Information and Privacy
Commissioner of Ontario
Commissaire à l'information et à la
protection de la vie privée de l'Ontario

Sous réserve de modifications

Discours de Patricia Kosseim, commissaire à l'information et à la protection de la vie privée de l'Ontario
Sommet sur les données synthétiques 2025
16 mai 2025

Valoriser les données tout en protégeant la vie privée

Introduction

- Bonjour. Merci de votre invitation, c'est un plaisir de me joindre à vous.

- Je suis ravie de me retrouver ici à Montréal, ma ville natale qui me rappelle de si beaux souvenirs de ma jeunesse.

- Ce sommet est une excellente occasion d'explorer comment la production de données synthétiques peut contribuer à résoudre des défis sociaux parmi les plus complexes de notre époque.

- Des technologies nouvelles et émergentes transforment le monde à un rythme effréné et d'une manière que nous n'aurions jamais imaginée.

- Les données et les technologies numériques font désormais partie intégrante de presque tous les aspects de notre quotidien.

- Elles mettent à notre portée les connaissances et les informations dont nous avons besoin pour améliorer considérablement notre santé, notre économie et le bien-être de notre société.

- Les progrès technologiques permettent désormais aux organisations de recueillir d'énormes quantités de données et d'exploiter tout leur potentiel pour favoriser l'innovation et améliorer la fourniture de biens, de services et de programmes.

- Les données sur les choses ne posent généralement pas de problème, à part les questions de propriété intellectuelle et de confidentialité, évidemment.

- Mais celles qui portent sur des gens suscitent de tout autres préoccupations.

- Les fuites de données font augmenter le risque d'atteinte au droit à la vie privée et minent la confiance du public en conséquence.

- Ce n'est pas d'hier que l'on a de la difficulté à accéder à ses renseignements personnels, surtout dans le secteur de la santé.

- Premièrement, il y a la question de l'identifiabilité, qui fait l'objet de critères très tranchés. En vertu des lois canadiennes sur la protection de la vie privée, dès qu'il est raisonnablement possible d'identifier un particulier à partir de données prises seules ou avec d'autres, ces données sont considérées comme des renseignements personnels et sont assujetties à toutes les exigences des lois sur la protection de la vie privée, qu'il est impossible d'assouplir en fonction du degré de risque.

- Il est donc souhaitable de dépersonnaliser les données en éliminant les identifiants directs et indirects, afin de les soustraire à l'application de la loi et d'en faire ce que l'on veut. Mais si les données ne sont pas dépersonnalisées correctement, en tenant compte du contexte, des particuliers pourraient être réidentifiés, ce qui pourrait se solder par une grave atteinte à la vie privée. Inversement, les données pourraient perdre toute leur utilité si elles sont dépersonnalisées à l'extrême.

- À cela s'ajoute un autre problème de longue date, celui du consentement. Obtenir le consentement pour recueillir et utiliser des renseignements identificatoires peut se révéler irréaliste quand un grand nombre de personnes sont en cause et qu'il est impossible de les joindre. Ou bien, s'il vous est possible d'obtenir le consentement, vous devez composer avec le biais de consentement : en effet, les personnes qui acceptent que l'on accède à leurs renseignements personnels sur la santé et les utilise présentent des caractéristiques différentes de celles qui refusent, ce qui introduit un biais dans vos données.

- Parfois, la loi vous permet d'utiliser des renseignements personnels sur la santé sans consentement sous réserve de l'approbation d'un tiers, comme une commission d'éthique de la recherche. Mais on sait que ce n'est pas une tâche facile. Il faut d'abord déterminer s'il est nécessaire d'obtenir d'approbation d'une telle commission. Et si c'est le cas, il faut ensuite composer avec des commissions multiples de plusieurs territoires de compétence, dont bon nombre n'ont ni la capacité ni le temps nécessaire pour vous donner une réponse très réfléchie en temps opportun.

La génération de données synthétiques

- Mais si vous pouviez éviter tout ça en utilisant des données synthétiques, qui ne portent sur aucune personne en particulier? Pourriez-vous profiter de tous les avantages des données sans risque pour la vie privée? C'est essentiellement pour répondre à cette question que nous sommes réunis ici aujourd'hui.

