Directives sur la dépersonnalisation des données

Préparé par le Groupe de travail sur la COVID-19 du Réseau Portage pour l'Association des bibliothèques de recherche du Canada (ABRC)

Kristi Thompson (Université Western)
Lucia Costanzo (Université de Guelph)
Erin Clary (Portage)
Beth Knazook (Portage)
Nick Rochlin (Université de Colombie-Britannique)
Felicity Tayler (Université d'Ottawa)
Jane Fry (Université Carleton)
Chantal Ripp (Université d'Ottawa)
Kathy Szigeti (Université de Waterloo)
Qian Zhang (Université de Waterloo)
Roger Reka (Université de Windsor)
Minglu Wang (Université York)
Rebecca Dickson (COPPUL)
Mark Leggott (RDC-DRC)
Melanie Parlette-Stewart (Portage)

SEPTEMBRE 2020

Portage Network
Canadian Association of Research Libraries
portage@carl-abrc.ca

www.carl-abrc.ca

Table of Contents

Directives sur la dépersonnalisation

Les directives qui suivent devraient vous aider à minimiser le risque lié à la divulgation lorsque vous partagez des données collectées auprès de participants humains. Le cas échéant, veuillez indiquer dans votre documentation et fichier README que vous utilisez l'une ou l'autre des techniques suivantes pour anonymiser vos données.[1] Dans un souci de transparence, vous devriez indiquer clairement la manière dont l'ensemble de données a été modifié pour protéger les participants.

Avant tout, vous devez savoir que toutes les données venant de participants humains n'ont pas à être dépersonnalisées ou dépouillées des identifiants directs et indirects. Passez en revue votre formulaire de consentement et préparez vos données pour partager seulement ce que les participants veulent bien partager. Si vous ne savez pas si vous devez ou non dépersonnaliser vos données, veuillez consultez le guide de Portage Puis-je partager mes données ? et le comité d'éthique de la recherche de votre établissement.[2] Si vous avez besoin d'aide pour choisir un dépôt pour vos données, veuillez consulter le guide de Portage intitulé Dépôts recommandés pour les données de recherche sur la COVID-19 ou demander aux bibliothécaires de votre établissement si d'autres formes de soutien sont disponibles.[3]

Si vous avez besoin d'une aide pour comprendre des termes utilisés dans ce document, veuillez consulter le Glossaire terminologique sur l'utilisation des données sensibles à des fins de recherche du Réseau Portage.[4] Pour en savoir plus, vous pouvez également consulter la Matrice de risque lié aux données de recherche avec des êtres humains ou Langage en matière de gestion de données de recherche pour le consentement éclairé.[5]

---

[1] Pour en savoir plus sur la création d'une documentation appropriée voir le Groupe de travail sur la COVID-19 du Réseau Portage. « Ensembles des documents requis pour les dépôts ». 23 septembre 2020, https://doi.org/10.5281/zenodo.4046707.

[2] Groupe de travail sur la COVID-19 du Réseau Portage, « Puis-je partager mes données ? », 23 septembre 2020, https://doi.org/10.5281/zenodo.4046659.

[3] Groupe de travail sur la COVID-19 du Réseau Portage, « Dépôts recommandés pour les données de recherche sur la COVID-19 », 23 septembre 2020, https://doi.org/10.5281/zenodo.4046685.

[4] Groupe d'experts sur les données sensibles du Réseau Portage, "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," 14 octobre 2020, https://doi.org/10.5281/zenodo.4088985.

[5] Groupe d'experts sur les données sensibles du Réseau Portage, "Boîte à outils pour les données sensibles — destiné aux chercheurs Partie 2 : Matrice de risque lié aux données de recherche avec des êtres humains," 19 octobre 2020, https://doi.org/10.5281/zenodo.4107118 et Groupe d'experts sur les données sensibles du Réseau Portage, "Boîte à outils pour les données sensibles — destiné aux chercheurs Partie 3 : Langage en matière de gestion de données de recherche pour le consentement éclairé," 19 octobre 2020, https://doi.org/10.5281/zenodo.4107185.

Recenser et retirer les identifiants directs

Les identifiants directs sont les identifiants au moyen desquels les participants d'une étude risquent d'être tout de suite repersonnalisés. À moins que les participants aient donné leur consentement explicite, ils doivent être retirés de toute version publiée de votre ensemble de données. La liste qui suit est basée sur diverses sources : les directives des principaux organismes de financement internationaux, la loi américaine *Health Insurance Portability and Accountability Act* (HIPAA) (trad. : loi sur la transférabilité et la responsabilité de l'assurance santé) et le *British Medical Journal*. Voir Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers (trad. : conseils sur la préparation des données cliniques brutes pour publication à l'intention des éditeurs de revues, des auteurs et des pairs examinateurs) et la List of 18 items considered under HIPAA to be identifiers (trad. : liste des 18 éléments considérés comme des identifiants en vertu de la HIPAA).[6]

Les identifiants directs sont :

- noms ou initiales, ainsi que les noms des proches ou membres du ménage ;
- adresses et identifiants de petites zones géographiques, tels que code postal ou code zip ;
- numéros de téléphone ;
- identifiants électroniques tels qu'adresses Web, adresses courriel, pseudonymes de médias sociaux ou adresses IP d'ordinateurs personnels ;
- numéros d'identification uniques, tels que numéro de dossier de l'hôpital, numéro d'assurance sociale, numéros de dossier d'essai clinique, numéros de compte, numéros de certificat ou de permis ;
- dates précises d'événements personnels, tels que mariage, admission à l'hôpital ou congé de l'hôpital, intervention médicale ;
- données multimédias : photographies, données audio ou vidéos intactes d'une personne ;
- identifiants biométriques, dont images de l'iris ou de la rétine, empreintes digitales ou vocales ;
- données génomiques humaines sauf si, après explication donnée sur le risque, les participants de l'étude ont donné leur consentement à ce que les données soient partagées ou que les données fassent l'objet d'une utilisation secondaire ;
- renseignements sur l'âge pour les personnes de plus de 89 ans.

Comment puis-je retirer cette information ?

Il est assez simple de retirer les identifiants directs de vos données. Vous pouvez enregistrer ces renseignements personnels dans un document, un tableur ou une base de données à part et lier le tout à

[6] Iain Hrynaszkiewicz, Melissa L. Norton, Andrew J. Vickers, et Douglas G. Altman, "Preparing Raw Clinical Data for Publication: Guidance for Journal Editors, Authors, and Peer Reviewers." *BMJ* 340 (29 janvier 2010): c181, https://www.bmj.com/content/340/bmj.c181; et Steve Alder, "What is Considered PHI Under HIPAA Rules?" *HIPAA Journal* (28 décembre 2017), https://www.hipaajournal.com/considered-phi-hipaa/.

d’autres points de données en vous servant d’une série de codes qui peuvent être retirés avant la publication. Vous pouvez aussi choisir de supprimer entièrement les points de données de personnalisation à la fin du projet. Retournez à vos formulaires de consentement pour choisir la manière de procéder. Si vous ne savez pas s’il est possible de simplement dissocier les données ou si elles doivent être détruites, consultez votre comité d’éthique de la recherche.

