



Un appel à un cadre éthique lors de l'utilisation des données des médias sociaux pour des applications d'intelligence artificielle dans la recherche en santé publique

Jean-Philippe Gilbert^{1*}, Victoria Ng², Jingcheng Niu³, Erin E. Rees²

Résumé

Les progrès de l'intelligence artificielle (IA), plus précisément du sous-domaine de l'apprentissage automatique, et ses applications aux données Internet de source ouverte, telles que les médias sociaux, progressent plus rapidement que la gestion des questions éthiques liées à leur utilisation dans la société. Un cadre éthique aide les scientifiques et les décideurs politiques à considérer l'éthique dans leurs domaines de pratique, à légitimer leur travail et à protéger les membres du public générateur de données. Une question centrale pour faire avancer le cadre éthique est de savoir si les gazouillis, les publications Facebook et autres données de médias sociaux de source ouverte générées par le public représentent ou non de données humaines. Le présent document a pour but de mettre en évidence les problèmes éthiques auxquels le secteur de la santé publique sera ou est déjà confronté lors de l'utilisation des données des médias sociaux dans la pratique. Les problèmes comprennent le consentement éclairé, la confidentialité, l'anonymisation et la recherche d'un équilibre entre ces problèmes et les avantages de l'utilisation des données des médias sociaux pour le bien commun. Les cadres éthiques actuels doivent fournir des orientations pour permettre de résoudre les problèmes découlant de l'utilisation des données issues des médias sociaux dans le secteur de la santé publique. Des discussions dans ce domaine devraient avoir lieu, alors que l'application des données de source ouverte est encore relativement nouvelle, et suivre la cadence, car d'autres problèmes découlent de l'évolution technologique en cours.

Cette œuvre est mise à la disposition selon les termes de la licence internationale Creative Commons Attribution 4.0



Affiliations

¹ Université Laval, Québec, QC

² Agence de la santé publique du Canada, Ottawa, ON

³ University of Toronto, Toronto, ON

*Correspondance :
jean-philippe.gilbert.5@ulaval.ca

Citation proposée : Gilbert J-P, Ng V, Niu J, Rees EE. Un appel à un cadre éthique lors de l'utilisation des données des médias sociaux pour des applications d'intelligence artificielle dans la recherche en santé publique. Relevé des maladies transmissibles au Canada 2020;46(6):191–6. <https://doi.org/10.14745/ccdr.v46i06a03f>

Mots-clés : éthique, recherche éthique, médias sociaux, intelligence artificielle

Introduction

Les progrès technologiques rapides de l'intelligence artificielle (IA), et plus particulièrement du traitement du langage naturel (TLN) utilisant des techniques d'apprentissage automatique, permettent un accès et une utilisation faciles des données massives de source ouverte. Le TLN permet aux ordinateurs d'analyser des ensembles de données du discours en langage naturel (c'est-à-dire, un texte non structuré pour une analyse quantitative).

En santé publique, l'épidémiologie numérique est devenue un nouveau domaine qui se concentre sur l'utilisation de données du secteur de la santé non public telles que les données Internet de source ouverte (e.g. Google Trends, les médias) et les données des réseaux sociaux (e.g. Twitter et Facebook), alors

que l'épidémiologie traditionnelle utilise les données collectées à des fins de soins de santé, telles que la déclaration des maladies à déclaration obligatoire par les professionnels de la santé, pour contribuer aux données de surveillance des cas de maladie.

