on work ● Automatic termination & reinstatement 	<ul style="list-style-type: none"> ● No warranties ● No license for Patent and trademark ● No license for publicity/personality ● Does not address other rights ● Liability Limitation 	<ul style="list-style-type: none"> ● Retain name, License, notices, and URI ● Indicate License ● Indicate modification ● Attribution done in reasonable manner ● Derivatives need to be BY-SA

Name	Type	Scope	Representations	Conditions
Open Data Commons Attribution License (ODB-By)	Contract <ul style="list-style-type: none"> •License grant applies to Copyrights •License grant applies to Database rights •Explicitly a contract •Selects law of jurisdiction of place of enforcement 	<ul style="list-style-type: none"> •Does not include Content of the database protected by copyright and moral rights •Structure of the database protected by sui generis right, <i>copyrights and moral rights</i> •No sublicensing •Automatic termination & reinstatement 	<ul style="list-style-type: none"> •No warranties •No license for Patent and trademark rights •Liability Limitation 	<ul style="list-style-type: none"> •Preserve notices •When copying Include notices in the database file <ul style="list-style-type: none"> ◦Unless not possible •Must include notice when publishing Produced Work publically
ODC Open Database License (ODbL)	Contract <ul style="list-style-type: none"> •License grant applies to Copyrights •License grant applies to Database rights •Explicitly a contract •Selects law of jurisdiction of place of enforcement 	<ul style="list-style-type: none"> •Does not include Content of the database protected by copyright and moral rights •Structure of the database protected by sui generis right, <i>copyrights and moral rights</i> •No sublicensing •Automatic termination & reinstatement 	<ul style="list-style-type: none"> •No warranties •No license for Patent and trademark rights •Liability Limitation 	<ul style="list-style-type: none"> •Copies and derivatives must be under this license (copyleft) or permitted license •Permitted license is <ul style="list-style-type: none"> ◦this license ◦a later version of ODbL ◦a “compatible” license •Preserve notices •When copying Include notices <i>in the database file</i> <ul style="list-style-type: none"> ◦Unless not possible •Must include notice when publishing Produced Work publically •“If you Publicly Use ... ”: <ul style="list-style-type: none"> ■a Derivative Database, or ■a Produced Work from a Derivative Database <ul style="list-style-type: none"> ◦“must offer recipients of the Derivative Database or Produced work” ◦“a copy in machine readable form” ◦a copy of the Derivative Database •No additional restrictive terms or technologies

Name	Type	Scope	Representations	Conditions
Community Data License Agreement - Permissive (CDLA-Permissive)	Contract <ul style="list-style-type: none"> ●License grant applies to Copyrights ●License grant applies to Database rights ●Waives moral rights ●Explicitly a contract 	<ul style="list-style-type: none"> ●Content protected by copyright and <i>sui generis</i> right ●Sublicensing allowed ●Automatic termination on “awareness” of violation & automatic reinstatement ●Litigation retaliation clause 	<ul style="list-style-type: none"> ●No warranties ●No license for Patent and trademark ●Represents no Private/Confidential data ●Liability Limitation 	<ul style="list-style-type: none"> ●Derivatives/Enhanced data may be under different terms ●Must provide copy of the agreement when copying ●Preserve notices “in Data” ●Indicate modification of files ●Preserve all credit or attribution to the Data Provider(s)
Community Data License Agreement - Sharing (CDLA-Sharing)	Contract <ul style="list-style-type: none"> ●License grant applies to Copyrights ●License grant applies to Database rights ●Waives moral rights ●Explicitly a contract 	<ul style="list-style-type: none"> ●Content protected by copyright and <i>sui generis</i> right ●Automatic termination on “awareness” of violation & automatic reinstatement ●Litigation retaliation clause 	<ul style="list-style-type: none"> ●No warranties ●No license for Patent and trademark ●Representation that there is no Private/Confidential data ●Liability Limitation 	<ul style="list-style-type: none"> ●Derivatives/Enhanced Data must published under this Agreement ●Must provide copy of the agreement when copying ●Preserve notices “in Data” ●Indicate modification of files ●Preserve all credit or attribution to the Data Provider(s) ●May not restrict public publication ●May not restrict publication in designated Ledgers ●No additional restrictions on Derivative works.