Recenser et évaluer les identifiants indirects ou les quasi-identifiants en fonction du risque et de l'utilité qui leur sont associés

Les identifiants indirects ou les quasi-identifiants sont des attributs (tels que les données démographiques) associés aux personnes qui, lorsqu'ils sont liés à d'autres sources de données, peuvent entacher la confidentialité. Les quasi-identifiants en eux-mêmes ne permettent pas nécessairement de personnaliser un sujet, mais ils peuvent le faire s'ils sont combinés à d'autres données. Par exemple, révéler la taille de la communauté d'origine d'un participant dans une zone géographique bien définie d'une étude pourrait permettre à quelqu'un de déduire plus précisément d'où vient ce participant. Une variable devrait être considérée comme un quasi-identifiant lorsqu'une personne pourrait vraisemblablement l'associer à des renseignements tirés d'une autre source. Voir les principes de l'anonymisation de l'International Household Survey Network et L'anonymisation et autres mesures de protection de la vie privée (directives en anglais) du Commissaire à l'information et à la protection de la vie privée de l'Ontario.[7]

Liste de quasi-identifiants possibles :

- identifiants géographiques du lieu d'origine (géographie du recensement, nom de la ville, indicateur rural ou urbain), lieu de naissance, lieu de traitement, lieu d'études ou toute autre donnée géographique associée à une personne ;
- identité selon le sexe ou le genre, orientation ;
- données ethniques, race, minorité visible ou statut d'autochtone
- statut d'immigration ;
- appartenance à des organisations ;
- utilisation de réseaux ou de services sociaux particuliers ;
- données socioéconomiques, telles que la profession ou le lieu de travail, le revenu ou le niveau de scolarité ;
- composition du ménage et de la famille, état civil, nombre d'enfants/de grossesses ;
- dossier criminel et autres renseignements susceptibles d'induire un rapprochement avec des dossiers publics ;
- données généralisées liées à une personne, par exemple l'âge, l'année d'obtention du diplôme, l'année d'immigration ;
- Certaines réponses complètes
    - Note : chacune des réponses doit être vérifiée. Par exemple, un commentaire tel que « *La bibliothèque devrait prolonger ses heures d'ouverture* » ne permet pas de personnaliser un sujet alors qu'une phrase commençant par « *À titre de président d'un groupe de recherche fréquentant la bibliothèque*… » le permettrait.

[7] "Anonymization Principles," International Household Survey Network, document obtenu le 4 août 2020, https://ihsn.org/node/137; Information and Privacy Commissioner of Ontario, *Deidentification Guidelines for Structured Data*, Information and Privacy Commissioner of Ontario, 8 juin 2016. https://www.ipc.on.ca/privacy-organizations/de-identification-centre/.

- *Certains* renseignements médicaux (p. ex. handicaps permanents ou conditions médicales rares) pourraient permettre de personnaliser un sujet ; une maladie temporaire ou une blessure est moins susceptible de le faire. Le test de validation consiste à se demander s'il s'agit d'une information susceptible de se trouver ailleurs et de ce fait d'être utilisée pour repersonnaliser le sujet.

Comment savoir quelles combinaisons de quasi-identifiants représentent un problème ?

1. Observer les combinaisons possibles

Tout d'abord, il faut commencer par examiner les variables démographiques dans l'ensemble de données et ensuite décrire une personne à un ami en mentionnant uniquement les valeurs de ces variables. Y a-t-il une chance que la personne soit reconnue ? Exemple : « *Je pense à une personne mariée de sexe féminin vivant à Toronto, détentrice d'un diplôme universitaire qui est âgée de 40 à 55 ans et dont le revenu se situe entre 60 000 et 75 000 dollars.* » Même s'il n'y a qu'une seule personne ayant ces attributs dans l'ensemble de données, les renseignements ne sont sans doute pas suffisants pour créer un risque SAUF si l'information contextuelle au sujet de l'ensemble de données circonscrit les possibilités. Par exemple, si vos données se rapportent à un groupe restreint et spécifique de personnes, disons les arbitres de l'association de hockey de l'Ontario, la liste des quasi-identifiants mentionnés dans l'exemple pourrait suffire à personnaliser un sujet. Il faut que la portée des quasi-identifiants soit évaluée dans le contexte de ce qui est connu ou qui peut être raisonnablement induit au sujet de la population de l'étude.

2. Évaluer ces combinaisons mathématiquement

La *K*-anonymity (trad. : *k*-anonymisation) est une approche permettant de démontrer mathématiquement qu'un ensemble de données a été anonymisé, la variable *k* étant un nombre entier choisi par le chercheur pour représenter un groupe de dossiers contenant la même information parmi tous les quasi-identifiants.[8] À l'intérieur d'un ensemble de données, un ensemble de « *k* » dossiers (p. ex. 3 ou 5) est appelé « classe d'équivalence ». La *k*-anonymisation est réussie s'il n'est pas possible de distinguer un dossier d'un autre dans sa classe d'équivalence. Par exemple, si vous donnez la valeur 5 à *k*, chaque dossier dans votre ensemble de données doit avoir l'ensemble exact de quasi-identifiants présents dans au moins 4 autres dossiers pour réussir la *k*-anonymisation.

La k-anonymisation ne fonctionne que pour estimer avec précision le risque si un ensemble de données est un échantillon complet d'une population en particulier. La *k*-anonymisation aura pour effet de surestimer considérablement le risque si un ensemble de données représente un sous-échantillon d'une population. Pour déterminer la valeur appropriée à attribuer à « *k* », vous seriez bien avisé(e) de tenir compte de ce qui suit :

---

[8] Khaled El Emam et Fida Kamal Dankar, "Protecting Privacy Using k-Anonymity," *Journal of the American Medical Informatics Association* 15, no. 5 (septembre 2008): 627–637, https://doi.org/10.1197/jamia.M2716.

- une variable « *k* » d'une valeur de 3 pourrait convenir pour les ensembles de données qui contiennent de petits échantillons d'une grande population ;
- une valeur plus grande (ou plus prudente) devrait être donnée à la variable « *k* » si un ensemble de données est un échantillon complet d'une population.

Gardez à l'esprit qu'un ensemble de données qui est un échantillon complet d'une population connue pourrait receler des facteurs de risque additionnels. Imaginons que tous les répondants d'une classe d'équivalence particulière aient donné la même réponse à une question d'une enquête : vous sauriez ainsi comment chaque personne appartenant à cette classe d'équivalence a répondu. Les répondants à une enquête se font généralement dire que leurs réponses resteront confidentielles, en plus que personne ne saura quelle rangée de données contient leurs réponses. Un ensemble de données « *k* – anonyme » qui est un échantillon complet ne garantit peut-être pas le respect de la confidentialité.

