ANONYMISATION DES DONNÉES DE MOBILITÉ

Une introduction pratique à la prise de décision responsable en matière d'anonymisation des données

NORD OUVERT

Janvier 2026

TABLE DES MATIÈRES

INTRODUCTION : DONNÉES DE MOBILITÉ ET CONFIDENTIALITÉ

La plupart, sinon la totalité, des organisations fournissant des services dans le domaine de la mobilité collectent certains renseignements personnels auprès de leurs clients. Que vous gériez une agence de transport, une flotte de vélos en libre-service ou une application de VTC, il est souvent essentiel pour le service de savoir qui sont vos utilisateurs et où ils se rendent.

Lorsque les renseignements personnels sont détenus et utilisés uniquement par l'organisation qui les a collectés, cela ne pose généralement pas de problèmes majeurs en matière de confidentialité, à condition que ces informations aient été collectées avec le consentement éclairé des personnes concernées, dans un but raisonnable (par exemple, pour traiter un paiement ou permettre la navigation) et que des mesures de protection appropriées soient mises en place pour protéger les données.[1]

Cependant, le paysage change. Il est de plus en plus courant que les organisations utilisent les renseignements personnels qu'elles ont collectés à d'autres fins ou qu'elles mettent leurs données en commun avec d'autres organisations. Ces initiatives, qui visent souvent à générer de nouvelles connaissances, à améliorer l'urbanisme ou à créer de nouveaux produits de données, introduisent de nouvelles complexités. Lorsque les données quittent leur « silo » d'origine ou sont réutilisées, les risques pour la vie privée changent, et notre approche de leur gestion doit également changer.

> L'anonymisation des données consiste à **réduire** les risques tout en garantissant que les données restent utiles.

Bien qu' il soit important de garder à l'esprit qu'il est impossible d'éliminer complètement ce risque, il existe toute une gamme d'outils à la disposition des organisations pour réduire celui-ci. Une approche holistique de l'atténuation des risques repose sur la définition de rôles, de politiques et de processus de gestion de la confidentialité, la mise en œuvre de mesures de sécurité appropriées pour protéger les renseignements personnels et l'utilisation de diverses technologies de renforcement de la confidentialité (TRC), notamment les techniques d'anonymisation et de dépersonnalisation des données.

L'un de ces outils est l'anonymisation des données. Celui-ci consiste à modifier les ensembles de données afin de rendre plus difficile l'identification des personnes.

(1) L-39.1 - Loi sur la protection des renseignements personnels dans le secteur privé

TERMES CLÉS ET DÉFINITIONS

Spectre d'identifiabilité

Lorsque l'on évalue le degré d'identifiabilité d'un ensemble de données, il est utile de prendre en compte un spectre d'identifiabilité.[2] Celui-ci définit trois états d'information :

- **Renseignements personnels**
  - **Informations identifiées :** informations qui, à elles seules, identifient directement une personne
  - **Informations identifiables :** informations pour lesquelles il existe une forte probabilité qu'elles puissent être associées à une personne identifiable.
- **Renseignements non personnels**
  - **Informations non identifiables :** informations pour lesquelles il n'existe aucune possibilité sérieuse qu'elles puissent être associées à une personne identifiable.

En général, plus les informations sont identifiables, plus elles sont utiles à des fins diverses. Cependant, cette utilité supplémentaire s'accompagne de risques accrus. L'avantage d'adopter une approche spectrale pour définir l'identifiabilité est qu'elle permet un large éventail d'utilisations innovantes des informations tout en tenant compte et en atténuant un niveau raisonnable de risque résiduel.

Anonymisation vs dépersonnalisation

La Loi 25 du Québec (Loi modernisant des dispositions législatives en matière de protection des renseignements personnels) établit une distinction importante :

Les renseignements dépersonnalisés sont des données dont les identifiants directs (tels que les noms) ont été supprimés, mais qui peuvent encore permettre d'identifier des personnes moyennant des efforts ou des informations supplémentaires. Par exemple, le remplacement du nom d'une personne par un code aléatoire est une forme de dépersonnalisation. En vertu de la loi, les données dépersonnalisées sont toujours considérées comme des renseignements personnels, ce qui signifie que toutes les obligations en matière de confidentialité continuent de s'appliquer.

Les informations anonymisées, en revanche, ont été traitées de manière si approfondie qu'il est pratiquement impossible d'identifier des personnes, que ce soit directement ou indirectement. Lorsqu'elles sont traitées correctement, les données anonymisées ne sont plus considérées comme des renseignements personnels, ce qui permet une utilisation et un partage beaucoup plus larges.

(2) Canadian Anonymization Network, « Spectrum of Identifiability », 2020, https://deidentify.ca/wp-content/uploads/2020/10/CANON-States-of-Data-One-Pager.pdf.

Menaces et risques liés aux données de mobilité

Les renseignements personnels qui identifient directement ou indirectement des individus comportent un risque inhérent d'utilisation abusive. En vertu de la législation québécoise sur la protection de la vie privée dans le secteur privé, les organisations qui collectent et détiennent des renseignements personnels sont tenues de prendre les mesures appropriées pour empêcher tout accès, utilisation et divulgation non autorisés.