Annexe 9 : Avantages et inconvénients des licences ouvertes pour les données

Licence	Avantages	Inconvénients
Creative Commons CCZero (CC0) (Déclaration de domaine public)	<p>Outil mondialement reconnu et déjà couramment utilisé dans le milieu de la recherche.</p> <p>Résumé juridique facilement compréhensible.</p> <p>Couvre la base de données et les données.</p> <p>Disponible dans un langage pour les machines.</p> <p>Lève tous les droits sur les données en ne créant aucune ambiguïté. Permet une réutilisation totale des données et à toutes fins.</p>	<p>Manque de contrôle sur la manière dont les données sont réutilisées.</p>
Open Data Commons Public Domain Dedication and License (PDDL) (Déclaration de domaine public)	<p>Résumé juridique facilement compréhensible.</p> <p>Couvre la base de données et/ou les données.</p> <p>Lève tous les droits sur les données en ne créant aucune ambiguïté. Permet une réutilisation totale des données et à toutes fins.</p>	<p>Manque de contrôle sur la manière dont les données sont réutilisées.</p> <p>Non disponible dans un langage pour les machines.</p> <p>Moins connue que CC0.</p>
Creative Commons Attribution 4.0 (CC-BY) (Licence d'attribution)	<p>Licence mondialement reconnue et déjà couramment utilisé dans le milieu de la recherche.</p> <p>Résumé juridique facilement compréhensible.</p> <p>Permet une réutilisation totale des données et à toutes fins.</p> <p>Couvre la base de données et les données.</p> <p>Disponible dans un langage pour les machines.</p>	<p>Manque de contrôle sur la manière dont les données sont réutilisées.</p> <p>Problème de l'empilement des attributions.</p> <p>Souvent, les données de la recherche ne sont pas protégées par le droit d'auteur, la licence n'aura pas d'effet dans ce cas-là.</p>
Creative Commons Attribution-ShareAlike 4.0 (CC-BY-SA) (Licence d'attribution copyleft)	<p>Licence mondialement reconnue et déjà courante dans le milieu de la recherche.</p> <p>Résumé juridique facilement compréhensible.</p> <p>Disponible dans un langage pour les machines.</p>	<p>Problème de l'empilement des attributions.</p> <p>Problème de l'incompatibilité des licences copyleft entre elles.</p> <p>Souvent, les données de la recherche ne sont pas protégées par le droit d'auteur, la licence n'aura pas d'effet dans ce cas-là.</p>
Open Data Commons Attribution License (ODB-By) (Licence d'attribution)	<p>Résumé juridique facilement compréhensible.</p> <p>Permet une réutilisation totale des données et à toutes fins.</p> <p>Permet de distinguer la base de données de son contenu.</p> <p>Distingue les bases de données dérivées et les contenus dérivés qui ne sont pas des bases de données, mais produits à partir de celles-ci (ex. graphiques, schémas, cartes).</p>	<p>Manque de contrôle sur la manière dont les données sont réutilisées.</p> <p>Besoin d'utiliser deux licences différentes pour la base de données et le contenu.</p> <p>Non disponible dans un langage pour les machines.</p> <p>Difficile à comprendre par des non-juristes.</p> <p>Problème de l'empilement des attributions.</p>

