

Direction des bibliothèques

Université 
de Montréal

Enjeux éthiques et juridiques du partage et de la diffusion des données en sciences sociales :

Réflexions issues de l'École d'été du CIQSS
(Montréal, 10 - 11 juin 2019)

Direction du soutien à la réussite, à la recherche et à l'enseignement
Direction des bibliothèques
Université de Montréal
Novembre 2020

Remerciements

Les Bibliothèques de l'Université de Montréal tiennent à remercier le Centre interuniversitaire québécois de statistiques sociales (CIQSS), la Bibliothèque de l'Université Laval, Véronique Dupéré et Isabelle Archambault, toutes deux professeures à l'École de psychoéducation de l'Université de Montréal, ainsi que toutes les personnes ayant participé – que ce soit comme organisateur, comme présentateur, comme participant dans la salle – à l'École d'été sur les enjeux éthiques et juridiques du partage et de la diffusion des données en sciences sociales qui s'est tenue le 10 et le 11 juin 2019 au Carrefour des arts et des sciences de l'Université de Montréal. Elles tiennent également à remercier le rédacteur, Thierry M. Laforce, qui a assuré la prise de notes durant l'activité et ensuite structuré et rédigé ce document, ainsi que tous ceux et celles qui l'ont relu avant sa diffusion. Nous tenons finalement à remercier le Conseil de recherches en sciences humaines du Canada qui a financé la réalisation de l'événement ainsi que la rédaction et la traduction de ces *Réflexions*.

Table des matières

Remerciements.....	2
Introduction.....	4
1. Données de recherche.....	4
1.1. Une définition	4
1.2. Le cycle de vie des données.....	5
2. Contexte : énoncé de principes des trois conseils	5
3. Des enjeux. Un point de vue éthique.....	6
3.1. Le partage de données	8
3.2. La valeur du consentement des participants	9
3.3. Les mégadonnées	10
3.3.1 Les « lacs de données »	12
3.3.2 Le rôle de l'intelligence artificielle	14
4. Des enjeux. Un point de vue légal.....	16
4.1. Qui sont les auteurs, qui sont les propriétaires des données?	17
4.2. Qui peut autoriser l'accès aux données de la recherche?.....	19
4.3. Un mot sur les renseignements personnels	20
Conclusion	20
Références.....	22

Introduction

Le Centre interuniversitaire québécois de statistiques sociales (CIQSS) en collaboration avec les Bibliothèques de l'Université de Montréal et la Bibliothèque de l'Université Laval a tenu une école d'été le 10 et le 11 juin 2019 au Carrefour des sciences de l'Université de Montréal. Lors de cet événement, une diversité de personnes qui font de la recherche, la soutiennent, où y participent ont assisté à des présentations portant sur certains des enjeux éthiques et juridiques du partage et de la diffusion des données en sciences sociales. Les présentations des conférenciers et les discussions qui suivirent ont servi à préparer ce document sans pour autant en limiter le contenu. Plusieurs présentations sont disponibles à l'adresse : <https://www.ciqss.org/evenement/ecole-d-ete/enjeux-ethiques-et-juridiques-du-partage-et-de-la-diffusion-des-donnees-aux>.

Le présent document a pour objectif de faire un survol des questions fort pertinentes abordées lors de cette École d'été du CIQSS de l'été 2019 en complément des fichiers de présentation mentionnés. Il n'a cependant pas la prétention de couvrir tout ce qui a été discuté lors de ces journées, ni celle d'affirmer que les perspectives présentées sont les seules existantes ou possibles sur ces enjeux critiques et complexes.

Le document est divisé en quatre sections. La première section est une introduction au concept général de données de recherche et au cycle de vie des données de recherche. La deuxième section établit le contexte dans lequel se trouve la recherche. Il y est notamment question de la publication prochaine de la Politique des trois organismes sur la gestion des données de recherche. Ensuite, les troisième et quatrième sections présentent des enjeux éthiques et juridiques du partage et de la diffusion des données en sciences sociales. Chacune de ces sections est précédée d'une courte mise en situation et se termine par quelques considérations et informations pratiques.

1. Données de recherche

1.1. Une définition

Les données de recherche prennent de nombreuses formes : « écrits, notes, chiffres, symboles, texte, illustrations, films, vidéos, enregistrements sonores, reproductions picturales, dessins, croquis ou autres représentations graphiques, manuels de procédures, formulaires, diagrammes, organigrammes, descriptions de matériel, fichiers de données, algorithmes de traitement ou d'analyse de

données, statistiques » (*notre traduction*) (1). Une définition des données de recherche les présente « comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche » (2).

Les données de recherche peuvent être numériques ou non; dans le cadre de ce document, il sera discuté essentiellement de données numériques.

1.2. Le cycle de vie des données

Plusieurs modèles de cycle de vie des données de recherche existent. Voici celui proposé par le Réseau Portage, une initiative de l'Association des bibliothèques de recherche du Canada (3).



