Boîte à outils pour les données sensibles — destiné aux chercheurs

Partie 1: Glossaire terminologique sur l'utilisation des données sensibles à des fins de recherche

Préparé par le Groupe d'experts sur les données sensibles (GEDS) du réseau Portage au nom de l'Association des bibliothèques de recherche du Canada (ABRC)

SEPTEMBRE 2020

Réseau Portage
Association des bibliothèques de recherche du Canada
portage@carl-abrc.ca

www.carl-abrc.ca

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Introduction

Le Groupe d'experts sur les données sensibles du Réseau Portage a créé une suite d'outils pour les chercheurs canadiens. Ces outils ont été créés pour aider les chercheurs à comprendre comment les données de recherche s'inscrivent dans le processus d'éthique de la recherche et pour aborder l'évolution des pratiques de gestion des données de recherche (GDR) telles que le partage et le stockage des données dans le contexte des cadres actuels d'éthique de la recherche.

Cet outil intitulé « Glossaire terminologique sur l'utilisation des données sensibles à des fins de recherche » fournit des définitions pour un certain nombre de termes couramment utilisés lors de discussions sur la gestion des données sensibles dans le contexte canadien. Ces définitions ont été établies à partir d'une analyse de l'environnement des politiques, des procédures et des formulaires de l'Université Brock, de l'Université Mount Saint Vincent, de Données de recherche Canada (DRC), de CANARIE, de l'Autorité d'éthique de la recherche en santé de Saint-Jean (Terre-Neuve-et-Labrador), de l'Université de l'Alberta, de l'Université d'Ottawa, de l'Université de Toronto et de l'Université Queen's. Les ressources supplémentaires qui ont été consultées incluent l'Énoncé de politique des trois Conseils : Éthique de la recherche avec des êtres humains (EPTC2), le Consortia Advancing Standards in Research Administration Information (CASRAI), le Règlement général sur la protection des données de l'Union européenne (RGPD), la Loi sur la protection des renseignements personnels sur la santé (LPRPS) de l'Ontario et l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO).

Le Groupe d'experts sur les données sensibles est composé d'un large éventail de membres des communautés de recherche, notamment des professionnels de l'éthique de la recherche, des représentants d'organismes de financement et des membres d'organisations autochtones, qui sont intéressés au domaine des données de recherche sensibles. Le groupe travaille ensemble à l'élaboration de conseils et d'outils pratiques pour la gestion des données sensibles au Canada.

Glossaire terminologique sur l'utilisation des données sensibles à des fins de recherche

Note : Compte tenu de la nature évolutive de la gestion de données de recherche, le présent document se veut dynamique. Lorsqu'il y a des divergences entre les définitions canadiennes et les définitions internationales, les définitions canadiennes auront préséance. Lorsqu'il y a des divergences entre deux définitions canadiennes à l'égard de la recherche avec des êtres humains, les définitions prévues à l'Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains (EPTC 2) auront préséance. Pour ajouter des termes ou proposer des modifications au présent document, veuillez contacter le reseau Portage.

Accès aux données : Le droit ou la possibilité d'utiliser ou de consulter les données conservées dans une base de données ou un dépôt/un environnement de recherche/ un environnement virtuel de recherche. L'accès aux données est un élément important dans la gestion des données de recherche.

Accès libre (AL) : Ensemble de principes et de pratiques qui assurent la distribution en ligne de résultats de recherche avec un minimum d'obstacles à l'accès ou sans obstacle à l'accès. Tous les types de contenu numériques peuvent être disponibles en accès libre : textes, données, logiciels, audio, vidéo et multimédias. Si les contenus textes sont les plus répandus en ce moment, de plus en plus de sources en font l'intégration d'images, de données et de code exécutable. Les sources AL peuvent aussi s'appliquer à du contenu non académique comme de la musique, des films et des romans.

Autoriser l'accès aux données : Le chercheur autorise le personnel (à titre individuel ou fonctionnel) à accéder aux données.

Chiffrement : Processus de transformation de l'information en un format différent ou de cryptage des données. Seule la personne possédant la clé de déchiffrement peut lire les données d'origine. Le chiffrement a pour objet d'assurer la confidentialité de données numérisées stockées sur un ordinateur ou un serveur et transmises par internet ou enregistrées sur des périphériques de stockage. Les données chiffrées sont brouillées et illisibles par toute personne qui ne possède pas la clé de déchiffrement, le code secret ou le mot de passe.