Licence	Avantages	Inconvénients
ODC Open Database License (ODbL) (Licence d'attribution copyleft)	Résumé juridique facilement compréhensible. Permet de distinguer la base de données de son contenu. Distingue les bases de données dérivées et les contenus dérivés qui ne sont pas des bases de données, mais produits à partir de celles-ci (ex. graphiques, schémas, cartes).	Besoin d'utiliser deux licences différentes pour la base de données et le contenu. Non disponible dans un langage pour les machines. Difficile à comprendre par des non-juristes. Problème de l'empilement des attributions. Problème de l'incompatibilité des licences copyleft entre elles.
Community Data License Agreement - Permissive (CDLA-Permissive) (Licence d'attribution)	Permet une réutilisation totale des données et à toutes fins. Couvre la base de données et les données. Distingue l'œuvre dérivée des résultats d'une analyse computationnelle, qui eux, ne sont pas soumis à la licence. Cela se révèle utile pour l'intelligence artificielle ou l'Internet des objets. Contient une clause qui déclare qu'un soin « raisonnable » a été apporté à s'assurer que les données sont libres de droits de propriété intellectuelle de tiers, ou d'obligations en termes de protection de la vie privée ou de la confidentialité. Cette garantie est utile pour la réutilisation des données.	Manque de contrôle sur la manière dont les données sont réutilisées. Difficile à comprendre par des non-juristes. Non disponible dans un langage pour les machines. Licence très récente et très peu connue. Problème de l'empilement des attributions.
Community Data License Agreement - Sharing (CDLA-Sharing) (Licence d'attribution copyleft)	Couvre la base de données et les données. Distingue l'œuvre dérivée des résultats d'une analyse computationnelle, qui eux, ne sont pas soumis à la licence. Cela se révèle utile pour l'intelligence artificielle ou l'Internet des objets. Contient une clause qui déclare qu'un soin « raisonnable » a été apporté à s'assurer que les données sont libres de droits de propriété intellectuelle de tiers, ou d'obligations en termes de protection de la vie privée ou de la confidentialité. Cette garantie est utile pour la réutilisation des données.	Difficile à comprendre par des non-juristes. Non disponible dans un langage pour les machines. Licence très récente et très peu connue. Problème de l'empilement des attributions. Problème de l'incompatibilité des licences copyleft entre elles.

Annexe 10 : Fiches info sur le droit d'auteur et les licences pour les données

Fiche info #1

Le droit d'auteur et les données en Suisse

1 Qu'est-ce qui est protégé par le droit d'auteur ?

Les œuvres, soit les créations de l'esprit qui ont un caractère individuel. Elles doivent être exprimées pour être protégées. Les **données** peuvent donc être protégées à plusieurs niveaux : celui de la donnée individuelle, de la base de données, des métadonnées ou encore des visualisations de données, pour autant que cela soit des créations de l'esprit au caractère individuel. Il est ainsi nécessaire d'évaluer ce qui est protégé au cas par cas, pour chaque élément qui pourrait être concerné par le droit d'auteur. **En revanche, les données scientifiques brutes** (par exemple données météorologiques, géographiques, formules mathématique, etc.) **ne sont pas protégées**.

2 Quels droits sont liés aux données ?

L'auteur possède des **droits patrimoniaux**, qui lui permettent de décider si, quand et de quelle manière son œuvre sera utilisée. Ceux-ci comportent par exemple, le droit de diffusion, de reproduction, de modification, etc. Ces droits peuvent être transférés. Les **droits moraux**, sont le second type de droits que possède l'auteur. Ils protègent sa personnalité et eux, sont irrévocables. Il s'agit par exemple du droit de décider de la première divulgation de son œuvre ou du droit d'en revendiquer la paternité.

3 Qui est titulaire des droits d'auteur ?

Normalement c'est l'auteur lui-même ou les coauteurs. Mais comme les droits patrimoniaux sont transférables, un tiers peut être titulaire de ces droits sur le contenu protégé. Par exemple, dans le monde de la recherche, il est courant que les rapports de travail entre un chercheur et l'institution qui l'emploie prévoient une cession (partielle ou non) des droits d'auteur patrimoniaux. Cet aspect doit donc être vérifié.

4 Comment réutiliser des données protégées ?

Plusieurs options sont possibles :

- **demande l'autorisation à l'auteur,**
- **respecter la licence (s'il y en a une),**
- **utiliser une exception légale.**

Si aucune de ces options n'est réalisable, il faudra alors renoncer à l'utilisation.

Il existe des **exceptions** pour :

- l'usage strictement personnel,
- l'usage à des fins pédagogiques,
- la reproduction à des fins d'information interne et de documentation,
- la reproduction à des fins de recherche scientifique.

L'exception pour la recherche scientifique autorise la reproduction d'œuvres lorsqu'elle est nécessaire à l'application d'un procédé technique, ce qui permet la pratique du *text and data mining*. Les autres exceptions peuvent aussi être évoquées dans le contexte de la recherche, mais elles sont assez limitées et ne permettent pas le partage des données voulu par le mouvement de l'Open Science. Mettre à disposition ses données sous une licence favorise ainsi leur ouverture et leur réutilisation.