2. Contexte : énoncé de principes des trois conseils

Les trois conseils subventionnaires canadiens – CRSH, CRSNG, IRSC – sont en voie de publier la *Politique des trois organismes sur la gestion des données de recherche*.

Préalablement à cette publication, les organismes ont mis à la disposition de la communauté de la recherche une ébauche s'appuyant sur leurs politiques existantes afin d'aider la communauté « à répondre aux exigences actuelles et futures des organismes en matière de gestion des données de recherche ». Cette ébauche demeure un outil pertinent pour mieux comprendre le rôle des chercheurs quant à la gestion des données de recherche.

Selon les trois conseils, la recherche subventionnée à l'aide de fonds publics devrait profiter à la société canadienne, c'est-à-dire que les données de recherche devraient être, si possible, du domaine public afin qu'elles servent à plus d'un projet de recherche (4). Leur objectif est de démocratiser l'accès aux résultats de recherche, de favoriser la progression des connaissances, éviter la duplication des efforts de recherche, d'encourager la réutilisation des résultats de travaux précédents et de mettre en valeur les réalisations des chercheurs canadiens (4).

Afin d'atteindre ces objectifs, les organismes subventionnaires comptent entre autres sur un plan de gestion des données élaboré dès le début du projet, qui explicitera comment seront organisées les données en cours de recherche et comment leur réutilisation sera encouragée.

Les données devraient être accompagnées de métadonnées, c'est-à-dire de données, de documentation, voire de documents, par exemple, les codes qui leur ont permis de dresser les conclusions de leur recherche. Les métadonnées permettent le repérage et la réutilisation des données.

Bien entendu, les données réutilisées devront être convenablement citées. « Les données sont importantes et constituent des produits légitimes de la recherche; elles doivent être reconnues comme telles » (4).

3. Des enjeux. Un point de vue éthique.

Le virage numérique de la recherche a influencé les perceptions à propos des données de recherche. Les considérations éthiques évoluent avec le contexte dans laquelle la recherche se réalise. Cette évolution peut se constater avec les changements qui ont été faits aux réglementations encadrant la recherche impliquant des êtres humains. Au fil des décennies, les politiques sont devenues de plus en plus précises afin de mieux encadrer les travaux des chercheurs :

- + Adoption du Code de Nuremberg en 1947;
- + Déclaration d'Helsinki en 1964;
- + Adoption du National Research Act par le Congrès américain en 1974;

- + Publication du Code déontologique de la recherche utilisant des sujets humains du Conseil de recherches en sciences humaines du Canada en 1977;
- + Publication des Lignes directrices concernant la recherche sur des sujets humains du Conseil de recherches médicales du Canada en 1987;
- + Première publication de l'Énoncé de politique des trois conseils : éthique de la recherche avec des êtres humains (1998);
- + Révision de l'EPTC en 2010, en 2014 et en 2018.

Avec chacun de ces jalons en éthique de la recherche, les institutions ont développé des politiques et des directives pour guider les chercheurs dans leurs projets impliquant des êtres humains et les organismes subventionnaires ont établi des cadres de référence s'appliquant aux projets qu'ils financent.

En ce sens, la politique actuellement en vigueur au Canada, soit l'EPTC 2, remet entre les mains des chercheurs et des institutions la responsabilité de protéger les données. Les institutions qui reçoivent des fonds de recherche des trois conseils signent d'ailleurs l'Entente sur l'administration des subventions et des bourses des organismes par les établissements de recherche. Les chercheurs doivent bien informer les participants lors du recrutement afin d'obtenir un consentement libre et éclairé. L'EPTC 2 prévoit des exceptions à cette obligation pour que les chercheurs puissent collecter des données dans certaines situations.

Le paradigme numérique de la recherche implique qu'il n'est plus seulement question des données de participants à une recherche. Il peut s'agir d'étudier des bases de données incluant un nombre d'individus sans précédent. Un plus grand nombre de données et de nouveaux outils pour les analyser permettent de tirer des conclusions ayant un impact sur un nombre important d'individus. Avec les données numériques vient la possibilité d'amalgame et de croisement. Une notion de respect de l'autonomie des participants est souvent opposée à une notion d'intérêts collectifs. Sans que ces deux notions soient en opposition, la numérisation des données a exacerbé une tension entre ces objectifs. Il s'agit d'une question complexe à laquelle il n'existe pas de réponse simple.

Une réflexion critique est essentielle et la communauté scientifique ne peut répondre seule à ces interrogations. La population dans laquelle elle évolue devra aussi s'interroger et se positionner. En tant que chercheur, il faut être conscient des impacts que peuvent avoir les projets sur les individus et sur les populations. Une réflexion sur les enjeux d'une étude particulière est nécessaire dès les débuts de sa conception jusqu'à la publication des résultats, et même plus.

3.1. Le partage de données

Un exemple

Un chercheur a collecté des données sur une population au courant de sa carrière. Les sondages et les entrevues qu'il a réalisés lui ont permis de bâtir une base de données substantielle qu'il a utilisée à différentes reprises pour publier des articles scientifiques. À l'aube de sa retraite, il constate que sa base de données contient encore suffisamment d'information pour permettre la publication d'autres articles. Comme il n'aura pas le temps de les réaliser lui-même, il confie les données à une collègue en espérant qu'elle les analysera à son tour. Est-ce que les normes actuelles permettent à cette chercheuse d'utiliser les données?