Connaissances traditionnelles/données sur les spécificités autochtones : Connaissances (savoirs) détenues par les peuples des Premières Nations, les Inuits et les Métis, qui constituent les peuples autochtones du Canada. Ces connaissances se rattachent à un lieu précis, sont généralement transmises oralement et sont fondées sur l'expérience de plusieurs générations. Elles sont déterminées par les terres, le

milieu, la région, la culture et la langue d'une communauté autochtone. Elles comprennent les connaissances provenant des générations passées, de même que les innovations et les nouvelles connaissances transmises aux générations futures.

Conservation numérique (ou archivage) : Gestion active au fil du temps d'un contenu numérique. Le processus inclut les activités suivantes : la sélection de contenus à préserver, la préparation et l'entretien de contenus dans un format ou dans un environnement propice à leur utilisation à long terme et la mise en place de stratégies pour assurer l'accessibilité des contenus à long terme.

Couplage : La combinaison de deux ou de plusieurs ensembles de données possédant des éléments en commun susceptibles de fournir de nouveaux renseignements ou de nouveaux ensembles de données. Le nombre croissant de bases de données et les progrès de la technologie permettant de coupler des bases de données créent de nouvelles possibilités pour la recherche, mais aussi de nouveaux risques d'atteinte à la vie privée. Il se peut notamment que le couplage de bases de données dépersonnalisées ou anonymisées permette de réidentifier des personnes. En vertu de l'article 5.7 de l'EPTC 2, les chercheurs doivent obtenir l'autorisation du CER avant de procéder au couplage de donnée provenant d'une recherche avec des êtres humains.

Cycle de vie des données : Tous les stades d'existence de données à partir de la collecte jusqu'à la destruction. La perspective du cycle de vie permet la gestion active de données au fil du temps, pour en assurer ainsi la sécurité, l'accessibilité et l'utilité.

Dépersonnalisation (brouillage) : Le fait de modifier les données particulières liées à la personne afin de réduire le risque de divulgation de son identité. Cela peut inclure le brouillage d'identificateurs directs (p. ex. nom, numéro de téléphone, coordonnées géographiques), la transformation (recodage, combinaison) ou la suppression d'identificateurs indirects qui pourraient être utilisés seuls ou en combinaison pour identifier une personne (p. ex. date d'anniversaire, coordonnées géographiques, dates d'événements clés). Lorsque la dépersonnalisation (brouillage) est exécutée convenablement, le risque de réidentification de données partagées ou publiées est atténué.

Dépôt de données : Consignation de données dans un dépôt reconnu (institutionnel, national, code source ouvert, dépôt par discipline ou pluridisciplinaire ou autre) en vue de la conservation et possiblement de l'utilisation éventuelle par d'autres chercheurs.

Dommages : Conséquences négatives qui peuvent se traduire en effets indésirables sur les particuliers, les groupes ou les organismes. Les conséquences négatives peuvent être d'ordre social, comportemental, psychologique, économique, écologique ou légal. Elles peuvent porter atteinte à la vie privée, à la sécurité, à la réputation ou au statut d'une personne ou d'un organisme.

Données : Faits, mesures, enregistrements, archives ou observations sur le monde. Les données peuvent exister dans tout format ou sur tout support. Elles peuvent prendre la forme d'écrits, de notes, de chiffres, de symboles, de textes, d'images, de films, de vidéos, d'enregistrements sonores, d'illustrations, d'esquisses, de dessins ou d'autres représentations graphiques, de reproductions, de manuels de procédures, de formulaires, de diagrammes, d'organigrammes de flux de travail, de descriptions d'équipement, de dossiers de données, d'algorithmes de traitement de données ou de registres statistiques.

Données anonymes : Données dépourvues d'identificateurs directs depuis la première collecte; le risque d'identification à l'aide d'identificateurs indirects est faible et même très faible. Notez que selon l'article 2.4 de l'EPTC 2 (2018), l'évaluation par le CER n'est pas exigée dans le cas de recherches qui se limitent exclusivement à l'utilisation secondaire de données anonymes des participants humains ou de matériel biologique humain anonyme dans la mesure où le processus de couplage de données, d'enregistrement ou de diffusion des résultats ne génère aucun renseignement identificatoire.