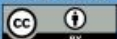
Fiches info sur le droit d'auteur et les licences créées dans le cadre du travail de master en Sciences de l'information à la Haute école de gestion de Genève intitulé *Données de la recherche : cadre juridique et licences*.

Publié sous licence CC-BY 4.0
DOI : 10.5281/zenodo.3967402
Le 17 août 2020

Fiche info #2

Différentes familles de licences pour les données

Il existe principalement trois familles de licences adaptées pour les données : les Creative Commons, les Open Data Commons et les Community Data License Agreement.

Creative Commons (CC)	Open Data Commons (ODC)	Community Data License Agreement (CDLA)
Émises par Creative Commons	Émises par The Open Knowledge Foundation	Émises par The Linux Foundation
	 Open Knowledge Foundation	
Sorties en 2001	Sorties en 2009	Sorties en 2017
Type : licences.	Type : licences doublées d'un contrat.	Type : licences doublées d'un contrat.
Portée : données et bases de données protégées par le droit d'auteur et le droit <i>sui generis</i> européen.	Portée : bases de données protégées par le droit d'auteur et le droit <i>sui generis</i> européen. Le contenu n'est pas concerné par les licences.	Portée : données et bases de données protégées par le droit d'auteur et le droit <i>sui generis</i> européen.
Famille composée de 6 licences et 2 outils de domaine public.	Famille composée de 2 licences et 1 outil de domaine public.	Famille composée de 2 licences.
Caractéristiques :	Caractéristiques :	Caractéristiques :
<ul style="list-style-type: none"> • Licences mondialement répandue et les plus connues. • Relativement aisées à comprendre. • Utilisées depuis longtemps dans le milieu de la recherche scientifique. • Disponibles dans un langage pour les machines. 	<ul style="list-style-type: none"> • Permettent de distinguer la base de données de son contenu. • Distinguent la base de donnée dérivée et les contenus dérivés qui ne sont pas de type base de données, mais produits à partir de celle-ci (par ex. graphiques, schémas, cartes). Ceux-ci ne sont pas soumis à la licence. 	<ul style="list-style-type: none"> • Pensées pour des grands volumes de données. • Ne s'appliquent pas aux résultats de l'analyse computationnelle des données sous licence. • Contiennent une clause qui déclare que les données sont libres de droits de tiers et d'obligations vis-à-vis de la protection de la vie privée ou de la confidentialité.
Crédits des logos : Creative Commons logo, https://creativecommons.org/about/downloads/ , sous licence CC-BY 4.0. Open Knowledge Foundation logo, https://en.wikipedia.org/wiki/File:Open_Knowledge_Foundation_logo_-_Landscape.png , sous licence CC-BY 4.0. The Linux Foundation and The Linux Foundation logo design are registered trademarks of The Linux Foundation. Linux is a registered trademark of Linus Torvalds. https://www.linuxfoundation.org/brand/		
Anouk Santos 	Fiches info sur le droit d'auteur et les licences créées dans le cadre du travail de master en Sciences de l'information à la Haute école de gestion de Genève intitulé <i>Données de la recherche : cadre juridique et licences</i> .	Publié sous licence CC-BY 4.0 DOI : 10.5281/zenodo.3967402 Le 17 août 2020

Fiche info #3

Les licences Creative Commons

CC-BY : Attribution



Toute exploitation de l'œuvre (partager, copier, reproduire, distribuer, communiquer, réutiliser, adapter) par tous moyens, sous tous formats et sous toutes licences sont possibles, y compris à des fins commerciales.

Obligations

Citer les auteurs de l'œuvre originale.
Indiquer si des modifications ont été effectuées.

CC-BY-SA : Attribution-ShareAlike



Toute exploitation de l'œuvre (partager, copier, reproduire, distribuer, communiquer, réutiliser, adapter) par tous moyens sont possibles, y compris à des fins commerciales, pour autant que la diffusion des nouvelles œuvres se fasse selon la même licence.

Obligations

Citer les auteurs de l'œuvre originale.
Indiquer si des modifications ont été effectuées.
Diffuser les nouvelles œuvres dérivées selon la même licence que celle de l'œuvre originale.