Dans le secteur de la mobilité, ces risques sont accrus par la nature très granulaire des données de localisation. Même de petits fragments d'informations, lorsqu'ils sont combinés au fil du temps, peuvent révéler des informations détaillées sur la vie des individus. Les différents types d'ensembles de données de mobilité présentent donc différents niveaux de risque pour la vie privée. Prenons deux types de données de mobilité courants : les enregistrements de trajets et les données des comptes clients.

Les enregistrements de trajets, par exemple, comprennent généralement des éléments tels que :

- un identifiant client ou de compte;
- les points de départ et d'arrivée;
- les horodatages;
- la distance ou la durée.

Bien que les enregistrements de trajet ne contiennent pas nécessairement de noms, la combinaison du lieu et de l'heure peut être très révélatrice, en particulier lorsque les trajets sont répétés et que des schémas se dessinent.

Les données relatives aux comptes clients comprennent souvent :

- les prénoms et noms;
- l'adresse et les coordonnées;
- les informations de paiement;
- l'historique des abonnements ou de l'utilisation.

Lorsque les données des comptes clients sont associées aux enregistrements de trajets, le risque de réidentification des personnes augmente considérablement.

Ensemble, ces caractéristiques donnent lieu à plusieurs menaces courantes pour la confidentialité des données de mobilité, notamment :[3]

- l'identification des personnes, en déduisant leur lieu de résidence ou de travail à partir de leurs habitudes de déplacement régulières;
- la prédiction de comportements, par exemple en identifiant les moments où le domicile d'une personne est susceptible d'être inoccupé;
- la divulgation d'informations sensibles, par exemple par le biais de visites répétées dans des établissements médicaux ou de soins;
- le recoupement de données, lorsque des informations concernant des déplacements apparemment « anonymes » sont mises en correspondance avec d'autres sources d'informations, telles que des publications sur les réseaux sociaux liées à des heures et des lieux spécifiques.

Compte tenu de ces risques, l'anonymisation est souvent perçue comme un levier permettant d'exploiter ou de partager les données de mobilité tout en protégeant la vie privée. Toutefois, déterminer s'il convient d'anonymiser les données, et de quelle manière, exige une réflexion qui dépasse le simple cadre de l'ensemble de données. Il est essentiel de comprendre **l'environnement des données** : c'est-à-dire la relation entre les données et le contexte dans lequel elles évoluent. Cela inclut les autres sources d'information avec lesquelles elles peuvent être croisées, les personnes qui y ont accès (ainsi que leurs rôles et responsabilités), ainsi que les systèmes techniques utilisés pour les traiter et les sécuriser.

La section suivante présente ce concept et explique comment l'évaluation de la situation des données permet de prendre des décisions éclairées en matière d'anonymisation.

(3) Henry Laville. « Rapport : Symposium « Anonymisation des données » » (OBVIA, 2024). https://doi.org/10.61737/XAFK3054.

UNE APPROCHE STRUCTURÉE DE LA PRISE DE DÉCISION EN MATIÈRE D'ANONYMISATION DES DONNÉES

L'anonymisation est à la fois un processus de gestion des risques et un exercice de prise de décision. Elle permet essentiellement de répondre à une question clé : *Devons-nous partager ou divulguer ces données, et si oui, sous quelle forme et dans quelles conditions?*

Il est essentiel de noter que l'on ne peut pas déterminer si les données peuvent être partagées ou divulguées en toute sécurité en examinant uniquement l'ensemble de données. Une approche structurée est nécessaire pour garantir que les décisions d'anonymisation reposent sur une évaluation holistique des risques découlant de l'interaction entre les données, les personnes, les cadres juridiques, les systèmes informatiques, la culture organisationnelle et les pratiques de gouvernance.

Ce cadre décisionnel comprend trois activités principales:[4]

- évaluer la situation des données (c'est-à-dire la relation entre les données et leur environnement);
- évaluer et contrôler les risques liés à la divulgation;
- gérer les impacts.

Ce guide se concentre principalement sur l'évaluation de la situation des données. Cette étape est fondamentale, car elle établit le contexte pour évaluer et contrôler les risques de divulgation et pour gérer les impacts qui peuvent résulter de ces risques.

Remarque sur les exemples :

Tout au long de cette section, nous ferons référence à deux scénarios réalistes pour démontrer comment le processus fonctionne dans la pratique :

Scénario A : Données sur les transports publics

L'agence de transport public de la ville de Verdeville Transit collecte des données sur la fréquentation afin de soutenir la planification continue des services et l'amélioration des opérations. L'objectif de l'agence est de mieux comprendre comment les habitants utilisent le réseau de transport afin d'ajuster les horaires, les itinéraires et les points de correspondance pour mieux répondre à la demande.

Les données recueillies à cette fin comprennent des informations telles que les temps d'attente, les arrêts et les itinéraires fréquemment utilisés, ainsi que les lieux de correspondance courants. Ces données sont analysées de manière agrégée afin d'identifier des tendances et des schémas plutôt que d'examiner le comportement individuel des usagers.