Données anonymisées (dépersonnalisées/brouillées) : Données dont les identificateurs directs ont été supprimés de manière irréversible; aucun code n'est conservé qui pourrait permettre le recouplage futur avec des identificateurs directs; le risque de réidentification de la personne à partir des identificateurs qui subsistent est faible et même très faible. Selon l'EPTC 2 (2018), l'utilisation secondaire de données anonymisées (dépersonnalisées) à des fins de recherche avec des êtres humains exige l'évaluation et l'approbation du Comité d'éthique de recherche (REB).

Données brutes : Données qui n'ont pas été interprétées en vue d'une utilisation pertinente. Les données brutes ont le potentiel de se transformer en « information », mais au préalable elles doivent être soumises à une extraction sélective et requièrent organisation, analyse et formatage avant d'être présentées. À moins d'être anonymes, les données brutes peuvent être considérées sensibles car elles ne sont pas dépersonnalisées.

Données comportant un niveau de risque élevé : Données qui nécessitent la mise en place de contrôles rigoureux contre la divulgation non autorisée, les pertes ou les modifications de données susceptibles de porter atteinte tant aux chercheurs qu'aux participants de recherche, qui peuvent être des particuliers, des groupes ou des organismes. Les données qui comportent un niveau élevé de risque comprennent, mais non de façon limitative, des informations concernant la race ou l'origine ethnique; les opinions politiques; les croyances religieuses ou croyances de toute autre nature similaire; l'affiliation syndicale; la santé ou la condition physique ou mentale; la vie sexuelle; et l'infraction ou l'infraction alléguée d'un délit par un participant.

Données comportant un niveau de risque faible : Données qui nécessitent la mise en place de contrôles contre les modifications de données non autorisées afin d'assurer l'intégrité des données plutôt que de prévenir les atteintes contre les chercheurs ou les participants de recherche. Les exemples comprennent, mais sans limitation, les données à accès libre composées de données entièrement dépersonnalisées ou anonymes, les formulaires de consentement en blanc, les fiches signalétiques et les données recueillies sur des sites Web destinés au public.

Données comportant un niveau de risque moyen : Données qui nécessitent la mise en place de contrôles rigoureux contre la divulgation non autorisée, les pertes ou les modifications de données ; dans la plupart des cas, ces données sont jugées confidentielles. La divulgation, la perte ou l'utilisation non autorisée de données confidentielles pourraient mettre en danger les participants. Des exemples de données confidentielles peuvent inclure les renseignements personnels identifiables (RPI), les renseignements personnels sur la santé (RPS) et les renseignements sur le crédit personnel (RCP).

Données confidentielles : Renseignements confiés à une personne, organisme ou personne morale responsable de maintenir le caractère privé de tels renseignements et de contrôler ou de restreindre l'accès à ceux-ci.

Données d'ordre administratif : Renseignements recueillis avant tout à des fins administratives et non à des fins de recherche. Sont compris dans cette catégorie les profils et les CV de chercheurs, la portée et l'impact de projets de recherche, les sources de financement, les citations et les résultats de recherche. Les renseignements de ce genre sont souvent recueillis par les services gouvernementaux et par d'autres organismes à des fins d'inscription, de transaction ou d'ouverture de dossiers liés à la prestation de services.

Données de recherche : Toute information recueillie, observée, générée ou créée soit en vue d'une recherche ou de la validation des résultats de recherche; elles peuvent être répliquées ou réutilisées.

Données légalement accessibles : Données légalement accessibles au public, sous réserve de la nomination d'un gérant/dépositaire légalement désigné (voir **gérant de données**) qui assure la confidentialité et la protection de droits propriétaires des données (p. ex., un coordinateur d'accès à l'information et confidentialité ou consignataire des données du recensement canadien).

Données non identifiables : Données qui, dès la première collecte, ne peuvent pas être utilisées pour identifier un particulier, pour distinguer une personne par rapport à une autre ou pour retracer des informations personnelles identifiables. À noter qu'il faut toujours procéder avec prudence si deux ensembles de données de ce genre doivent être couplés, car le couplage risque de produire des données identifiables.