Le code détaillé à l'Annexe 1 peut être utilisé dans votre progiciel de statistique pour créer des classes d'équivalence basées sur les quasi-identifiants dans votre ensemble de données et pour en faire la liste selon leur taille. Si une classe d'équivalence comporte moins de membres que la valeur « *k* » choisie, utilisez les techniques de réduction des données ci-dessous pour diminuer encore plus le risque lié à l'ensemble de données.

Pour un complément d'information sur la *k*-anonymisation, consultez la page mesurer le risque de divulgation de l'International Household Survey Network (IHSN) et la section 2.2.2 sur l'anonymisation garantie du cadre décisionnel d'anonymisation du UK Anonymisation Network.[9]

3. Contrôler le risque lié à l'ensemble de données en appliquant une technique de réduction des données

Des fréquences univariées et des tableaux à deux variables peuvent être utilisés pour définir des petites[10] catégories de quasi-identifiants. Il est possible d'appliquer des techniques de réduction des données pour atténuer le risque après que ces petits groupes ont été définis. Il existe trois types de réduction de données facilement applicables :

1. Le type de réduction le plus simple consiste à complètement **retirer les variables risquées** de l'ensemble de données. Cette option convient pour les variables présentant un risque relativement élevé dont la valeur pour la recherche n'est pas jugée élevée. (Par exemple, dans

[9] "Measuring the Disclosure Risk," International Household Survey Network, document obtenu le 4 août 2020, https://ihsn.org/anonymization-risk-measure; et Mark Elliot, Elaine Mackey, Kieron O'Hara, et Caroline Tudor, *The Anonymisation Decision-Making Framework*. UK Anonymisation Network (UKAN), University of Manchester, 2016, https://ukanon.net/ukan-resources/ukan-decision-making-framework/.

[10] Terme « petites » à relativiser; au premier passage, les groupes plus petits que 5 % de l'ensemble de données ou contenant moins de 20 dossiers pourraient être pris en compte.

certains ensembles de données, la géographie pourrait être jugée relativement moins importante que l'ethnicité ou la langue.)

2. Le deuxième type de réduction est le **recodage global** ou l'agrégation des valeurs observées dans un ensemble de classes défini, consistant par exemple à transformer une variable avec l'âge en une variable avec des catégories d'âge de 10 ans ou à fixer la limite supérieure d'une catégorie de revenu élevé à « 100 000 $ et plus ».
3. Le troisième type de réduction pour les cas inhabituels est la **suppression locale**. Par exemple, l'état civil d'un très jeune répondant marié pourrait être marqué « manquant » comme solution de rechange au recodage global de la variable « âge » qui par ailleurs ne serait pas risquée dans un grand groupe.

Après chaque exercice de réduction des données, répétez le test de validation décrit plus haut et vérifiez les classes d'équivalence jusqu'à ce que tous les groupes soient plus grands que la valeur « *k* » sélectionnée.

Pour un complément d'information, notamment au sujet de types de réduction de données plus complexes, consultez la section 2.5 sur les solutions au regard de l'anonymisation du cadre décisionnel d'anonymisation du UK Anonymisation Network.[11]

Comment évaluer la sensibilité des variables préservant l'anonymat dans l'ensemble de données?

Les renseignements préservant l'anonymat sont les mesures et les réponses à une enquête qui ne risquent pas d'être associées à des personnes en particulier. Des opinions, des rangs, des échelles ou des mesures temporaires telles que la fréquence cardiaque au repos après une séance de méditation ou le nombre de fois où une personne a pris un petit déjeuner dans une semaine en sont des exemples.

Il se peut que ce genre de renseignements soit aussi très sensible. L'information pouvant stigmatiser ou discriminer une personne, telle que le dossier criminel, les pratiques sexuelles, l'usage de substances interdites, l'état de santé mentale et l'équilibre psychologique ainsi que d'autres renseignements médicaux de nature délicate, fait augmenter le risque associé à l'ensemble de données et devrait être examinée au moment de décider si on peut ou non rendre les données publiques. Vous pourriez choisir de supprimer ou de modifier ces variables pour produire une version de données moins sensible.

[11] Elliot, Mackey, O'Hara, et Tudor, *The Anonymisation Decision-Making Framework*.

Facteurs à considérer pour la dépersonnalisation des données qualitatives

Les données qualitatives décrivent les qualités ou les attributs observables mais non nécessairement mesurables. Les données de ce genre sont collectées par le truchement d'entretiens, d'enquêtes ou d'observations et peuvent se présenter sous forme de transcriptions, de notes, d'enregistrements audio ou vidéo, d'images et de documents textes. À l'instar des données quantitatives, ces données peuvent contenir des identifiants directs comme des noms, des dates et des lieux de naissance, d'autres lieux et même des photos. Il est possible de jumeler ces identifiants directs à des identifiants indirects ou des quasi-identifiants, tels que des renseignements médicaux, des renseignements sur les études, les finances et l'emploi, pour retracer ou trouver l'identité d'une personne.

La marche à suivre pour retirer les renseignements de personnalisation dans une vidéo, un entretien audio ou une transcription orale est très différente de celle utilisée pour dépersonnaliser un dossier médical. D'abord, il est plus difficile de le faire à l'aide d'un programme. Des notes ou de l'information audiovisuelle extrêmement détaillées exigent souvent que quelqu'un lise ou regarde le contenu très attentivement.

Conseils d'ordre général

- Évitez en premier lieu de demander de l'information de personnalisation.
    - Il est plus facile de modifier l'information au point de saisie que de retirer l'information après qu'elle a été enregistrée.
    - Si vous avez besoin de ce genre d'information à l'étape de la recherche, essayez de la saisir dans les premières minutes d'un entretien ou d'un enregistrement ; il sera ainsi facile de la modifier rapidement. Vous pouvez aussi transcrire l'information dans un document séparé qui peut être retiré d'un dossier personnel.
- Intégrez la dépersonnalisation dans le processus du consentement éclairé.
    - Faites en sorte que les participants de l'étude soient au courant de l'utilisation que vous prévoyez faire de leurs données et qu'ils sachent que cette information peut être rendue anonyme pour assurer leur protection. Indiquez clairement dans votre formulaire de consentement dans quelle mesure cette information sera rendue anonyme (c'est-à-dire quels sont les éléments qui seront remplacés ou retirés). Bien que les identifiants directs puissent être éliminés (nom, adresse, date de naissance, etc.), il pourrait rester de subtils indices sur l'identité dans l'enregistrement ou la transcription.
    - Convenez à l'avance avec les participants des genres d'information de personnalisation qui peuvent être révélés dans un entretien. Par exemple, le participant pourrait ne pas vouloir que le nom de son employeur soit divulgué. Il est plus facile de procéder ainsi que de retirer l'information après le fait.
    - Gardez à l'esprit que ce ne sont pas toutes les données qui doivent être dépersonnalisées ou anonymisées. Dans certains cas, vous pourriez enregistrer des récits extrêmement personnels et vous devriez vous souvenir du droit du participant à raconter son histoire dans ses propres mots. Certains participants pourraient avoir un intérêt personnel à ce que leur identité reste connue.

