Utiliser la *protection intégrée de la vie privée* pour innover au moyen des données massives sans compromettre la confidentialité

Deloitte.

Ann Cavoukian, Ph.D.
Commissaire à l'information
et à la protection de la vie privée
Ontario, Canada

David Stewart
Leader national, Analytique avancée
Deloitte

Beth Dewitt
Directrice, Service des risques
d'entreprise
Deloitte

Le 10 juin 2014

REMERCIEMENTS

Les auteurs souhaitent remercier Megan Brister, Leader nationale, Protection de la vie privée; et Michelle Chibba, directrice des politiques et des projets spéciaux, CIPVP. Nous remercions également Catherine Thompson, conseillère en politiques et en réglementations; David Weinkauf, agent chargé de l'analyse 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.

Commissaire à l'information
et à la protection de la vie privée de
Ontario, Canada

2, rue Bloor Est
Bureau 1400
Toronto (Ontario)
M4W 1A8
CANADA

416-326-3333
1-800-387-0073
Télécopieur : 416-325-9195
ATS (Téléimprimeur) : 416-325-7539
Site Web : www.ipc.on.ca
Privacy by Design: www.privacybydesign.ca

Utiliser la protection intégrée de la vie privée pour innover au moyen des données massives sans compromettre la confidentialité

TABLE DES MATIÈRES

Avant-propos

L'argument voulant que la protection de la vie privée nuise à l'innovation au moyen des données massives relève d'une approche périmée à somme nulle. Il s'agit d'une fausse croyance nous obligeant à faire d'inutiles compromis entre les avantages des données massives et la protection des renseignements personnels qu'elles contiennent. C'est plutôt l'inverse qui est vrai : la protection de la vie privée stimule l'innovation et force les innovateurs à faire preuve de créativité pour trouver des solutions servant plusieurs fonctionnalités. Nous devons donc cesser de croire qu'elle nuit à l'innovation et adopter un paradigme positif conciliant l'innovation au moyen des données massives et la protection de la vie privée.

Saviez-vous qu'on ne peut tirer des résultats de qualité des données massives sans protéger la vie privée? En effet, le contexte est un facteur essentiel dans le domaine des données massives. Quand on a découvert que la capacité de Google Flu à prédire la propagation de la grippe avait été surestimée, on a invoqué ne pas avoir recueilli suffisamment d'informations sur ce qui avait motivé les sujets à faire une recherche sur la grippe dans Google. Les données recueillies directement auprès d'une personne qui en est avisée et y consent sont invariablement de meilleure qualité aux fins d'analyse.