CC-BY-NC : Attribution-NonCommercial



Toutes diffusions de l'œuvre originale (partager, copier, reproduire, distribuer, communiquer, réutiliser, adapter) par tous moyens, sous tous formats et sous toutes les licences sont possibles, pour autant que l'exploitation de l'œuvre ou des œuvres dérivées ne se fasse pas à des fins commerciales.

Obligations

Citer les auteurs de l'œuvre originale.
Indiquer si des modifications ont été effectuées.

Interdiction

Tirer profit (gain ou plus-value commerciale) de l'œuvre ou des œuvres dérivées.

CC-BY-ND : Attribution-NoDerivatives



L'œuvre originale peut être partagée, copiée, reproduite, distribuée ou communiquée, y compris à des fins commerciales, par tous moyens, sous tous formats et sous toutes licences pour autant que l'œuvre soit diffusée sans modification et dans son intégralité.

Obligation

Citer les auteurs de l'œuvre originale.

Interdiction

Effectuer une diffusion partielle, une modification, une adaptation ou une traduction de l'œuvre.

CC-BY-NC-SA :



Attribution-NonCommercial-ShareAlike

Toutes diffusions de l'œuvre originale (partager, copier, reproduire, distribuer, communiquer, réutiliser, adapter) par tous moyens sont possibles, pour autant que l'exploitation de l'œuvre ou des œuvres dérivées ne se fasse pas à des fins commerciales et que la diffusion de l'œuvre se fasse sous la même licence.

Obligations

Citer les auteurs de l'œuvre originale.
Indiquer si des modifications ont été effectuées.
Diffuser les nouvelles œuvres dérivées selon la même licence que celle de l'œuvre originale.

Interdiction

Tirer profit (gain ou plus-value commerciale) de l'œuvre ou des œuvres dérivées.

CC-BY-NC-ND :



Attribution-NonCommercial-NoDerivatives

Toutes diffusions de l'œuvre originale (partager, copier, reproduire, distribuer, communiquer) par tous moyens, sous tous formats et sous toutes les licences sont possibles, pour autant que l'exploitation de l'œuvre ne se fasse pas à des fins commerciales et que l'œuvre soit diffusée sans modification et dans son intégralité.

Obligation

Citer les auteurs de l'œuvre originale.

Interdictions

Tirer profit (gain direct ou plus-value commerciale) de l'œuvre ou des œuvres dérivées.
Effectuer une diffusion partielle, modification, adaptation ou traduction de l'œuvre.

CC0 : Public Domain Dedication



L'œuvre originale peut être librement réutilisée, améliorée, modifiée, y compris à des fins commerciales, sans qu'il soit obligatoire de citer les auteurs de l'œuvre originale, ni de diffuser la nouvelle œuvre sous la même licence. L'éthique scientifique veut néanmoins que les sources soient mentionnées lorsque des données sont réutilisées.

Descriptions des licences issues de : FAVRE, Janelise et GERMOND, Tania, 2018. Augmenter la visibilité et l'impact d'une publication scientifique en maîtrisant le droit d'auteur. [oai100.snf.ch](https://oai100.snf.ch/en ligne) [en ligne]. Août 2018. [Consulté le 26 juin 2020]. Disponible à l'adresse : <https://oai100.snf.ch/wp-content/uploads/2019/02/Augmenter-la-visibilite-C3%A9-et-l-impact-d-une-publication-scientifique-en-ma...pdf>. Sous licence CC-BY 4.0. Les phrases ont parfois été raccourcies et quelques mots modifiés.

Anouk Santos



Fiches info sur le droit d'auteur et les licences créées dans le cadre du travail de master en Sciences de l'information à la Haute école de gestion de Genève intitulé *Données de la recherche : cadre juridique et licences*.

Publié sous licence CC-BY 4.0
DOI : 10.5281/zenodo.3967402
Le 17 août 2020

Fiche info #4

Des licences ouvertes pour des données ouvertes

Selon l'*Open Definition* de l'Open Knowledge Foundation, communément admise, **les données sont ouvertes** si elles peuvent être :

librement utilisées, modifiées et partagées par tout le monde et dans n'importe quel but.