Directives sur la dépersonnalisation

- Utilisez des pseudonymes et modifiez des éléments de personnalisation pour protéger l'anonymat.
    - S'il est possible de changer le nom de la personne, son lieu de résidence ou sa profession sans compromettre l'ensemble de données, l'anonymat pourrait être protégé. Sachez toutefois que l'utilité de l'ensemble de données pourrait en être affectée puisque la perception d'un prochain chercheur au regard de la condition socioéconomique ou du comportement de la personne interrogée pourrait s'en trouver altérée.
- Si nécessaire, retirez des blocs de texte sensible ou modifiez des parties des données audiovisuelles.
    - Une partie de la recherche pourrait devoir être rédigée. Alors, faites preuve de prudence si vous utilisez des techniques de « recherche et remplacement » parce qu'il est facile de remplacer la mauvaise information.
    - Les voix des enregistrements audio pourraient devoir être masquées par une altération du ton.
    - Les figures paraissant dans des données visuelles pourraient devoir être pixellisées.
- Restreignez l'accès.
    - Ce n'est pas souhaitable, mais certains ensembles de données perdront de leur utilité si tous les identifiants sont retirés. Il pourrait être possible que des chercheurs voulant un accès secondaire aux données demandent que des recherches soient effectuées par l'équipe de recherche de départ, qui pourrait ensuite partager les résultats dans la mesure où ils ne révèlent pas d'information confidentielle ou qu'ils peuvent être dépersonnalisés.

Pour un complément d'information, voir le guide d'anonymisation des données qualitatives du UK Data service.[12]

[12] "Anonymisation: Qualitative Data," UK Data Service, dernière modification le 30 juin 2020, https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative.aspx.

Facteurs à considérer au regard des médias sociaux, des images médicales et des données génomiques

Données collectées des médias sociaux ou des plateformes de réseau social (p. ex. Twitter, Facebook)
Même si l'information hébergée dans les sites des réseaux sociaux est en accès et en consultation libres, cela ne signifie pas nécessairement qu'elle peut être redistribuée librement. Un bon nombre de plateformes prévoient des conditions d'utilisation auxquelles vous devrez vous plier, et les personnes qui utilisent l'une ou l'autre des plateformes s'attendent sans doute à ce que leur vie privée soit protégée. Certaines plateformes exigent des utilisateurs qu'ils s'enregistrent avant que le contenu soit visible et d'autres pourraient imposer des conditions interdisant la collecte de données, le moissonnage des données ou la réédition du contenu ailleurs.

Voici quelques questions que vous devrez vous poser avant de déposer les données des médias sociaux :

- Le sujet à l'étude pourrait-il être jugé sensible ?
- Vos données pourraient-elles conduire à une stigmatisation de l'auteur du contenu ou à une discrimination à son endroit ?
- La population à l'étude est-elle vulnérable ?
- Quelles sont les attentes des utilisateurs de la plateforme en ce qui a trait à la protection de la vie privée ?
- Est-il possible ou raisonnable d'obtenir un consentement éclairé ?
- Les données peuvent-elles ou devraient-elles être anonymisées ?
- Les conditions d'utilisation de la plateforme vous autorisent-elles à redistribuer le contenu ?

Prenons l'exemple de *Twitter*. Cette plateforme permet à l'auteur de contenu de garder le contrôle sur ses micromessages (« *twitts* »). Dans ses politiques, *Twitter* indique que seuls les identifiants numériques des micromessages et les identifiants d'utilisateur devraient être redistribués.[13] Si, après avoir soupesé les questions ci-dessus, vous décidez de déposer votre ensemble de données, les micromessages doivent d'abord être « déshydratés » (distillés jusqu'à leurs identifiants) en se servant d'un outil comme twarc de DocNow.[14] Pour pouvoir faire une utilisation secondaire des données, l'utilisateur final devra réhydrater les identifiants des micromessages à l'aide de REST API de Twitter ou un outil externe tel que Hydrator de DocNow.[15] Le contenu ne sera pas envoyé pour les micromessages qui ont été depuis supprimés.

Vous trouverez des directives approfondies dans les ressources suivantes :

---

[13] "Developer Terms: More About Restricted Uses of the Twitter APIs," Twitter, document obtenu le 4 août 2020, https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases.
[14] Documenting the Now, "DocNow/twarc." GitHub, document obtenu le 4 août 2020, https://github.com/docnow/twarc.
[15] Documenting the Now, "DocNow/hydrator." GitHub, document obtenu le 4 août 2020, https://github.com/DocNow/hydrator.

- Zeffiro et Brodeur, Social Media Research Data Ethics and Management (trad. : la gestion des données de recherche tirées des médias sociaux et les questions d'éthique) (diapositives d'un atelier présenté à l'université McMaster)[16]
- Comité d'éthique de la recherche de l'université Ryerson, Guidelines for Research Involving Social Media (trad. : lignes directrices sur la recherche basée sur les médias sociaux)[17]
- Mannheimer et Hull, Sharing Selves: Developing an Ethical Framework for Curating Social Media Data) (trad. : élaborer un cadre éthique pour la curation des données des médias sociaux à l'ère de l'exposition du moi)[18]

La Social Media Archives Toolkit, (trad. : boîte à outils des archives des médias sociaux) de l'université de l'État de la Caroline du Nord, qui contient des directives sur les legal and ethical implications (trad. : implications d'ordre juridique et éthique) de partager les données des médias sociaux ainsi qu'une bibliographie annotée fournissant des ressources supplémentaires.[19]

Images médicales

Avant d'archiver les images médicales, retirez les identifiants directs que vous n'avez pas l'autorisation explicite de partager, tels que le nom, l'identifiant du patient et les dates paraissant dans l'en-tête des images ou les métadonnées incorporées, et caviardez les pixels de l'image qui contiennent des renseignements personnels. Les images cérébrales doivent être défigurées à l'aide d'un outil prévu à cette fin, tel que PyDeface.[20]

Les ressources suivantes fournissent des directives approfondies sur la dépersonnalisation des fichiers DICOM :

- La page De-identification Overview. (trad. : aperçu de la dépersonnalisation) de The Cancer Imaging Archive (TCIA)[21]

---

[16] Andrea Zeffiro and Jay Brodeur, "Social Media Research Data Ethics and Management." L'atelier a été presenté le 5 avril 2018, Sherman Centre for Digital Scholarship, McMaster University, http://hdl.handle.net/11375/25327.

[17] Ryerson University Research Ethics Board, "Guidelines for Research Involving Social Media," Ryerson University, novembre 2017, https://www.ryerson.ca/content/dam/research/documents/ethics/guidelines-for-research-involving-social-media.pdf.

[18] Sara Mannheimer et Elizabeth Hull, "Sharing Selves: Developing an Ethical Framework for Curating Social Media Data," International Journal of Digital Curation 12, no. 2 (18 avril 2018), https://doi.org/10.2218/ijdc.v12i2.518.