L'accès aux données est strictement limité. L'ensemble de données est stocké sur les serveurs internes de l'agence, et seuls trois membres du personnel autorisés sont autorisés à y accéder dans le cadre de leurs fonctions. Les données sont utilisées exclusivement à des fins d'analyse interne et ne sont pas partagées en dehors de l'organisation.

(4) Cette approche s'appuie sur le cadre décisionnel en matière d'anonymisation élaboré par le Réseau britannique d'anonymisation (UKAN).

Scénario B : Programme municipal de vélos en libre-service

Le programme municipal de vélos en libre-service City Bike Share, un partenariat public-privé, collecte des données sur les trajets dans le cadre de ses activités quotidiennes. Le programme souhaite rendre une partie de ces données accessibles au public afin de mettre en évidence l'activité cycliste dans la ville et de soutenir les efforts plus larges visant à améliorer les infrastructures cyclables et la planification de la mobilité.

L'objectif de la publication de ces données est de permettre aux acteurs publics travaillant dans le domaine de la mobilité, notamment les chercheurs universitaires, les journalistes, les fonctionnaires, les organisations de la société civile et les analystes du secteur privé, de mieux comprendre les habitudes d'utilisation du vélo. Les membres du public peuvent utiliser l'ensemble de données seul pour rendre compte de l'activité de partage de vélos ou le combiner avec d'autres ensembles de données pour obtenir des informations sur les tendances générales en matière de transport.

Les enregistrements des trajets sont publiés sur un portail de données ouvertes et peuvent être téléchargés sans restriction d'accès. Comme ces données sont destinées à être rendues publiques et peuvent être combinées avec d'autres sources d'information, elles doivent être préparées de manière à minimiser le risque d'identifier les cyclistes individuels tout en préservant leur valeur analytique.

Activité 1 : Évaluer la situation des données

Cette activité vous aidera à cerner et à définir les questions pertinentes pour évaluer votre environnement de données. Vous devrez synthétiser et décrire de manière systématique les données, vos objectifs de traitement et les enjeux qui en découlent. Une évaluation rigoureuse de cet environnement constitue le fondement de la prochaine étape.

Tâches spécifiques :

- Décrire l'environnement de données;
- Connaître vos données;
- Analyser les cas d'usage;
- Définir le cadre juridique;
- Évaluer les enjeux de consentement et vos obligations éthiques.

Décrire l'environnement de données

L'environnement de données décrit la manière dont un ensemble de données s'inscrit dans le monde réel et interagit avec son milieu. Cet environnement comprend les personnes qui accèdent aux données, les systèmes qui les stockent et les traitent, les autres ensembles de données avec lesquels ils peuvent être combinées, ainsi que les règles et les ententes qui régissent leur utilisation.

Dans la pratique, les données restent rarement au même endroit. Selon le cas d'usage, elles peuvent passer par plusieurs systèmes ou équipes, et parfois franchir les frontières organisationnelles ou entrer dans le domaine public. Chaque mouvement ou transformation introduit de nouvelles considérations en matière de confidentialité.

La cartographie des flux de données à partir du point de collecte permet de rendre ces dynamiques tangibles. Elle vous offre une visibilité sur le parcours des données, leurs transformations et l'émergence de risques potentiels. Tout aussi important, cet exercice permet de définir clairement l'imputabilité et les responsabilités à chaque étape du cycle de vie des données.

Questions clés à se poser :

- Qui contrôle ou gère actuellement ces données?
- Qui les recevra ou y aura accès ensuite?
- Les données seront-elles utilisées uniquement par le personnel interne, partagées avec des partenaires ou rendues publiques?
- Où les données seront-elles stockées ou accessibles (par exemple, sur des serveurs internes sécurisés, des plateformes infonuagiques ou des sites Web publics)?
- Quels sont les contrôles de gouvernance en place, tels que les contrats, les restrictions d'accès ou les accords de partage de données?

Dans le scénario Verdeville Transit, les données sur la fréquentation sont transférées d'une base de données interne chiffrée vers un poste de travail sécurisé destiné aux analystes. L'accès est limité à un petit nombre d'employés autorisés, et l'environnement est fermé et bien contrôlé.

Dans le scénario City Bike Share, les données relatives aux trajets sont transférées d'une base de données interne vers un environnement de traitement, puis finalement publiées sur un portail de données ouvertes. Cet environnement ouvert permet à quiconque d'accéder aux données et de les réutiliser, ce qui augmente considérablement leur exposition et les risques.

Connaître vos données

Une fois l'environnement de données bien défini, l'étape suivante consiste à examiner de près les données elles-mêmes. Il s'agit de déterminer à qui elles se rapportent, de préciser leur format, d'identifier les types de variables qu'elles contiennent et de relever les propriétés qui influent sur leur degré de sensibilité.

Personnes concernées

Les personnes concernées sont les personnes auxquelles les données se rapportent, telles que les usagers des transports en commun ou les utilisateurs de vélos en libre-service. Comprendre qui elles sont et comment elles interagissent avec le service permet de définir des attentes quant à l'utilisation appropriée des données.

Questions clés à poser :

- Qui sont les personnes représentées dans cet ensemble de données?
- Le service est-il facultatif ou essentiel de leur point de vue?
- Quelles attentes peuvent-elles raisonnablement avoir quant à l'utilisation de leurs données?

