Deloitte.

Gagner sur tous les tableaux

Protéger la vie privée à l'ère de l'analytique

Remerciements:

Les auteurs souhaitent remercier Megan Brister, Leader nationale, Protection de la vie privée; et Michelle Chibba, directrice, Politiques et projets spéciaux, IPC. Nous remercions également Catherine Thompson, conseillère en réglementation et politiques; David Weinkauf, chef des politiques et des technologies de l'information, CIPVP; ainsi que Michelle Gordon, Daniel Horovitz et Sylvia Kingsmill, de Deloitte, pour leur contribution à la rédaction du présent document.

sommaire

Les données massives, synonymes d'innovation

L'univers des données évolue de façon spectaculaire. Les entreprises, les gouvernements et les autres types d'organisation créent de la valeur en transformant l'information recueillie chaque jour en perspectives utiles. Deux grandes tendances sont au cœur de ce virage :

- Le nombre de données a explosé. Tous les deux jours, on crée dans le monde autant de nouvelles données que depuis le début de la civilisation jusqu'en 2003. *Tous les deux jours*. Voilà pourquoi on les appelle *données massives*.
- Avec les progrès réalisés en analytique des données, de grands ensembles de données structurées et non structurées peuvent être traités à très haute vitesse.

Notre capacité à faire des liens entre les données, à repérer des tendances et à personnaliser les interactions pour obtenir des résultats optimaux a atteint un degré que nous n'aurions jamais cru possible.

Et c'est bien là le problème.

La vie privée repose sur des renseignements personnels

L'analytique des données est si puissante qu'elle permet de combiner des ensembles de données pour en déduire le mode de vie d'une personne, ses habitudes de consommation, son utilisation des réseaux sociaux et bien d'autres renseignements encore – même si aucun de ces ensembles de données ne révèle à lui seul ces renseignements personnels.

Il n'est donc pas étonnant qu'on s'inquiète de l'incidence des données massives sur la protection de la vie privée. On craint que les protections fondamentales que nous tenions jadis pour acquises ne soient maintenant menacées par la vitesse, l'exactitude et le volume des données ainsi que par la façon dont elles peuvent être manipulées. Certains croient que notre perception de la confidentialité doit changer et que l'impératif de l'innovation et de la valeur contenue dans les données doit être privilégié par rapport à nos concepts traditionnels.

Or, le nécessaire compromis entre protection de la vie privée et innovation est une idée périmée qui ne saurait nous être utile. Nous croyons qu'il est tout à fait possible de protéger la vie privée tout en utilisant l'analytique des données afin d'acquérir de nouvelles connaissances et d'innover pour réaliser de vrais progrès.

Tout comme la technologie est à l'origine de l'analytique des données, elle peut aussi nous aider à résoudre les problèmes de protection de la vie privée qui en résulte.

Nous pouvons gagner sur tous les tableaux

La protection intégrée de la vie privée (PIVP) a pour but de concilier le besoin d'assurer une protection robuste des données et la volonté d'exploiter le potentiel de l'innovation fondée sur les données. Élaborée à la fin des années 1990 par Ann Cavoukian, Ph. D., commissaire à l'information et à la protection de la vie privée de l'Ontario, la PIVP intègre la protection de la vie privée directement dans les spécifications de conception des technologies, des pratiques d'affaires et de l'infrastructure en réseau.

Plusieurs options technologiques s'appuyant sur le cadre de la PIVP pour protéger la vie privée tout en facilitant l'analytique des données sont à la disposition des organisations, notamment :

- **Minimisation des données** : Aucun renseignement permettant d'identifier une personne n'est recueilli, à moins qu'un objectif spécifique et impérieux ne le justifie, ce qui élimine pratiquement tous les risques d'entrave à la vie privée dès le début.
- **Anonymisation** : On élimine d'un ensemble de données tous les renseignements pouvant servir à identifier une personne, que ce soit directement ou indirectement, par des associations avec d'autres ensembles de données.
- **Contrôles d'accès par les utilisateurs** : Ensemble de processus par lesquels on accorde ou on refuse les demandes spécifiques d'obtention de l'information; habituellement combinés à d'autres politiques en matière de sécurité.

Les données massives ne sont pas près de disparaître, mais cela ne signifie pas que nous devons sacrifier la protection de la vie privée ou cesser d'innover. Grâce à une planification rigoureuse et à l'application de techniques et de principes tels que ceux qui font partie de la protection intégrée de la vie privée, les entreprises peuvent utiliser les données pour combler leurs besoins tout en protégeant les renseignements personnels que ces données contiennent.

