Boîte à outils pour les données sensibles — destiné aux chercheurs

Partie 2: Matrice de risque lié aux données de recherche avec des êtres humains

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 « Matrice de risque lié aux données de recherche avec des êtres humains » a été conçu pour aider les chercheurs à déterminer le degré de risque lié aux données de recherche avec les êtres humains et à prendre des décisions par rapport à la gestion, au stockage et à l'accès ou l'utilisation future convenable de ces données. Le risque peut être déterminé en tenant compte des trois facteurs suivants : 1. Le caractère identifiable des données au moment de leur collecte et de leur stockage ; 2. la vulnérabilité des sujets des données, soit en tant qu'individu, collectivité ou population ; et 3. la sensibilité des données quant à leur susceptibilité à causer des torts, soit physiques, psychologiques/émotionnels, sociaux et légaux, entre autres. Cette matrice est destinée à être utilisée de concert avec la partie 1 de cette boîte à outils, Glossaire terminologique sur l'utilisation des données sensibles à des fins de recherche.

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.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Définitions des degrés de risque	Données disponibles publiquement, pour lesquelles les participants n'ont pas d'attente raisonnable en matière de respect de la vie privée, quel que soit leur caractère sensible ou identifiable. Les données sont collectées sans aucune information qui pourrait raisonnablement identifier des personnes ou des groupes. Les données ne contiennent aucun renseignement confidentiel ou sensible. Les sujets liés aux données ne sont pas vulnérables dans le contexte de la recherche et ne subiraient aucun tort en cas de fuite.	Tous les identificateurs ont été dépersonnalisés afin que les données déposées ne contiennent aucune information qui pourrait raisonnablement identifier des personnes ou des groupes. Les données pourraient contenir de l'information collectée initialement comme confidentielle ou sensible. Les sujets liés aux données ne sont pas vulnérables dans le contexte de la recherche et ne subiraient aucun tort en cas de fuite.	Les renseignements demeurent identifiables ou une nouvelle identification est possible ou probable. Les données contiennent de l'information confidentielle ou sensible. Les sujets liés aux données pourraient être vulnérables dans le contexte de la recherche et pourraient subir des torts en cas de fuite.	L'obtention des données est régie par une entente (formelle ou informelle) avec la personne détentrice des données, interdisant la rétention ou l'utilisation future de celles-ci. Les renseignements demeurent identifiables ou une nouvelle identification est possible ou probable. Les données contiennent de l'information confidentielle ou sensible. Les sujets liés aux données sont vulnérables dans le contexte de la recherche et subiraient des torts en cas de fuite.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Consentement éclairé	Avis indiquant que les données seront disponibles pour une utilisation future.	Avis indiquant que les données seront disponibles pour une utilisation future. Une option de désistement au stockage devrait être considérée.	Avis indiquant que les données seront disponibles pour une utilisation future. Demande de permission pour le partage ou le stockage de données clairement indiquée dans le formulaire ou le processus de consentement. Offrir des options par rapport aux domaines de recherches futures, si possible.	Confidentialité maintenue aussi longtemps que les données sont disponibles. Données partagées uniquement avec les membres de l'équipe de recherche.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Collecte de données	Données disponibles publiquement, p. ex. trouvées en ligne, dans des archives publiques ou collectées par observation en milieu naturel. Les chercheurs ne connaissent pas les identités des participants ou des sujets liés aux données. Les méthodes ne devraient pas comporter d'interaction directe avec les participants aux recherches. Celles-ci incluent habituellement des sondages, des questionnaires et de la recherche par observation. Aucun identificateur direct ou indirect n'est collecté.	Les chercheurs peuvent connaître les identités des participants ou des sujets liés aux données et peuvent avoir promis de protéger leur confidentialité par le biais du consentement éclairé. Les méthodes de collecte de données varient beaucoup et peuvent comporter de l'interaction directe avec les participants aux recherches. Des identificateurs directs ou indirects peuvent être collectés. **La majorité des recherches avec les humains sont dans cette catégorie.**	Les chercheurs peuvent connaître les identités des participants ou des sujets liés aux données et peuvent avoir promis de protéger leur confidentialité par le biais du consentement éclairé. Les méthodes de collecte de données varient beaucoup et peuvent comporter de l'interaction directe avec les participants aux recherches. Des identificateurs directs peuvent être collectés. Des identificateurs indirects collectés peuvent également suffire pour rendre les participants identifiables.	Les chercheurs peuvent connaître les identités des participants ou des sujets liés aux données et ont promis de protéger leur confidentialité par le biais du consentement éclairé. Les méthodes de collecte de données varient beaucoup et peuvent comporter de l'interaction directe avec les participants aux recherches. Des identificateurs directs peuvent être collectés, mais des identificateurs indirects collectés peuvent également suffire pour rendre les participants identifiables.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Analyse/ gestion des données	Aucune restriction quant à l'analyse des données disponibles publiquement. L'analyse de données doit respecter le document ou le libellé du consentement éclairé et le protocole autorisé par le CÉR.	