Rapport

Symposium « Anonymisation des données »

Henry Laville

Avril 2024

Crédits

Ce rapport a été rédigé par **Henry Laville**, doctorant en droit, à l'Université de Montréal (UdeM) et l'Université Sorbonne Paris Nord, sous la supervision de **Vincent Gautrais**, professeur titulaire à l'UdeM et titulaire de la Chaire L.R. Wilson, et coresponsable de l'axe Droit, cyberjustice et cybersécurité de l'Obvia, ainsi qu'**Antoine Legrain**, professeur à Polytechnique Montréal et membre chercheur associé à l'Obvia.

Les organisateurs de l'évènement souhaitent chaleureusement remercier les conférenciers pour leur présentation, Dr. Khalel El Emam, professeur à l'Université d'Ottawa et titulaire de la Chaire de recherche du Canada (niveau 1) en IA médicale, Ulrich Aïvodji, professeur à l'École de technologie supérieur (ETS), Vincent Dionne, responsable du Service Données massives de la Société de Transport de Laval (STL), Stéphanie Pham-Dang, responsable du Service de gestion des données de recherche de l'UdeM, Me Antoine Guilmain, associé et codirecteur du groupe de pratique national Cybersécurité et protection des données chez Gowling WLG, Henry Laville, candidat au doctorat à l'Université de Montréal et l'Université Sorbonne Paris Nord, Alice Friser, professeure au département des sciences administratives de l'Université du Québec en Outaouais (UQO), Ledy Rivas Zannou, professeur au département de droit de l'UQO et, Maya Cachecho, professeure adjointe à l'UdeM.

Nous remercions tout particulièrement les partenaires de l'évènement, l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (Obvia), la Fondation Barreau du Québec, la Chaire L.R. Wilson, le Centre de recherche en droit public (CRDP) ainsi que le Groupe d'études et de recherche en analyse des décisions (GERAD).

Enfin, nous remercions les organisateurs de l'évènement, les professeurs Vincent Gautrais, Antoine Legrain, Ledy Rivas Zannou, ainsi qu'Henry Laville.

Produit avec le soutien financier des Fonds de recherche du Québec

Fonds de recherche Québec

ISBN : 978-2-925138-51-8
DOI : 10.61737/XAFK3054

Table des matières

Sommaire

Actualité de l'anonymisation des données – Le 22 septembre 2022 dernier sont entrées en vigueur les premières dispositions de la Loi 25 sur la protection des renseignements personnels au Québec. Les obligations relatives à l'anonymisation des renseignements personnels sont entrées en vigueur l'année suivante, le 22 septembre 2023. Ces obligations ont soulevé (et soulèvent toujours) de nombreuses questions parmi les destinataires des normes, mais aussi au sein des communautés de pratiques et du monde académique.

Un évènement interdisciplinaire – L'ambition du Symposium « Anonymisation des données[1] », organisé le 29 avril 2024 à l'Université de Montréal, en partenariat avec la Chaire L.R. Wilson, le CRDP, l'OBVIA, la Fondation du Barreau du Québec et le GERAD, était de réunir dans une perspective interdisciplinaire des professionnels et des chercheurs ayant une expertise technique ou juridique sur l'anonymisation des données. Leurs présentations ainsi que leurs échanges ont donné lieu à un panorama complet sur l'état des techniques d'anonymisation, des besoins et des défis rencontrés par les acteurs du secteur en sus des enjeux juridiques et politiques de la gouvernance des données.

L'anonymat : un idéal – Ce rapport présente les lignes de force qui se sont dégagées à l'occasion de cette journée de discussion, dont quatre qui méritent d'être mentionnées dès à présent. La première est que le risque 0 de réidentification n'existe pas. Pr. Khalel El Emam a souligné le caractère non-binaire de l'anonymat en mettant en évidence qu'il s'agissait d'une question de méthodes et de seuils déterminés par les meilleures pratiques du secteur. Pr. Ulrich Aïvodji a poursuivi en présentant les différentes méthodes d'anonymisation des données tout en soulignant les limites de chacune d'entre-elles.

L'anonymisation : une question de contexte – À la suite de ces rappels théoriques, les intervenants des secteurs économique et académique ont insisté sur caractère hautement contextuel de l'anonymisation des données au sein des organisations. En parlant de sa pratique dans une société de transport du public, Vincent Dionne a souligné les défis que son organisation rencontrait dans l'anonymisation des données de mobilité collectées pour la conception des trajets en raison de leur très grande diversité. Il a insisté sur l'équation anonymisation/utilité des données à laquelle est confrontée son organisation pour continuer à remplir ses missions. À chacune des données de recherche collectées, Stéphanie Pham-Dang a présenté les différentes solutions mises en place par une université pour en garantir la protection.

Gouvernance actuelle de l'anonymisation : entre incertitudes et solutions – Les dispositions légales relatives à l'anonymisation des données soulèvent de nombreuses questions tant dans leur interprétation que dans leur application. Après être revenu sur les dispositions législatives et règlementaires sur l'anonymisation dans la Loi 25, Me Antoine Guilmain en a souligné les spécificités ainsi que les limites. Henry Laville a poursuivi en présentant le rôle et les limites des Normes nationales du Canada comme instrument pour anonymiser les données.