Données ouvertes : Données disponibles en accès libre. Ce sont des données structurées qui sont accessibles, lisibles à la machine, utilisables, intelligibles et partageables gratuitement. Les données ouvertes sont gratuites et réutilisables. Elles se prêtent à la modification et à la distribution collective, sous réserve au maximum d'attribution-ShareAlike.

Données sensibles : Informations qui doivent être protégées contre l'accès non autorisé ou la divulgation. Elles peuvent inclure:

- Renseignements personnels
- Renseignements personnels sur la santé
- Dossiers scolaires
- Dossiers clients
- Informations financières
- Renseignements d'ordre criminel
- Renseignements géographiques (p. ex. localisations détaillées d'espèces en péril)
- Renseignements personnels confidentiels
- Renseignements jugés confidentiels; données confiées à une tierce personne, à une organisation ou à une entité dans l'intention d'en préserver la confidentialité en interdisant ou en limitant les droits d'accès
- Renseignements protégés en vertu de toute politique interdisant l'accès non autorisé

Les données sensibles comprennent tout renseignement concernant une personne physique, une organisation ou une entité identifiées ou identifiables.

Entrepôt de données : Une localisation où les chercheurs peuvent consigner leurs données en vue de stockage et de gestion. Les entrepôts de données peuvent être sujets à des exigences particulières en matière de sujet, de domaine de recherche, de format et de réutilisation des données. D'autres entrepôts sont conçus pour recevoir une grande diversité de données. Les documents consignés à l'entrepôt de données font idéalement l'objet de curation, dont la responsabilité incombe au gérant des données qui doit assurer l'authenticité, la découvrabilité et l'accessibilité à moyen terme.

EPTC 2: *L'Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains – EPTC 2* est une politique commune des trois organismes de recherche fédéraux : le Conseil de recherches en sciences humaines du Canada (CRSH), le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et les Instituts de recherche en santé du Canada (IRSC), aussi appelés « les Organismes ». La Politique exprime l'engagement constant des Organismes envers la population canadienne à promouvoir la conduite éthique de la recherche avec des êtres humains

au sein d'institutions canadiennes admissibles au financement des trois conseils précités. Cette politique s'applique à toutes les recherches avec des participants humains menées au nom et/ou sous la direction de l'institution.

Gérant de données : La personne responsable de la définition des données (c.-à-d. qui définit les traits particuliers d'éléments contenus dans la base de données) et de l'autorisation d'accès, surtout l'accès aux données ou la divulgation aux tierces personnes ; les gérants de données en collaboration avec les enquêteurs principaux fournissent du soutien pour :

(a) la collecte et l'intégration de données ou la réutilisation de données existantes;

(b) le contrôle de qualité des données;

(c) la description du flux de travail/processus scientifique;

(d) la fourniture de métadonnées conformes aux normes; et

(e) la soumission et la production de données.

Après consultation et communication avec les enquêteurs principaux, les gérants de données sont responsables de:

(a) la préservation de données et de produits de données; et

(b) la mise à disposition de modèles (p. ex., services Web, NetCDF, etc.) à des fins de découverte et d'intégration des données.

Gestion de données : Processus dynamique impliquant la mise en œuvre de normes et de pratiques exemplaires au cours du cycle de recherche afin de maximiser l'efficacité et la réutilisabilité de données et de sous-produits connexes, y compris, mais non de façon limitative, les normes et les pratiques ayant trait à la gestion de données, la planification, les métadonnées, le stockage, l'émission de licences, l'éthique, la découverte, la conservation et la réutilisation.

Gestionnaire des données : Un informaticien ou un organisme spécialisé en informatique responsable de l'infrastructure informatique destinée à accueillir et à protéger des données conformément aux politiques et aux pratiques relatives à la gouvernance des données.

Logiciel de sondage en ligne : Logiciel utilisé par des chercheurs pour recueillir des données sur les répondants à l'aide de questionnaires en ligne, dont les formulaires sont habituellement reliés à une base de données dans laquelle les réponses sont consignées et dont l'analyse est effectuée à l'aide des logiciels statistiques du produit.

Il incombe tant aux chercheurs qu'aux participants de bien connaître les politiques en matière de confidentialité et de sécurité des données mises au point par les fournisseurs de ces logiciels.