Les chercheurs et les décideurs reconnaissent le potentiel des données épidémiologiques numériques pour améliorer l'alerte précoce des menaces pour la santé publique (1–3). Odlum et Yoon (4) ont utilisé le TLN pour évaluer les données de Twitter et ont indiqué que les gazouillis liés à Ebola ont augmenté dans les jours précédant l'alerte officielle de l'épidémie d'Ebola de 2014 en Afrique. Yousefinaghani et coll. (5) ont montré que 75 % des notifications d'épidémie de grippe aviaire en temps réel étaient repérables sur Twitter; un tiers des notifications d'épidémie ont



été signalées sur Twitter plus tôt que les rapports officiels. Ces observations soutiennent l'utilisation de volumes de données Twitter pour prédire la survenue d'épidémies, de même que pour prévoir le nombre de cas prévus. Cela a également été démontré avec les données de Google Trends (1,6). En outre, le peaufinement des données des médias sociaux par une répartition en diverses catégories pertinentes à la maladie, en utilisant le TLN pour classer les gazouillis en types de symptômes (e.g. fièvre, vomissements), ou en concentrant l'analyse sur des termes de recherche spécifiques de Google Trends, contribue à augmenter la précision des prévisions d'occurrence d'épidémie et des estimations des prévisions.

La recherche qui utilise des données de participants humains nécessite une approbation éthique. Un processus d'examen mené par un organisme gouvernemental ou un comité universitaire indépendant des chercheurs permet d'évaluer si l'utilisation de ces données garantit la sécurité, la dignité et les droits des participants. Les chercheurs doivent démontrer au comité d'éthique de la recherche (CER) que leur étude minimise les préjudices aux participants et respecte leur autonomie, génère et maximise les avantages (e.g. pour la société, la science, les participants) et qu'ils agissent avec intégrité, équité et transparence pour tous les intervenants (e.g. les participants, les bénéficiaires de la recherche). Cependant, dans le cadre d'un examen systématique de l'utilisation de Twitter pour la recherche en santé, seulement 32 % des études ont obtenu l'approbation éthique (7).

Il s'agit d'un exemple de technologie évoluant plus rapidement que la politique, dans la mesure où la disponibilité de nouvelles sources de données, telles que celles des médias sociaux, a dépassé la nécessité d'évaluer l'éthique de leur utilisation. Cela a donné lieu à des études comportant des démarches éthiques douteuses, ce qui entache tous les domaines qui utilisent des données massives. Un exemple est l'étude « Tastes, Ties, and Time » en 2007, où les chercheurs ont publié un ensemble de données anonymisé d'un groupe d'étudiants universitaires et métadonnées contenant des renseignements sur l'ensemble de données; l'ensemble de données était identifiable à partir de métadonnées (8). De même, en 2012, des preuves de contagion émotionnelle en ligne ont été recherchées, sans consentement préalable, en manipulant le flux de nouvelles Facebook de milliers de personnes pour voir si cela modifiait les sentiments dans les publications de ces personnes (9).

Dans cet article, nous explorons les problèmes liés aux cadres éthiques traditionnels en relation avec la recherche fondée sur l'IA, en particulier dans le domaine de la santé publique et de l'épidémiologie numérique. Nous présentons ensuite des cadres éthiques qui permettent aux scientifiques et aux décideurs politiques d'utiliser les données des médias sociaux et leurs applications.

Éthique contemporaine

Dans la science contemporaine, les chercheurs ont besoin d'une approbation éthique pour pouvoir utiliser des données humaines. Ce critère est le principal problème de la recherche fondée sur les données massives. Il soulève une question apparemment simple : Un message ou un gazouillis représente-t-il des données humaines ou des données textuelles? (10). Plusieurs questions et points de vue découlent de cette question et donnent lieu à un débat nécessaire étant donné que la popularité de l'utilisation des données issues des médias sociaux augmente dans plusieurs domaines scientifiques, y compris l'épidémiologie numérique.

Actuellement, les études qui utilisent les données des médias sociaux sont généralement perçues comme n'entrant pas dans le cadre de l'évaluation des comités d'éthique, car ces données ne sont généralement pas considérées comme des données humaines (11,12). De nombreux chercheurs, décideurs et praticiens supposent qu'ils peuvent utiliser des données de source ouverte, par exemple des gazouillis, des messages publics sur Facebook, des photos publiques sur Instagram et des questions Google Trends, dont l'accès n'exige pas de mot de passe (8,13). Cependant, pour de nombreux utilisateurs des médias sociaux, publier publiquement ne revient pas à donner leur consentement pour que le message soit utilisé à des fins de recherche (8,11,12). Cette question n'est pas couverte par les mécanismes d'examen éthique existants (14).