[19] "Social Media Archives Toolkit," North Carolina State University Libraries, document obtenu le 4 août 2020, https://www.lib.ncsu.edu/social-media-archives-toolkit.

[20] Certains dépôts pourraient vous aider pour la défiguration ou vous recommander des outils pour le faire, par exemple, l'International Neuroimaging Data-Sharing Initiative (INDI) peut aider les chercheurs qui planifient de partager leurs données sur la plateforme INDI. Pour en savoir plus, voir le Data Contribution Guide (trad. : guide sur la manière de verser des données) document obtenu le 31 août 2020, http://fcon_1000.projects.nitrc.org/indi/indi_data_contribution_guide.pdf. See also, Omer Faruk Gulban, Dylan Nielson, Russ Poldrack, John Lee, Chris Gorgolewski, Vanessasaurus, et Satrajit Ghosh, "Poldracklab/pydeface: V2.0.0." 31 octobre 2019. http://doi.org/10.5281/zenodo.3524401.

[21] Kirby, Justin. "Submission and De-identification Overview." The Cancer Imaging Archive (TCIA), University of Arkansas for Medical Sciences, 27 avril 2020,

    - Voir en particulier la liste complète des marqueurs DICOM jugés non sûrs dans le tableau 1 : Table 1 - DICOM Tags Modified or Removed at the source site (trad. : marqueurs DICOM modifiés ou supprimés du site Web).
- Le protocole de dépersonnalisation de l'International Covid-19 Open Radiology Database (RICORD) de la Radiological Society of North America (RSNA)[22]
- La norme DICOM donne des directives sur la dépersonnalisation de l'information donnée en en-tête. L'Appendix E : Attribute Confidentiality Profiles (trad. : l'annexe E : profils de confidentialité des attributs) de DICOM 15 : Security and System Management Profiles (trad. : Profils de gestion de système et de sécurité) en particulier pourrait être utile.[23]
    - Ces profils se veulent un moyen d'arriver au juste milieu entre la nécessité de protéger la vie privée et la nécessité de conserver de l'information pour que les données restent utiles.
    - S'il faut conserver des identifiants, vous devriez en principe avoir mentionné le profil que vous comptez utiliser dans votre demande au comité d'éthique de la recherche, et votre formulaire de consentement devrait indiquer clairement quels renseignements seront partagés.

La dépersonnalisation de fichiers DICOM peut se faire à l'aide d'un programme basé sur un logiciel servant à retirer les identifiants de l'en-tête.

- TCIA recommande le logiciel Clinical Trial Processor (CTP) (trad. : le traitement des essais cliniques) développé par RSNA.[24]
- La Covid-19 Open Radiology Database RSNA (RICORD) recommande un autre logiciel de la RSNA appelé « *Anonymizer* » (« anonymiseur ») ; elle a publié des instructions sur la manière de l'installer et de l'utiliser (install and use it). [25] Ce logiciel exécute le protocole de dépersonnalisation (de-identification protocol) de la RICORD.[26]

---

https://wiki.cancerimagingarchive.net/display/Public/Submission+and+De-identification+Overview.

[22] "RSNA International Covid-19 Open Radiology Database (RICORD) De-identification Protocol," Radiological Society of North America, International COVID-19 Open Radiology Database, document obtenu le 10 août 2020, https://www.rsna.org/-/media/Files/RSNA/Covid-19/RICORD/RSNA-Covid-19-Deidentification-Protocol.pdf.

[23] Medical Imaging & Technology Alliance, DICOM Standards Committee, "DICOM Part 15: Security and System Management Profiles." *DICOM Standard* (Arlington, VA: National Electrical Manufacturers Association), document obtenu le 4 août 2020, https://www.dicomstandard.org/current/.

[24] "Clinical trial processor (CTP)," Radiological Society of North America, Medical Imaging Resource Community (MIRC), document obtenu le 4 août 2020, https://www.rsna.org/research/imaging-research-tools.

[25] "RSNA COVID-19 DICOM Data Anonymizer," Radiological Society of North America, International COVID-19 Open Radiology Database, document obtenu le 10 août 2020, https://www.rsna.org/-/media/Files/RSNA/Covid-19/RICORD/RSNA-Anonymizer-Program-Instructions.pdf.

[26] "RSNA International Covid-19 Open Radiology Database (RICORD) De-identification Protocol," Radiological Society of North America, International COVID-19 Open Radiology Database.

- Il existe bien d'autres options non commerciales, dont l'outil DicomCleaner™.[27]
- Comme avec tous les logiciels de dépersonnalisation, les résultats pourront varier ; vous devriez donc confirmer que les renseignements personnels ont été retirés avant de partager vos images. À noter ce qui suit :
    - Les fournisseurs ou les utilisateurs finaux pourraient ne pas avoir toujours utilisé les éléments DICOM de manière conforme à la norme.
    - Les éléments relatifs à la vie privée ou les marqueurs privés pourraient avoir été utilisés pour sauvegarder de l'information personnelle, et l'utilisation de ces marqueurs pourrait ne pas avoir été bien définie dans les documents du fournisseur.

Données génomiques et autres prélèvements biomédicaux

La séquence ADN de chaque personne étant unique, les matériels biologiques ne pourront jamais être tout à fait anonymes. Avant d'archiver ces données ou de les verser dans une biobanque, vous voudrez bien relire votre formulaire de consentement. En principe, tout au long du processus conduisant au consentement vous aurez :

- informé les participants sur la manière dont leurs données seront utilisées, analysées, sauvegardées et partagées ;
- recensé les éléments d'information qui seront sauvegardés à côté des données ;
- communiqué au participant le niveau de protection de la vie privée ou de confidentialité qu'il peut attendre et indiqué qui pourrait accéder aux données ;
- précisé si les données/prélèvements seront conservés au Canada ou à l'étranger ;
- reconnu la possibilité que les données soient utilisées à des fins commerciales ;
- expliqué clairement les risques de divulgation.

Vous trouverez plus d'information dans l'EPTC (Énoncé de politique des trois conseils) 2 (2018), chapitre 12 : Matériel biologique humain et matériel lié à la reproduction humaine (en particulier les sections A et D), et chapitre 13 : Recherche en génétique humaine.[28] Voir également Canada : will privacy rules continue to favour open science? (trad. : Canada : les règles de confidentialité continueront-elles de favoriser la science ouverte ?) de Thorogood (2018).[29]

La page web sur la confidentialité en génomique des National Institutes of Health (NIH) donne un aperçu de certains des avantages et des risques liés à la communication de renseignements génétiques.[30] Pour

---

[27] Clunie, David A., "DicomCleaner™," PixelMed Publishing™, document obtenu le 16 juillet 2020, http://www.dclunie.com/pixelmed/software/webstart/DicomCleanerUsage.html.

[28] « Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains – TCPS 2 (2018) », Gouvernement du Canada, dernière modification le 19 février 2020. https://ethics.gc.ca/fra/policy-politique_tcps2-eptc2_2018.html.