Types de données

Les données peuvent se présenter sous diverses formes, allant de microdonnées détaillées (au niveau individuel) à des résumés hautement agrégés. Le niveau de granularité a une incidence directe sur le risque d'atteinte à la vie privée.

Questions clés à poser :

- Cet ensemble de données est-il constitué d'enregistrements individuels (microdonnées) ou de statistiques agrégées?
- Les mêmes objectifs analytiques pourraient-ils être atteints en utilisant des données plus agrégées?

Types de variables

La plupart des ensembles de données sur la mobilité contiennent un mélange de types de variables. Les identifiants directs permettent d'identifier une personne à eux seuls, comme un nom ou un numéro d'assurance sociale. Les identifiants indirects ne permettent pas d'identifier une personne à eux seuls, mais peuvent le faire lorsqu'ils sont combinés à d'autres informations, telles que des lieux, des dates ou des heures. Les variables cibles sont des attributs sensibles ou révélateurs que quelqu'un pourrait vouloir déduire à propos d'un individu.

Questions clés à poser :

- L'ensemble de données comprend-il des identifiants directs?
- Quels sont les identifiants indirects présents, tels que des lieux précis ou des horodatages?
- L'ensemble de données comprend-il des attributs sensibles qui pourraient être révélateurs même si les identités ne sont pas explicites?

Propriétés des données

Les propriétés des ensembles de données peuvent augmenter ou réduire le risque de divulgation, mais à ce stade, elles ne sont utilisées que comme indicateurs généraux pour signaler les domaines qui nécessitent une analyse plus approfondie ultérieurement, et ne remplacent pas une évaluation complète des risques. Les propriétés clés pertinentes des données peuvent inclure :

- **Qualité des données :** des données précises et de haute qualité sont plus utiles, mais peuvent augmenter le risque de divulgation, tandis que de petites erreurs peuvent involontairement rendre l'identification plus difficile.
- **Âge des données :** les données plus anciennes présentent généralement moins de risques, car la localisation, les habitudes et la situation des

personnes changent avec le temps, même si elles peuvent également devenir moins précises.

- **Niveau de détail :** des données très granulaires augmentent la probabilité que des individus puissent être identifiés au sein d'un ensemble de données.
- **Données hiérarchiques ou groupées :** les données qui relient des individus au sein de groupes ou de lieux communs peuvent augmenter le risque, car les combinaisons au niveau du groupe peuvent être uniques.
- **Données horodatées ou longitudinales :** les données collectées au fil du temps augmentent le risque en révélant des schémas ou des changements de comportement uniques.
- **Couverture de la population par rapport à celle de l'échantillon :** les données à l'échelle de la population présentent un risque accru, car la présence d'un individu y est certaine. Contrairement à un échantillon aléatoire, cette certitude de présence facilite grandement les tentatives de réidentification.

Questions clés à poser :

- Quelle est la précision des données, en particulier en ce qui concerne le temps et le lieu?
- Les données sont-elles récentes?
- L'ensemble de données couvre-t-il l'ensemble de la population ou seulement un échantillon?

Analyser les cas d'usage

Il est essentiel d'avoir une compréhension claire et commune des cas d'usage. Lorsque vous savez exactement pourquoi les données sont utilisées et par qui, il devient beaucoup plus facile de décider quelles données sont vraiment nécessaires et quel niveau de risque est acceptable.

Questions clés à poser :

- Pourquoi voulons-nous partager, publier ou réutiliser ces données?
- Qui est censé les utiliser, par exemple les équipes internes, les partenaires, les chercheurs ou le public?
- Quels types d'analyse ou de réutilisation l'environnement de données prend-il réellement en charge?

Dans la plupart des cas, le cas d'usage sera l'un des quatre types suivants, selon la manière dont la réutilisation est définie.

1. Partage de données avec un tiers;
2. Diffusion des données (publication ou divulgation);
3. Utilisation continue des données au-delà d'une période de conservation préalablement définie;
4. Réutilisation par la même organisation des données à des fins autres que celles pour lesquelles elles ont été collectées.

Dans le scénario Verdeville Transit, le cas d'usage est l'analyse interne visant à améliorer la prestation de services, avec un accès limité et aucun partage externe.

Dans le scénario City Bike Share, le cas d'utilisation est la diffusion publique, avec l'espoir que de nombreux utilisateurs différents puissent combiner les données avec d'autres sources.

Définir le cadre juridique

Tant que l'anonymisation n'est pas finalisée, les données demeurent des renseignements personnels. Cela signifie que vous devez comprendre et respecter la législation sur la protection de la vie privée applicable à votre organisation, selon que celle-ci est un organisme public ou une organisation privée.

Bien que les organismes publics et les organisations privées soient soumis à des lois différentes en matière de protection de la vie privée, les deux doivent se conformer au Règlement sur l'anonymisation des renseignements personnels lorsqu'ils anonymisent des données. Le règlement exige que les organisations évaluent périodiquement les risques de réidentification. Les organisations doivent également documenter leur démarche en tenant un registre des décisions et des méthodes d'anonymisation employées.