De plus, la facilité d'accès aux données des réseaux sociaux (en l'absence de règles éthiques et en utilisant la capture rapide de données à l'aide de l'IA) signifie que le nombre de points de données est souvent beaucoup plus important que dans les ensembles de données épidémiologiques traditionnels. Par conséquent, les décisions concernant l'utilisation et les répercussions des données des médias sociaux peuvent potentiellement affecter plus de personnes (14). Par exemple le nombre de personnes réidentifiées par accident ou par malveillance dans une base de données Twitter n'est limité que par les ressources utilisées pour compiler et analyser la base de données, ce qui est bien inférieur aux systèmes de surveillance traditionnels (14).

Consentement éclairé

Le consentement éclairé, tel qu'il existe dans l'éthique contemporaine, est mal adapté aux données des médias sociaux. Premièrement, il est presque impossible d'obtenir le consentement éclairé des personnes dont les données contribuent à l'épidémiologie numérique, car les ressources sont souvent insuffisantes pour contacter un nombre aussi élevé de personnes, lesquelles peuvent vivre n'importe où (15).

Deuxièmement, pour obtenir un consentement éclairé, les scientifiques doivent confirmer l'identité des utilisateurs des médias sociaux (16). Il n'y a aucun moyen de s'assurer que la personne qui se cache derrière le profil de médias sociaux est bien celle qu'elle prétend être ou de confirmer si le message sur



les réseaux sociaux n'a pas été généré par un agent numérique (c.-à-d., un « robot » responsable de messages sur les réseaux sociaux générés par ordinateur). En raison de cette complication, certains chercheurs considèrent que le consentement aux conditions et modalités de service d'une plateforme de médias sociaux, que les utilisateurs doivent donner pour utiliser la plateforme, est un substitut au consentement éclairé (16). Cependant, les utilisateurs ne lisent souvent pas les conditions et modalités de service ou ne les comprennent pas bien (17–19); celles-ci ne stipulent pas non plus les conditions dans lesquelles les données seront utilisées pour la recherche, ce qui remet en question la légitimité et l'intégrité de l'utilisation des conditions et modalités de service comme substitut d'un consentement éclairé. De nombreux « participants » à l'épidémiologie numérique ne savent pas que leurs données ont été collectées ou utilisées (20).

Problèmes de confidentialité et d'anonymisation

Nous dépendons de plus en plus de la technologie pour structurer et analyser les données qui prolifèrent dans nos sociétés numériques. L'exploration de données aide les chercheurs à trouver des patrons de données complexes et peu intuitifs. Cependant, les méthodes d'exploration de données peuvent également révéler des renseignements confidentiels à partir de données de médias sociaux apparemment inoffensives, par exemple des affiliations politiques (12,21). De plus, Wang et coll. (22) ont déclaré être en mesure de déterminer l'orientation sexuelle des personnes en traitant des photos de personnes tirées d'un site de rencontres.

Un ensemble de données anonymisé est l'exigence minimale pour protéger l'identité des sujets en sciences sociales (23) ou en épidémiologie traditionnelle (20). Selon la Règle commune, également connue sous le nom de « 45 CFR 46 Subpart A », le principal règlement pour la recherche humaine du Department of Health and Human Services des États-Unis (24), 17 identificateurs doivent être supprimés pour considérer un ensemble de données anonymisé. Il s'agit notamment du nom, du lieu de résidence, de toutes les dates à l'exception de l'année et des identifiants biométriques (25). Les Instituts de recherche en santé du Canada (IRSC), le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et le Conseil de recherches en sciences humaines du Canada (CRSH) précisent des identifiants similaires (26). Cependant, la suppression des 17 identificateurs que préconise la Règle commune ne suffit souvent pas à garantir l'anonymisation d'un ensemble de données. En effet, les données des médias sociaux sont très complexes (c.-à-d. qu'elles ont une grande dimensionnalité). De nombreux attributs non traditionnels peuvent permettre l'identification, comme la ré-identification par l'évaluation de la structure des réseaux sociaux (c.-à-d., les connexions humaines) à partir de multiples plateformes de médias sociaux (15,27). Les progrès liés aux algorithmes d'IA et à la puissance de calcul pour extraire des informations et évaluer les modèles signifient qu'il n'est plus