[29] Adrian Thorogood, "Canada: Will Privacy Rules Continue to Favour Open Science?" Human Genetics 137: 595–602 (16 juillet 2018), https://doi.org/10.1007/s00439-018-1922-z.

[30] "Privacy in Genomics," National Human Genome Research Institute, 24 février 2020, https://www.genome.gov/about-genomics/policy-issues/Privacy.

avoir un exemple de la manière dont des renseignements génétiques ont été utilisés pour trouver l'identité des participants d'une étude, voir la page sur identification des génomes personnels par inférence du nom de famille, ou l'étude dans l'éditorial de la revue *Nature* de 2013 sur la confidentialité génétique.[31]

Pour un complément d'information sur l'éthique et le consentement en génomique, voir les ressources de la boîte à outils sur la réglementation et l'éthique, telles que la politique sur la confidentialité et la protection des données et la politique sur le consentement.[32]

[31] Gymrek, Melissa, Amy L. McGuire, David Golan, Eran Halperin, and Yaniv Erlich. "Identifying Personal Genomes by Surname Inference." *Science* 339, no. 6117 (18 janvier 2013): 321-324. https://doi.org/10.1126/science.1229566; and "Genetic privacy" [Editorial], *Nature* 493 (24 janvier 2013): 451, https://doi.org/10.1038/493451a.
[32] Global Alliance for Genomics & Health. *Genomic Toolkit: Regulatory & Ethics Toolkit.* Toronto, ON: Global Alliance for Genomics and Health, document obtenu le 20 juillet 2020, https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/.

Annexe 1 : Code de vérification de la *K*-Anonymisation

-- Stata --

```
* Stata code for checking k-anonymity

* Kristi Thompson, May 2020


* create the equivalence groups

egen equivalence_group= group(var1 var2 var3 var4 var5)

* create a variable to count cases in each equivalence group

sort equivalence_group

by equivalence_group: gen equivalence_size =_N

* list the ID numbers of equivalence groups containing 3 or fewer cases

tab equivalence_group if equivalence_size < 3, sort

* list the values of the quasi-identifiers for each small equivalence
class.

list var1 var2 var3 var4 var5 if equivalence_group == 1
```

--- R --

```
# R code for checking k-anonymity

# Carolyn Sullivan and Kristi Thompson, May 2020


# install plyr, a useful data manipulation package.

install.packages("plyr")

# Load the library.

library('plyr')
```

```
datafile <- " location of the data file - csv format -  "

# Read the csv  file.

df <- read.csv (datafile)


# Figure out what equivalence classes there are, and how many cases in
each equivalence class.

dfunique <- ddply(df, .(var1, var2, var3, var4, var5), nrow)

dfunique <- dfunique[order(dfunique$V1),]

View(dfunique)
```

Un code à utiliser avec le logiciel SPSS est fourni à l'annexe B du cadre décisionnel d'anonymisation.[33]

[33] Elliot, Mackey, O'Hara, et Tudor, *The Anonymisation Decision-Making Framework*.

Annexe 2 : Progiciels de dépersonnalisation gratuits

Un bon nombre de ces outils ont une approche hiérarchique pour la dépersonnalisation des données, ce qui signifie que vous devrez définir à l'avance les généralisations possibles pour les quasi-identifiants dans l'ensemble de données ; le programme recherchera les solutions possibles et recommandera une série de généralisations à utiliser pour bien réussir l'anonymisation. Cette approche pourrait se révéler utile pour les ensembles de données comportant un grand nombre de quasi-identifiants ou dans les cas où plusieurs ensembles de données comportant des quasi-identifiants semblables doivent être recensés. Pour les ensembles de données de petite taille, il pourrait être plus simple d'avoir recours à un logiciel de statistique. Les progiciels mentionnés ici présentent tous des difficultés d'utilisation et imposent une courbe d'apprentissage plutôt exigeante. Amnesia et l'interface utilisateur graphique avec sdcMicro sont sans doute les outils les plus conviviaux.

Outils recommandés :

- Amnesia
    - Ce logiciel est disponible en version en ligne et en version de bureau ; cependant, téléverser les données de nature délicate dans un site Web d'un tiers n'est généralement pas recommandé. Si c'est possible, installer le logiciel localement (Windows ou Linux uniquement).
    - Amnesia accepte la *k*-anonymisation et la k^m-anonymisation (une approche d'anonymisation offrant une certaine flexibilité lorsque l'ensemble de données contient un nombre élevé de quasi-identifiants : il permet d'effectuer des combinaisons de *m* quasi-identifiants qui paraîtront au moins *k* fois dans les données publiées).
    - Quelques limites : il ne permet pas actuellement de préciser les valeurs manquantes ; la documentation manque de fini ; par exemple, la définition des hiérarchies n'est pas évidente.
    - Ce logiciel conviendra parfaitement pour les données cliniques ou les données qui ne proviennent pas d'une enquête.
- sdcMicro
    - Logiciel en R pour le contrôle statistique de la divulgation des microdonnées (anonymisation). Ce logiciel peut lire un grand nombre de types de données (p. ex. csv, sav, dta, sas7bdat, xlsx) et peut être utilisé dans les systèmes d'exploitation Windows, Linux ou Mac. Il exécute le code muArgus.
    - Une interface utilisateur graphique est disponible, et une vignette avec directives appelée « Using the interactive GUI - sdcApp » (trad. : utiliser l'IUG interactive – sdcApp) accessible depuis la page de renvoi de sdcMicro dans le dépôt CRAN *Comprehensive R Archive Network*).[34]

[34] "Using the interactive GUI – sdcApp, The Comprehensive R Archive Network (CRAN), document obtenu le 31 août 2020, https://cran.r-project.org/web/packages/sdcMicro/vignettes/sdcMicro.html.

- Il faut savoir que le chargement d'ensembles de données volumineux prend du temps et que le temps de calcul d'ensembles de données complexes est long. [35]
- La section des directives sur le contrôle statistique de la divulgation (CSD) pour les microdonnées sur SDC with sdcMicro in R (trad. : CSD avec sdcMicro en R) pourrait vous fournir plus d'indications sur l'installation et l'utilisation du logiciel *sdcMicro*. Voir aussi sdcMicro GUI manual Documentation (trad. : documentation sur l'IGU de sdcMicro). [36]

Autres outils utiles :