Questions clés à poser :

- Quelle loi sur la protection de la vie privée s'applique à notre organisation?
- Avons-nous clairement défini un objectif sérieux et légitime pour l'anonymisation?
- Une personne compétente a-t-elle été désignée pour superviser le processus d'anonymisation?
- Les identifiants directs ont-ils été supprimés avant de procéder à l'analyse des risques?
- Documentons-nous les techniques et les décisions dans un registre d'anonymisation?

Évaluer les enjeux de consentement et vos obligations éthiques

La conformité juridique n'est pas synonyme de gain et de maintien de la confiance du public. Même lorsque les données sont anonymisées, les organisations ont des responsabilités éthiques envers les personnes dont elles ont collecté les données.

Le niveau de confiance des individus repose sur les informations que ceux-ci reçoivent lors de la collecte et sur leur niveau de dépendance au service.. Les données de mobilité, en particulier, peuvent révéler des schémas sensibles que les individus ne mesurent pas toujours pleinement.

Questions clés à poser :

- Dans quelle mesure cette utilisation des données correspond-elle aux attentes initiales des utilisateurs?
- Sommes-nous clairs et transparents sur la manière dont les données de mobilité sont analysées et partagées?
- Cette utilisation des données pourrait-elle affecter de manière disproportionnée certains groupes, tels que les personnes handicapées?
- Les utilisateurs ont-ils un choix significatif ou existe-t-il des déséquilibres de pouvoir dus à la nature du service?
- Cette utilisation des données sert-elle clairement l'intérêt public?

Même lorsque la loi ne l'exige pas, le fait de proposer des options de désinscription lorsque cela est possible et de revoir régulièrement les pratiques avec les parties prenantes de la communauté peut contribuer à garantir que l'utilisation des données reste responsable, compréhensible et conforme aux valeurs publiques.

Ce guide est accompagné d'une **liste de contrôle conçue pour vous aider à évaluer votre environnement de données**. Vous pouvez l'utiliser pour poser les bonnes questions sur votre situation en matière de données.

PROCHAINES ÉTAPES

Une fois que vous avez évalué votre situation en matière de données, vous êtes en bien meilleure position pour décider de la marche à suivre. La phase suivante consiste à comprendre le niveau de risque de divulgation associé à vos données et à appliquer les mesures techniques appropriées pour réduire ce risque.

Cette phase se divise en deux activités principales. Tout d'abord, vous évaluez les risques de divulgation de manière plus détaillée. Ensuite, vous sélectionnez et appliquez des approches techniques pour contrôler ces risques. Ces étapes s'appuient directement sur la compréhension du contexte développée dans l'activité 1 et ne doivent pas être abordées isolément.

Activité 2 : Évaluer et contrôler les risques de divulgation

L'objectif de cette activité est de rassembler les processus que vous utiliserez pour mesurer et gérer les risques de divulgation associés à votre situation en matière de données. À ce stade, le but n'est pas de trouver une seule technique d'anonymisation « correcte », mais de sélectionner une combinaison d'approches adaptées à votre contexte spécifique.

Évaluer le risque de divulgation

Une fois que vous avez déterminé qu'il existe un risque de divulgation non négligeable, vous devez examiner de plus près comment ce risque pourrait se concrétiser. Cela implique d'examiner les données à la lumière de leur environnement, de leurs utilisateurs prévus et des façons dont elles pourraient raisonnablement être combinées avec d'autres informations.

Le risque ne se mesure pas de manière abstraite. Il dépend des personnes susceptibles d'accéder aux données, de ce qu'elles pourraient en faire et des efforts nécessaires pour identifier les individus. Un ensemble de données communiqué en interne à une petite équipe de confiance présente des risques très différents de ceux liés à sa publication publique.

Questions clés à se poser :

- Une personne pourrait-elle être identifiée dans cet ensemble de données, même si aucun nom n'y figure?
- Cet ensemble de données pourrait-il être associé à d'autres données disponibles afin d'identifier des personnes?
- Quelqu'un pourrait-il déduire des informations nouvelles ou sensibles sur des individus à partir de ces données?
- Quels sont les acteurs les plus susceptibles de tenter une réidentification et de quelles capacités disposent-ils réellement?
- Quelles seraient les conséquences potentielles en cas de réidentification ?

Répondre à ces questions vous aide à passer d'une perception générale du risque à une compréhension plus claire des principales vulnérabilités.

Contrôler le risque de divulgation

Une fois les risques de divulgation cartographiés, vous pouvez choisir les méthodes de traitement les plus adaptées. Cette étape, souvent confondue avec l'anonymisation elle-même, n'est efficace que si elle s'appuie sur une analyse préalable rigoureuse du contexte et des menaces.

Approches courantes en matière d'anonymisation

Il n'existe pas de méthode unique pour anonymiser les données. Dans la pratique, l'anonymisation implique généralement de combiner plusieurs techniques, appliquées différemment selon les variables d'un même ensemble de données. L'objectif est de réduire le risque à un niveau très bas tout en préservant autant que possible la valeur analytique.