possible d'avoir des bases de données anonymes (28,29). De nombreux exemples dans la littérature scientifique démontrent ce problème en réidentifiant un ensemble de données anonymisé, puis publié (12,21).

Le bien commun

Le bien commun prend ses racines dans la vision utilitariste de l'éthique. Dans cette vision, le bien commun que la recherche peut générer est considéré par rapport au préjudice potentiel pour les personnes. Un certain niveau de préjudice peut être toléré s'il en résulte une « moralité positive ». Dans le contexte des médias sociaux, le préjudice est principalement une atteinte à la vie privée (30). Les gens sont plus disposés à sacrifier leur droit à la vie privée s'ils estiment que l'utilisation de leurs données bénéficiera au bien commun (31,32). Pour les utilisateurs de médias sociaux les plus enthousiastes dans l'étude de Mikal et coll. (31), [traduction] « c'est cool quand c'est des choses [...] comme la grippe, parce que c'est ainsi que [les décideurs en santé publique] savent comment acheminer les vaccins. » De même, pour les utilisateurs des médias sociaux de l'étude de Golder et coll. (32), cela [traduction] « pourrait donner une voix aux patients et à d'autres groupes, permettre de découvrir les vrais problèmes actuels et d'améliorer les soins aux patients ». Les facteurs qui influencent la disposition des personnes à partager leurs données pour le bien commun comprennent le type de recherche et les affiliations des chercheurs (c.-à-d., l'université, l'entreprise, le gouvernement) (32–34).

En fin de compte, bien que la majorité des personnes soient d'accord avec le concept de bien commun, il n'y a pas de seuil convenu pour lequel une atteinte à la vie privée peut et doit être tolérée pour la recherche en santé publique.

Nouveaux cadres éthiques

De nouveaux cadres qui répondent aux nouveaux défis éthiques concernant l'utilisation de l'IA pour la recherche ont été proposés par l'Association of Internet Researchers (AoIR) (35) et Zook et coll. (36) (**tableau 1**).

Suivre un cadre peut aider à légitimer la recherche pour la population (37). Étant donné que le cadre de l'AoIR (35) est accepté dans la littérature scientifique, l'Association étant l'une des organisations les plus citées en termes d'éthique et de données massives, les scientifiques pourraient vouloir utiliser ce cadre plutôt que le cadre moins connu de Zook et coll. Cependant, le cadre de Zook et coll. (36) est moins restrictif et plus facile à suivre.

De nombreux points de ces lignes directrices sont déjà des considérations que les scientifiques de la santé publique doivent aborder (e.g. la protection de la population vulnérable, les préjudices potentiels de l'étude, le processus d'anonymisation). Les scientifiques de la santé publique utilisent déjà fréquemment des données hautement confidentielles. La principale différence entre les données des médias sociaux et les données



Tableau 1 : Cadres éthiques proposés

Auteurs	Lignes directrices
AoIR (35)	<ol style="list-style-type: none"> 1. Protéger les populations vulnérables 2. Évaluer les dommages potentiels des études de recherche au cas par cas 3. Considérer les données d'humains comme étant humaines 4. Équilibrer les droits de toutes les parties concernées (c.-à-d. le droit à la vie privée du sujet et le droit de faire des recherches pour le scientifique) 5. La variabilité temporelle des considérations éthiques doit être résolue lorsqu'elle se produit 6. Discuter des problèmes éthiques avec des professionnels qualifiés lorsqu'ils surviennent
Zook et coll. (36)	<ol style="list-style-type: none"> 1. Reconnaissez que les données sont des personnes et peuvent causer du tort 2. Reconnaissez que la confidentialité est plus qu'une valeur binaire 3. Protégez contre la réidentification de vos données 4. Pratiquez le partage éthique des données 5. Tenez compte des forces et des limites de vos données; « grand » ne signifie pas automatiquement « mieux » 6. Discutez des choix éthiques difficiles 7. Développez un code de conduite pour votre organisation, communauté de recherche ou industrie 8. Concevez vos données et systèmes de façon à assurer leur vérifiabilité 9. Interagissez avec les conséquences plus larges des données et des pratiques d'analyse 10. Sachez quand enfreindre ces règles