Maintenant qu'il est possible d'échanger d'importantes quantités de données rapidement et relativement facilement, les limites du partage des données sont de moins en moins structurelles. Les chercheurs doivent s'interroger sur les normes et les règlements de chaque organisation, de chaque pays, de chaque communauté qui doivent être respectés. Les chercheurs doivent évaluer l'impact que le partage aura sur les individus dont proviennent les données. Un partage de données devrait toujours se faire en tenant compte des risques de divulgation ou d'inférence sur les individus impliqués dans la recherche. Toutefois, il peut être difficile pour un chercheur d'identifier toutes les variables posant un risque d'identification. Certaines sont simples à reconnaître, mais d'autres peuvent être écartées sans anticiper le risque que représente leur croisement avec d'autres variables. Une expertise est à développer en ce sens et la responsabilité ne peut reposer uniquement sur les épaules des chercheurs.

Des initiatives ont été lancées dans certaines disciplines afin que les données soient rendues disponibles lors de la publication d'un article scientifique. Cela entraîne à son tour des enjeux de partage : comme mentionné précédemment, les données partagées ne doivent pas causer de tort aux individus dont elles proviennent. De plus, les chercheurs n'ont pas toujours les connaissances techniques pour rendre disponibles leurs données dans un format répondant aux besoins du partage.

Un premier pas pour favoriser le partage peut être la description convenable des données et le partage cette description avec leur communauté scientifique. Un document accompagnant les publications pourrait indiquer comment accéder aux données brutes ou, si elles ne sont pas accessibles, indiquer comment le chercheur y a eu accès et quelles limites restreignent l'accès.

Quelques considérations

- + Considérez les intérêts des participants à la recherche dont vous avez les données avant de les transmettre :
 - Est-ce que votre jeu de données peut les identifier?
 - Un croisement de données peut-il mener à une identification?
 - Ont-ils consenti au partage de leurs données avec une tierce partie?
- + Vérifiez si l'encadrement au sujet de l'entreposage des données diffère entre les collaborateurs d'un échange de données.

Informations pratiques

L'*Open Science Framework* (<https://osf.io/>) est une plateforme d'entreposage et de partage des données de la recherche. Dans une optique de recherche ouverte, l'OSF permet aux chercheurs d'accéder aux données de recherche et de partager les leurs dans un environnement sécuritaire.

3.2. La valeur du consentement des participants

Un exemple

Un groupe de chercheurs a mis sur pied une banque de tissus utilisée pour séquencer le génome des participants. Leur objectif est de faire avancer la recherche en médecine personnalisée. En mettant sur pied leur banque, ils ont constaté que le séquençage du génome pouvait mener à des découvertes fortuites, soit des informations que ni eux ni les participants ne s'attendaient à obtenir. Le consentement initial des participants prévoyait une divulgation automatique de ces découvertes, mais elles sont parfois faites plusieurs années après la signature du formulaire de consentement. Est-ce que la décision initiale du participant de recevoir ou non cette information tient toujours malgré le passage des années?

Inspiré de Thorogood et coll. (5).

Le principe du consentement libre et éclairé est au cœur de la recherche faite avec des participants humains. Ce respect de l'autonomie est primordial pour assurer le respect de la dignité des participants, mais aussi pour maintenir le lien de confiance qui existe entre la population et les chercheurs. Lorsque des données sont conservées pendant de nombreuses années, il faut se demander si le consentement obtenu au début de la recherche est encore valide après un, deux, trois, cinq, voire dix ans.

Ces éléments sont généralement abordés lors de la discussion initiale menant au consentement. Le chercheur et le participant tiennent alors pour acquis que cette

décision perdurera dans le temps jusqu'à ce que le chercheur détruise les données. Ce dernier a donc la responsabilité de respecter les engagements qu'il a pris auprès du participant longtemps après la fin de la collecte des données. La réflexion du chercheur au sujet du sort des données ne prend pas fin après l'obtention de l'autorisation du comité d'éthique de la recherche. Bien entendu, l'importance de la réflexion doit être proportionnelle à la sensibilité des données collectées et aux risques encourus par les participants.

Quelques considérations

- + Le consentement libre et éclairé d'un participant est requis en continu :
 - Identifiez des moments clés lors desquels le consentement doit être réaffirmé;
 - Ne faites pas de supposition au sujet du consentement d'un participant : dans le doute, contactez-le;
 - Prévoir des consentements plus larges pour des usages secondaires, le cas échéant, en sachant que le métaconsentement est possible au Canada, mais pas un consentement du genre *blanket consent* à l'américaine.
- + Des communications régulières permettent de tenir le participant informé au sujet du projet auquel il participe. Considérez des moyens de garder vos participants informés de l'utilisation que vous faites de leurs données.