Gouvernance future de l'anonymisation : pistes d'amélioration – Enfin, les intervenants ont présenté des solutions pour améliorer la gouvernance de l'anonymisation des données. Pr. Alice Friser a présenté l'instrument de la participation publique pour améliorer l'acceptabilité sociale d'une innovation technologique. Pr. Ledy Rivas Zannou a, quant à lui, souligné l'importance de la participation dans l'élaboration des normes relatives à l'anonymisation des données. Enfin, Pr. Maya Cachecho a insisté sur les besoins des organisations Fintechs et de leurs clients en matière de gouvernance.

1 Page de l'évènement, en ligne <https://www.chairelrwilson.ca/actualites/2024/symposium-anonymisation-des-donnees/> (consulté le 12 juin 2024)

1. L’anonymisation des données : contexte, théorie et pratiques

1.1 Définition et fonction de l'anonymat

Au-delà de la binarité - L'anonymisation des données est un concept qui ne se laisse pas enfermer dans une représentation binaire, où une donnée serait ou ne serait pas anonymisée. Dans sa présentation intitulée « The Definition of Data Anonymity[2] », Khalel El Emam pose une définition opérationnelle de l'anonymisation en revenant sur quelques méthodes ayant fait leurs preuves ces dernières années.

Spectre de l'identifiabilité des données - Une donnée anonymisée est une donnée non identifiable. Or, l'identifiabilité s'envisage comme un spectre au sein duquel il est possible de mesurer un seuil d'identifiabilité, soit un niveau de risque de réidentification. Plusieurs méthodes et seuils établis fixent ce qu'est un risque acceptable. Le risque zéro de réidentification n'existant pas, une donnée ne peut jamais être totalement anonymisée.

Risques de réidentification - Le risque de réidentification s'apprécie comme une combinaison de risques liés à la fois aux données et au contexte. Le risque lié aux données peut être atténué en utilisant des techniques telles que le masquage des données ou encore l'anonymisation par k-anonymat. Le risque contextuel peut être contrôlé en mettant en œuvre des contrôles de sécurité, des contrôles de confidentialité et des contrôles contractuels.

Appréciation des risques en fonction de la nature des données - Le risque de réidentification ne s'apprécie pas de la même manière s'il s'agit de données publiques ou non. En présence de données privées, le risque contextuel demeure, et il ainsi possible d'agir sur les contrôles de sécurité, les contrôles de confidentialité et les contrôles contractuels pour réduire le risque de réidentification. En revanche, dans le cas de données publiques, il n'y a pas de risque de contextuel, et le seul moyen pour réduire le risque de réidentification est d'agir directement sur les données.

Méthodes, seuils et types de variables - Les méthodes et les seuils d'évaluation de ces risques sont fixés dans de nombreux standards à travers le monde[3]. Les méthodes varient en fonction du type de variables sur lesquels on agit. Les variables des jeux de données sont classifiées en trois types :

- Les identifiants directs (le nom, le prénom, l'adresse mail, etc.);
- Les quasi-identifiants (localisation, etc.), et;
- Les variables sensibles.

Données anonymisées vs. pseudonymisées : une question de seuil - L'anonymisation des identifiants directs diminue le risque de réidentification, mais les données sont toujours considérées comme des données personnelles. On parle alors de données pseudonymisées. Les attaques de réidentification opèrent, en effet, le plus souvent sur les quasi-identifiants qui permettent alors de réidentifier la personne. Afin d'obtenir des données dé-identifiées/anonymisées, il faut encore réduire le risque de réidentification. La question du seuil à partir duquel une donnée pseudonymisée devient une donnée dé-identifiée/anonymisée se pose alors. Les normes ISO proposent des seuils[4].

2 El Emam, Khaled, « The Definition of Data Anonymity », 2024, en ligne : pour le support <https://www.chairelrwilson.ca/files/sites/36/2024/04/ElEmamK.pdf> (consulté le 12 juin 2024) et pour la présentation <https://www.youtube.com/watch?v=GqbS_6pSXVc> (consulté le 12 juin 2024

3 El Emam, Khaled, « The Definition of Data Anonymity », 2024, 4, en ligne <https://www.chairelrwilson.ca/files/sites/36/2024/04/ElEmamK.pdf> (consulté le 12 juin 2024)

4 El Emam, Khaled, « The Definition of Data Anonymity », 2024, 8, en ligne <https://www.chairelrwilson.ca/files/sites/36/2024/04/ElEmamK.pdf> (consulté le 12 juin 2024)

Donnée anonymisée : une définition idéale ? - Pour Khalel El Emam, une définition idéale de l'anonymat devrait inclure certains éléments. L'information ne devrait pas être utilisée pour identifier un individu et le processus d'anonymisation devrait être effectué par une personne possédant les connaissances et l'expertise appropriées. Le processus d'anonymisation devrait aussi suivre des principes scientifiques et statistiques généralement acceptés. Le risque de réidentification devrait être très faible. Par ailleurs, le risque devrait être évalué du point de vue du destinataire. Enfin, les méthodes et les résultats du processus d'anonymisation devraient être documentés.

1.2 Méthodes d'anonymisation et application aux données de mobilité

Anonymisation : une question de contexte - La conférence intitulée « Publication de données préservant la vie privée – Cas des données de mobilité[5] » met en évidence le caractère profondément contextuel de la question de l'anonymisation des données à partir de l'exemple des données de mobilité. Après avoir abordé dans une perspective technique et critique l'environnement et les méthodes d'anonymisation des données, Ulrich Aïvodji présente les enjeux spécifiques entourant l'anonymisation des données de mobilité.