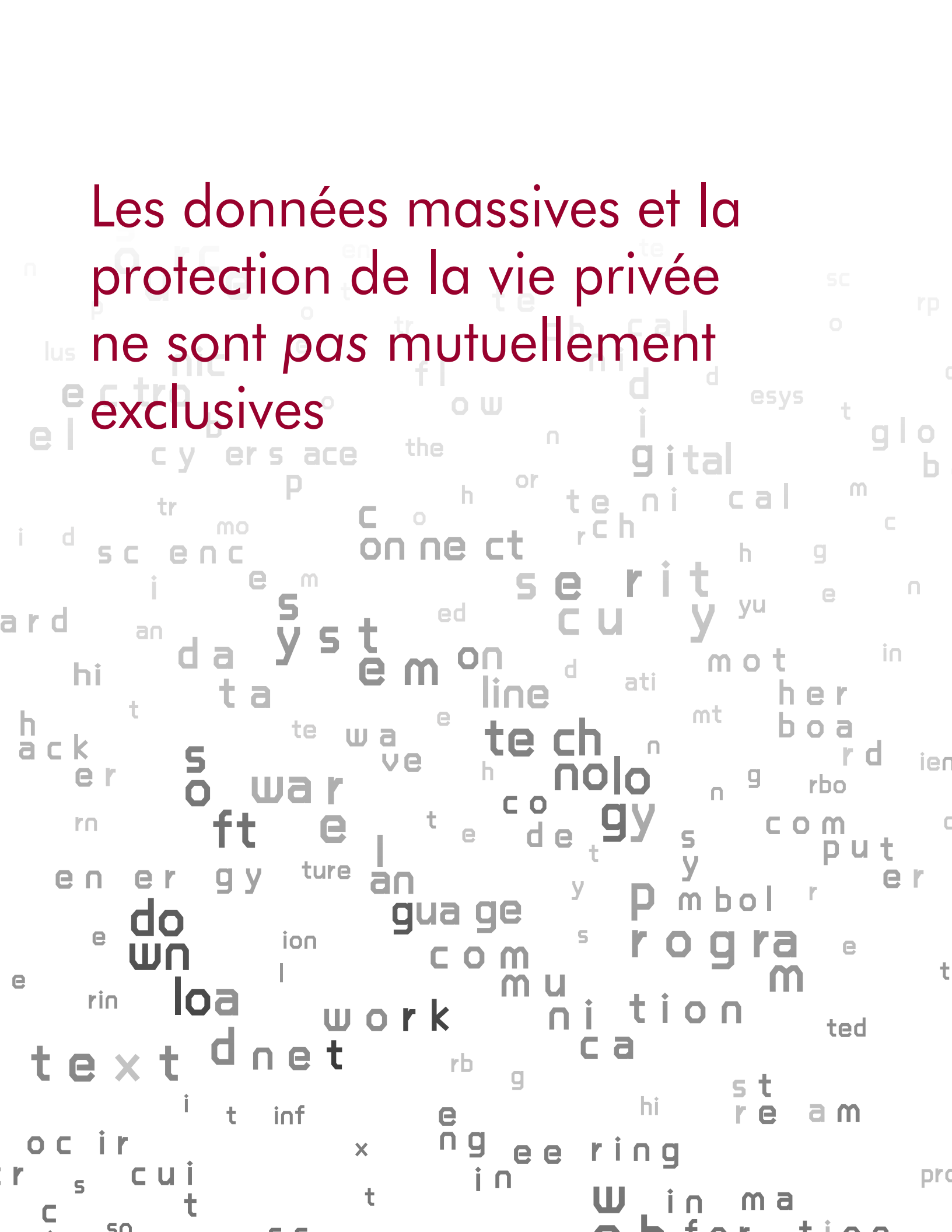
L'utilisation d'outils de protection de la vie privée dans les données massives permet de protéger les renseignements personnels, mais aussi d'analyser ces renseignements. Ces outils comprennent l'anonymisation, l'agrégation des données et de nouvelles technologies telles que la protection différentielle et les données synthétiques, qui seront décrites dans le présent document. Même dans les cas où des algorithmes sont utilisés pour trouver des liens au sein de grands ensembles de données, la minimisation des données doit également être considérée comme un outil de préservation des renseignements permettant d'identifier une personne; c'est l'outil qui nous aide à trouver l'aiguille, mais *sans* la botte de foin.

La protection de la vie privée est aussi importante que les données massives. Des outils permettent de protéger systématiquement les renseignements personnels tout en profitant des avantages des données massives. Ensemble, nous pouvons faire en sorte que les données massives et la protection de la vie privée puissent cohabiter dans un monde où tout le monde est gagnant.

Ann Cavoukian, Ph. D.

Commissaire à l'information
et à la protection de la vie privée
Ontario, Canada

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



Les données figurent parmi les actifs les plus précieux d'une entreprise. Les renseignements que l'on obtient en appliquant diverses techniques analytiques peuvent procurer aux décideurs des perspectives essentielles pour élaborer des stratégies, favoriser la croissance et le rendement opérationnel et gérer les risques. Les données deviennent de plus en plus l'oxygène de l'entreprise moderne.