- Contrairement à vous, je ne suis pas une experte des données synthétiques. Mais je suis très favorable aux technologies d'amélioration de la confidentialité (TAC), comme la génération de données synthétiques, entre autres.

- Quand j'ai commencé mon mandat en tant que commissaire il y a près de cinq ans, je nous ai donné la vision de devenir un organisme de réglementation moderne et efficace qui exerce une influence concrète.

- Pour réaliser cette vision, mon personnel et moi devons adopter de nouvelles idées et des approches innovantes afin de favoriser l'utilisation responsable des données de façon positive pour notre santé, notre économie et notre société.

- En tant qu'organismes de réglementation, nous devons répondre à l'innovation par l'innovation, et trouver des solutions pragmatiques aux défis modernes liés aux données afin de conserver notre pertinence pour les organisations que nous supervisons.

- Nous avons besoin de nouvelles perspectives sur les méthodes et techniques à adopter pour protéger la vie privée dans les secteurs public et privé.

- D'après un sondage mené récemment par le Commissariat à la protection de la vie privée du Canada, les trois quarts des Canadiennes et Canadiens sont moins ou beaucoup moins enclins à communiquer des renseignements personnels aux organisations qu'il y a cinq ans.

- 78 % ont refusé de fournir des renseignements personnels à des organisations en raison d'inquiétudes liées à la protection de la vie privée, et 41 % ont cessé de faire appel à une organisation ayant subi une atteinte à la vie privée.

- Nous savons tous qu'il peut être difficile de conjuguer innovation et protection de la vie privée.

- Cependant, les technologies d'amélioration de la confidentialité, comme les données synthétiques, constituent une solution prometteuse pour réduire les risques pour la vie privée tout en profitant de tout le potentiel des données.

Le pouvoir des TAC

- En janvier dernier, mon bureau a tenu son événement annuel à l'occasion de la Journée de la protection des données. Il avait pour thème le pouvoir des technologies d'amélioration de confidentialité, les TAC.

- Un enregistrement de cet événement est accessible sur la chaîne YouTube du CIPVP. Si ce sujet vous intéresse, je vous recommande fortement de le regarder.

- Nous avons eu la chance d'accueillir des experts technologiques de renom des secteurs public et privé, des universités, du droit et de groupes de défense de la vie privée pour une discussion approfondie et fascinante.

- Cet événement portait sur l'utilisation des TAC pour favoriser la recherche et l'innovation en permettant aux organisations d'utiliser des données précieuses tout en protégeant les renseignements personnels.

- Les panélistes ont évoqué différentes technologies d'amélioration de la confidentialité : la dépersonnalisation, la génération de données synthétiques, l'apprentissage fédéré et la confidentialité différentielle, leur fonctionnement, leurs risques et avantages et leurs similitudes et différences.

- Nous nous sommes familiarisés avec des exemples d'utilisation de ces TAC dans le secteur public, le secteur privé et le secteur de la santé, et nous avons souligné que pour réussir, les organisations doivent chercher avant tout à susciter la confiance du public.

- Nous avons également appris que certains territoires de compétence, au lieu de considérer les renseignements comme étant personnels ou non, sans nuances, commencent à envisager leur nature identificatoire comme se situant sur un continuum.

- En vertu de modifications aux lois sur la protection de la vie privée qui ont été adoptées récemment ou sont envisagées au Canada et ailleurs, on commence à réglementer les renseignements personnels en fonction du risque d'identifiabilité au lieu de diviser les renseignements en catégories strictes qui ne sont plus adaptées au monde numérique complexe d'aujourd'hui.

- Khaled El Emam, chercheur résident du CIPVP et principal organisateur du sommet d'aujourd'hui, a participé à cet événement.

- Il a expliqué que des ensembles de données synthétiques produits à partir de renseignements personnels peuvent sembler réels sur le plan de leurs propriétés et tendances statistiques générales, mais qu'ils varient suffisamment au niveau individuel de sorte qu'ils ne révèlent pas l'identité des gens sur lesquels ils sont fondés. De plus en plus, ces ensembles de données artificiels sont générés à grande échelle pour entraîner des grands modèles de langage et faire avancer des recherches importantes dans le domaine de la santé.