Se protéger des adversaires - Les méthodes d'anonymisation des données permettent de se protéger d'adversaires souhaitant accéder à des *enregistrements* de données. Un adversaire est défini comme : « toute entité qui cherche à récupérer les données personnelles d'une personne concernée, sans son consentement explicite, pour établir un profil ou déduire ses données privées[6] ».

Enregistrement de données et attributs - Un enregistrement est composé de données dont il est possible de distinguer les plus souvent trois types d'attributs. Les identifiants explicites (nom, prénom, NAS, enregistrements biométriques, etc.) permettent d'identifier directement les individus, alors que les quasi-identifiants (code postal, date de naissance, sexe, etc.) combinés entre eux permettent potentiellement de réidentifier un sujet. Les attributs sensibles (maladie, salaire, religion, etc.) et les attributs non sensibles constituent la dernière catégorie.

Attaques de réidentification - Un jeu de données qui a été anonymisé peut subir plusieurs attaques visant à réidentifier les données. Quatre de ces attaques sont présentées : le couplage d'enregistrement, le couplage d'attribut, le couplage de table ainsi que l'attaque probabiliste.

Méthodes d'anonymisation - Différentes méthodes d'anonymisation permettent de lutter contre ces attaques. L'agrégation permet de combiner plusieurs enregistrements en un seul afin de rendre plus difficile l'inférence d'informations spécifiques à un individu. La pseudonymisation remplace les identifiants par des pseudonymes. Le k-anonymat est une méthode qui protège contre le couplage d'enregistrement en généralisant ou en supprimant un certain nombre de quasi-identifiants afin que pour chacun de ces derniers un nombre *k* d'enregistrements partage ces mêmes identifiants. La *l*-diversité permet de réduire les risques de réidentification dans ces cas en agissant sur la distribution des valeurs au sein des valeurs sensibles dans l'enregistrement des données.

5 Aïvodji, Ulrich, « Publication de données préservant la vie privée – Cas des données de mobilité », 2024, en ligne : pour le support <https://www.chairelrwilson.ca/files/sites/36/2024/04/AivodjiU.pdf> (consulté le 12 juin 2024) et pour la présentation <https://www.youtube.com/watch?v=zd8lRnyzHLk> (consulté le 12 juin 2024)

6 Aïvodji, Ulrich, « Publication de données préservant la vie privée – Cas des données de mobilité », 2024, 4, en ligne : pour le support <https://www.chairelrwilson.ca/files/sites/36/2024/04/AivodjiU.pdf> (consulté le 12 juin 2024)

Limites des méthodes - Or, ces méthodes subissent certaines limites. L'agrégation ne permet pas de se protéger de la réidentification d'un sujet à partir des informations auxiliaires dont dispose l'adversaire. La combinaison de deux jeux de données anonymisés peut permettre à un adversaire de réidentifier un sujet malgré les méthodes de pseudonymisation mises en place. Enfin, le *k*-anonymat ne permet pas de se protéger contre le couplage d'enregistrement, tandis que la *l*-diversité ne protège pas le jeu de données contre le couplage d'attribut.

Paradigme de la confidentialité différentielle - Aussi, les besoins de protection contre les attaques de réidentification et de trouver un compromis entre vie privée/utilité des données a mené la création d'un nouveau paradigme où les jeux de données ne sont plus publiés à l'exception des résultats des opérations réalisées sur ces jeux de données. Afin de s'assurer qu'il ne soit pas possible de réidentifier un individu à partir des résultats publiés, l'outil théorique de la confidentialité différentielle a été créé.

Attributs des données de mobilité - Les données de mobilité doivent être protégées, car elles ont un fort pouvoir d'inférence et elles sont collectées en très grande quantité. Un jeu de données de mobilité contient généralement trois catégories d'attributs : des identifiants (nom, prénom, numéro de téléphone, etc.), des coordonnés spatio-temporels (coordonnées GPS, heure et date) ainsi que des attributs additionnels (sexe, âge, etc.).

Menaces pesant sur les enregistrements de données de mobilité - Plusieurs menaces pèsent sur un jeu de données de mobilité. Un adversaire peut ainsi identifier les points d'intérêt, prédire les déplacements, apprendre la sémantique des localisations et des mouvements, dé-anonymiser des données géolocalisées, chaîner un individu dans différentes bases de données, reconstruire un réseau social d'un sujet ainsi que prédire des attributs démographiques.

Protéger les données de mobilité - Les données de mobilité peuvent être protégées du risque de réidentification par différentes méthodes d'anonymisation. L'agrégation spatio-temporelle permet de réaliser une généralisation du couple localisation et temps. On peut aussi intervenir sur la manière dont les échantillons sont conçus. La méthode de « Mix-Zones » permet d'éviter de chaîner le trajet d'un individu. Le *k*-anonymat géographique permet d'étendre la méthode du *k*-anonymat aux données spatio-temporelles en veillant à ce que pour chaque unité de temps, un individu soit dans une zone partagée par au moins *k*-1 autres individus. La confidentielle différentielle est aussi une approche lorsqu'est publié que certains résultats (points d'intérêt, etc.). Enfin, la création de trajets synthétiques en générant des représentations au sein desquels du bruit sera introduit est une méthode envisagée.