Les données massives favorisent la rupture

La quantité de données créées par les personnes, les appareils mobiles et les entreprises au moyen des recherches dans Internet, des médias sociaux, des données de localisation GPS, des transactions boursières, etc. augmente à une vitesse exponentielle. On dénombre actuellement 9,6 milliards d'appareils connectés à Internet[1] et 1,3 milliard de connexions mobiles à large bande[2] dans le monde. Tous les deux jours, notre utilisation de ces appareils crée environ 5 exaoctets (10^{18}) de données, soit l'équivalent de toutes les données créées par les humains depuis les débuts de la civilisation jusqu'en 2003.[3]

À l'ère des données massives, les données figurent parmi les actifs les plus précieux d'une entreprise. Leur analyse peut procurer des perspectives essentielles pour élaborer des stratégies, favoriser la croissance et le rendement opérationnel et en plus de gérer les risques.

Naturellement, les entreprises veulent exploiter pleinement le potentiel des données pour en extraire de la valeur. Elles cherchent des façons d'utiliser les données pour prendre des décisions plus avisées leur permettant d'offrir de meilleurs services à leurs clients, d'améliorer l'efficience de leurs processus et de tirer de meilleurs résultats.

Et elles le peuvent.

Les progrès rapides observés récemment permettent maintenant de traiter de grandes quantités de données structurées et non structurées à très haute vitesse. L'analytique des données accélère l'innovation et révolutionne les modèles d'affaires traditionnels. Par exemple :

- Les détaillants adaptent leur mise en marché aux préférences et aux comportements d'achat de leurs clients.
- Les entreprises de services financiers donnent des conseils et recommandent des produits avant même que les clients ne sachent qu'ils les désirent.
- Les organismes de soins de santé améliorent les diagnostics, les traitements et la gestion de la santé publique.
- Les gouvernements rendent leurs données accessibles au grand public afin d'accroître la transparence du gouvernement et d'encourager la participation du public.
- Dans certains secteurs, les concurrents partagent leurs données afin de relever des défis communs dans des domaines comme la fraude, la cybersécurité et le rendement des pratiques de santé et de sécurité.

De nos jours, l'analytique des données permet aux organisations de faire des liens, de repérer des tendances, de prédire des comportements et de personnaliser leurs interactions à un degré qu'elles n'auraient jamais cru possible il y a une décennie seulement.

Et c'est là que réside le problème.

[1] Internet connected devices approaching 10 billion, to exceed 28 billion by 2020, [En ligne], IMS Research, octobre 2012 http://www.cellular-news.com/story/Reports/56702.php.
[2] Communiqué de presse, GSMA Research Demonstrates that Mobile Industry is creating a Connected Economy, [En ligne], GSMA, 27 février 2012, http://www.gsma.com/newsroom/gsma-research-demonstrates-that-mobile-industry-is-creating-a-connected-economy/.
[3] M.G. Siegler, Eric Schmidt, Every 2 Days We Create As Much Information As We Did Up to 2003, [En ligne], TechCrunch, 4 août 2010, http://techcrunch.com/2010/08/04/schmidt-data/.

Un grand bouleversement entraînant de grands risques

De façon particulière, les organisations doivent se tenir à l'affût des menaces d'accès non autorisé aux données, surtout aux renseignements personnels. Ces menaces comprennent une atteinte à la réputation, une poursuite, des sanctions réglementaires et la perturbation des activités internes, sans oublier une diminution de la fidélité des clients se traduisant par une perte de revenus et de profits.

Toutefois, le plus grand risque que posent les données massives est celui de créer des associations automatiques entre des données qui, en apparence, ne permettent pas d'identifier une personne, mais qui peuvent en fournir un portrait général.

De puissantes solutions analytiques permettent de lier des ensembles de données pour connaître le mode de vie d'une personne, ses habitudes de consommation, son utilisation des réseaux sociaux et bien plus encore, et ce, même si aucun de ces ensembles ne révèle à lui seul ces renseignements personnels. Par exemple, un numéro de téléphone ou un code postal peuvent être combinés à d'autres données pour repérer le lieu où une personne vit et travaille; une adresse IP ou de courriel peut servir à déterminer ses habitudes et ses réseaux sociaux.