- ARX
  - Outil d'anonymisation de source publique utilisable avec Windows, Linux et Mac. Il accepte les bases de données SQL, et les fichiers xlsx et csv, et est doté d'une interface utilisateur graphique.
  - Il prend en charge divers modèles de confidentialité (privacy models) y compris la *k*-anonymisation et les variantes ℓ*-diversity* (diversité), *t-closeness* (proximité), β*-Likeness* (similarité) et plus. [37]
  - Il permet aux utilisateurs finaux de catégoriser, généraliser et transformer les données de manières plus complexes et d'en fixer les limites inférieure et supérieure.
  - Le chargement des ensembles de données volumineux et le calcul pour les ensembles de données volumineux ou complexes pourraient prendre du temps.
- mu-Argus
  - Logiciel d'application des techniques de contrôle statistique de la divulgation. Le programme suit une approche hiérarchique pour la dépersonnalisation des données.
  - Le fichier JAR devrait être exécutable dans Windows ou Mac OS.
  - Un testeur a démontré qu'il était difficile d'obtenir les données chargées et correctement définies. Il a affirmé que le programme devrait améliorer sa documentation sur la création des hiérarchies.
- Boîte à outils d'anonymisation de l'université du Texas à Dallas
  - Elle prend en charge six différentes méthodes d'anonymisation et trois définitions de la vie privée, y compris *k*-anonymisation, ℓ*-diversity* (diversité) et t-closeness (proximité).
  - Les algorithmes peuvent être appliqués directement à un ensemble de données ou être utilisés en tant que fonctions de bibliothèque dans d'autres applications.
  - Il s'agit d'un ensemble de routines Java. Les curateurs de données qui préfèrent faire leur programme statistique en Java pourraient le trouver utile.

---

[35] "Computation time," SDC with sdcMicro in R: Setting Up Your Data and more, SDC Practice Guide, 2019, https://sdcpractice.readthedocs.io/en/latest/sdcMicro.html#computation-time.

[36] "Statistical Disclosure Control (SDC): An Introduction," SDC Practice Guide, 2019, https://sdcpractice.readthedocs.io/en/latest/SDC_intro.html; et Thijs Benschop et Matthew Welch. *Statistical Disclosure Control for Microdata: A Practice Guide for sdcMicro,* International Household Survey Network, document obtenu le 10 août 2020, https://sdcpractice.readthedocs.io/en/latest/index.html.

[37] "Privacy Models," ARX – Data Anonymization Tool, document obtenu le 31 août 2020, https://arx.deidentifier.org/overview/privacy-criteria/.

Annexe 3 : Services tarifés de dépersonnalisation

Voici quelques services tarifés que les chercheurs pourraient choisir pour leur dépersonnalisation :

- d-wise (bureaux américains et européens)
    - Offre de services d'anonymisation gratuits à ceux et celles qui travaillent sur le vaccin contre la COVID-19.[38]
    - Offre de services d'anonymisation gratuits aux chercheurs qui déposent dans Vivli les données de chaque participant des essais cliniques sur la COVID-19.[39]
- Le Consortium interuniversitaire de recherche politique et sociale (ICPSR) dont les archives se trouvent à l'université du Michigan.
    - Si vous voulez que l'ICPSR effectue une analyse de vos données, vous devrez acheter la trousse de curation professionnelle. Le coût est basé sur le nombre de variables et la complexité des données. Communiquez avec le service des acquisitions de l'ICPSR à l'adresse deposit@icpsr.umich.edu pour obtenir plus de renseignements (information obtenue dans la FAQ d'Open ICPSR dans les sections sur la tarification et les données à usage restreint).[40]
- Privacy Analytics (entreprise d'Ottawa)
    - Cette entreprise peut examiner des ensembles de données dans le cadre de ses Data Privacy Validation Services (trad. : services de validation de la confidentialité des données).[41]
    - Sa méthode est basée sur la norme de dépersonnalisation selon la détermination experte de la loi sur la transférabilité et la responsabilité de l'assurance maladie des États-Unis.
    - Pour en savoir davantage sur leurs services, remplissez le formulaire qui se trouve dans le bas de leur page Web de « Certification ».[42]

---

[38] "d-wise Offers Free Transparency Services Accelerating COVID-19 Vaccine Research," Cision PRWeb, 10 mars 2020, https://www.prweb.com/releases/d_wise_offers_free_transparency_services_accelerating_covis_19_vaccine_research/prweb16970368.htm.

[39] "d-wise offers anonymization services available on Vivli COVID-19 portal," Center for Global Clinical Research Data, 13 avril 2020, https://www.prweb.com/releases/d_wise_offers_free_transparency_services_accelerating_covis_19_vaccine_research/prweb16970368.htm.

[40] "FAQs," OpenICPSR, document obtenu le 31 août 2020,, https://www.openicpsr.org/openicpsr/faqs.

[41] "Clinical Trial Transparency Services," Privacy Analytics, document obtenu le 31 août 2020,, https://privacy-analytics.com/clinical-trial-transparency/ctt-services/.

[42] "Double-check your data and leverage it with confidence," Privacy Analytics, document obtenu le 31 août 2020,, https://privacy-analytics.com/health-data-privacy/health-data-services/expert-data-opinion-services/.

Ressources

1. Amnesia https://amnesia.openaire.eu/
2. ARX https://arx.deidentifier.org/overview/
3. d-wise https://www.d-wise.com/de-identification-services
4. mu-Argus https://github.com/sdcTools/muargus
5. Inter-university Consortium for Political and Social Research (ICPSR) https://www.openicpsr.org/openicpsr/
6. Privacy Analytics https://privacy-analytics.com/services/certification/
7. sdcMicro https://cran.r-project.org/web/packages/sdcMicro/index.html
8. The University of Texas at Dallas Anonymization Toolbox http://cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php

Références

1. Alder, Steve. “What is Considered PHI Under HIPAA Rules?” *HIPAA Journal*, 28 décembre 2017. https://www.hipaajournal.com/considered-phi-hipaa/.
2. Benschop, Thijs, et Matthew Welch. “Statistical Disclosure Control for Microdata: A Practice Guide for sdcMicro.” International Household Survey Network. Document obtenu le 10 août 2020. https://sdcpractice.readthedocs.io/en/latest/index.html.
3. Clunie, David A. “DicomCleaner™.” PixelMed Publishing™.” Document obtenu le 16 juillet 2020. http://www.dclunie.com/pixelmed/software/webstart/DicomCleanerUsage.html.
4. Documenting the Now. “DocNow/hydrator.” GitHub. Document obtenu le 4 août 2020. https://github.com/DocNow/hydrator.
5. Documenting the Now. “DocNow/twarc.” GitHub. Document obtenu le 4 août 2020. https://github.com/docnow/twarc.
6. El Emam, Khaled, et Fida Kamal Dankar. “Protecting Privacy Using k-Anonymity.” *Journal of the American Medical Informatics Association* 15, no. 5 (septembre 2008): 627–637. https://doi.org/10.1197/jamia.M2716.
7. Elliot, Mark, Elaine Mackey, Kieron O’Hara, et Caroline Tudor. *The Anonymisation Decision-Making Framework.* UK Anonymisation Network (UKAN). University of Manchester. 2016. https://ukanon.net/ukan-resources/ukan-decision-making-framework/.
8. “Genetic privacy.” [éditorial]. *Nature* 493 (24 janvier 2013): 451. https://doi.org/10.1038/493451a.
9. Global Alliance for Genomics & Health. *Genomic Toolkit: Regulatory & Ethics Toolkit.* Toronto, ON: Global Alliance for Genomics and Health. Document obtenu le 20 juillet 2020. https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/.
10. Gouvernement du Canada (Instituts de recherche en santé du Canada, Conseil de recherche en sciences naturelles et en génie du Canada et Conseil de recherche en sciences humaines du Canada). « Énoncé des trois conseils : Éthique de la recherche avec des êtres humains — TCPS 2 (2018) ». Dernière modification le 19 février 2020. https://ethics.gc.ca/fra/policy-politique_tcps2-eptc2_2018.html.

