Anonymisation des données personnelles : un enjeu de taille, notamment en matière de santé

Alors qu'une deuxième vague de coronavirus menace, le gouvernement croit en l'utilité des données « pseudonymisées » de son application mobile StopCovid malgré le peu d'utilisateurs. Mais le respect de la vie privée ne suppose-t-il pas une « anonymisation » ? Le dilemme se pose dans la santé.

Par Olivia Roche, avocate, et Prudence Cadio, avocate associée, cabinet LPA-CGR avocats

La crise sanitaire liée au covid-19 et le développement concomitant des outils de surveillance de l'évolution de l'épidémie ont mis en lumière les enjeux liés à l'anonymisation des données à caractère personnel et, en particulier, des données de santé. Souvent présenté par la Commission nationale de l'informatique et des libertés (Cnil) comme un moyen indispensable pour préserver la vie privée des personnes, le procédé d'anonymisation aboutit cependant nécessairement à une perte d'informations, parfois contestée par les professionnels de santé.

Pseudonymisation ou anonymisation ?

Les recommandations publiées le 19 mai dernier par la Cnil (1) à ce sujet et les débats entourant l'application mobile StopCovid – mise à disposition par le gouvernement dans le cadre de sa stratégie globale de « déconfinement progressif » – permettent de mieux appréhender ces problématiques. Si cette application mobile ne dispose d'aucune information directement identifiante comme le nom ou le prénom, elle n'est pas pour autant « anonyme » au sens de la règlementation relative à la protection des données (2). La confusion entre « données pseudonymes » et « données anonymes » demeure répandue, alors que le règlement général européen sur la protection des données (RGPD) – en vigueur depuis le 25 mai 2018 – a entériné la distinction entre ces deux notions en son considérant 26.

Comme le rappelle la Cnil, la pseudonymisation consiste à traiter les données personnelles de façon à ce que celles-ci ne puissent plus être attribuées à une personne concernée sans avoir recours à des informations complémentaires. De manière plus concrète, ce processus consiste par exemple à remplacer des données personnelles directement identifiantes telles que le nom ou le prénom par des données indirectement identifiantes telles qu'un alias, un numéro ou un code. La pseudonymisation constitue ainsi un outil utile pour conserver des données tout en préservant la vie privée des personnes, puisque celles-ci ne sont plus directement identifiantes. Néanmoins, l'opération de pseudonymisation étant réversible, il est possible de réidentifier ou identifier indirectement les personnes sur la base de ces données. En conséquence, les « données pseudonymes » demeurent des « données personnelles » auxquelles s'applique l'ensemble des exigences de la règlementation sur la protection des données personnelles. Au contraire, les « données anonymisées » au sens du RGPD exclut toute possibilité de réidentification des personnes. Il s'agit d'appliquer un procédé aux données personnelles pour rendre toute individualisation et toute indentification, directe ou indirecte, impossible et ce de manière irréversible et définitive.

Cette distinction constitue un enjeu central, puisque les « données anonymes » ou « rendues anonymes », lesquelles, contrairement aux « données pseudonymes », ne sont pas ou plus des « données personnelles », ne sont pas soumises aux exigences du RGPD. Ce règlement européen indique, en effet, expressément qu'il ne s'applique *« pas au traitement de telles informations anonymes, y compris à des fins statistiques ou de recherche »* (3). A cet égard, il faut veiller à distinguer que, lorsqu'un procédé d'anonymisation est appliqué, c'est bien uniquement le résultat obtenu – les « données anonymisées » – qui peut être exclu du champ d'application matériel du RGPD mais non les données à caractère initialement collectées. De même, le processus d'anonymisation constitue un « traitement » qui, effectué sur des données personnelles, n'échappe pas en tant que tel aux exigences de la règlementation sur la protection des données à caractère personnel. Quelles sont au juste les techniques d'anonymisation ?

Randomisation et généralisation

Dans un avis en date du 10 avril 2014, le groupe dit de l'Article 29 – ce « G29 » ayant été remplacé depuis l'entrée en application du RGPD par le Comité européen de la protection des données (4) – proposait trois critères pour s'assurer que des données personnelles faisaient bien l'objet d'un procédé d'anonymisation et non de pseudonymisation : l'individualisation (il doit être impossible d'isoler un individu dans l'ensemble de données), la corrélation (il ne doit pas être possible de

Notes

(1) - https://lc.cx/Cnil19-05-20

(2) - https://lc.cx/FaqStopCovid

(3) - RGPD, Considérant 26 : https://lc.cx/DP-16

(4) - https://edpb.europa.eu/edpb_fr

(5) - https://lc.cx/G29Anonymisation2014

(6) - https://lc.cx/Cnil-FNMF-2004

relier deux ensembles distincts de données concernant un même individu) et l'inférence (il doit être impossible de déduire une information sur un individu). Pour éliminer toute possibilité d'identification, la Cnil rappelle que deux grandes techniques d'anonymisation sont possibles. La « randomisation » qui consiste à rendre moins précises les données, par exemple en permutant certaines informations dans l'ensemble de données tout en conservant la répartition globale. La seconde technique dite de « généralisation » consiste quant à elle à modifier l'échelle ou l'ordre de grandeur des données (par exemple en ne conservant que l'année de naissance au lieu de la date précise) afin d'éviter l'individualisation ou la corrélation avec d'autres ensembles de données. Ces méthodes d'anonymisation doivent cependant être réévaluées régulièrement car les techniques et possibilités de réidentification évoluent rapidement, à mesure des avancées technologiques.