- J'ai également reçu Khaled à mon balado *L'info, ça compte* pour un épisode sur les données synthétiques intitulé *Les données synthétiques : vraies ou fausses?*. Si vous voulez l'écouter, il s'agit du premier épisode de la deuxième saison de notre balado *L'info, ça compte*, accessible partout où vous téléchargez vos balados.

Avancées internationales sur le plan des TAC

- Dans le monde, des autorités de protection des données appuient l'adoption de technologies d'amélioration de la confidentialité comme moyens raisonnables de réduire les risques pour la vie privée.

- Cependant, en tant qu'organismes de réglementation, pour vraiment joindre le geste à la parole et favoriser l'adoption de TAC, nous devons fournir des conseils pratiques aux organisations de toute taille afin de réduire l'incertitude, de susciter la confiance et d'inspirer les innovateurs à aller de l'avant.

- Heureusement, un nombre croissant d'organismes de réglementation, au pays et à l'étranger, publient des outils et des documents d'orientation pour favoriser l'adoption responsable de TAC.

- À Singapour, l'Infocomm Media Development Authority a publié un document d'orientation sur la génération de données synthétiques pour aider les organisations à comprendre les techniques et les utilisations éventuelles de cette technologie.

- Cet organisme a également mis sur pied un bac à sable de technologies d'amélioration de la confidentialité dans lequel il est possible d'élaborer, de mettre à l'essai et de valider en toute sécurité des TAC novatrices, pendant une période limitée, sous sa supervision.

- Cet organisme de Singapour jumelle également des organisations à des fournisseurs de TAC, accorde des subventions pour la mise en œuvre de projets pilotes et fournit un soutien réglementaire pour assurer la conformité.

- Le bac à sable de TAC de Singapour, et d'autres bacs à sable réglementaires, sont décrits dans un rapport de recherche publié récemment en collaboration avec mon bureau.

- Ce rapport est intitulé *Étude sur la mise en place d'un bac à sable réglementaire pour la protection de la vie privée en Ontario*; Teresa Scassa, Ph. D., et Elif Nur Kumru de l'Université d'Ottawa en sont les autrices.

- Vous pouvez le télécharger à partir du Carrefour de la recherche et de l'innovation de notre site Web pour en savoir plus.

- Une autre initiative internationale portant sur la génération de données synthétiques a été lancée par la commission de protection des renseignements personnels de Corée du Sud.

- Cette commission a publié un document d'orientation sur les méthodes et procédures à suivre pour créer et utiliser des données synthétiques tout en restant conforme aux lois sur la protection de la vie privée.

- Et le bureau du commissaire à l'information du Royaume-Uni a publié un document d'orientation sur les TAC à l'intention des responsables de la protection des données.

- Ce document explique les divers types de TAC, y compris les données synthétiques. Il reconnaît que certaines des vulnérabilités des données réelles, comme les biais, peuvent se retrouver dans les données synthétiques à moins d'être retirées des données d'origine.

- Il précise également que l'évaluation des risques de réidentification associés aux données synthétiques est toujours à l'étude.

Les TAC en Ontario

- Chez nous, je suis particulièrement fière d'affirmer que l'Ontario a fait figure de pionnier dans le domaine des technologies d'amélioration de la confidentialité.

- Dès 1995, le CIPVP a présenté pour la première fois le concept de technologie d'amélioration de la confidentialité dans un rapport rédigé conjointement avec le prédécesseur de l'Autorité de protection des données des Pays-Bas.

- Ce rapport décrit comment des TAC peuvent être intégrées dans la conception de nouvelles technologies de l'information afin d'accroître le recours à des données anonymisées pour des opérations quotidiennes.

- Depuis, nous avons poursuivi notre travail dans ce domaine, en publiant des documents d'orientation, en rendant des ordonnances et des décisions et en réclamant des modifications et cadres réglementaires.

- Par exemple, en 2016, nous avons publié nos lignes directrices sur la dépersonnalisation des données structurées primées à l'échelle internationale.