1.3 L'anonymisation des données massives dans une société de transport : défis et solutions

Anonymisation : un équilibre coût/utilité - Les enjeux soulevés par l'anonymisation des données de mobilité au sein d'une société de transport explicitent la problématique que rencontrent les organisations lorsqu'elles doivent choisir un point d'équilibre entre coût des mesures de protection des données et conservation de l'utilité des données pour les missions de l'organisation. Dans sa présentation intitulée « Données de mobilité et loi 25 : Les enjeux pour la planification des transports[7] », Vincent Dionne brosse un aperçu de l'utilité des données de mobilité pour la Société de Transport de Laval avant d'aborder le défi de l'anonymisation de celles-ci.

Collecte des données de mobilité - La Société de Transport de Laval est un exploitant d'autobus, qui contiennent de nombreux capteurs d'informations (compteurs de passager, télémétrie, compteurs par caméra; priorité aux feux; système d'aide à l'exploitation et information voyageur, boîte de perception, paiement par carte de crédit, positions GPS et Validation par carte à puce) qui récoltent de nombreuses données. Ces données sont utilisées pour offrir un service optimal et efficient à destination du public.

Loi 25 et données de mobilité - La Loi 25 soulève de nombreux défis pour le transporteur, notamment celui de déterminer comment tirer un maximum de valeur des données de mobilité, tout en s'assurant de masquer les facettes qui touchent la vie privée.

Données inutiles pour une société de transport - Les renseignements personnels des clients n'intéressent pas directement le transporteur pour la planification ou l'opération des transports publics. Il n'a pas besoin, en effet, du nom, de l'adresse, du dossier médical ou encore du numéro de carte de crédit des clients. Dans tous ces cas, si l'anonymisation était requise, elles seraient aisément mises en place compte tenu de la nature de ces informations.

Définition des besoins d'une société de transport - Dans le cadre de ses activités, le transporteur a besoin de générer des itinéraires totalement désagrégés pour pouvoir planifier, caractériser la demande, modéliser, mesure les impacts des projets ou encore diffuser l'information. Ses sources d'information sont multiples et comprennent notamment : des enquêtes origines-destination (OD), l'usage des cartes à puce et le GPS.

EFVP dans une société de transport - Afin de se mettre en conformité avec la Loi 25, le transporteur doit se poser certaines questions sur ces données. Il faut déterminer d'une part l'utilité de la collecte des données en déterminant les besoins qui la sous-tendent. D'autre part, il faut savoir si l'on peut atteindre ces objectifs à l'aide de données anonymisées ou dépersonnalisées. Enfin, il faut trouver un équilibre entre le niveau d'anonymisation (ou de dépersonnalisation) et la quantité d'information à conserver afin d'assurer la mission de l'organisation.

7 Dionne, Vincent, « Données de mobilité et loi 25 : Les enjeux pour la planification des transports », 2024, en ligne : pour le support <https://www.chairelrwilson.ca/files/sites/36/2024/04/DionneV.pdf> (consulté le 12 juin 2024) et pour la présentation <https://www.youtube.com/watch?v=zd8lRnyzHLk> (consulté le 12 juin 2024)

Défis soulevés par l'enquête OD – L'enquête OD permet de collecter tous les cinq ans par sondage des informations chez 4 % de la population afin de réaliser des itinéraires totalement désagrégés. Ces informations comprennent à la fois des données sur le ménage, sur les personnes et sur leurs déplacements. Il est possible de réidentifier indirectement un individu grâce aux informations qui permettent de localiser le domicile. Si l'anonymisation de ces données est possible, la question du niveau d'anonymisation se pose.

Défis soulevés par la validation des cartes à puce – La validation des cartes à puce constitue une autre source de données pour le transporteur grâce à l'identification des arrêts de montée et des arrêts de descente. Ces données permettent de mesurer à la fois la fréquentation des transports en commun, mais aussi la séquence d'embarquements. Contrairement à l'enquête OD, ces données permettent une analyse quotidienne du comportement des usagers, mais l'identification indirecte est beaucoup plus complexe, puisque les lieux exacts de domicile et d'activités ne sont pas connus Il est par contre possible de générer synthétiquement un lieu de domicile, l'arrêt de descente, le lieu d'activité, les motifs du transport ainsi que le segment de la clientèle intéressée.

Défis du coût d'implémentation de l'anonymisation – Pour le transporteur, l'anonymisation soulève la question du potentiel de l'utilisation future de ces données. La détermination des besoins, le coût d'implémentation des techniques d'anonymisation et la question du stockage des données sont autant d'enjeux auxquels le transporteur a à faire face maintenant.

1.4 L'anonymisation des données des recherches en milieu académique : défis et solutions

Anonymisation : des solutions adaptées – Le cas des données de recherche met en évidence les besoins de solutions adaptées pour protéger ces données. La présentation de Stéphanie Pham-Dang, intitulée « L'anonymisation des données de recherche[8] », revient sur les défis que rencontre l'Université de Montréal dans la gestion des données de recherche avant de présenter les solutions institutionnelles existantes.

Définition de la gestion des données de recherche – La gestion des données de recherche (GDR) en milieu académique « comprend les processus utilisés tout au long du cycle de vie d'un projet de recherche pour orienter la collecte [...], le stockage, le partage et la préservation des données de recherche[9] ».