Voici d'autres exemples de risque :

- **La divulgation non autorisée, la perte ou le vol** de données représente clairement une menace pour la vie privée, et cette menace est naturellement plus grande quand les données massives contiennent des renseignements centralisés permettant d'identifier une personne. Dans les cas extrêmes, la divulgation non autorisée de renseignements personnels peut constituer une menace pour la sécurité publique.

- **Le *nudging*** est l'utilisation de données permettant d'identifier une personne pour brosser son portrait et analyser, prédire et modifier son comportement. Par exemple, une personne qui craint les pénuries recevra automatiquement une publicité disant « jusqu'à épuisement des stocks », alors qu'une autre qui a tendance à suivre les autres recevra une publicité disant « meilleur vendeur ». Bien que le concept du *nudging* gagne en popularité, il peut être perçu comme envahissant.

- **L'impartition** de l'analytique des données peut rendre la gestion de la responsabilité plus difficile.

- **L'utilisation secondaire des données** pose d'autres problèmes. De façon générale, les entreprises ne peuvent utiliser les renseignements personnels d'une personne que pour les fins énoncées au moment où l'information a été recueillie avec le consentement de cette personne. L'utilisation de ces renseignements à des fins d'analytique pourrait constituer une utilisation secondaire de ces renseignements et donc une entrave à la vie privée, sauf si la personne consent à cette utilisation secondaire.

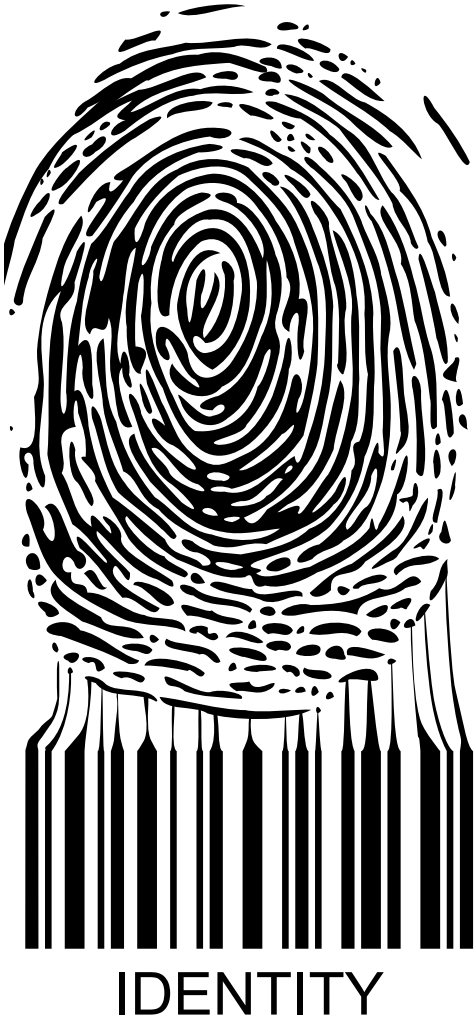
Le problème le plus crucial concerne la *vie privée*.

La vie privée, c'est personnel

On entend par *protection de la vie privée* le droit ou la capacité d'une personne à exercer un contrôle sur la collecte, l'utilisation et la divulgation de ses renseignements personnels par d'autres parties. Presque tous les renseignements pouvant être liés à une personne identifiable peuvent être personnels.

Les gens sont de plus en plus inquiets à ce sujet :

- 93 % se préoccupent de la protection de leur vie privée en ligne
- 45 % ne confient aucun renseignement personnel à des entreprises
- 89 % évitent de faire affaire avec des entreprises qu'ils soupçonnent de ne pas protéger la vie privée



Cependant, les données ne permettent pas toutes d'identifier une personne, et les renseignements non personnels ne sont pas tous les mêmes.

- On entend par **information anonymisée** les dossiers desquels on a retiré ou rendu illisible suffisamment de renseignements personnels pour qu'il n'y ait aucune raison de croire qu'elle puisse servir à identifier une personne.

- **L'information agrégée** englobe les éléments d'information dont les valeurs ont été générées par un calcul de toutes les unités individuelles. Par exemple, pour élaborer de nouvelles stratégies thérapeutiques, les chercheurs médicaux cherchent parfois à repérer des tendances dans des données agrégées sur les patients, mais ils n'ont aucun moyen de lier ces données à une personne en particulier.

- **L'information confidentielle non personnelle** désigne de l'information qui a souvent beaucoup de valeur et d'importance pour les entreprises, notamment les plans d'affaires et les recherches exclusives ou d'autres éléments de propriété intellectuelle. La divulgation ou la perte de ces renseignements confidentiels peut être très préoccupante pour les entreprises, mais elle ne constitue pas une entrave à la vie privée, car elle n'implique pas le traitement de renseignements personnels.