Abréviation : AoIR, Association of Internet Researchers

traditionnelles est la façon dont les données sont accessibles; l'intention initiale pour laquelle les données sont produites; et la capacité limitée des utilisateurs de médias sociaux à fournir un consentement éclairé. Les données représentent toujours des humains, et peuvent entraîner des conséquences involontaires telles que l'identification de la personne qui est à l'origine du contenu de médias sociaux. Les scientifiques de la santé publique ont l'obligation de protéger les personnes à l'origine de leurs données tout en équilibrant cela avec le bien commun; il est extrêmement difficile de se mettre d'accord sur cette décision subjective.

Discussion

À mesure que la technologie progresse rapidement et que davantage de recherches sont effectuées sur l'IA et les données des médias sociaux, un cadre éthique établi est essentiel pour empêcher une mauvaise utilisation des données des médias sociaux dans les applications de santé publique. Les chercheurs en santé publique, en informatique et en éthique doivent se réunir pour développer un cadre qui aidera les scientifiques à mener une recherche responsable. De façon générale, les cadres

existants ont été développés pour être utilisés dans tous les domaines scientifiques. Les décisions liées à la santé publique peuvent toutefois avoir des répercussions importantes sur la population, allant jusqu'à restreindre la liberté de circulation des personnes en cas de maladie hautement infectieuse, par exemple (20).

Le CER est un élément important du processus visant à garantir que la recherche respecte le cadre éthique. Ce qui est associé à l'utilisation de données de médias sociaux de source ouverte est le fait que les gens ne savent pas que leurs données sont utilisées ou n'ont pas la possibilité de consentir à leur utilisation. Ainsi, le CER fournit les moyens de défendre la sécurité, la dignité et les droits des participants comme stipulé dans le cadre éthique.

Le CER et le cadre éthique sont également nécessaires pour remédier aux limites des données des médias sociaux. De nombreuses plateformes de médias sociaux sont disponibles, et la prédominance de leur utilisation peut varier selon l'emplacement. Par exemple, Twitter et Facebook sont largement utilisés dans les pays occidentaux, mais interdits en République populaire de Chine; le gouvernement chinois autorise l'utilisation de Sina Weibo et de WeChat comme équivalents respectifs de Twitter et Facebook. De plus, les données démographiques d'utilisation peuvent varier selon les applications. Les générations plus âgées ont tendance à utiliser Twitter et Facebook, tandis que les générations plus jeunes ont tendance à utiliser Snapchat, Instagram et TikTok. C'est ce que l'on appelle la fossé numérique (38). Certains profils peuvent être sous-représentés (e.g. les enfants et les personnes âgées), selon les plateformes de médias sociaux.

Conclusion

Les questions éthiques liées à l'utilisation des données des médias sociaux pour les applications d'IA dans la recherche en santé publique se concentrent sur la question de savoir si ces données sont considérées comme humaines. Les cadres éthiques actuels sont inadéquats pour la santé publique. Pour éviter toute utilisation abusive des données des médias sociaux, nous soutenons que le fait de considérer les données des médias sociaux comme humaines faciliterait un processus de CER qui garantit la sécurité, la dignité et les droits des fournisseurs de données des médias sociaux. Nous sommes d'avis en outre qu'il faut accorder davantage d'attention à l'équilibre entre le bien commun et l'atteinte à la vie privée. La collaboration entre les chercheurs en éthique et les épidémiologistes numériques est nécessaire pour développer des comités d'éthique, des directives et pour superviser la recherche dans le domaine.