La quantité de données générées par les personnes, les appareils connectés à Internet et les entreprises augmente à une vitesse exponentielle. Par exemple, les entreprises de services financiers, de vente au détail et de soins de santé génèrent de grandes quantités de données lors de leurs interactions avec des fournisseurs, des patients, des clients et des employés. D'autres données sont créées à l'extérieur de ces entreprises au moyen des recherches dans Internet, des médias sociaux, des données de localisation GPS des appareils mobiles, des transactions boursières et bien plus encore.

Au sein du cycle d'information traditionnel, la génération de données est un processus continu. Des données brutes sont recueillies puis analysées, transformées et reliées à d'autres ensembles de données brutes. Ce traitement permet d'appliquer les connaissances existantes à la fois aux perspectives issues de l'analytique et aux nouvelles perspectives créées par le processus même de transformation des données pour obtenir, en fin de compte, l'*information*. Des experts traitent cette nouvelle information, et leur interprétation donne des *renseignements stratégiques*. La diffusion de ces renseignements permet ensuite d'établir une nouvelle série de priorités d'affaires.

On dénombre actuellement 9,6 milliards d'appareils connectés à Internet[1], 1,3 milliard de connexions mobiles à large bande[2] et 1,2 zettaoctet (10^{21}) de données transmises sur les réseaux IP mondiaux chaque année[3]. 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[4]. Cette conjoncture est actuellement désignée comme la révolution des données ou l'époque des « données massives ».

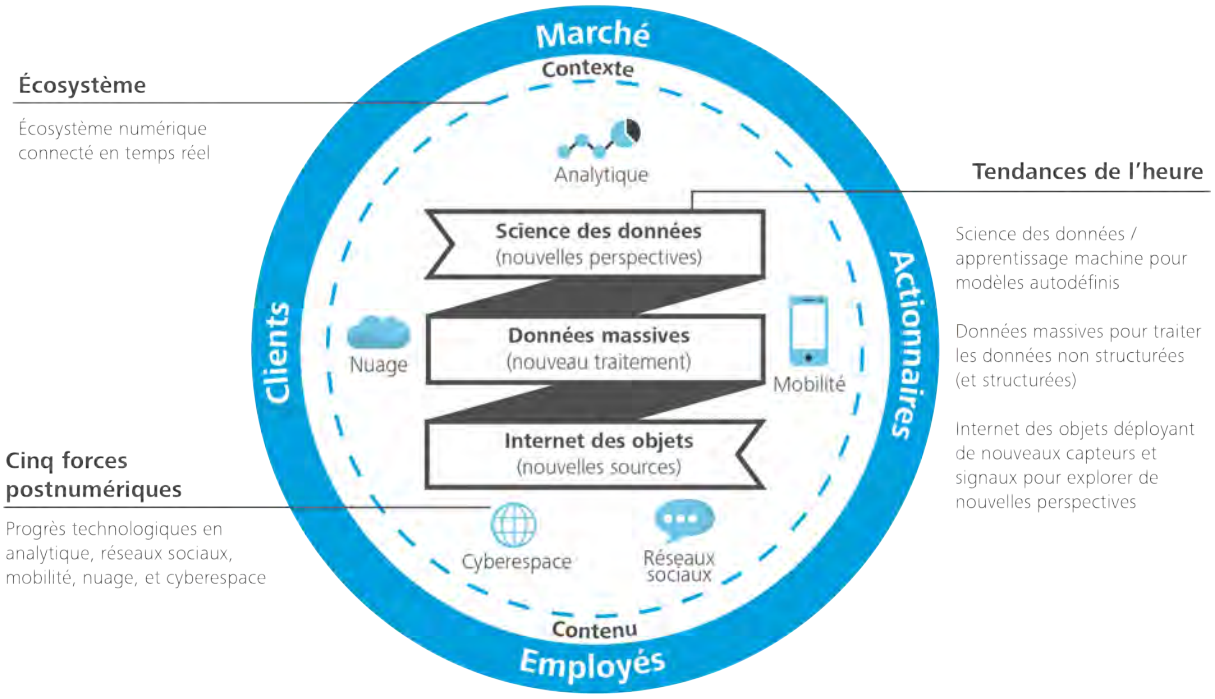
Le terme « données massives » est utilisé pour désigner les très grands ensembles de données contenant divers types de renseignements. Ces ensembles ont favorisé la création d'une nouvelle génération de technologies et d'architectures d'information qui permettent de les traiter à haute vitesse et d'en extraire de la valeur en les analysant au moyen de plateformes distribuées. Dans l'usage courant, le terme « données massives » désigne à la fois ces grands ensembles de données et le processus permettant d'analyser plusieurs « silos » de données et d'en extraire de la valeur.