DESCRIPTION	CONSIDÉRATIONS CLÉS	EXEMPLE
Suppression Suppression complète des variables afin qu'elles ne soient plus disponibles dans l'ensemble de données. Cela inclut généralement les identifiants directs et les identifiants indirects à haut risque.	› Cette variable est-elle essentielle pour le cas d'utilisation? › Sa suppression réduirait-elle considérablement le risque?	› Dataset includes: Name, Address, Age, Colour. › Action: Remove Name and Address.
Généralisation Remplacement des valeurs précises par des plages ou des catégories plus larges. Cela réduit l'unicité tout en préservant les modèles utiles.	› Les valeurs précises peuvent-elles être remplacées par des plages? › Quel est le niveau de précision réellement nécessaire pour l'analyse?	› Distance : « 10-50 km » au lieu de « 18 km ». › Emplacement : « Montréal » au lieu de « 4388, rue Saint-Denis ».
Suppression Retrait des lignes correspondant à des valeurs aberrantes ou à des individus ayant trop peu de pairs (unicité). Cette action élimine les profils atypiques qui sont les plus faciles à réidentifier par recoupement.	› Cet enregistrement se démarque-t-il trop? › Sa suppression altère-t-elle la qualité globale des données?	› Données de trajet : suppression d'un trajet effectué à 3 h du matin dans une région isolée, car il s'agit d'un cas unique.
Pseudonymisation Remplacement des identifiants par d'autres valeurs, telles que des codes aléatoires ou des hachages. (Remarque : il s'agit d'une mesure de sécurité, et non d'une anonymisation complète).	› Une analyse longitudinale (suivi dans le temps) est-elle nécessaire? › L'environnement est-il strictement contrôlé?	› L'identifiant utilisateur : « John Smith » est remplacé par « User_8923 ».

Activité 3 : Gérer les impacts

Même après l'anonymisation des données, vos responsabilités ne s'arrêtent pas là. L'activité suivante vise à garantir que les risques demeurent faibles au fil du temps, que les parties prenantes sont correctement informées et que vous êtes prêt à réagir en cas de problème.

Tâches spécifiques :

- Maintenir la confiance des parties prenantes;
- Prévoir les mesures à prendre en cas de problème;
- Surveiller la situation des données.

Maintenir la confiance des parties prenantes

Impliquer les parties prenantes dès le début et communiquer clairement sur la manière dont les données sont utilisées peut réduire considérablement l'impact de tout problème futur. Lorsque les individus comprennent pourquoi les données sont partagées et comment les risques sont gérés, ils sont moins susceptibles d'être surpris ou de se sentir lésés en cas d'incident.

Questions clés à poser :

- Avons-nous expliqué, dans un langage simple, comment et pourquoi les données sont anonymisées?
- Les utilisateurs savent-ils que leurs données peuvent être analysées ou partagées sous cette forme?
- Les parties prenantes ont-elles la possibilité de poser des questions ou de faire part de leurs préoccupations?

La transparence est particulièrement importante pour les données de mobilité, qui peuvent être perçues comme très personnelles même lorsqu'elles sont anonymisées.

Prévoir les imprévus

Malgré les précautions, un incident de confidentialité reste possible. Votre organisation doit disposer d'un plan de réponse clair, adapté à votre structure et aux types de données traitées. Si les politiques de gestion de crise varient en fonction de l'organisation et du type de données détenues, elles doivent toutefois couvrir plusieurs domaines essentiels. Tenez compte de différents scénarios d'incident de confidentialité lorsque vous élaborez votre politique.

Questions clés à poser :

- **Gestion des incidents :** quelles sont les prochaines étapes immédiates? Quels sont les rôles des membres du personnel et qui est responsable de la prise de décision?
- **Notification :** qui doit être informé et quand? Par exemple, lorsque le personnel découvre qu'une fuite de données a eu lieu, qui doit-il informer et comment?
- **Examen :** comment allez-vous déterminer ce qui n'a pas fonctionné et quelles mesures peuvent être mises en place pour atténuer le risque de récidive?
- **Communication :** comment communiquerez-vous avec vos parties prenantes au sujet de l'incident? Incluez vos obligations légales en matière de communication avec les parties prenantes, le public et le gouvernement.

Remarque : en vertu de la Loi 25 du Québec, en cas d'incident de confidentialité impliquant des renseignements personnels, les organisations sont généralement tenues d'informer la Commission d'accès à l'information du Québec ainsi que les personnes concernées si la violation présente un « risque de préjudice grave ».

Surveillance et maintenance

L'anonymisation n'est pas une opération ponctuelle. Au fil du temps, les nouvelles technologies, les nouveaux ensembles de données ou les modifications législatives peuvent accroître les risques de réidentification.