Impacts sur la vie privée

A cet égard, dans son avis « Techniques d'anonymisation » (5), le G29 indiquait déjà que *« le résultat de l'anonymisation, en tant que technique appliquée aux données à caractère personnel, devrait être, dans l'état actuel de la technologie aussi permanent qu'un effacement, c'est-à-dire qu'il devrait rendre impossible tout traitement de données à caractère personnel »*. En effet, des données publiées comme « anonymes » à un instant T peuvent, grâce par exemple à une nouvelle technique développée par un tiers, redevenir indirectement identifiantes. Leur publication à titre de « données anonymes », sans veiller au respect du RGPD, pourrait ainsi constituer une violation de données. Si ces procédés d'anonymisation permettent de conserver et de réutiliser des données pour des durées étendues tout en assurant le respect des droits et libertés des personnelles, reste la question de l'utilité de données anonymes, notamment dans le secteur de la recherche scientifique et médicale. Comme le démontrent les débats entourant les traitements de données personnelles mis en œuvre par le biais de l'application mobile StopCovid, l'intérêt scientifique des données anonymes – qui ont perdu tout caractère individualisant – est plus limité.

Depuis toujours la problématique d'anonymisation est très présente dans le secteur de la santé. En effet, les données relatives à la santé des personnes constituent à la fois des données personnelles particulièrement risquées en termes d'impacts sur la vie privée, mais elles constituent également un enjeu important dans le cadre de la recherche scientifique et médicale. Par exemple, dès 2004, la Cnil s'était prononcée sur la volonté de la Fédération nationale de la mutualité française (FNMF) – regroupant 540 mutuelles adhérentes dont 266 mutuelles santé – d'avoir accès sous un format anonymisé à des données figurant sur des feuilles de soins électroniques. Ce traitement devait être mis en œuvre à des fins statistiques pour étudier l'impact d'un remboursement en fonction du service médical rendu pour les médicaments. La Cnil avait autorisé le traitement en donnant des précisions et recommandations strictes sur les modalités d'anonymisation des données, les mesures de sécurité et le respect des droits des personnes concernées (6).

Dans cette lignée, fin avril 2020, la Cnil s'est prononcée favorablement à l'application mobile StopCovid déployée par le gouvernement dans le cadre de sa stratégie de déconfinement progressif – sous réserve que les données personnelles collectées soient traitées sous un format pseudonymisé, puis supprimées de 15 jours ou 6 mois selon les catégories. *« La* [Cnil] *prend acte de ce que l'article 4 du projet de décret* [décret du 29 mai 2020 publié au *J.O.* du 30 mai dernier (7), ndlr] *prévoit une conservation des clés et des identifiants associés aux applications pendant la durée de fonctionnement de l'application StopCovid et au plus tard six mois à compter de la fin de l'état d'urgence sanitaire, et une conservation des historiques de proximité des personnes diagnostiquées ou testées positives pendant quinze jours à compter de leur émission »* (8).

Cependant, le 21 juin dernier, le Conseil scientifique covid-19 a publié un avis afin d'indiquer qu'il considérait essentiel d'appliquer l'option prévue par l'article 2 du projet de loi organisant la sortie de l'état d'urgence sanitaire permettant de conserver les données personnelles collectées par StopCovid pour une durée plus longue. Ce conseil scientifique (9) précise en outre que ces données devraient être conservées *« sous une forme pseudonymisée et non simplement anonymisée, de façon à ce que les données d'un même individu, non-identifiantes, puissent tout de même être reliées entre elles (ex : documentation d'une ré-infection), ou chaînées avec des données d'autres bases »* (10). Cette position illustre parfaitement les enjeux liés à l'articulation entre exploitation des données, durées de conservation et anonymisation ou pseudonymisation.

Risque de seconde vague

La position de la Cnil et l'arbitrage qui sera opéré entre respect de la vie privée et le nécessaire suivi de l'épidémie de covid-19, en particulier avec le risque d'une seconde vague, permettra d'étayer davantage les critères d'application et la distinction entre pseudonymisation et anonymisation. Au 23 juin 2020, soit en trois semaines d'existence de StopCovid, seuls quatorze cas à risque de contamination ont été détectés par l'application mobile. A cette date de premier bilan, elle a été téléchargée 1,9 million de fois, mais désinstallée 460.000 fois. Le gouvernement se dit néanmoins convaincu de son utilité, surtout en prévision de cette seconde vague. @

Notes

(7) - https://lc.cx/DécretStopCovid29-05-20

(8) - https://lc.cx/Cnil-StopCovid-2020

(9) - https://lc.cx/CSCI9

(10) - https://lc.cx/CSCI9-Note21-06-20