1. *Internet connected devices approaching 10 billion, to exceed 28 billion by 2020*, [En ligne], IMS Research, octobre 2012 [http://imsresearch.com/press-release/Internet_Connected_Devices_Approaching_10_Billion_to_exceed_28_Billion_by_2020&cat_id=113&type=LatestResearch].

2. Communiqué de presse, *Mobile Industry Set to Grow from US$ 1.5 Trillion in 2011 to US$ 1.9 Trillion in 2015*, [En ligne], GSMA, 27 février 2012 [http://www.gsma.com/newsroom/gsma-research-demonstrates-that-mobile-industry-is-creating-a-connected-economy].

3. *Global Cloud Index*, [En ligne], Cisco, 2012-2017 [http://www.cisco.com/c/en/us/solutions/service-provider/global-cloud-index-gci/index.html].

4. Siegler MG, 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].



Les données massives jouent un rôle essentiel dans ce que Deloitte appelle « l'entreprise numérique », qui est notre vision évolutive de la direction que les entreprises prendront au cours des prochaines années. Grâce à l'innovation dans le domaine des appareils mobiles, nous avons déjà d'incroyables technologies entre nos mains. Les réseaux sociaux permettent aux gens de se connecter par des moyens qu'ils n'auraient jamais imaginés. Le nuage réduit considérablement les coûts associés aux infrastructures de matériel et de données, tandis que les capacités de stockage de données sont de plus en plus imposantes et de moins en moins coûteuses. L'analytique des données peut maintenant interpréter de grandes quantités de renseignements pour produire des perspectives exploitables.

Analytique des données : favoriser la rupture[5]

Naturellement, les entreprises veulent exploiter pleinement le potentiel des données pour en extraire de la valeur. Elles cherchent ardemment 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 de leurs stratégies.

Les progrès rapides observés récemment dans les domaines de la vitesse de traitement des données et des algorithmes analytiques permettent maintenant de traiter de grandes quantités de données structurées et non structurées à très haute vitesse. De nos jours, l'analytique des données permet aux entreprises 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.

L'analytique des données accélère l'innovation et révolutionne les modèles d'affaires traditionnels. Elle permet aux détaillants d'adapter avec précision leurs produits et leurs services aux préférences et aux comportements d'achat de leurs clients. Elle aide les entreprises de services financiers à donner des conseils et à recommander des produits de façon proactive. Elle aide aussi les organismes de soins de santé à améliorer les diagnostics, les traitements et la gestion de la santé publique. Dans certains secteurs, les concurrents partagent leurs données afin de relever des défis communs dans des domaines comme la fraude, la cybersécurité et la performance des pratiques de santé et de sécurité.