Cette définition des données ouvertes a une implication directe sur les licences. Il n'y a plus que deux types de restrictions qui sont acceptables : l'obligation d'attribution à l'auteur original et/ou une licence de type copyleft (partage à l'identique).

Une licence ouverte peut donc soit :

- ✓ être de domaine public ;
- ✓ demander l'attribution ;
- ✓ être de type copyleft ;
- ✓ demander l'attribution et être de type copyleft.

Dès lors, il faut signaler que toutes les licences Creative Commons ne sont pas adaptées aux données de la recherche, les clauses interdisant la modification ou l'utilisation commerciale n'étant pas ouvertes.

Les différentes licences ouvertes adéquates pour les données de la recherche :

	Domaine public	Attribution	Attribution et copyleft
Creative Commons	CC0	CC-BY	CC-BY-SA
Open Data Commons	Public Domain Dedication and License (PDDL)	ODC-By	ODC-ODbL
Community Data License Agreement		CDLA-Permissive	CDLA-Sharing

Fiche info #5

L'ouverture des licences Creative Commons

Le plus de libertés



Déclaration de domaine public (CC0)

Mettre un jeu de données au plus près du domaine public autorise toutes les réutilisations possibles et règle les doutes sur l'existence ou non de droits d'auteur sur les données. Cette ouverture totale favorise la génération de nouvelles connaissances. Et puisque l'éthique scientifique veut que l'on cite ses sources, l'origine du jeu de données sera généralement indiquée quand même.

Licence la plus ouverte



Attribution (BY)

Il arrive souvent que les données scientifiques ne soient pas protégées par le droit d'auteur : demander à être crédité n'aura alors aucun effet. De plus, lorsque les attributions s'empilent au fur et à mesure des combinaisons de jeux de données, surtout pour ceux avec de nombreux contributeurs, indiquer les crédits peut devenir difficile à gérer.

Licences concernées : CC-BY et toutes les autres, sauf la CC0

Licences ouvertes mais problématiques



Partage à l'identique (SA)

Bien qu'imposer le partage à l'identique force l'ouverture des données, l'incompatibilité des licences de type copyleft entre elles fait qu'il est presque impossible de combiner des données sous des licences copyleft différentes dans un nouveau jeu dérivé. Cette situation limite ainsi leur réutilisation.

Licences concernées : CC-BY-SA et CC-BY-NC-SA



Pas d'utilisation commerciale (NC)

Ce qu'est une utilisation commerciale est difficilement définissable. En l'occurrence, cette clause pourrait empêcher quelqu'un d'utiliser un jeu de données s'il entend éventuellement publier les résultats de sa recherche.

Licences concernées : CC-BY-NC, CC-BY-NC-SA et CC-BY-NC-ND

Licences non ouvertes



Pas de modifications (ND)

L'utilisation d'une clause ND empêchera quelqu'un de combiner un jeu de données avec d'autres, et donc de le réutiliser pleinement. Toutes les réutilisations possibles doivent être autorisées pour qu'un jeu de données soit réellement en Open Access.

Licences concernées : CC-BY-ND et CC-BY-NC-ND

Le moins de libertés

Sources : CREATIVE COMMONS UK, 2017, Fact Sheet on Creative Commons & Open Science, DOI : 10.5281/zenodo.840652, sous licence CC-BY 4.0. SHADDIM, 2016, Creative commons license spectrum, *Wikimedia Commons* [en ligne], 29 février 2016, https://commons.wikimedia.org/w/index.php?title=File:Creative_commons_license_spectrum.svg&oldid=424766672. Sous licence CC0.