Stockage, partage et réutilisation des données de recherche – La question de la gestion des données sensibles de recherche se pose à trois stades : au niveau du stockage des données actives durant la recherche, au niveau du partage des données finales une fois la recherche terminée en vue de leur réutilisation des données de recherche repérées dans un dépôt de données, ainsi qu'un niveau de leur préservation à long terme selon des règles archivistiques de conservation.

8 Pham-Dang, Stéphanie, « L'anonymisation des données de recherche », 2024, en ligne : pour le support <https://www.chairelrwilson.ca/files/sites/36/2024/04/PhamDangS.pdf> (consulté le 12 juin 2024) et pour la présentation <https://www.youtube.com/watch?v=zd8lRnyzHLk>

9 Alliance de recherche numérique du Canada, « Gestion des données de recherche », en ligne < https://alliancecan.ca/fr/services/gestion-des-donnees-de-recherche> (consulté le 12 juin 2024)

Niveau macro de la gouvernance des données de recherche – La gouvernance des données de recherche s'appréhende à deux niveaux. Au niveau macro, la gouvernance des données de recherche est encadrée par plusieurs textes : la politique des trois organismes subventionnaires fédéraux sur la gestion des données de recherche (2021), la politique des trois conseils sur l'éthique de la recherche avec des êtres humains (EPTC 2, 2022), la stratégie institutionnelle pour la gestion des données de recherche (2023) ainsi que la Loi 25 au Québec (2023).

Niveau micro de la gouvernance des données de recherche – Au niveau micro, les politiques des éditeurs et les solutions offertes par les institutions académiques participent à la gouvernance des données de recherche. Les premières prévoient souvent une obligation de communication et d'accessibilité des données de recherche utilisées pour les articles soumis. Les secondes offrent des solutions de stockages aux chercheurs en fonction du degré de sensibilité de celles-ci. Or, si les données de recherche sont anonymisées, alors la publication des données finales est rendue beaucoup plus simple, lors du partage dans des dépôts de données de recherche, tels que Borealis UdeM.

Sensibilité des données de recherche – La sensibilité d'une donnée de recherche varie en fonction du contexte de la recherche. Certaines variables peuvent impacter le niveau de sensibilité d'une donnée, tels que le sujet de la recherche, le lieu où se déroule la recherche, les types de données, etc.

Étapes de l'anonymisation des données de recherche – Trois étapes rythment l'anonymisation des données de recherche. Il s'agit d'une part de déterminer et de supprimer ou masquer les identifiants directs. Ensuite, il s'agit de déterminer les identifiants indirects (âge, date de naissance, profession, revenu, géographie, religion, origine, etc.) et d'en supprimer certaines variables. Enfin, il faut examiner plusieurs paramètres tels que la fréquence de certaines informations permettant potentiellement de réidentifier une personne ou certaines variables textuelles susceptibles de révéler des informations personnelles.

Solutions d'anonymisation des données de recherche – Il existe plusieurs solutions d'anonymisation des données de recherche. Certains services conseils sur mesure à l'UdeM existent pour certains types de données (notamment les données en santé). Pour des projets faisant appel à d'autres types de données, il n'y a pas de solution unique : il existe des solutions ponctuelles d'anonymisation des données de recherche impliquant des solutions web nécessitant une formation de base pour utiliser ces outils d'aide développés par diverses parties prenantes de l'écosystème de la recherche scientifique. Outre ces outils en ligne, il est aussi possible d'agir au niveau du contrôle d'accès des données et de la méthodologie de recherche.

2. L’anonymisation des données comme instrument de protection des renseignements personnels : définition, limites et solutions

2.1 L'anonymisation des renseignements personnels dans la Loi 25

Anonymisation et botté en touche de la Loi 25 - La place réservée à l'anonymisation au sein de la Loi 25 soulève de nombreuses questions juridiques. Dans sa présentation sur « Les origines et les enjeux de l'anonymisation dans la Loi 25[10] », Antoine Guilmain constate que la Loi botte en touche sur la définition de l'anonymisation des renseignements personnels en renvoyant à un règlement le soin d'en définir les modalités. Cette manière de faire se distingue des autres pays qui ont plutôt opté pour une approche où le régulateur détermine un code de bonne conduite qui est ensuite mis en place par une communauté de pratique.

Anonymisation comme alternative à la destruction - Le projet de règlement sur l'anonymisation publiée à la Gazette soulève de nombreuses questions. Tout d'abord, l'anonymisation n'est pas envisagée en tant que telle, mais comme une alternative à la destruction des renseignements personnels. Cela signifie que les cas d'utilisation de l'anonymisation sont sensiblement plus limités que ceux prévus au niveau du projet de loi C-27 au fédéral ou encore ceux du RGPD. Il n'y a donc pas un régime général prévu pour l'anonymisation, mais simplement un régime spécifique alternatif à la destruction à un moment précis du cycle de vie des données.

Anonymisation comme processus balisé - Ensuite, le règlement fixe un processus lourd pour parvenir à l'anonymisation des données[11] : établir les finalités d'anonymisation, superviser par une personne compétente, retirer les renseignements personnels, préanalyser les risques de réidentification, établir les techniques d'anonymisation, analyser les risques de réidentification, mettre à jour l'analyse des risques de réidentification et maintenir un registre d'anonymisation.