Le secteur public explore également le potentiel de l'analytique des données en créant des initiatives de « gouvernement ouvert » et de « données ouvertes ». Ces initiatives rendent certaines données massives accessibles au grand public, souvent pour la première fois[6]. Un des objectifs est d'accroître la transparence du gouvernement et d'encourager la participation du public. Les gouvernements espèrent aussi que les citoyens et les entreprises pourront utiliser ces données pour acquérir de nouvelles connaissances et innover[7]. Selon la stratégie « Canada numérique 150 » lancée par le gouvernement canadien en avril 2014, « le Canada sera l'un des chefs de file mondiaux de l'application des données volumineuses afin de changer nos idées sur les soins de santé, la recherche-développement et les multiples activités des entreprises et des gouvernements et de modifier nos façons de faire dans ces domaines[8] ».

Pour Deloitte, l'entreprise numérique est celle qui maîtrisera l'art du possible, exploitera le potentiel des entrepôts de données et utilisera les connaissances qu'elle en tirera pour faire preuve d'audace, innover et remettre en question notre façon conventionnelle de faire des affaires, et ce, tout en protégeant constamment la vie privée.

---

5. Tech Trends 2014: *inspiring disruption*, [En ligne], Deloitte [https://www.deloitte.com/assets/Dcom-Luxembourg/Local%20Assets/Documents/Whitepapers/2014/dtt_en_wp_techtrends_10022014.pdf].

6. Par exemple, le gouvernement ontarien offre un accès à des données hydrographiques et démographiques, des données sur les forêts et les transports ainsi que d'autres données dans le cadre de son programme de données ouvertes (http://www.ontario.ca/government/government-ontario-open-data).

7. Cavoukian, Ann. P*rivacy and Government 2.0: The Implications of an Open World*, [En ligne], mai 2009 [http://www.ipc.on.ca/english/Resources/Discussion-Papers/Discussion-Papers-Summary/?id=874].

8. « Canada numérique 150 », [En ligne], gouvernement du Canada [http://www.ic.gc.ca/eic/site/028.nsf/fra/accueil].

Protection de la vie privée et renseignement *personnels*

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. Malgré certaines différences territoriales, les renseignements personnels (également appelés renseignements permettant d'identifier une personne) englobent toute information, enregistrée ou sous toute autre forme que ce soit, concernant une personne identifiable. Presque tous les renseignements pouvant être liés à une personne identifiable peuvent être personnels, qu'ils soient biographiques, biologiques, généalogiques, historiques, transactionnels, géographiques, relationnels, computationnels, professionnels ou liés à la réputation. Pour déterminer si certaines données entrent dans la catégorie des renseignements personnels, on doit tenir compte du contexte. S'il est raisonnablement possible d'identifier une personne, que ce soit directement, indirectement ou par la manipulation ou la mise en correspondance de données, le respect de la vie privée doit être pris en compte.

Cependant, les données ne permettent pas toutes d'identifier une personne et ne soulèvent donc pas toutes des enjeux de respect de la vie privée. Il est important de bien distinguer les différentes formes de renseignements non personnels :

- On entend par *information anonymisée* les dossiers desquels on a retiré ou rendu illisible suffisamment de renseignements personnels pour que l'information restante ne permette pas d'identifier une personne et qu'il n'y ait aucune raison de croire qu'elle puisse servir à cette fin[1].
- L'*information agrégée* englobe les éléments d'information dont les valeurs ont été générées par un calcul de toutes les unités individuelles. Pour élaborer de nouvelles stratégies thérapeutiques, les chercheurs médicaux utilisent parfois des données agrégées sur les patients, par exemple le pourcentage de patients prenant une certaine association de médicaments qui ont eu des effets indésirables, 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 pour les plans d'affaires, les prévisions de revenus, 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 – Deloitte conseille d'ailleurs souvent ses clients sur la façon de prévenir de telles pertes –, mais elle ne constitue pas une entrave à la vie privée, car elle n'implique pas le traitement de renseignements *personnels*. Deloitte maintient le plus haut degré de protection pour toutes les données de ses clients, qu'elles constituent ou non des renseignements personnels ou confidentiels.