Certains types d'information ne sont pas facilement assimilables. C'est le cas des métadonnées, qui sont des renseignements au sujet d'autres renseignements, par exemple la durée des appels et les données sur l'utilisation fonctionnelle générée par les téléphones mobiles. L'enchevêtrement complexe des associations révélées par les métadonnées peut constituer une entrave beaucoup plus importante à la vie privée que le simple fait d'accéder aux communications d'une personne.

Les données massives et la protection de la vie privée ne sont pas mutuellement exclusives

Naturellement, les avis divergent à ce sujet. Pour certains, l'analytique des données massives menace les protections fondamentales de la vie privée, mais pour d'autres, nos exigences en matière de vie privée nous empêchent de récolter les fruits de l'analytique avancée. Aucun de ces arguments ne résout ce dilemme. Nous devons trouver une nouvelle solution qui tiendra compte des intérêts et des objectifs de toutes les parties et créera une situation gagnante pour tous.

L'argument voulant que la protection de la vie privée nuise à l'innovation est aussi périmé et faible que celui voulant que la vie privée doive être sacrifiée au profit de l'innovation. En fait, c'est tout le contraire : la protection de la vie privée stimule l'innovation. Elle force les innovateurs à faire preuve de créativité pour trouver des solutions permettant la multifonctionnalité.

Nous estimons qu'il est parfaitement possible de protéger la vie privée à l'ère des données massives tout en utilisant l'analytique des données pour acquérir de nouvelles connaissances et innover afin de propulser une entreprise vers l'avenir. À notre avis, les approches de protection de la vie privée fondées sur la conformité ont tendance à mettre l'accent sur les entraves à la vie privée qui ont déjà eu lieu. Nous recommandons plutôt aux entreprises d'intégrer des stratégies de protection de la vie privée directement dans leurs technologies, leurs stratégies d'affaires et leurs processus opérationnels afin de prévenir les entraves avant qu'elles ne se produisent.

Heureusement, il existe déjà un cadre leur permettant d'accomplir cela.

Protection intégrée de la vie privée

Une des approches de protection proactive de la vie privée parmi les plus reconnues est la protection intégrée de la vie privée (PIVP). Ce cadre a été élaboré à la fin des années 1990 par Ann Cavoukian, Ph. D., commissaire à l'information et à la protection de la vie privée de l'Ontario, en réponse aux effets toujours croissants des technologies de l'information et des communications et des grands systèmes de données en réseau.

La PIVP consiste à intégrer des mesures protectrices directement dans les systèmes informatiques, les pratiques d'affaires et l'infrastructure en réseau, fournissant ainsi un compromis qui procure un juste équilibre entre la nécessité d'innover et de préserver un avantage concurrentiel et celle de protéger la vie privée. (Voir l'encadré.)

La mise en œuvre de ce cadre peut occasionner des changements aux structures de gouvernance, aux objectifs opérationnels et stratégiques, aux rôles et responsabilités, aux politiques, aux systèmes d'information et flux de données, aux processus décisionnels, aux relations avec les parties prenantes, et même à la culture d'entreprise. Le concept de protection intégrée de la vie privée n'est pas un feu de paille, puisqu'il a été adopté par de nombreuses autorités des secteurs public et privé aux États-Unis, dans l'Union européenne et ailleurs dans le monde. Elles comprennent notamment la Maison-Blanche aux États-Unis, la Federal Trade Commission, le département de la Sécurité intérieure, le Government Accountability Office, la Commission européenne, le Parlement européen et le Groupe de travail « Article 29 » ainsi que d'autres organismes publics ailleurs dans le monde qui ont adopté des lois de protection de la vie privée. De plus, les autorités internationales de protection de la vie privée et des données ont endossé à l'unanimité la protection intégrée de la vie privée en tant que norme internationale dans ce domaine.

La protection intégrée de la vie privée est une façon particulièrement efficace d'intégrer la confidentialité dans l'ADN d'une entreprise en vue d'instaurer des activités d'analytique des données qui favorisent l'innovation sans compromettre la confidentialité des renseignements personnels.