Déclaration des auteurs

J. P. G. — Rédaction – ébauche originale, administration de projet, conceptualisation



- V. N. — Rédaction – révision et édition, conceptualisation, supervision
N. J. — Rédaction – révision et édition, conceptualisation, supervision
E. E. R. — Rédaction – révision et édition, conceptualisation, supervision

Conflit d'intérêts

Aucun.

Remerciements

Les auteurs tiennent également à remercier S. de Montigny, N. Barrette et P. Gachon pour leurs commentaires.

Financement

Ce travail est soutenu par l'Agence de la santé publique du Canada.

Références

1. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;457(7232):1012–4. [DOI](#) [PubMed](#)
2. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneau P; Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. *PLoS One* 2013;8(3):e57252. [DOI](#) [PubMed](#)
3. Salathé M. Digital epidemiology: what is it, and where is it going? *Life Sci Soc Policy* 2018;14(1):1–5. [DOI](#) [PubMed](#)
4. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? *Am J Infect Control* 2015;43(6):563–71. [DOI](#) [PubMed](#)
5. Yousefinaghani S, Dara R, Poljak Z, Bernardo TM, Sharif S. The assessment of Twitter's potential for outbreak detection: avian influenza case study. *Sci Rep* 2019;9(1):18147. [DOI](#) [PubMed](#)
6. Rangarajan P, Mody SK, Marathe M. Forecasting dengue and influenza incidences using a sparse representation of Google trends, electronic health records, and time series data. *PLOS Comput Biol* 2019;15(11):e1007518. [DOI](#) [PubMed](#)
7. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. *Am J Public Health* 2017;107(1):e1–8. [DOI](#) [PubMed](#)
8. Zimmer M. "But the data is already public": on the ethics of research in Facebook. *Ethics Inf Technol* 2010;12(4):313–25. [DOI](#)
9. Jouhki J, Lauk E, Penttinen M, Sormanen N, Uskali T. Facebook's emotional contagion experiment as a challenge to research ethics. *Media Commun* 2016;4(4):75–85. [DOI](#)
10. Buchanan E. Considering the ethics of big data research: A case of Twitter and ISIS/ISIL. *PLoS One* 2017;12(12):e0187155. [DOI](#) [PubMed](#)
11. Fiesler C, Proferes N. "Participant" perceptions of Twitter research ethics. *Social Media Soc* 2018;4(1):14. [DOI](#)
12. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. *PLoS One* 2018;13(10):e0204937. [DOI](#) [PubMed](#)
13. Gehner M, Oughton D. Ethical challenges in social media engagement and research: considerations for code of engagement practices. *J Radiol Prot* 2016;36(2):S187–92. [DOI](#) [PubMed](#)
14. Lee EC, Asher JM, Goldlust S, Kraemer JD, Lawson AB, Bansal S. Mind the scales: harnessing spatial big data for infectious disease surveillance and inference. *J Infect Dis* 2016;214 suppl_4:S409–13. [DOI](#) [PubMed](#)
15. Lipworth W, Mason PH, Kerridge I, Ioannidis JP. Ethics and epistemology in big data research. *J Bioeth Inq* 2017;14(4):489–500. [DOI](#) [PubMed](#)
16. Webb H, Jirotnka M, Stahl BC, Housley W, Edwards A, Williams ML, Procter R, Rana OF, Burnap P. The Ethical Challenges of Publishing Twitter Data for Research Dissemination. Proceedings of the 2017 ACM Web Science Conference. 2017, New York: Assoc Computing Machinery; 339–48.
17. Fiske ST, Hauser RM. Protecting human research participants in the age of big data. *Proc Natl Acad Sci USA* 2014;111(38):13675–6. [DOI](#) [PubMed](#)
18. Saqr M. Big data and the emerging ethical challenges. *Int J Health Sci (Qassim)* 2017;11(4):1–2. [PubMed](#)
19. Hesse A, Glenna L, Hinrichs C, Chiles R, Sachs C. Qualitative research ethics in the big data era. *Am Behav Sci* 2019;63(5):560–83. [DOI](#)
20. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. *PLOS Comput Biol* 2015;11(2):e1003904. [DOI](#) [PubMed](#)
21. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. *Annu Rev Public Health* 2018;39:95–112. [DOI](#) [PubMed](#)



22. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *J Pers Soc Psychol* 2018;114(2):246–57. [DOI](#) [PubMed](#)
23. Phillips A, Barry P, Shabani M. Research ethics review for the use of anonymized samples and data: A systematic review of normative documents. *Account Res* 2017;24(8):483–96. [DOI](#) [PubMed](#)
24. Shade J, Coon H, Docherty AR. Ethical implications of using biobanks and population databases for genetic suicide research. *Am J Med Genet B Neuropsychiatr Genet* 2019;180(8):601–8. [DOI](#) [PubMed](#)
25. Rothstein MA. Is deidentification sufficient to protect health privacy in research? *Am J Bioeth* 2010 Sep;10(9):3–11. [DOI](#) [PubMed](#)
26. Instituts de recherche en santé du Canada. Conseil de recherches en sciences humaines Conseil de recherches en sciences naturelles et en génie du Canada Instituts de recherche en santé du Canada, Énoncé de politique des trois conseils Éthique de la recherche avec des êtres humains – EPTC 2 (2018). Ottawa (ON): Gouvernement du Canada; 2018 déc. <https://ethics.gc.ca/fra/documents/tcps2-2018-fr-interactive-final.pdf>
27. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. *Swiss Med Wkly* 2018;148:w14571. [DOI](#) [PubMed](#)
28. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. *PLoS One* 2011;6(12):e28071. [DOI](#) [PubMed](#)
29. de Montjoye YA, Radaelli L, Singh VK, Pentland AS. Identity and privacy. Unique in the shopping mall: on the reidentifiability of credit card metadata. *Science* 2015;347(6221):536–9. [DOI](#) [PubMed](#)
30. Herron M, Sinclair M, Kernohan WG, Stockdale J. Ethical issues in undertaking internet research of user-generated content: a review of the literature. *Evid Based Midwifery* 2011;9(1):9–15.
31. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. *BMC Med Ethics* 2016;17(1):22. [DOI](#) [PubMed](#)
32. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. *J Med Internet Res* 2017;19(6):e195. [DOI](#) [PubMed](#)
33. Beninger K, Fry A, Jago N, Lepps H, Nass L, Silvester H. Research using social media; users' views. London (UK): NatCen Social Research; 2014. https://www.researchgate.net/publication/261551701_Research_using_Social_Media_Users'_Views
34. Golder S, Scantlebury A, Christmas H. Understanding public attitudes toward researchers using social media for detecting and monitoring adverse events data: multi methods study. *J Med Internet Res* 2019;21(8):e7081. [DOI](#) [PubMed](#)
35. Markham A, Buchanan E. Ethical decision-making and internet research: recommendations from the AoIR Ethics Working Committee (Version 2.0). Association of Internet Researchers; 2012.
36. Zook M, Baracas S, Boyd D, Crawford K, Keller E, Gangadharan SP, Goodman A, Hollander R, Koenig BA, Metcalf J, Narayanan A, Nelson A, Pasquale F. Ten simple rules for responsible big data research. *PLOS Comput Biol* 2017;13(3):e1005399. [DOI](#) [PubMed](#)
37. Shilton K, Sayles S. "We aren't all going to be on the same page about ethics:" Ethical practices and challenges in research on digital and social media. In: Bui TX, Sprague RH, editors. Proceedings of the 49th Annual Hawaii International Conference on System Sciences. 2016, IEEE Computer Soc: Los Alamitos. 1909–18.
38. Hargittai E. Digital Na(t)ives? Variation in internet skills and uses among members of the "Net Generation". *Sociol Inq* 2010;80(1):92–113. [DOI](#)