11. Groupe d'experts sur les données sensibles du Réseau Portage. “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.” 14 octobre 2020. https://doi.org/10.5281/zenodo.4088985.
12. Groupe d'experts sur les données sensibles du Réseau Portage. “Boîte à outils pour les données sensibles — destiné aux chercheurs Partie 2 : Matrice de risque lié aux données de recherche avec des êtres humains.” 19 octobre 2020. https://doi.org/10.5281/zenodo.4107118.
13. Groupe d'experts sur les données sensibles du Réseau Portage. “Boîte à outils pour les données sensibles — destiné aux chercheurs Partie 3 : Langage en matière de gestion de données de recherche pour le consentement éclairé.” 19 octobre 2020. https://doi.org/10.5281/zenodo.4107185.
14. Groupe de travail sur la COVID-19 du Réseau Portage, « Dépôts recommandés pour les données de recherche sur la COVID-19 », 23 septembre 2020, https://doi.org/10.5281/zenodo.4046685.
15. Groupe de travail sur la COVID-19 du Réseau Portage. « Ensembles des documents requis pour les dépôts ». 23 septembre 2020, https://doi.org/10.5281/zenodo.4046707.
16. Groupe de travail sur la COVID-19 du Réseau Portage, « Puis-je partager mes données ? », 23 septembre 2020, https://doi.org/10.5281/zenodo.4046659.
17. Gulban, Omer Faruk, Dylan Nielson, Russ Poldrack, John Lee, Chris Gorgolewski, Vanessasaurus, and Satrajit Ghosh. “Poldracklab/pydeface: V2.0.0.” 31 octobre 2019. http://doi.org/10.5281/zenodo.3524401.
18. Gymrek, Melissa, Amy L. McGuire, David Golan, Eran Halperin, and Yaniv Erlich. “Identifying Personal Genomes by Surname Inference.” *Science* 339, no. 6117 (18 janvier 2013): 321-324. https://doi.org/10.1126/science.1229566.
19. Hrynaszkiewicz, Iain, Melissa L. Norton, Andrew J. Vickers, and Douglas G. Altman. “Preparing Raw Clinical Data for Publication: Guidance for Journal Editors, Authors, and Peer Reviewers.” *BMJ* 340 (29 janvier 2010): c181. https://www.bmj.com/content/340/bmj.c181.
20. Information and Privacy Commissioner Ontario. *Deidentification Guidelines for Structured Data*. Information and Privacy Commissioner of Ontario. 8 juin 2016. https://www.ipc.on.ca/privacy-organizations/de-identification-centre/.
21. International Household Survey Network. “Anonymization Principles.” Document obtenu le 4 août 2020. https://ihsn.org/node/137.
22. International Household Survey Network. “Measuring the Disclosure Risk.” Document obtenu le 4 août 2020. https://ihsn.org/anonymization-risk-measure.
23. International Neuroimaging Data-Sharing Initiative (INDI). *Data Contribution Guide*. Document obtenu le 4 août 2020. http://fcon_1000.projects.nitrc.org/indi/indi_data_contribution_guide.pdf.
24. Kirby, Justin. “Submission and De-identification Overview.” The Cancer Imaging Archive (TCIA), University of Arkansas for Medical Sciences. 27 avril 2020. https://wiki.cancerimagingarchive.net/display/Public/Submission+and+De-identification+Overview.
25. Mannheimer, Sara, and Elizabeth Hull. “Sharing Selves: Developing an Ethical Framework for Curating Social Media Data.” *International Journal of Digital Curation* 12, no. 2 (18 avril 2018). https://doi.org/10.2218/ijdc.v12i2.518.
26. Medical Imaging & Technology Alliance, DICOM Standards Committee. “DICOM Part 15: Security and System Management Profiles.” In DICOM Standard. Arlington, VA: National Electrical Manufacturers Association. Document obtenu le 4 août 2020. https://www.dicomstandard.org/current/.

27. Moore, Stephen M., David R. Maffitt, Kirk E. Smith, Justin S. Kirby, Kenneth W. Clark, John B. Freymann, Bruce A. Vendt, Lawrence R. Tarbox, et Fred W. Prior. “De-identification of Medical Images with Retention of Scientific Research Value.” *RadioGraphics* 35, no. 3 (13 mai 2015). https://doi.org/10.1148/rg.2015140244.
28. National Human Genome Research Institute. “Privacy in Genomics.” 24 février 2020. https://www.genome.gov/about-genomics/policy-issues/Privacy.
29. North Carolina State University Libraries. “Social Media Archives Toolkit.” Document obtenu le 4 août 2020. https://www.lib.ncsu.edu/social-media-archives-toolkit.
30. Radiological Society of North America, International COVID-19 Open Radiology Database. “RSNA International Covid-19 Open Radiology Database (RICORD) De-identification Protocol.” Document obtenu le 10 août 2020. https://www.rsna.org/-/media/Files/RSNA/Covid-19/RICORD/RSNA-Covid-19-Deidentification-Protocol.pdf.
31. Radiological Society of North America, International COVID-19 Open Radiology Database. “RSNA COVID-19 DICOM Data Anonymizer.” Document obtenu le 10 août 2020. https://www.rsna.org/-/media/Files/RSNA/Covid-19/RICORD/RSNA-Anonymizer-Program-Instructions.pdf.
32. Radiological Society of North America, Medical Imaging Resource Community (MIRC). “Clinical trial processor (CTP).” Document obtenu le 4 août 2020. https://www.rsna.org/research/imaging-research-tools.
33. Ryerson University Research Ethics Board. “Guidelines for Research Involving Social Media.” Ryerson University. novembre 2017. https://www.ryerson.ca/content/dam/research/documents/ethics/guidelines-for-research-involving-social-media.pdf.
34. Thorogood, Adrian. “Canada: Will Privacy Rules Continue to Favour Open Science?” *Human Genetics* 137: 595–602 (16 juillet 2018). https://doi.org/10.1007/s00439-018-1922-z.
35. Twitter. “Developer Terms: More About Restricted Uses of the Twitter APIs.” Document obtenu le 4 août 2020. https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases.
36. UK Data Service. “Anonymisation: Qualitative Data.” Dernière modification le 30 juin 2020. https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative.aspx.
37. Zeffiro, Andrea et Jay Brodeur. “Social Media Research Data Ethics and Management.” L’atelier a été presenté le 5 avril 2018. Sherman Centre for Digital Scholarship. McMaster University. http://hdl.handle.net/11375/25327.