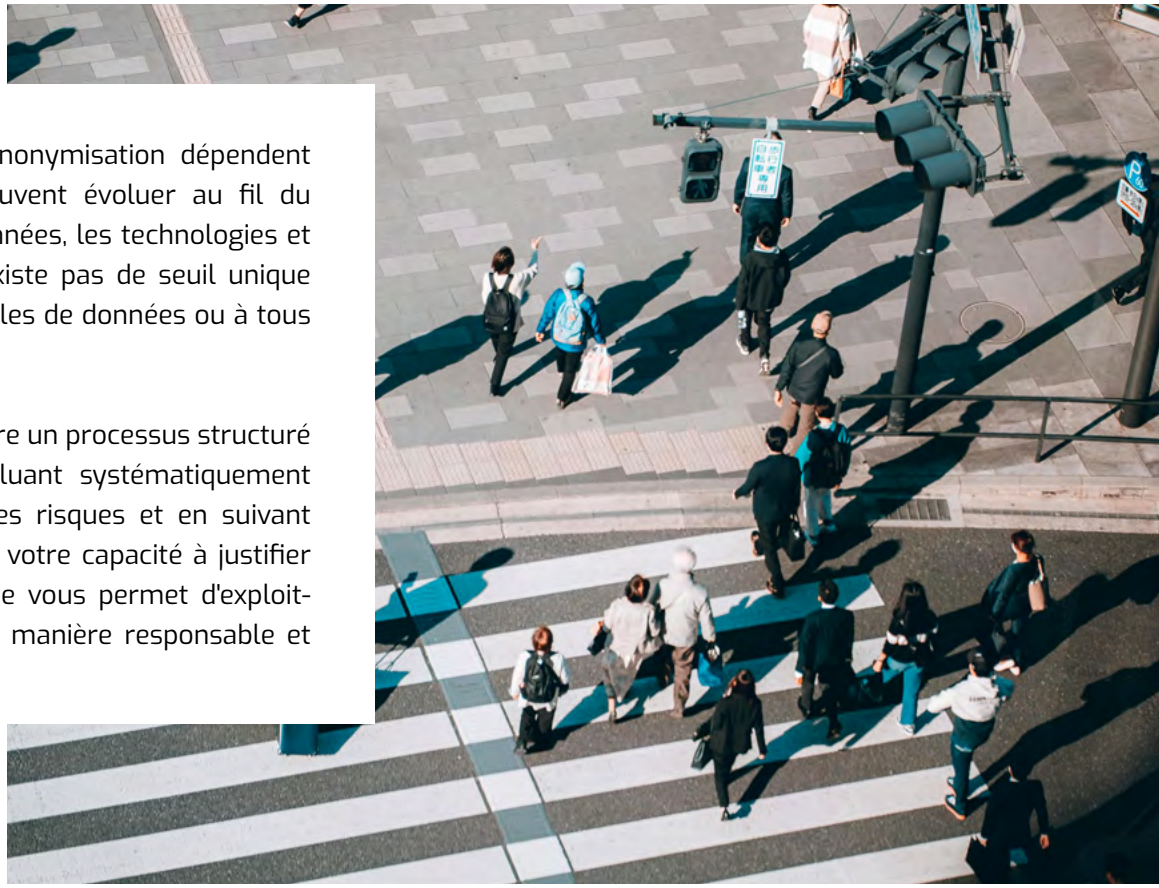
Les organisations doivent disposer d'un plan pour examiner régulièrement leur situation en matière de données et leurs décisions d'anonymisation.

Questions clés à se poser :

- Disposons-nous d'un processus permettant de suivre les changements susceptibles d'avoir une incidence sur les risques, tels que les nouveaux ensembles de données ou les nouvelles techniques d'analyse?
- À quelle fréquence réévaluons-nous les ensembles de données anonymisées?
- Tenons-nous un registre des données qui ont été partagées ou divulguées?
- Nos systèmes et nos canaux de communication fonctionnent-ils toujours comme prévu?

Les systèmes qui prennent en charge l'anonymisation de vos données doivent être entretenus et les moyens permettant aux parties prenantes de communiquer avec vous au sujet des données doivent rester actifs.

CONCLUSION



Les décisions relatives à l'anonymisation dépendent toujours du contexte et peuvent évoluer au fil du temps, à mesure que les données, les technologies et les attentes changent. Il n'existe pas de seuil unique applicable à tous les ensembles de données ou à tous les cas d'usage.

Le plus important est de suivre un processus structuré et bien documenté. En évaluant systématiquement le contexte, en contrôlant les risques et en suivant les impacts, vous consolidez votre capacité à justifier vos décisions. Cette approche vous permet d'exploiter la valeur des données de manière responsable et transparente.

Commencez dès aujourd'hui

Pour commencer à documenter votre démarche, vous pouvez utiliser la liste de contrôle jointe à ce guide. Cet outil vous aidera à poser les questions pertinentes pour évaluer votre contexte et déterminer comment l'anonymisation peut protéger la vie privée tout en préservant l'utilité des données.

Si vous avez besoin d'aide pour mettre en œuvre un cadre décisionnel en matière d'anonymisation ou pour adapter ces lignes directrices à votre contexte spécifique, Nord Ouvert peut vous aider grâce à son service d'accompagnement ciblé.

OUTIL : LISTE DE CONTRÔLE POUR L'ÉVALUATION DE LA SITUATION DES DONNÉES

Décrire l'environnement de données

- [ ] Qui contrôle ou gère actuellement ces données?
- [ ] Qui les recevra ou y aura accès ensuite?
- [ ] Les données seront-elles utilisées uniquement par le personnel interne, partagées avec des partenaires ou rendues publiques?
- [ ] Où les données seront-elles stockées ou accessibles (par exemple, sur des serveurs internes sécurisés, des plateformes infonuagiques ou des sites Web publics)?
- [ ] Quels sont les contrôles de gouvernance en place, tels que les contrats, les restrictions d'accès ou les accords de partage de données?