- Nous continuons de parfaire notre expertise dans ce domaine, et nous avons entamé la mise à jour de notre document sur la dépersonnalisation en tenant compte des progrès technologiques réalisés au cours des dix dernières années.

- Nous collaborons avec Khaled El Emam et nous consultons des experts d'un éventail de secteurs pour recueillir leurs commentaires. Nous comptons publier notre document d'orientation mis à jour sur la dépersonnalisation au cours des prochaines semaines.

- Suivez-nous sur les médias sociaux pour en savoir plus sur ce document et sa date de publication.

- En plus de publier des documents d'orientation, notre bureau fournit des conseils et des recommandations sur les textes de loi, programmes, politiques de confidentialité et pratiques de gestion de l'information des pouvoirs publics.

Règlement pris en application de la LPRPS sur la dépersonnalisation

- Sur le plan législatif, la *Loi de 2004 sur la protection des renseignements personnels sur la santé* de l'Ontario, la LPRPS, a été modifiée en 2020 afin de permettre la prise d'un règlement qui définirait la dépersonnalisation (appelée « anonymisation » dans la loi) en établissant des exigences plus précises.

- Aucun règlement n'a encore été adopté à ce jour.

- Par souci de certitude et de clarté, nous avons recommandé au ministère un libellé précis pour un règlement possible concernant la dépersonnalisation des renseignements personnels sur la santé.

- Il s'agissait pour nous d'une occasion de fournir aux dépositaires de renseignements sur la santé des règles claires afin de favoriser l'échange plus responsable de données.

- Plus précisément, nous avons formulé les recommandations suivantes :

- Le processus de dépersonnalisation des renseignements personnels sur la santé doit :

    - être effectué par une personne qualifiée dans ce domaine, justifiant de connaissances et d'une expertise appropriées, ou sous sa supervision;

    - s'appuyer sur des pratiques exemplaires généralement reconnues;

    - évaluer le risque de réindentification;

    - comprendre des étapes qui sont raisonnables dans les circonstances pour s'assurer que le risque de réidentifier un particulier est très faible.

- S'il faut divulguer des renseignements dépersonnalisés à une personne en particulier, une entente écrite doit être conclue interdisant à cette personne de tenter de réidentifier les données et l'obligeant à prendre des mesures raisonnables pour s'assurer que le risque de réidentification demeure très faible.

- Les données dépersonnalisées doivent être réévaluées périodiquement en tenant compte de certains facteurs prescrits pour s'assurer que le risque de réidentification demeure très faible.

- Certains aspects du processus de dépersonnalisation doivent être documentés, et un avis raisonnable doit être fourni aux particuliers dont les renseignements personnels sur la santé sont dépersonnalisés, précisant les fins générales auxquelles ils pourraient être utilisés ou divulgués.

- Nous avons cherché à aligner nos recommandations sur un règlement adopté récemment au Québec en application de la loi 25, la *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*, qui établit des exigences précises en matière d'anonymisation des données, une première au Canada.

- Nous avons également cherché à nous aligner sur nos exigences établies dans notre guide d'examen et d'approbation des personnes prescrites et des entités prescrites.

- Ces organismes prescrits comprennent certains des plus grands dépositaires de données sur la santé en Ontario, qui tiennent des registres de renseignements personnels sur la santé pour l'amélioration de soins de santé, ou qui analysent ou compilent des renseignements statistiques à l'égard de la gestion, de l'évaluation, de la surveillance ou de la planification du système de santé.

- Ces organismes comprennent Santé Ontario, l'ICES, l'Institut canadien d'information sur la santé (ICIS) et Action Cancer Ontario.

- Nos recommandations vont également dans le sens des normes d'intégration des données établies en vertu de notre loi s'appliquant au secteur public qui permet à des services prescrits d'intégration des données d'établir des liens entre des données émanant de différents ministères et qui autorise l'accès à des ensembles de données dépersonnalisées pour :

    - la gestion ou l'affectation de ressources;
    - la planification de la prestation de programmes et de services fournis ou financés par le gouvernement de l'Ontario;
    - l'évaluation de ces programmes et services.