Métadonnées : Signifient littéralement « données à propos de données »; elles servent à définir ou à décrire les caractéristiques d'autres données le but étant de faciliter la compréhension des données et des processus ayant un lien avec les données. **Les métadonnées commerciales comprennent** les définitions et les noms commerciaux de domaines thématiques, d'entités et d'attributs, de types de données d'attribut et d'autres caractéristiques d'attribut, de descriptions de gamme, de noms de domaine valides et leurs définitions. **Les métadonnées techniques** comprennent les noms de bases de données physiques, les caractéristiques des colonnes et les caractéristiques d'autres objets dans la base de données, y compris les moyens utilisés pour le stockage des données. **Les métadonnées opérationnelles** définissent et décrivent les caractéristiques d'autres composantes du système (processus opérationnels, règles d'affaires, programmes, emplois, outils, etc.). **Les métadonnées sur la gestion des données** concernent les gérants de données, le processus de gestion et l'imputabilité en matière de responsabilité.

Partage de données : Pratique qui consiste à rendre disponibles les données à des fins de découverte et de réutilisation. Pour en assurer l'accès, les données doivent être consignées dans un dépôt ou être publiées par tout autre moyen. Le partage peut faire l'objet de limitations ou autres conditions particulières, particulièrement lorsque les données sont sensibles, sous réserve d'exigences ou lorsqu'elles sont protégées à titre de propriété exclusive.

Périphérique de stockage portable : Tout dispositif ou support facilement transportable sur lequel les données peuvent être stockées. Cette définition n'est pas limitée aux périphériques conçus à cette fin comme les CD/DVD, disques durs amovibles et clés USB, mais peut s'appliquer aux ordinateurs portables, tablettes électroniques, téléphones intelligents, PDA ou tout autre dispositif informatique. Les périphériques de stockage portables peuvent être connectés ou non connectés à Internet et les mesures de sécurité applicables seront variables selon le contexte.

Personne identifiable : Toute personne susceptible d'être identifiée directement ou indirectement, notamment par référence à un numéro d'identification ou par un ou plusieurs éléments qui sont spécifiquement liés à son identité physique, physiologique, mentale, économique, culturelle ou sociale. Même si on parle de « personnes » identifiables, les mêmes considérations sont applicables aux communautés, aux organismes et aux personnes morales.

Plan de gestion de données (PGD) : Énoncé formel décrivant comment les données de recherche seront gérées et documentées au cours d'un projet de recherche. Les

éléments de base suivants sont normalement inclus dans le plan de gestion de données : les métadonnées, les politiques régissant l'accès aux données, le partage, l'échange, la réutilisation et la répartition des données ; ainsi que les mesures visant le stockage, la conservation ou la destruction de données.

Pseudonymisation : Technique d'authentification ou utilisation d'un nom fictif (pseudonyme) pour désigner une personne, un groupe ou un lieu spécifique afin de supprimer tout lien; le lien ne peut plus être établi en l'absence de renseignements supplémentaires. Dans le domaine de la recherche, la pseudonymisation est une sorte de dépersonnalisation pour protéger l'identité des participants et des organismes impliqués dans la recherche.

Publication de données : La divulgation de données de recherche, de métadonnées associées, de documentation d'appui et de code logiciel (dans les cas où les données brutes ont fait l'objet de traitement ou de manipulation) en vue d'une réutilisation ou d'une analyse, de telle manière qu'elles deviennent découvrables sur le Web et puissent être citées de manière unique et constante.

Renseignements disponibles au public : Tout matériel documentaire, enregistrement ou publication archivée existante qui peut comprendre ou non des renseignements identificatoires et qui

(a) n'a pas de restrictions d'utilisation ou de distribution et ne comporte aucune attente raisonnable en matière de respect de la vie privée, particulièrement à l'égard de données de recherche avec participants humains; ou

(b) qui peut être divulgué publiquement en vertu de mécanismes précis établis par règlement sous réserve de la nomination d'un gérant/dépositaire désigné en conformité avec les lois sur l'accès à l'information et la protection de la vie privée, lequel doit assurer la confidentialité et la protection de droits propriétaires des données (p. ex., un coordinateur d'accès à l'information et confidentialité ou consignataire des données du recensement canadien).