Informations pratiques

La littérature scientifique en éthique regorge de propositions pour améliorer l'obtention du consentement libre et éclairé des participants. Avant de lancer une recherche approfondie, vous pouvez compléter le didacticiel de l'EPTC-2 (http://www.ger.ethique.gc.ca/fra/education_tutorial-didacticiel.html) ou celui du Ministère de la Santé et des Services sociaux (<http://ethique.msss.gouv.qc.ca/didacticiel/>). Ils traitent de bien plus que du processus de consentement, mais cela vous donnera une base solide pour poursuivre vos recherches par la suite.

3.3. Les mégadonnées

Un exemple

Depuis 1955, le New South Wales State Emergency Service (NSW SES) utilise diverses sources d'informations (données météorologiques historiques et quotidiennes, images satellites, réseaux sociaux, photos et vidéos disponibles en

ligne, etc.) pour planifier et organiser les secours en cas de catastrophes naturelles en Australie. En 2009, l'infrastructure informatique a complètement été modernisée afin de favoriser le transfert d'information entre les acteurs impliqués dans les secours. Cette infrastructure permet d'intégrer l'information provenant de plusieurs sources pour identifier les risques potentiels que courent les régions exposées à des événements naturels. En combinant les sources d'information, le NSW SES arrive à prévoir les impacts d'une catastrophe et planifie ainsi ses plans d'intervention en fonction des mégadonnées qu'il interprète.

Tiré de Wamba et coll., 2015 (6).

L'expression mégadonnées, ou « Big Data » en anglais, fait référence à l'accumulation rapide et substantielle de données au sujet d'une multitude de faits et à leur conservation dans des bases de données tout aussi impressionnantes. Elles diffèrent des données traditionnellement collectées en recherche. Outre leurs caractéristiques précédemment mentionnées, elles sont aussi difficilement analysables à l'aide des outils statistiques usuels (7). Les mégadonnées sont définies par 5 « V » :

- Volume : la quantité de données collectées;
- Variété : leurs différentes natures, formats et structures;
- Vitesse : la rapidité de leur collecte et de leur entreposage;
- Valeur : la possibilité de générer des profits;
- Véracité : la qualité, la précision et la fiabilité des données (8).

En recherche, le recours aux mégadonnées peut soulever des enjeux éthiques importants entre autres au niveau de l'identification des individus et de leur vie privée, mais aussi au niveau de la propriété des données. D'abord, certains vont différencier les données anonymes – en supposant un anonymat encore possible – des données rendues anonymes, des données dépersonnalisées. Les premières ont été collectées sans qu'aucun identificateur direct y ait été associé. Les secondes sont des données dont les marqueurs identificatoires ont été retirés afin de les rendre anonymes. Les troisièmes sont des données dont le marqueur identificatoire a été remplacé par un code. Une donnée sensible quant à elle peut se définir par la valeur morale pour l'individu ou l'institution qui la partagent avec le chercheur. Les mesures de sécurité que le chercheur met en place pour protéger les données dont il fait l'étude doivent être proportionnelles à la sensibilité des données.

Comment alors protéger les individus desquels proviennent les données? Les technologies de protection de la vie privée (*Privacy Enhancing Technologies* – PETs

en anglais) peuvent offrir aux utilisateurs et aux chercheurs un meilleur contrôle sur les données qui sont partagées. Par exemple, le système de communication par messagerie texte qu'utilise un individu peut être chiffré afin que même le développeur de l'application ne puisse accéder au contenu des messages. Les fondements derrière ces technologies sont 1) que les données collectées soient minimisées et 2) que l'utilisateur soit souverain des données qu'il partage. De cette façon, les données qui sont collectées et entreposées sont limitées à celles que l'utilisateur accepte de partager.

Quelques considérations

- + Développez un plan de protection des données confidentielles avant de commencer la collecte des données;
 - o Il faut savoir si les données analysées sont anonymes, sensibles ou personnelles : la protection ne sera pas la même.

3.3.1 Les « lacs de données »

Un exemple

Le CHUM a mis sur pied le Centre d'intégration et d'analyse des données médicales (CITADEL) afin de structurer l'ensemble des données produites et entreposées par le centre hospitalier. Ce centre récupère les données de soins des patients et les organise afin de les rendre utilisables par des chercheurs intéressés. Avec les autorisations adéquates de la Direction des services professionnels et du Bureau d'éthique de la recherche, les chercheurs peuvent communiquer directement avec l'équipe de CITADEL pour obtenir un jeu de données. C'est l'équipe de CITADEL qui est chargée d'assurer la confidentialité des données transmises aux chercheurs à partir du lac qu'elle opère.

Pour plus d'information : <https://www.chumontreal.qc.ca/crchum/plateformes-et-services>

Les « lacs de données » sont des bases de données dans lesquelles sont entreposées des quantités importantes de données brutes, structurées ou semi-structurées. Ils sont utilisés par les institutions collectant les données et par celles offrant un espace d'entreposage aux chercheurs analysant les données. Lorsqu'un chercheur les utilise pour entreposer ou analyser des données, il doit réfléchir à l'impact que cela aura sur les individus concernés. Est-ce que l'information peut porter atteinte à la dignité? Est-ce qu'elle peut mener à son identification? Est-ce qu'elle concerne un groupe restreint de la population? Est-ce que le consentement donné par le participant tenait compte de ce type d'entreposage?