- Mentionnons comme exemple le service interministériel d'intégration des données du ministère de la Santé de l'Ontario.

- Enfin, nous voulions nous assurer que notre règlement recommandé était aligné sur nos enquêtes et décisions antérieures sur la dépersonnalisation dans le secteur de la santé, dont deux décisions rendues en vertu de la LPRPS :

Décisions rendues par le CIPVP en vertu de la LPRPS concernant la dépersonnalisation

- La première, la Décision 243 en vertu de la LPRPS, portait sur un groupe de chercheurs de l'Université de Toronto et leur base de données de recherche UTOPIAN, qui contient des données dépersonnalisées sur des patients qui ont été extraites de dossiers médicaux électroniques de nombreux médecins de première ligne participants.

- Un groupe de médecins a déposé une plainte anonyme alléguant, entre autres choses, que :

    - le processus de dépersonnalisation employé par les chercheurs était inadéquat, et qu'il y avait un risque de divulgation de renseignements éventuellement identificatoires contenus dans la base de données à d'autres chercheurs voulant accéder aux données des dossiers médicaux électroniques.

- Notre enquête n'a pas révélé de lacune dans les processus de dépersonnalisation. Cependant, nous avons recommandé à l'université :

    - de faire preuve de plus de transparence et d'une communication plus ouverte avec les dépositaires pour gagner leur confiance dans leurs travaux de recherche;

    - d'effectuer une évaluation des risques de réidentification en suivant les pratiques exemplaires des lignes directrices sur la dépersonnalisation de données structurées du CIPVP.

- Nous avons également constaté que la plateforme de recherche était demeurée en service après l'expiration de l'approbation de la commission d'éthique de la recherche, ce qui était très préoccupant.

- En définitive, s'il est vrai que la recherche est essentielle pour améliorer la qualité des soins et l'efficacité de notre système de santé, les citoyens doivent avoir l'assurance que l'on protège leurs renseignements personnels sur la santé.

- La seconde affaire, la Décision 175 en vertu de la LPRPS, comportait une enquête sur la vente de données dépersonnalisées par un groupe de cliniques médicales à une entreprise tierce.

- Nous avons lancé cette enquête après avoir pris connaissance de cette situation dans les médias.

- Trois leçons sont à tirer de cette enquête :

- Premièrement, nous avons conclu qu'en vertu de la LPRPS, l'utilisation de renseignements personnels sur la santé en vue de les dépersonnaliser est autorisée sans le consentement du particulier concerné lorsque des conditions précises énoncées dans la loi sont respectées.

- Deuxièmement, les dépositaires de renseignements sur la santé doivent faire preuve de transparence quant à leurs pratiques relatives aux renseignements. S'ils vendent des données dépersonnalisées à un tiers à quelque fin que ce soit, y compris pour des recherches en santé, ils doivent le mentionner clairement et explicitement dans un avis au public.

- Troisièmement, tout contrat de vente entre un dépositaire de renseignements sur la santé et un tiers doit comprendre des mesures de contrôle adéquates de la protection de la vie privée et de la sécurité afin que les données dépersonnalisées remises au tiers restent dépersonnalisées.

Contrats avec des fournisseurs externes

- Depuis la Décision 175 en vertu de la LPRPS, nous avons publié un document d'orientation sur les contrats à l'intention des institutions publiques, y compris les organismes de soins de santé.

- Je recommande aux personnes intéressées de consulter ce document intitulé *La protection de la vie privée et l'accès à l'information dans les contrats du secteur public avec des fournisseurs externes*.

- Il contient des conseils pratiques pour relever les facteurs touchant l'accès à l'information et la protection de la vie privée dont il faut tenir compte dans ses contrats avec des fournisseurs externes.

- Il comprend également des pratiques exemplaires et des recommandations pour faire preuve de diligence et de responsabilité à toutes les étapes du processus d'approvisionnement : planification, appel d'offres, sélection du fournisseur, contrat, gestion du contrat et résiliation.

- Comme je le répète souvent, une institution peut externaliser ses services ou le traitement de ses données, mais elle ne peut se soustraire à ses obligations.