Renseignements identificatoires : Renseignements qui permettent d'identifier une personne, un organisme ou une personne morale ou qui risquent vraisemblablement de permettre de faire cette identification lorsque, en certaines circonstances ils sont utilisés seuls ou en combinaison avec d'autres renseignements.

Renseignements d'identification directe : Renseignements permettant d'identifier une personne en particulier par des identificateurs directs (p. ex. nom, numéro d'assurance sociale ou numéro d'assurance maladie).

Renseignements d'identification indirecte : Renseignements qui peuvent vraisemblablement permettre d'identifier une personne par une combinaison d'identificateurs indirects (p. ex. date de naissance, lieu de résidence ou caractéristiques personnelles distinctives).

Renseignements personnels : Renseignements concernant une personne, y compris, mais non de façon limitative :

- nom, adresse ou numéro de téléphone,
- race, nationalité ou origine ethnique, couleur, croyances religieuses ou politiques et affiliations à une association,
- âge, genre, orientation sexuelle, état civil ou familial,
- numéro d'identifiant, symbole ou autre signe attribué à une personne, symbole ou adresse protocole Internet (adresse IP),
- empreintes digitales, type sanguin ou traits héréditaires,
- renseignements sur la condition de santé ou antécédents médicaux, y compris les conditions de santé physique ou mentale,
- renseignements sur l'éducation, les antécédents criminels, la condition financière ou statut d'emploi ou expérience professionnelle,
- croyances, opinions, évaluations ou commentaires de la personne.

* À noter que l'évaluation concernant le risque d'identification des renseignements se fera dans le contexte d'un projet de recherche spécifique.

Renseignements personnels identifiables (RPI) : Semblables aux identificateurs directs, les RPI font référence à tout renseignement qui peut être utilisé pour identifier spécialement une personne, une organisation ou une entité (aux présentes la « personne ») afin de contacter ou de retracer celle-ci, ou tout renseignement combiné avec d'autres sources de renseignement pour arriver au même objectif. Les RPI comprennent, mais non de façon limitative, le nom ou d'autres traits identificatoires de la personne comme la date d'anniversaire, l'adresse ou le géocodage. Les renseignements encodés à l'aide d'identificateurs personnels uniques peuvent être identificatoires si le détenteur des renseignements détient également la liste maîtresse ou la clé créant le lien entre les personnes et les identificateurs personnels uniques (IPU). Les renseignements peuvent également être identificatoires à ce niveau à cause de différents éléments d'information connus à l'égard de cette personne. L'identité de la personne peut également être établie à partir de données agrégées, s'il n'y a qu'un nombre limité de personnes qui figurent dans cette catégorie. La capacité identificatoire est liée aux caractéristiques uniques des renseignements, à la quantité de renseignements détenus susceptibles d'être combinés ainsi qu'aux habiletés et technologies que le détenteur des renseignements pourrait déployer en vue de telles combinaisons.

Renseignements personnels sur la santé : Les renseignements identificatoires d'une personne de source orale ou de source écrite sont les suivants :

(a) renseignements concernant la santé physique ou mentale d'une personne, y compris les renseignements au sujet de l'historique de santé de la famille,

(b) renseignements concernant la prestation de soins de santé, incluant la personne qui en assure la prestation,

(c) renseignements qui font partie d'un plan de service pour la personne au sens de La Loi sur les soins à domicile et les services communautaires, 1994,

(d) renseignements concernant les versements à la personne ou le droit de la personne aux soins de santé et à la prestation de soins de santé,

(e) renseignements concernant le don par la personne de son corps ou d'une de ses substances corporelles ou tout renseignement provenant des résultats de tests ou d'examens effectués sur une partie du corps ou une substance corporelle de celui-ci,

(f) le numéro d'assurance maladie de la personne ou

(g) les renseignements identifiant son décisionnaire remplaçant.

À noter que l'évaluation concernant le risque d'identification des renseignements se fera dans le contexte d'un projet de recherche spécifique.

Renseignements primaires : Renseignements de première main recueillis par le chercheur (c.-à-d. directement des participants) concernant directement la question principale de la recherche.