La confidentialité des individus est sans contredit au cœur des enjeux des lacs de données. En regroupant un grand nombre d'informations au sein d'un seul espace et en donnant accès à cet espace à des chercheurs, une institution s'expose à des bris de confidentialité. Plus un lac contient de variables au sujet d'un individu, plus il est facile de l'identifier par un croisement de variables. Comment alors assurer la protection de la confidentialité? Un des moyens est de retirer les variables pouvant mener à l'identification avant de les partager. Ainsi, au lieu de transmettre aux chercheurs l'ensemble des variables d'une population, retirer celles qui, une fois croisées, pourraient mener à la réidentification des individus permet de protéger la confidentialité.

Un individu, qu'il soit participant à une recherche ou utilisateur d'un service, peut-il alors choisir le sort de ses données? À l'ère de l'informatisation, il est difficile de retirer son consentement à la collecte de données alors que l'utilisation d'outils technologiques est liée à la collecte de données au sujet des utilisateurs. Ce n'est toutefois pas le cas de la recherche. Un chercheur peut choisir le sort des données qu'il collecte et il peut informer les participants qui accepteront de se joindre à son étude. Il a donc la responsabilité de réfléchir à l'impact qu'aura cette décision pour les participants. Il ne s'agit pas simplement de les informer, mais aussi de considérer différentes méthodes de collecte et d'entreposage des données. Cette décision doit aussi faire le cas d'une réflexion continue tout au long du projet de recherche et même après.

Quelques considérations

- + Vérifiez si les variables que vous croisez en provenance d'un lac de données peuvent mener à l'identification d'individus;
- + Réfléchissez à l'origine des données que vous obtenez d'un lac de données;
 - Qui sont les individus derrière ces données? Des participants à un projet de recherche? Des patients? Des utilisateurs d'un service public ou privé?
 - Est-ce que l'utilisation que vous faites des données se fait dans l'intérêt de ces individus?
- + Préparez des mesures de protection des données avant leur réception;
- + Envisagez le sort des données collectées avant le début de votre projet.

Informations pratiques

Le *Center for Internet Security* se spécialise dans la protection des données des organisations. Bien qu'il vende des produits de cybersécurité, il publie aussi des informations qui peuvent être intéressantes pour des chercheurs. L'une de ses ressources résume la confection d'un plan de protection des données. Elle est disponible ici : <https://www.cisecurity.org/blog/how-to-create-a-data-protection-plan/>.

3.3.2 Le rôle de l'intelligence artificielle

Un exemple

Un chercheur développe un algorithme qui collecte des données au sujet d'utilisateur d'une application dédiée à la recherche. L'algorithme utilise l'apprentissage machine pour ne conserver que les données qui sont particulièrement pertinentes pour le chercheur. Une fois collectées, les données sont ensuite organisées par le même algorithme afin qu'elles soient présentées au chercheur sous une forme qu'il pourra utiliser pour un projet de recherche.

Inspiré de Determann, 2018, p. 19 (9).

Après la collecte et l'entreposage vient l'analyse des mégadonnées. Les outils traditionnels de la recherche ne permettent pas d'exploiter au maximum le fruit de la collecte de mégadonnées. En fait, elles sont si nombreuses et si différentes en taille et en format, qu'un chercheur peut difficilement s'y intéresser sans s'équiper d'outils modernes. Un des outils de l'intelligence artificielle (IA) est l'apprentissage machine (*machine learning*). En plus d'être sous la responsabilité des chercheurs qui les utilisent, il est attendu de ces outils qu'ils soient :

- loyaux envers les utilisateurs (ils se comportent tels que déclarés par le concepteur);
- équitables et non discriminatoires;
- transparents;
- conformes aux lois en vigueur.

Compte tenu de la capacité d'entreposage et d'analyse maintenant disponible, le rapport de la recherche avec les individus évolue. Dans ce nouveau rapport, le respect de la dignité humaine reste essentiel. Les possibilités d'inférences sont de plus en plus grandes et les avancées en IA permettent un très grand nombre d'analyses. Dans ce contexte, il faut trouver un équilibre entre les bénéfices sociaux et économiques offerts par l'IA et le respect de la vie privée des individus. On peut

penser ici aux [principes d'Asilomar](#) et à la [Déclaration de Montréal pour un développement responsable de l'intelligence artificielle](#).

L'IA permet d'analyser une grande quantité de données. Qui plus est, l'apprentissage machine permet même à un chercheur de guider les premières analyses pour ensuite continuer avec un minimum de surveillance humaine. Parmi les enjeux de tels outils, il y a une possible opacité de l'apprentissage machine, c'est-à-dire qu'entre le moment où l'analyse est lancée et les résultats obtenus, il peut exister une sorte de « boîte noire ».