Les sept principes de la protection intégrée de la vie privée

1. Prendre des mesures proactives et non réactives, prévoir et prévenir les incidents d'atteinte à la vie privée *avant* qu'ils ne se produisent.
2. Les renseignements personnels doivent systématiquement être protégés au sein des systèmes informatiques ou dans le cadre des pratiques internes. La vie privée d'un particulier est protégée même si ce dernier ne pose aucun geste.
3. La protection de la vie privée est intégrée dans la conception et l'architecture des systèmes informatiques et des pratiques d'affaires; elle n'y est pas greffée après coup.
4. La protection de la vie privée tient compte de tous les intérêts et objectifs légitimes en cause selon un paradigme à somme positive.
5. La protection de la vie privée persiste pendant toute la période de conservation des données.
6. Tous les intervenants sont assurés que, sans égard aux pratiques ou aux technologies employées, le système fonctionne conformément aux promesses et aux objectifs établis, sous réserve d'une vérification indépendante.
7. Les concepteurs et les utilisateurs doivent privilégier les intérêts des particuliers en prévoyant notamment des mesures strictes et implicites de protection de la vie privée, des exigences appropriées quant aux avis et des fonctions habilitantes et conviviales, axées sur l'utilisateur.

Stratégies et outils pour protéger les renseignements personnels

Plusieurs options technologiques s'appuyant sur le cadre de la PIVP pour protéger la vie privée tout en facilitant l'analytique des données sont à la disposition des organisations, notamment :

Minimisation des données – réduire la quantité de données recueillies
En vertu de cette stratégie, aucun renseignement permettant d'identifier une personne n'est recueilli à moins qu'un objectif spécifique et impérieux ne le justifie, ce qui élimine efficacement les risques d'entrave à la vie privée dès le début.

Anonymisation – rendre les personnes moins faciles à identifier
Par cette stratégie, on élimine d'un ensemble de données tous les renseignements pouvant servir à identifier une personne, que ce soit directement ou indirectement, par des associations avec d'autres ensembles de données.

Contrôles d'accès par les utilisateurs – restreindre l'accès
Grâce à un ensemble de processus, on accorde ou on refuse les demandes spécifiques d'obtention de l'information; ces processus sont habituellement combinés à d'autres politiques en matière de sécurité visant la protection des renseignements personnels.

Certaines technologies faisant l'objet de recherches sont très prometteuses et pourraient permettre la coexistence de la protection de la vie privée et de l'utilisation des données. Elles comprennent notamment les suivantes :

Protection différentielle
On intègre de façon aléatoire des « bruits » aux résultats de recherche dans un ensemble de données afin de fournir une garantie mathématique que la présence de toute personne dans cet ensemble sera masquée. Le logiciel évalue les risques d'entrave à la vie privée d'une requête et détermine ensuite le niveau de « bruit » qu'il faut introduire dans le résultat avant de le présenter.

Données synthétiques
Tant que le nombre de personnes incluses dans un ensemble de données est suffisamment grand, il est possible de générer un ensemble de données synthétiques entièrement composé de personnes « fictives » ou dont l'identité a été modifiée, mais qui conservent les propriétés statistiques de l'ensemble de données original tout en fournissant la garantie mathématique de « bruit » associée à la protection différentielle.

Innovation *plus* **protection de la vie privée**

L'ère des données massives n'est pas près de disparaître. La capacité d'utiliser des données pour relier l'information, repérer des tendances et personnaliser les interactions pour obtenir des résultats optimaux a atteint un degré extraordinaire de perfectionnement. Les organisations continueront d'utiliser l'analytique des données pour atteindre leurs objectifs stratégiques; les plus avisées s'inspireront de la nécessité de protéger la vie privée pour stimuler leur créativité et leur capacité d'innovation – en plus de l'intégrer dans leurs systèmes pour garantir des résultats de qualité.

Grâce à une planification rigoureuse et à l'application de techniques et de principes, les entreprises peuvent utiliser les données pour faire progresser leurs affaires tout en protégeant les renseignements personnels.

Nous pouvons vraiment **gagner sur tous les tableaux**.

Pour plus d'information à ce sujet, veuillez communiquer :

Dr. Ann Cavoukian, Ph.D.
Commissaire à l'information et à la protection de la vie privée
+1 416-326-3333
info@ipc.on.ca

David Stewart
Leader national, Analytique avancée
Deloitte
+1 416-775-7484
davstewart@deloitte.ca

Beth Dewitt
Directrice, Service des risques d'entreprise
Deloitte
+1 416-643-8223
bdewitt@deloitte.ca

www.deloitte.ca

Conçu et produit par le Service de conception graphique de Deloitte, Canada. 14-2185T