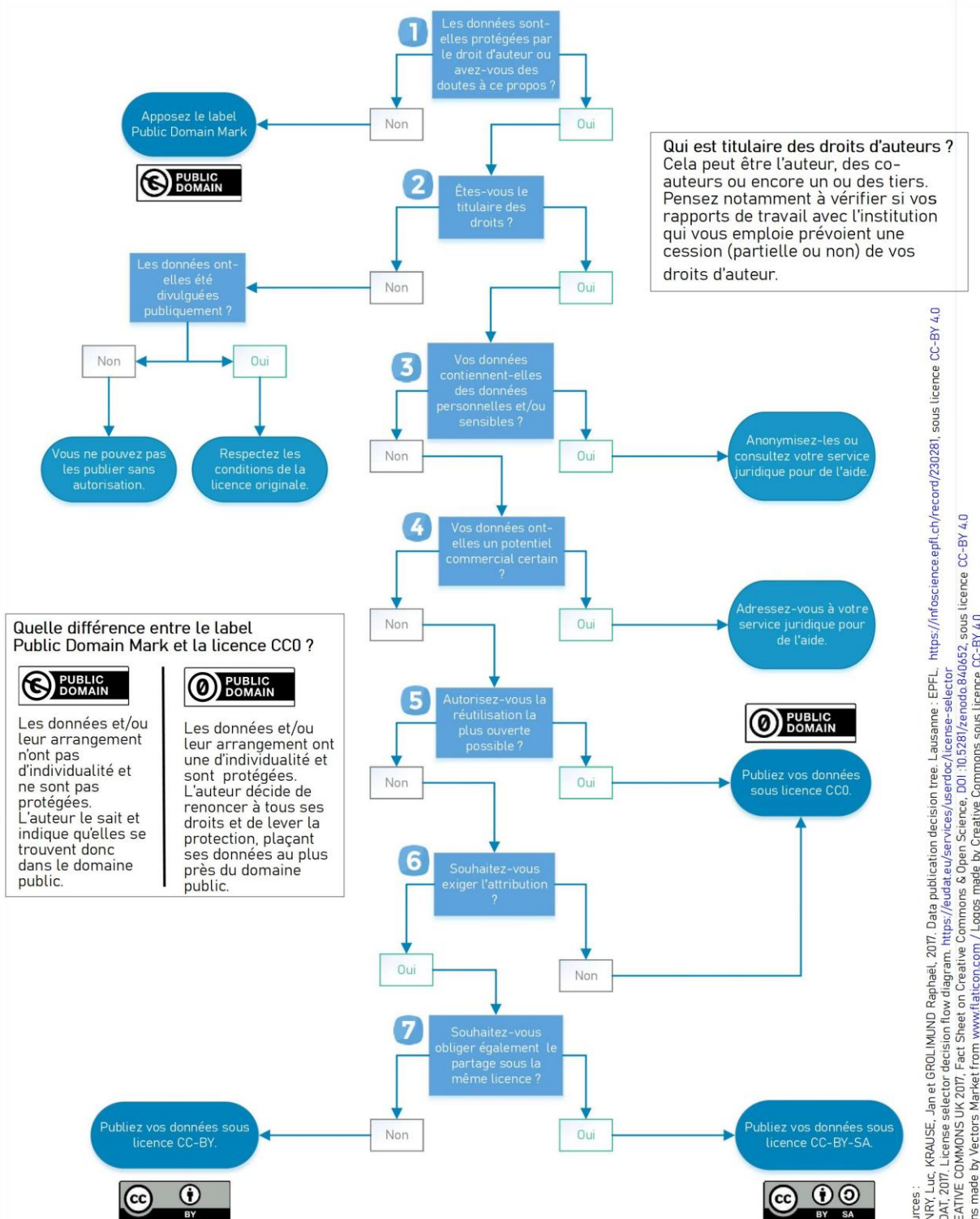


Fiches info sur le droit d'auteur et les licences créées dans le cadre du travail de master en Sciences de l'information à la Haute école de gestion de Genève intitulé *Données de la recherche : cadre juridique et licences*.

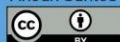
Publié sous licence CC-BY 4.0
DOI : 10.5281/zenodo.3967402
Le 17 août 2020

Annexe 11 : Arbre de décision pour le choix d'une licence pour les données

Arbre de décision : Quelle licence pour la publication des données de recherche ?



Anouk Santos



Fiches info sur le droit d'auteur et les licences créées dans le cadre du travail de master en Sciences de l'information à la Haute école de gestion de Genève intitulé *Données de la recherche : cadre juridique et licences*.

Publié sous licence CC-BY 4.0
DOI : 10.5281/zenodo.3967402
Le 17 août 2020