Cette boîte noire peut limiter la capacité des chercheurs à détecter la discrimination de programmation. Des mesures techniques et organisationnelles doivent être mises en place afin d'éviter cette discrimination. Par exemple, un échantillonnage des données avant l'analyse peut être fait afin que soient évaluées des données à leur entrée. Ou encore, l'algorithme d'analyse peut être modifié afin qu'il tienne compte de la discrimination et qu'il la corrige au courant de l'analyse.

Ces questions en sont à leurs balbutiements et il est encore difficile d'évaluer leur impact sur l'analyse des données en recherche. Bien qu'il soit possible d'analyser des données sous de nombreux angles, nous pouvons choisir de restreindre les options afin de respecter les individus dont proviennent les données. Ce n'est pas parce qu'un chercheur détient les données et les outils pour les analyser qu'il doit le faire. Il s'agit d'un questionnement important auquel il faut réfléchir, mais difficile à répondre pour un seul chercheur.

Il faut être conscient des impacts sur les individus. Le croisement de variable, la réutilisation de données à d'autres escients et les risques de divulgations massives peuvent porter atteinte aux individus ayant contribué à la collecte des données.

Pour conclure cette section sur les enjeux éthiques du partage de données, il est important de rappeler que tous les individus n'ont pas accès au monde numérique de façon égale. Ce phénomène se nomme « fracture numérique » et il fait référence à l'inégalité d'accès et de contribution à l'information, à la connaissance et aux réseaux (10). Des projets de recherche ancrés dans les technologies de l'information et des communications excluent d'emblée certains individus. L'EPTC-2 demande à ce que les avantages et les inconvénients de la recherche soient équitablement répartis entre les individus et les groupes d'une communauté. La communauté scientifique doit donc considérer cette fracture lorsqu'elle développe des projets de recherche. Il n'est évidemment pas du ressort des chercheurs d'équiper ces individus et de les initier au monde numérique. Ils

doivent tout de même être conscients de cette réalité et des limites qu'elle impose aux projets de recherche.

Quelques considérations

- + L'IA permet des analyses humainement impossibles, mais c'est un humain qui choisit quelles analyses faire;
- + Identifiez les biais du programmeur et de l'équipe derrière une IA. L'apprentissage machine est d'abord conçu par un humain : il doit réduire ses propres biais pour éviter que l'IA les répète;
- + Utilisez des IA qui vous permettent d'expliquer les résultats et les analyses obtenus;
- + Réfléchissez aux impacts d'une analyse d'un jeu de données pour la communauté ciblée.
- + Les outils que vous utilisez pour collecter des données peuvent exclure des individus de votre recherche. Identifiez ces individus et voyez si vous pouvez les inclure autrement.

Informations pratiques

Depuis 2017, l'Institut québécois d'intelligence artificielle rassemble des chercheurs de renom pour favoriser la collaboration et le partage de connaissances dans le développement de l'intelligence artificielle. Leurs activités peuvent être suivies ici : <https://mila.quebec/>.

4. Des enjeux. Un point de vue légal.

Les données numériques font maintenant partie du portrait de la recherche; une certaine souplesse en favorise jusqu'à un certain point la réutilisation. Cet avantage est aussi la source de bien des maux de tête. Le spectre des parties prenantes va des individus qui ont un intérêt à ce que leurs données personnelles soient protégées jusqu'aux entreprises souhaitant utiliser les données qu'elles collectent à des fins commerciales (11).

La déclaration de principes des organismes subventionnaires établit des responsabilités pour les acteurs suivants : les chercheurs, les communautés de chercheurs, les établissements de recherche et les bailleurs de fonds de recherche (4).

Les chercheurs doivent élaborer un plan de gestion des données, se conformer aux exigences de l'organisme subventionnaire et aux normes professionnelles,

reconnaître et citer les ensembles des données utilisées et s’informer des normes de leur discipline.

Les communautés des chercheurs doivent établir les normes de gestion de données qui s’appliqueront à leur communauté, et identifier des répertoires et des plateformes qui pourront être utilisés par les chercheurs.

Les établissements de recherche doivent sensibiliser et appuyer les chercheurs en ce qui concerne la gestion des données. L’élaboration de politiques de gestion de données conformes aux obligations éthiques, juridiques et commerciales autant provinciales que nationales est fondamentale. Chaque établissement administrant des fonds des trois conseils doit se doter d’une stratégie institutionnelle de gestion des données de recherche. Les établissements doivent reconnaître l’importance des données comme résultat de recherche.

Les bailleurs de fonds de recherche reconnaissent également l’importance des données comme résultat de recherche. Ils doivent élaborer des politiques et des exigences favorisant la gestion responsable des données et inclure des considérations en matière de gestion des données lors du processus d’évaluation des demandes de financement.