Certains types d'information ne sont pas facilement assimilables à des renseignements personnels ou non personnels. C'est le cas des métadonnées, qui sont des renseignements générés par nos appareils de communication et nos fournisseurs de services de communication lorsque nous utilisons des téléphones, des ordinateurs, des tablettes ou d'autres dispositifs informatiques connectés à un réseau filaire ou sans fil. Les métadonnées sont essentiellement des renseignements *au sujet* d'autres renseignements; dans notre exemple, ce sont des renseignements sur nos communications[2]. Le contexte est un facteur essentiel à prendre en considération pour déterminer si des renseignements sont personnels, et il est particulièrement important dans le cas des métadonnées. 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.

1. Voir *Guide to Protecting the Confidentiality of Personally Identifiable Information* (PII), National Institute of Standards and Technology, avril 2010, p. E-1.

2. Cavoukian A. *A Primer on Metadata: Separating Fact from Fiction*, [En ligne], juillet 2013, IPC [http://www.ipc.on.ca/images/Resources/metadata.pdf].

Analytique des données, innovation et protection de la vie privée : qui gagne?

À mesure que les entreprises adoptent l'analytique des données pour obtenir de nouvelles perspectives, les instances de réglementation, les législateurs, les groupes d'intérêt et les citoyens s'inquiètent des répercussions de ces activités sur la vie privée, qui vont de l'usage abusif ou de la divulgation non autorisée de renseignements personnels à la surveillance axée sur les données. Les protections fondamentales qui sont offertes aux personnes dont les renseignements personnels sont traités, et que nous tenions jadis pour acquises, par exemple les avis, les consentements, les spécifications des objectifs et les limites, sont de plus en plus menacées par la nature même de l'analytique des données massives. Certains croient que notre perception de la confidentialité doit changer, et que les exigences de consentement, de spécification des objectifs et de limites d'utilisation font obstacle à l'analytique des données massives[9]. Or, ces arguments représentent une approche périmée à somme nulle. 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.

Le fait de transférer la responsabilité relative aux renseignements personnels des personnes aux entreprises ne suffit pas à résoudre le problème. Cette solution constituerait même une forme de « paternalisme[10] » par lequel les entreprises pourraient déterminer « ce qui convient le mieux » aux personnes, alors que ces personnes ne pourraient pas participer aux discussions concernant l'utilisation ou l'usage abusif de leurs renseignements personnels. Si l'histoire de la protection de la vie privée nous apprend quelque chose, c'est que la perte du contrôle qu'une personne exerce sur ses renseignements personnels multiplie le nombre de cas d'usage abusif de ces renseignements, elle ne le diminuerait pas.

En fait, l'imposition de limites inadéquates et l'approche paternaliste pourraient conduire à ce que les défenseurs de la protection de la vie privée craignent le plus : une surveillance de masse systématique, des pratiques de profilage étendues et détaillées, une plus grande asymétrie de l'information, un déséquilibre des pouvoirs et, finalement, diverses formes de discrimination. La dilution des exigences relatives aux avis et aux consentements affaiblit les protections essentielles de la vie privée, et la réduction des limites imposées sur les objectifs spécifiés, la collecte et l'utilisation des renseignements personnels minimise la responsabilité plutôt que de la renforcer.

Les exigences relatives à la protection de la vie privée ne font pas obstacle à l'innovation ou à la réalisation des avantages sociétaux de l'analytique des données massives; elles peuvent même favoriser l'innovation et l'obtention de résultats profitables pour tous. En utilisant des technologies d'amélioration de la confidentialité telles que des techniques et des outils puissants d'anonymisation, et en appliquant des procédures appropriées d'évaluation des risques de réidentification, il est possible d'assurer un degré élevé de protection de la vie

9. H. Cate Fred, Peter Cullen et Viktor Mayer-Schönberger, *Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines*, décembre 2013.