Tension entre la Loi 25 et le projet de règlement - Le projet de règlement vient contredire la loi. Si la Loi 25 prévoit que l'anonymisation doit être irréversible, en revanche, le règlement fixe un standard moins haut d'un risque résiduel de réidentification qui doit être faible. Le projet de règlement vient en quelque sorte corriger la loi, car l'irréversibilité est dans les faits impossible à atteindre.

Au-delà des renseignements personnels - Enfin, la loi et le projet de règlement dénotent une tendance globale à vouloir encadrer le sort des renseignements non personnels en exigeant la détermination de « fins sérieuses et légitimes » pour utiliser les renseignements anonymisés. Cette tendance s'observe aussi au fédéral.

10 Guilmain, Antoine, « Les origines et les enjeux de l'anonymisation dans la Loi 25 », 2024, en ligne : pour la présentation <https://www.youtube.com/watch?v=LmDm9P2RD3g&t=1s> (consulté le 12 juin 2024

11 Guilmain, Antoine et Justin Boileau, « Règlement sur l'anonymisation de renseignements personnels au Québec : Modélisation et commentaires », 2024, en ligne <https://gowlingwlg.com/getmedia/8dc32bde-5f0e-422d-b2ec-eab3dc3cd84b/Diagramme-reglement-sur-l-anonymisation-Final.pdf.xml> (consulté le 12 juin 2024)

2.2 Un instrument pour l'anonymisation des renseignements personnels : les Normes nationales du Canada

Anonymisation et normes – Trouver des solutions pour anonymiser les données conformément aux lois en vigueur n'est pas chose aisée. Pourtant, les Normes nationales du Canada pourraient jouer un rôle dans ce domaine. Dans une présentation intitulée « Les Normes nationales du Canada : un instrument de protection des renseignements personnels ?[12] », Henry Laville aborde la question de comment identifier les normes qui permettent l'anonymisation des données. En effet, si la Loi 25 a établi le « Quoi faire ? » en matière d'anonymisation, elle reste évasive sur le « Comment » anonymiser les données. Or, les Normes nationales du Canada constituent un instrument pertinent dont il ne faut cependant pas surestimer la portée.

Normalisation et normes techniques – Les Normes nationales du Canada sont avant tout des normes techniques issues de la normalisation. La normalisation se définit comme une « activité propre à établir, face à des problèmes réels ou potentiels, des dispositions destinées à un usage commun et répété, visant à l'obtention du degré optimal d'ordre dans un contexte donné[13] ». La norme technique est, quant à elle, un « document, établi par consensus et approuvé par un organisme reconnu, qui fournit, pour des usages communs et répétés, des règles des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné[14] ».

Système national des normes – Au niveau du Canada, les Normes nationales du Canada sont présentes dans le Système national des normes (SNN) qui est le système « visant à élaborer, promouvoir et appliquer des normes volontaires au Canada[15] ».

Acteurs canadiens de la normalisation – Le SNN est composé de deux types d'acteurs. Le Conseil canadien des normes est l'organisme qui vient chapeauter l'activité de normalisation au niveau du Canada, et il accrédite les Organismes d'élaboration des normes (OEN) qui ont pour mission de produire les normes volontaires répondant aux intérêts et aux besoins canadiens. Ces OEN évoluent dans des secteurs très différents (air conditionné, plomberie, gaz, agriculture, bâtiment, assurance, industrie, etc.), et deux d'entre eux ont un rôle particulier dans le domaine de la gouvernance des données (l'Association canadienne de normalisation et l'Institut des normes de gouvernance numérique).

Identification des normes nationales du Canada – L'identification des Normes nationales du Canada se réalise de deux manières. Les Normes nationales du Canada se définissent avant tout comme des normes élaborées « par un OEN respectant les exigences et lignes directrices du CCN relatives : a) à l'accréditation des OEN; b) aux adoptions[16] ». Parmi les OEN, certains sont déclarés admissibles à l'auto-déclaration de conformité, tandis que d'autres ne le sont pas. Les premiers sont habilités à labéliser eux-mêmes les normes volontaires en Normes nationales du Canada, tandis que les seconds doivent soumettre les normes à labéliser au Conseil canadien des normes qui décidera si le label doit être attribué ou non.

12 Laville, Henry, « Les Normes nationales du Canada : un instrument de protection des renseignements personnels ? », 2024, en ligne : pour le support <https://www.chairelrwilson.ca/files/sites/36/2024/04/LavilleH.pdf> (consulté le 12 juin 2024) et pour la présentation < https://www.youtube.com/watch?v=LmDm9P2RD3g&t=1s> (consulté le 12 juin 2024)

13 Organisation internationale de normalisation, *Normalisation et activités connexes – Vocabulaire général*, 8 éd., 2004, 4

14 Organisation internationale de normalisation, *Normalisation et activités connexes – Vocabulaire général*, 8 éd., 2004, 12

15 *Loi sur le Conseil canadien des normes*, L.R.C. (1985), ch. S-16, art. 3.1

16 Conseil canadien des normes, *Exigences et lignes directrices – Accréditation des organisations d'élaboration de normes*, 2019, 8

Portées et limites des Normes nationales du Canada – La portée des Normes nationales du Canada est mise en perspective avec ses limites. Ces normes sont non-contraignantes, c'est-à-dire que les organisations ne peuvent être contraintes de les respecter. Elles ont avant tout un effet incitatif pour les acteurs économiques qui, grâce au label, ont l'information selon laquelle ces normes constituent le standard identifié et en vigueur au niveau du Canada. La majorité de ces normes sont payantes, ce qui peut représenter un coût substantiel pour de nombreuses organisations désireuses de les acquérir pour s'y conformer, mais il s'agit aussi d'une garantie d'indépendance des OEN qui trouvent leur source de financement dans l'activité de vente de normes. Ces normes sont censées représenter les « intérêts et besoins canadiens », mais la définition de ces intérêts et besoins canadiens est opérée par les organismes de normalisation eux-mêmes sans contrôle *a priori*. Enfin, la représentativité des OEN est un enjeu : le fonctionnement de ces organisations demeure opaque, et il est difficile de déterminer la représentation des différents intérêts de la société (privé, public, consommateurs, etc.).