4.1. Qui sont les auteurs, qui sont les propriétaires des données?

Un exemple

Une organisation recense des statistiques à propos des habitations d’une ville nord-américaine. Sur son site, elle rend disponibles des informations telles que la taille des maisons et des terrains, le nombre de pièces et d’étages, la valeur des propriétés telles qu’évaluées par la ville, l’année de construction, etc. L’organisation indique dans ces règles d’utilisation que les données présentées sont la propriété des collaborateurs les ayant collectées. Un chercheur utilise le site comme source de données afin de dresser le portrait de l’évolution des habitations de cette ville au fil des décennies. Enfreint-il la loi sur le droit d’auteur?

Cas présenté par Dominique Lapierre lors de l’école d’été du CIQSS

Les lois et les règlements québécois et canadiens accordent des intérêts, des droits d’accès et des restrictions à l’accès aux données (9). Pour une réponse plus complète, il faut se tourner vers la propriété intellectuelle, qui est divisée en deux branches : la propriété industrielle et le droit d’auteur. Cette seconde branche inclut le droit d’auteur pour les œuvres littéraires, dramatiques, musicales et artistiques. Si la *Loi sur le droit d’auteur*, LRC (1985), ch. C-42, ne protège pas les données brutes, elle régit certaines compilations de données.

Sous le droit d'auteur, il existe un équilibre entre la protection des droits de l'auteur d'une œuvre et les droits des utilisateurs. Une protection numérique comme un mot de passe peut toutefois devenir un moyen pour l'auteur de limiter l'utilisation de son œuvre (11).

En plus de la *Loi sur le droit d'auteur*, chaque établissement de recherche peut avoir ses propres règles en matière de propriété des données de la recherche. Par exemple, à l'Université Laval, le règlement sur la propriété intellectuelle stipule que :

L'Université est propriétaire d'un fonds documentaire constitué par un membre de l'Université ou par un groupe de membres de l'Université lorsque le membre ou le groupe de membres a utilisé le nom ou le temps ou les services ou les locaux de l'Université, ou bénéficié d'une subvention d'un commanditaire exigeant que le contrat ou la subvention soit entériné par l'Université.

Ces dispositions peuvent prendre la forme de règlements universitaires, mais elles peuvent aussi être enchâssées dans les conventions collectives des chercheurs ou encore dans les ententes intervenant entre les établissements et les organismes subventionnaires. Un chercheur devrait donc s'informer auprès de son établissement de recherche pour connaître précisément comment est encadrée la propriété des données de recherche dans son institution.

Dans un autre cadre légal, celui de l'Europe, un règlement général sur la protection des données est entré en vigueur au mois de mai 2018 (12).

Quelques considérations

- + Des données peuvent être assujetties au droit d'auteur : identifiez le propriétaire des données que vous utilisez avant de débiter votre projet de recherche;
- + Utilisez une protection numérique comme un mot de passe pour limiter l'utilisation par autrui d'un jeu de données tant que vous n'êtes pas prêt à le partager;
- + Consultez les règlements de votre établissement!

Informations pratiques

Votre établissement de recherche a fort probablement mis en place une politique d'attribution de la propriété intellectuelle. Il peut s'agir du Conseil universitaire, du Bureau de l'éthique de la recherche ou de toute autre instance chargée de superviser la recherche au sein de l'établissement. Contactez la personne responsable de cette instance afin de connaître la politique en vigueur.

4.2. Qui peut autoriser l'accès aux données de la recherche?

Un exemple

Dans le cadre de sa maîtrise, une étudiante a collecté et entreposé des données dans le laboratoire de recherche de sa directrice de recherche. Elle est présentement l'unique utilisatrice de ces données et elle travaille activement sur son mémoire de maîtrise. À la suite d'un conflit avec sa directrice de recherche, elle perd accès au laboratoire. Le code d'accès a changé; elle ne peut plus accéder aux données qu'elle analysait. À qui appartiennent les données?

Cas présenté par Dominique Lapierre lors de l'école d'été du CIQSS.

Au-delà de la question de la propriété des données, il est peut-être pertinent de se questionner sur l'accès aux données. Qui peut autoriser l'accès aux données? Chaque plateforme d'entreposage des données a des conditions d'utilisation. Certaines sont situées directement sur le site où les chercheurs travaillent, alors que d'autres se trouvent carrément sur un autre territoire juridique. Comment alors déterminer qui autorise l'accès? L'utilisation d'un verrou numérique, par exemple un mot de passe, est une forme de protection de l'accès aux données. Le responsable du verrou est nécessairement en position d'autorité par rapport aux utilisateurs désirant l'accès. Cette approche reconnaît que pour être accessible, une donnée doit être déposée sur un support physique. Ce support devient alors un objet pouvant servir à déterminer qui est propriétaire, ou du moins responsable, des données (12). Généralement, être responsable de l'accès équivaut aussi à être responsable des frais associés à l'entreposage des données, à leur transfert et à leur sécurité. Il faut décider des règles qui seront appliquées pour autoriser l'accès aux données, établir des mécanismes pour assurer le respect de ces règles et entretenir les données pour qu'elles demeurent accessibles. De nombreux chercheurs dépendent d'une organisation responsable de l'accès. Il faut alors s'en remettre aux règlements institutionnels pour dénouer des conflits qui pourraient survenir.