Les identificateurs directs doivent être remplacés par un code de liage (p. ex. : pseudonyme, code alphanumérique) et séparés physiquement ou électroniquement de la liste principale. Les formulaires de consentement doivent être conservés séparément des données de recherche. Les données identificatoires doivent seulement être accessibles aux membres de l'équipe de recherche.	Les identificateurs directs doivent être remplacés par un code de liage (par exemple : pseudonyme, code alphanumérique) et séparés physiquement ou électroniquement de la liste principale. Les formulaires de consentement doivent être conservés séparément des données de recherche. Les identificateurs indirects devraient être codés si possible. Les données ne doivent pas être accessibles ou analysées dans un espace public où quelqu'un pourrait voir les données sur un appareil ou par un autre moyen.	Les identificateurs directs doivent être remplacés par un code de liage (par exemple : pseudonyme, code alphanumérique) et séparés physiquement ou électroniquement de la liste principale. Les formulaires de consentement doivent être conservés séparément des données de recherche. Les identificateurs indirects devraient être codés si possible. Les données doivent seulement être accessibles aux membres de l'équipe de recherche, selon les indications dans le protocole autorisé, et l'accès/l'analyse doit se faire dans un environnement sécurisé.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Stockage (stockage actif) et sécurité des données	Tous les appareils de stockage, les types de partage de fichiers et les services infonuagiques sont permis, incluant les services infonuagiques publics et institutionnels. Les données doivent être sauvegardées de manière conséquente par rapport à leur degré de risque.	Les données identificatoires doivent être stockées sur des appareils protégés par mot de passe dans un lieu sécuritaire approprié. S'il est nécessaire de rendre les données accessibles par internet, elles doivent être cryptées. Les services infonuagiques publics ne doivent pas être utilisés, à moins qu'ils soient la seule option possible. S'ils sont utilisés, les fichiers et l'accès doivent être cryptés et protégés par mot de passe. Les services infonuagiques privés offerts par les établissements de recherche ou jugés sécuritaires peuvent être utilisés. Les données doivent être sauvegardées de manière conséquente par rapport à leur degré de risque.	Toutes les données doivent être stockées sur des appareils cryptés protégés par mot de passe dans un lieu sécuritaire approprié. Si les données doivent être accessibles par internet, elles doivent être cryptées. Les services infonuagiques publics sont strictement interdits. Les services infonuagiques privés offerts par les établissements de recherche ou jugés sécuritaires peuvent être utilisés si le CÉR les autorise. Les données doivent être sauvegardées de manière conséquente par rapport à leur degré de risque.	Toutes les données doivent être stockées sur un ordinateur/site centralisé indépendant crypté, protégé par mot de passe et situé dans un lieu sécuritaire approprié. Les données doivent être sauvegardées de manière conséquente par rapport à leur degré de risque.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Mobilité/ partage des données	Les données peuvent être partagées par courriel ou services infonuagiques tels que les services infonuagiques publics..	Les fichiers cryptés et protégés par mot de passe peuvent être partagés par courriel et services infonuagiques ou sites collaboratifs autorisés par l'établissement.	Les données réservées doivent seulement être partagées entre les membres de l'équipe de recherche selon les indications dans le protocole autorisé. Les fichiers doivent être cryptés et protégés par mot de passe.	Les données sont uniquement disponibles sur un ordinateur/site centralisé indépendant, crypté et protégé par mot de passe. Les fichiers ne doivent pas être copiés ou partagés. L'accès aux données doit se limiter aux personnes autorisées et identifiées explicitement dans le protocole du CÉR et au plus petit nombre possible de personnes.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Stockage des données et accès à celles-ci (y compris l'utilisation secondaire)	Les données doivent être déposées avec un accès sans restriction dans un délai raisonnable, compte tenu de la publication des articles originaux. L'utilisation secondaire des données ne nécessite pas l'autorisation du CÉR.	Les données des participants/sujets qui se désistent doivent être séparées des données à déposer. Les données dépersonnalisées doivent être déposées avec un accès sans restriction dans un délai raisonnable compte tenu de la publication des articles originaux, le besoin de reproduire la recherche et la durée appropriée des données pour leur réutilisation. L'utilisation secondaire des données dépersonnalisées nécessite actuellement l'autorisation du CÉR.	Les données des participants/sujets qui se désistent doivent être séparées des données à déposer. Les données dépersonnalisées doivent être déposées avec un accès restreint évalué par les gardiens des données. Les données peuvent être séparées en ensembles selon les utilisations auxquelles les participants ont consenti par le biais du consentement éclairé (par exemple : utilisation seulement pour cette étude, seulement pour les études dans le même domaine ou pour n'importe quelle utilisation). L'utilisation secondaire des données nécessite l'autorisation du CÉR.	Les données ne doivent pas être déposées ailleurs que dans le lieu convenable pour les besoins de stockage et d'accès des membres de l'équipe de recherche.

	Risque faible	Risque moyen	Risque élevé	Risque extrême
Rétention et destruction des données	Les données peuvent être retenues indéfiniment à des fins de découverte, d'accès et d'archivage.	Les données peuvent être retenues indéfiniment à des fins de découverte, d'accès et d'archivage.	Les données peuvent être retenues indéfiniment à des fins de découverte, d'accès et d'archivage selon le protocole autorisé par le CÉR.	Les données doivent être détruites le plus tôt possible selon le protocole autorisé par le CÉR.