10. Cavoukian Ann, Alexander Dix et Khaled El Emam. *The Unintended Consequences of Privacy Paternalism*, [En ligne], IPC, mars 2014 [http://www.ipc.on.ca/images/Resources/pbd-privacy_paternalism.pdf].

privée tout en garantissant un niveau de qualité des données approprié pour leur utilisation secondaire dans le cadre de l'analytique des données massives.

Dans certains cas, les principes de protection de la vie privée peuvent même *améliorer* l'information extraite des données massives. Dans l'exemple de Google Flu, des études subséquentes ont révélé que « les estimations de Google concernant la propagation de maladies d'allure grippale représentaient près du double des données réelles[11] ». Cela s'explique par l'absence de contexte, un élément clé de la protection de la vie privée. Les estimations de Google n'avaient pas pris en compte *pourquoi les* internautes cherchaient de l'information sur la grippe : « Avaient-ils la grippe? Connaissaient-ils quelqu'un qui avait la grippe? Voulaient-ils savoir comment éviter la grippe? » Quand on inclut chaque participant directement dans la collecte de données, le contexte de l'information devient considérablement plus précis.

11. Harford Tim. *Big data: are we making a big mistake?*, [En ligne], *Financial Times Magazine*, 28 mars 2014 [http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2yaNDIgbN].

Analytique des données : favoriser la rupture

Tech Trends 2014: Inspiring Disruption[1], cinquième rapport annuel de Deloitte sur l'univers en constante évolution des technologies, met l'accent sur les tendances perturbatrices qui transforment les entreprises, les gouvernements et la société en général. Les technologies de l'information sont encore dominées par cinq forces : l'analytique, la mobilité, les réseaux sociaux, le nuage et le cyberespace. Les éléments perturbateurs sont des domaines qui peuvent créer un bouleversement positif et durable des capacités des TI, des opérations d'affaires et même des modèles d'affaires.

- Aux entreprises qui doivent trouver des façons d'améliorer leur capacité de perception et de réaction, l'analytique cognitive offre une solution puissante qui comble l'écart entre l'espoir suscité par les données massives et la réalité du processus décisionnel. L'analytique cognitive est appelée à prendre de l'importance : les règles prédéfinies et les requêtes structurées seront augmentées grâce à l'intelligence artificielle, l'apprentissage machine et le traitement du langage naturel afin de générer des hypothèses à partir des données massives.
- L'adoption du pouvoir des foules permettra aux entreprises de trouver de façon dynamique des compétences spécialisées correspondant à leurs besoins en étendant leurs recherches à toutes les personnes et tous les lieux. Elles pourront utiliser le savoir collectif des masses pour réaliser diverses tâches allant de la saisie et du codage de données à l'analytique avancée, en passant par le développement de produits. La possibilité qu'il y ait un effet perturbateur sur les coûts est un facteur qui justifie à lui seul les premières expérimentations, mais ce concept pourrait aussi avoir des répercussions plus importantes sur les capacités d'innovation des entreprises
- Le contenu et les actifs sont de plus en plus numériques, et sont accompagnés d'éléments audiovisuels et interactifs. Ils sont utilisés sur de multiples plateformes : appareils mobiles, réseaux sociaux et Internet, ainsi qu'en magasin ou sur le terrain. L'engagement numérique consiste à créer une façon uniforme, intéressante et contextuelle de personnaliser, d'offrir et même de monétiser l'expérience globale de l'utilisateur, surtout si les produits de base sont augmentés ou remplacés par des éléments numériques de propriété intellectuelle
- La technologie vestimentaire se présente sous de nombreuses formes, que ce soit des lunettes, des montres ou des macarons et bracelets intelligents. Son potentiel est énorme. La technologie sans fil et à affichage tête haute peut nous aider à transformer notre façon de travailler, de prendre des décisions et de transiger avec les employés, les clients et les partenaires. Les produits technologiques vestimentaires s'appliquent à des situations où la sécurité, la logistique et même l'étiquette limitent l'utilisation d'ordinateurs portables et de téléphones intelligents. À l'heure actuelle, ils sont populaires sur le marché de consommation de masse, mais nous prévoyons que les entreprises les adopteront et les transformeront.