Connaître vos données

Personnes concernées

- [ ] Qui sont les personnes représentées dans cet ensemble de données?
- [ ] Ces personnes ont-elles la possibilité de choisir d'utiliser, ou non, le service par lequel leurs données sont collectées?
- [ ] Quelles attentes peuvent-elles raisonnablement avoir quant à l'utilisation de leurs données?

Types de données

- [ ] Cet ensemble de données est-il constitué d'enregistrements individuels (microdonnées) ou de statistiques agrégées?
- [ ] Les mêmes objectifs analytiques pourraient-ils être atteints en utilisant des données plus agrégées?

Types de variables

- [ ] L'ensemble de données comprend-il des identifiants directs?
- [ ] Quels identifiants indirects sont présents, tels que des emplacements précis ou des horodatages?
- [ ] L'ensemble de données comprend-il des attributs sensibles qui pourraient être révélateurs même si les identités ne sont pas explicites?

Propriétés des données

- [ ] Quel est le degré de précision des données, en particulier en ce qui concerne l'heure et l'emplacement?
- [ ] Les données sont-elles récentes?
- [ ] L'ensemble de données couvre-t-il l'ensemble de la population ou seulement un sous-ensemble?

Analyser les cas d'usage

- [ ] Pourquoi voulons-nous partager, publier ou réutiliser ces données?
- [ ] Qui est susceptible de les utiliser, par exemple les équipes internes, les partenaires, les chercheurs ou le grand public?
- [ ] Quels types d'analyse ou de réutilisation l'environnement de données prend-il réellement en charge?
  - Partage des données avec une autre partie;
  - Diffusion des données (publication ou divulgation);
  - Utilisation continue des données au-delà d'une période de conservation préalablement définie;
  - Réutilisation par la même organisation des données à des fins autres que celles pour lesquelles elles ont été collectées.

Définir le cadre juridique

- [ ] Quelle loi sur la protection de la vie privée s'applique à notre organisation?
- [ ] Avons-nous clairement défini un objectif sérieux et légitime pour l'anonymisation?
- [ ] Une personne compétente a-t-elle été désignée pour superviser le processus d'anonymisation?
- [ ] Les identifiants directs ont-ils été supprimés avant de procéder à l'analyse des risques?
- [ ] Documentons-nous les techniques et les décisions dans un registre d'anonymisation?

Évaluer les enjeux de consentement et vos obligations éthiques

- [ ] Dans quelle mesure cette utilisation des données correspond-elle aux attentes initiales des utilisateurs?
- [ ] Sommes-nous clairs et transparents sur la manière dont les données de mobilité sont analysées et partagées?
- [ ] Cette utilisation des données pourrait-elle affecter de manière disproportionnée certains groupes, tels que les personnes handicapées?
- [ ] Les utilisateurs ont-ils un choix significatif ou existe-t-il des déséquilibres de pouvoir dus à la nature du service?
- [ ] Cette utilisation des données sert-elle clairement l'intérêt public?

À propos de Nord Ouvert

Renforcer la confiance dans les données, pour le bien commun

Nord Ouvert est une organisation à but non lucratif qui se consacre à l'avancement du bien commun. Aux côtés des gouvernements et des organisations à vocation civique de toutes tailles, nous fournissons une expertise en matière de données afin d'éclairer la prise de décision, de stimuler l'innovation, d'améliorer les services publics et les services offerts par la société civile, et de relever les défis les plus urgents de la société.

Notre travail consiste à renforcer la capacité des organisations, à prendre de meilleures décisions concernant la gestion de leurs données afin qu'elles soient utiles, exploitables, sécurisées et dignes de confiance tout au long de leur cycle de vie. Chez Nord Ouvert, nous combinons une expertise approfondie en matière de données avec une approche multidisciplinaire. Notre équipe est composée d'urbanistes, d'ingénieurs en logiciel, d'organisateurs communautaires, de scientifiques des données, de spécialistes en vérification cybernétique et des technologies de l'information, de sociologues, de géographes et de juristes spécialisés dans les technologies, ce qui nous permet d'apporter des perspectives diverses à chaque projet.

Nord Ouvert fait partie de Montréal en commun, un projet mené par la Ville de Montréal dans le cadre du Défi des villes intelligentes, réalisé avec le soutien financier du gouvernement du Canada.

opennorth.ca/fr

À propos du Défi des villes intelligentes et de Montréal en commun

Montréal en Commun est une communauté d'innovation pilotée par la Ville de Montréal dont les partenaires expérimentent des solutions en accès à l'alimentation, en mobilité et en réglementation municipale dans un désir de repenser la ville. Les projets sont mis en œuvre grâce au prix octroyé à la Ville de Montréal par le Gouvernement du Canada dans le cadre du Défi des villes intelligentes.

Auteurs: John Griffin et Steven Coutts

Canada Montréal DÉFI DES VILLES INTELLIGENTES Montréal en commun

Anonymisation des données de mobilité