Norme nationale du Canada et protection des renseignements personnels – Les Normes nationales du Canada sont présentées dans un tableau synthétique identifiant et classifiant les différentes normes techniques intéressant la gouvernance des données en général, et l'anonymisation des données en particulier.

3. L’anonymisation des données dans la gouvernance des renseignements personnels : enjeux et perspectives

3.1 L'acceptabilité sociale d'une innovation technologique et le rôle de la participation publique

Gouvernance : une question de légitimité – Au centre de la gouvernance d'une innovation technologique, nous retrouvons la question de la légitimité. La présentation de Alice Friser, intitulée « Optimiser l'acceptabilité sociale d'une innovation technologique : le rôle de la participation publique[17] », met en évidence le caractère central de la participation publique comme composante de l'acceptabilité sociale au sein d'un projet de déploiement d'une nouvelle technologie.

Définition de l'acceptabilité sociale – Définie comme l' « assentiment de la population à un projet ou à une décision résultant du jugement collectif que ce projet ou cette décision est supérieur aux alternatives connues, y compris le statu quo[18] », l'acceptabilité sociale suppose une connaissance des enjeux des nouvelles technologies par la population afin que celle-ci puisse en discuter et prendre part aux délibérations publiques.

Facteurs de l'acceptabilité sociale – L'acceptabilité sociale des innovations technologiques peut être optimisée en agissant notamment sur quatre facteurs : l'efficience, la pertinence, la confiance et l'équité. Du point de vue de l'efficience, la technologie doit notamment être opérationnelle, abordable, sécuritaire, réglementaire et écologique. Pour être pertinente, la technologie doit proposer de répondre à un problème identifié par la majorité de la population et être en alignée par rapport aux valeurs et à l'idée du progrès. L'acceptabilité sociale suppose aussi de la confiance dans la technologie (ex. fuite des données) et de l'équité (ex. : profilage des données, fermeture de guichets physiques). La mise en place d'une application de mobilité servicielle dépend de ces quatre facteurs.

Acceptabilité sociale et anonymisation des données – L'anonymisation des données soulève un triple problème du point de vue de l'acceptabilité sociale. Les difficultés quant au succès de l'anonymisation constituent un enjeu au niveau de l'efficience. L'enjeu de la confiance se pose aussi lorsqu'il s'agit d'aborder la maîtrise des données. Enfin, les technologies de profilage constituent un enjeu pour l'équité.

Définition de la participation publique – La participation publique est une méthode qui consiste à demander au public son opinion afin de connaître ses valeurs, ses besoins, ses attentes et ses craintes afin de maximiser l'utilité de la technologie et son acceptation dans un milieu donné. La participation publique vise à la fois à informer le public sur les enjeux de la gouvernance des données et elle crée de la confiance dans le projet en question.

Cas d'usage de la participation publique – Le recours à la participation publique est recommandé dans trois circonstances. Lorsque le public s'intéresse au projet, la participation publique permet de l'informer. Lorsque les décideurs n'ont pas de connaissance de leur public, de ses attentes et de ses craintes par rapport au projet, la participation publique est un moyen pour obtenir ces différentes informations. Enfin, lorsqu'il y a un manque de confiance de la part du public dans le projet, la participation publique permet de créer cette confiance.

17 Friser, Alice, « Optimiser l'acceptabilité sociale d'une innovation technologique : le rôle de la participation publique », 2024, en ligne : pour le support <https://www.chairelrwilson.ca/files/sites/36/2024/04/FriserA.pdf> (consulté le 12 juin 2024) et pour la présentation <https://www.youtube.com/watch?v=c53qRmdyve8> (consulté le 12 juin 2024

18 Gendron, Corinne, « Penser l'acceptabilité sociale : au-delà de l'intérêt, les valeurs» (2014) 2014 :11 Éthique et relations publiques : pratiques, tensions et perspectives 117 – 129, 124, en ligne < https://journals.openedition.org/ricsp/584> (consulté le 12 juin 2024)

Mise en œuvre de la participation publique – De manière non exhaustive, six facteurs permettent de mettre en place un dispositif de participation publique : le choix d'un sujet de discussion approprié (choix du sujet; choix de l'échelle du sujet); le choix d'un dispositif participatif adapté; la mobilisation du public en amont du projet avant de prendre des décisions; outiller le public; demander au public de s'exprimer sur les impacts négatifs anticipés; tenir compte de ces éléments pour développer l'innovation.