Renseignements secondaires : Renseignements recueillis par une personne autre que le chercheur dans un but ne concernant pas directement la question principale de la recherche. Les renseignements secondaires sont des données préexistantes qui sont réutilisées dans un nouveau contexte qui est différent du contexte dans lequel les données ont été recueillies à l'origine.

Ressources destinées au public : Une ressource qui accepte les demandes anonymes d'accès à partir de toute adresse publique de protocole Internet, c'est-à-dire des ressources externes accessibles au public.

Risque : Éventualité de dommages ou d'atteintes qui peuvent être subis par des personnes, des communautés ou des organisations. Le risque est fonction de l'ampleur et de la gravité des dommages et de la probabilité de sa survenance chez

les participants ou les tiers. En ce qui concerne le risque, il faut prendre en considération les dommages immédiats, les dommages différés et les dommages en aval.

Risques/dommages en aval : Conséquences négatives imprévues et inattendues au moment de la prise de décision, qui peuvent se traduire en effets indésirables sur les particuliers, les groupes ou les organismes. Au moment de la collecte de données sensibles, et en vue de leur conservation, entretien et réutilisation, les risques de dommages immédiats et en aval doivent être pris sérieusement en considération.

Sécurité des données : Moyens utilisés pour sauvegarder les données pour que le chercheur puisse y accéder de manière appropriée et selon ses besoins (disponibilité des données), de manière à ce qu'elles ne subissent aucune altération ou modification (intégrité des données), qu'elles demeurent confidentielles (confidentialité des données) et qu'elles soient soigneusement préservées et détruites (rétention/destruction).

Services infonuagiques : Une méthode de partage et d'enregistrement de données qui sont hébergées sur des serveurs à distance et accessibles par Internet. Les services infonuagiques sont hébergés, exploités et gérés par un fournisseur de services infonuagiques à partir d'un serveur d'archivage. Il existe des services infonuagiques publics et privés. L'utilisation de services infonuagiques comporte toujours un élément de risque, mais les risques associés aux serveurs publics ou privés sont d'un autre ordre. Les principales différences concernent la localisation, le contrôle du serveur ou des vulnérabilités potentielles.

- **Les services infonuagiques privés** assurent le stockage sur des serveurs internes à l'aide d'infrastructures locales; aucun autre organisme n'a accès aux services infonuagiques privés. La gestion et l'accès aux données sur des serveurs privés internes sont contrôlés par l'institution ou l'organisme responsable de l'infrastructure. Les données sont stockées sur l'intranet de l'institution et sont habituellement protégées par un pare-feu. La gestion, la conservation et la mise à jour de données relèvent de l'institution. Les services infonuagiques privés offrent souvent un niveau de sécurité plus élevé parce qu'il y a peu de mise en commun de ressources. Cependant, plusieurs établissements ne disposent pas de l'infrastructure et/ou du personnel pour héberger les services infonuagiques privés.

- **Les services infonuagiques publics** sont fournis par des entreprises qui offrent aux utilisateurs des services de stockage gratuits ou moyennant des frais sur des serveurs à distance gérés par le fournisseur. Le fournisseur est responsable de la gestion et de la conservation des données hébergées dans son centre de données. Si différents clients ont le droit d'utiliser les services infonuagiques

publics, les données qui appartiennent aux clients et les applications que les clients font exécuter sur le nuage sont cachées aux autres utilisateurs du service infonuagique. À titre d'exemple de services infonuagiques publics, on peut nommer Google Drive, DropBox, iCloud et OneDrive (personnel). Il convient de signaler aux institutions postsecondaires canadiennes que les services infonuagiques publics sont offerts sur des serveurs qui échappent au contrôle de l'institution. Les serveurs peuvent se trouver dans n'importe quel pays et, à ce titre, seront soumis aux lois du pays en question. Même si les services de nuage publics sont assortis de mesures de sécurité et de protection de la vie privée, ils demeurent vulnérables par l'étendue de leurs infrastructures et par la multitude de points d'accès à travers lesquels les utilisateurs non autorisés peuvent essayer de dérober des données. Dans certains contextes, les services infonuagiques privés sont moins vulnérables à de telles attaques.

Suppression : Le processus de destruction de données conservées sur des disques durs, sur des périphériques mobiles ou sur d'autres supports électroniques afin de les rendre illisibles, irrécupérables, inaccessibles ou inutilisables.