Quelques considérations

- + Nommez un responsable de l'accès aux données.
 - Ce n'est pas nécessairement le propriétaire des données;
 - Dans certains cas, ce sera l'organisation qui héberge vos données qui en sera responsable.

4.3. Un mot sur les renseignements personnels

Qu'en est-il des renseignements personnels? Les renseignements personnels sont en effet un type de données qui peut être recueilli et utilisé dans le cadre de la recherche. La nuance qui suit est importante. Une donnée existe qu'elle soit utilisée ou non en recherche. Même s'il n'est pas propriétaire de ces données, un individu est tout de même en mesure de consentir à ce qu'un chercheur les utilise. Le chercheur ne devient pas propriétaire des données, même s'il doit s'assurer de la protection de l'accès aux données (R. c. Stewart [1988] 1 RCS 963). Le chercheur est alors responsable de la protection des renseignements personnels qu'il aura recueillis dans le cadre de sa recherche. Au Québec, la Commission d'accès à l'information (CAI) est, par force de loi, propriétaire morale des renseignements personnels et veille à leur protection. Lorsqu'un chercheur souhaite utiliser des données recueillies sans le consentement explicite des individus (p. ex. par le croisement de bases de données), la CAI peut se prononcer en leur nom. Elle peut ainsi autoriser un chercheur à recevoir communication de renseignements personnels aux fins de recherche.

Il est aussi important de savoir que les données confidentielles doivent être protégées par le chercheur qui les a colligées et qui souhaite les utiliser dans le cadre de sa recherche. C'est donc dire qu'il doit mettre en place des dispositions raisonnables pour éviter qu'elles soient consultées ou utilisées par un tiers non autorisé. Ces dispositions peuvent prendre la forme de protection par mot de passe, d'entreposage des données dans une enclave physique accessible seulement à certaines personnes autorisées, etc.

Conclusion

Le présent document se voulait d'abord et avant tout une ressource informationnelle pour les chercheurs souhaitant explorer quelques enjeux éthiques et juridiques du partage et de la diffusion des données en sciences sociales. Il est loin d'être exhaustif, mais il traite de certains enjeux discutés lors de l'école d'été du CIQSS qui s'est tenue le 10 et le 11 juin 2019. Les questions sur le partage des données, la valeur du consentement et les mégadonnées qui ont été

présentées ne sont que la pointe d'un iceberg qui tend à grossir avec le développement de nouvelles technologies et de nouveaux moyens de collecte et de préservation des données. Le cadre éthique et légal déterminant les droits et les responsabilités en ce qui concerne les données est en évolution. Prochainement, les trois Conseils subventionnaires canadiens publieront une politique sur la gestion des données de recherche. Les chercheurs et les établissements de recherche pourront alors s'y référer pour mieux encadrer leur recherche. D'ici là, tout un chacun doit rester à l'affût des enjeux que peuvent soulever leurs recherches.

Références

1. CASRAI Dictionary [En ligne]. Consortia Advancing Standards in Research Administration Information. Data [cité le 25 juin 2019]. Disponible: <https://dictionary.casrai.org/Data>
2. Secrétariat général de l'OCDE. Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics. Paris: Organisation de coopération et de développement économiques; 2007.
3. Réseau Portage. Sans titre. Document inédit; 2019.
4. Déclaration de principes des trois organismes sur la gestion des données numériques. Canada: Instituts de recherche en santé du Canada, Conseil de recherches en sciences naturelles et en génie, Conseil de recherches en sciences humaines; 2016.
5. Thorogood A, Joly Y, Knoppers BM, Nilsson T, Metrakos P, Lazaris A, et al. An implementation framework for the feedback of individual research results and incidental findings in research. BMC Medical Ethics. 2014;15(1):88.
6. Wamba SF, Akter S, Edwards A, Chopin G, Gnanzou D. How big data can make big impact: findings from a systematic review and a longitudinal case study. INT J PROD ECON. 2015;165:234-46.
7. Mégadonnées [En ligne]. Québec: Gouvernement du Québec; 2017 [cité le 10 juillet 2019]. Disponible: http://www.granddictionnaire.com/ficheOqlf.aspx?ld_Fiche=26507313
8. Bourany T. Les 5V du big data. Regards croisés sur l'économie. 2018;23(2):27-31.
9. Determann L. No one owns data. Hastings LJ. 2018;70:1.
10. Cliche D. La ville intelligente au service du bien commun – Lignes directrices pour allier l'éthique au numérique dans les municipalités au Québec. Québec: Commission de l'éthique en science et en technologie; 2017.
11. Scassa, T. Who Can Own Data in a Data Economy? CIGI [En ligne]. 2018 [cité le 25 juin 2019]. Disponible: www.cigionline.org/articles/who-can-own-data-data-economy.
12. Boerding A, Culik N, Doepke C, Hoeren T, Juelicher T, Roettgen C, et al. Data ownership – a property rights approach from a European perspective. J Civ L Stud. 2018;11:323.