3.2 La gouvernance participative de l'anonymisation des données

Gouvernance : le problème de l'effectivité – Les questions de gouvernance ne se résument pas aux enjeux de légitimité : l'effectivité d'une gouvernance est une composante toute aussi essentielle, surtout dans le domaine de l'anonymisation des données. Dans sa présentation intitulée « La gouvernance participative de l'anonymisation des données[19] », Ledy Rivas Zannou constate un problème dans la Loi 25, où il identifie deux niveaux de gouvernance dont il en interroge l'effectivité. En effet, la gouvernance législative soumet des normes obligatoires à destination des organisations. En ce sens, la gouvernance par délégation législative s'exprime suivant une double temporalité : d'une part, une délégation aux normes techniques (politiques internes et normes individuelles) et d'autre part, une délégation réglementaire.

Effectivité de la gouvernance – Or, la gouvernance par délégation législative aux normes techniques n'est pas effective. Ces normes sont en effet dénuées de force obligatoire, c'est-à-dire que l'État ne peut contraindre les organisations à se conformer à ces normes contrairement à la loi. Les normes techniques portent cependant en elles des incitatifs. Par ailleurs, ces normes ne reflètent pas nécessairement les intérêts de toutes les parties.

Solution : élaborer une gouvernance participative – La délégation législative de la Loi 25 peut s'analyser comme une invitation faite par le législateur à la participation pour créer et/ou identifier les normes déléguées. La mise en place de comités qui représenteraient différents intérêts catégoriels de la société est une solution à envisager. Si les différentes institutions créées par le RGPD européen sont riches d'enseignement, elles ne peuvent être reprises comme telles au niveau canadien et québécois. Les spécificités culturelles ne sont en effet pas les mêmes (en raison du fédéralisme notamment). Or, au niveau québécois, la Commission d'accès à l'information (CAI) pourrait être un de ces comités, pour autant que l'on détermine les moyens à lui donner.

19 Rivas Zannou, Ledy, « La gouvernance participative de l'anonymisation des données », 2024, en ligne pour la présentation < https://www.youtube.com/watch?v=c53qRmdyve8> (consulté le 12 juin 2024)

3.3 L'anonymisation des données et le cas des FinTechs

Gouvernance : le cas des Fintechs – Dans sa présentation intitulée « Anonymisation des données : Le cas des Fintechs[20] », Maya Cachecho présente dans une perspective sociologique, les enjeux de l'anonymisation des données dans le cas des Fintechs. Entreprises qui évoluent dans le secteur de l'innovation technologique applicable aux services financiers et bancaires, les Fintechs offrent différents services aux consommateurs tels que l'investissement et le paiement en ligne, ou encore la gestion d'épargne et le prêt. L'équipe de Maya Cachecho étudie le cas des Fintechs tant du point de vue des organisations que de celui des consommateurs.

Données et Fintechs – Du point de vue des Fintechs, les données représentent une ressource importante qu'elles ne souhaitent pas détruire. La conservation des données permet à ces organisations d'évaluer la performance de leurs services ainsi que la satisfaction de leur clientèle, mais les données constituent aussi une ressource économique par leur revente.

Fintechs et gouvernance des données – La position des Fintechs à l'égard de la gouvernance des données est contrastée. Premièrement, la nouvelle Loi 25 pose de nouvelles obligations qui peuvent être lourdes à mettre en place. Deuxièmement, les Fintechs souhaiteraient un cadre législatif qui leur permette de se rapprocher du RGPD européen. Troisièmement, ces organisations constatent qu'il est difficile de mettre en place le nouveau cadre législatif, notamment en raison des ambiguïtés, des zones grises et de l'absence de définitions claires dans la Loi 25. Ensuite, de nombreuses obligations législatives de la Loi 25 sont entrées en vigueur en 2023, ce qui a rendu difficile la mise en conformité avec le nouveau cadre législatif. Cinquièmement, les Fintechs soulignent les difficultés qu'il peut y avoir à se conformer à la législation lorsque l'on est une petite entreprise.

Fintechs et enjeux soulevés par l'anonymisation des données – Pour ces organisations, l'anonymisation des données n'est qu'une obligation parmi d'autres, qui, en fonction du contexte, ne pose pas de difficulté à mettre en œuvre ou, au contraire, soulèvent des enjeux. La question de l'irréversibilité de l'anonymisation en est un, et les Fintechs espèrent que le règlement sur l'anonymisation des données apportera des réponses sur ce point.

Fintechs et Loi 25 – Les Fintechs conçoivent l'application de la Loi 25 essentiellement au travers de la notion de consentement et de clarté des politiques de confidentialité. Les Fintechs ont des craintes par rapport aux conséquences d'éventuelles poursuites et aux montants des sanctions.

Consommateurs et gouvernance des données – Du point de vue des consommateurs, l'étude met en évidence le peu d'intérêt qu'ont les consommateurs pour le sort de leurs renseignements personnels par rapport à leur souci quant à la protection de leur capital investi dans ces plateformes de Fintechs. Plusieurs facteurs expliquent ce point : certains consommateurs préfèrent échanger leurs données plutôt que de se voir refuser l'accès à un service, d'autres n'ont pas connaissance des politiques de confidentialité, d'autres encore se fient à la réputation de l'organisation, voire certains d'entre eux ne se soucient pas de leur vie privée en ligne. La question demeure de savoir comment faire pour que le consommateur prenne conscience de ses droits à la vie privée.

20 Cachecho, Maya, « Anonymisation des données : le cas des Fintechs », 2024, en ligne pour la présentation <https://www.youtube.com/watch?v=c53qRmdyve8> (consulté le 12 juin 2024)

obvia

obvia.ca