Perspectives d'avenir

- Avant de conclure, j'aimerais vous faire part de certaines réflexions.

- Les technologies d'amélioration de la confidentialité sont des outils prometteurs pour tirer profit des données tout en protégeant la vie privée.

- Nous devons donc rendre les TAC accessibles, sécuritaires et faciles à adopter dans le but d'atténuer les risques pour la vie privée.

- Les TAC ne peuvent plus être considérées comme des outils technologiques complexes et inaccessibles que seules les grandes organisations sophistiquées peuvent se permettre d'utiliser.

- Ce sont plutôt des outils de protection essentiels qui doivent être accessibles à toutes les organisations, sans égard à leur taille, pour protéger la vie privée dans un monde axé sur les données.

- Tant les organismes de réglementation que les utilisateurs des données doivent relever ce défi afin de susciter auprès des citoyens et des consommateurs la confiance qui les motivera à participer de façon concrète au monde numérique.

- J'ai parlé du rôle des organismes de réglementation.

- Nous devons être ouverts à de nouvelles méthodes et techniques de protection de la vie privée afin de trouver des solutions pragmatiques aux défis modernes en matière de données et de demeurer pertinents pour les organisations que nous supervisons.

- En tant qu'organismes de réglementation, nous ne pouvons pas toujours jouer les rabat-joie. Nous devons nous adapter et répondre à l'innovation par l'innovation.

- Il revient à nous de fournir des conseils utiles et d'élaborer des lignes directrices uniformes pour favoriser l'adoption responsable de TAC par les organisations, et donner à celles-ci la latitude, la certitude et la stabilité dont elles ont besoin pour innover avec confiance.

- Et si vous me le permettez, j'ai quelques suggestions à vous faire, que bon nombre d'entre vous appliquent déjà au cours de l'événement d'aujourd'hui.

- Premièrement, en tant que producteurs et utilisateurs de données, il est important d'expliquer la génération de données synthétiques en termes simples que tout le monde peut comprendre. Comme l'a dit Einstein, « si vous ne pouvez expliquer quelque chose simplement, c'est que vous ne l'avez pas bien compris ». Les exemples pratiques sont très utiles pour décrire les avantages des données synthétiques en termes concrets afin de démystifier le processus et de favoriser l'adoption des outils, en sachant qu'il n'existe pas de solution universelle.

- Deuxièmement, faites preuve de transparence et expliquez très clairement les risques. Ces risques comprennent la divulgation de l'identité, des attributs et des associations. N'essayez pas de les minimiser. Élaborez des indicateurs pour

quantifier les risques de réidentification et cherchez sans relâche à les minimiser. Pour citer Einstein encore une fois, il a déjà dit à son sujet : « Ce n'est pas que je suis si intelligent, c'est que je reste plus longtemps avec les problèmes ».

- Et troisièmement, je vous recommande d'accepter l'invitation des autorités de protection des données de les consulter concernant vos approches novatrices et inédites.

- Que ce soit dans le contexte d'un bac à sable réglementaire officiel ou d'une consultation informelle sur les politiques, c'est cet engagement et cette collaboration qui nous aideront à comprendre nos positions respectives et à trouver les solutions créatives et pragmatiques dont nous avons besoin.

- Progressivement, par la collaboration, nous pouvons trouver des moyens pour les organisations de toutes sortes d'adopter des TAC pour contribuer à relever des défis parmi les plus complexes auxquels nous faisons face dans notre système de santé, notre économie et notre société.

- La confiance est la clé du succès, et pour la gagner, il faut être transparent et responsable.

- Les personnes qui savent que leurs données délicates sur leur santé sont traitées de façon responsable sont plus susceptibles de demander des services, d'adopter des produits et de contribuer à la recherche.

- Dans ce contexte, la conformité ne doit pas être considérée comme un obstacle, mais plutôt comme un catalyseur de meilleurs résultats.

- En investissant dans le pouvoir des TAC, nous pouvons profiter des avantages des données tout en protégeant la vie privée, et ouvrir la voie à de nouvelles perspectives pour tous.

- Merci.