D'ici 2020, il y aura plus de 40 zettaoctets de données dans le monde, dont la grande majorité ne sera pas structurée[2], provenant d'innovations technologiques récentes comme Internet, la connectivité mobile, l'informatique en nuage et les réseaux sociaux. De nouvelles technologies et procédures ont été créées pour analyser sérieusement toutes ces données aux fins d'affaires. Les processus d'affaires courants recueillent et analysent maintenant des données par des méthodes que personne n'aurait imaginées il y a quelques décennies seulement. Toutes ces forces ont un effet perturbateur parce qu'elles ont changé et continuent de changer notre façon de faire des affaires. Cependant, en plus de favoriser l'innovation, elles accroissent aussi les risques potentiels pour la protection de la vie privée.

1. https://www.deloitte.com/assets/Dcom-Luxembourg/Local%20Assets/Documents/Whitepapers/2014/dtt_en_wp_techtrends_10022014.pdf

2. John Gants et David Reinsel. *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, IDC, décembre 2012.

Risques d'entrave à la vie privée associés aux données massives

Le plus grand risque que posent les données massives pour la protection de la vie privée est celui de créer des associations automatiques entre des données qui, en apparence, ne permettent pas d'identifier une personne. Ces associations peuvent fournir un portrait général d'une personne, ce qui était jadis inconcevable car les identifiants étaient répartis dans diverses bases de données.

Les données massives permettent de relier facilement des éléments de données clés qui associent les personnes à des choses; les données ordinaires deviennent ainsi de l'information sur une personne identifiable et révèlent des détails sur son mode de vie et ses habitudes. 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.

Étant donné l'immense capacité des données massives à créer des liens entre les données, certains proposent que certaines données soient considérées comme des « superdonnées » ou du « supercontenu[12] ». Ces superdonnées sont situées un cran au-dessus des autres données dans le contexte des données massives, car l'utilisation de l'une d'entre elles, qui en soi ne révèle pas grand-chose sur une personne, peut créer de nouvelles associations qui augmentent de façon exponentielle jusqu'à ce que la personne soit identifiée. Chaque nouvelle transaction dans un système de données massives décuplerait cet effet et propagerait l'identification telle une contagion.

Quand un ensemble de données massives comprend des renseignements permettant d'identifier une personne, cela pose de nombreux risques d'entrave à la vie privée[13]. Par exemple, un grand ensemble de données permettant d'identifier une personne pourrait faire l'objet d'une divulgation non autorisée, d'une perte ou d'un vol; plus l'ensemble de données est grand, plus il est vulnérable à un usage abusif. Lorsqu'il y a divulgation non autorisée de données, l'incidence sur la protection de la vie privée est beaucoup plus importante, car l'information est centralisée et contient plus d'éléments de données. Dans les cas extrêmes, la divulgation non autorisée de renseignements personnels peut constituer une menace pour la sécurité publique. Qui plus est, bien que le concept du « nudging », qui consiste à pousser les consommateurs vers un choix, gagne en popularité, l'utilisation de données permettant d'identifier une personne pour brosser son portrait et analyser, prédire et modifier le comportement humain peut être perçue comme invasive[14].

12. Cameron Kim. *Digital Enlightenment Yearbook 2013: The Value of Personal Data*, « Afterword », p. 293.

13. Pour bien comprendre la question des risques d'entrave à la vie privée et des stratégies d'atténuation de ces risques pour tous les ensembles de données contenant des renseignements personnels, on recommande aux entreprises de réaliser une évaluation des risques d'entrave à la vie privée.

14. Le « nudging » est une technique consistant à exploiter les tendances irrationnelles de l'être humain (« biais cognitif ») afin de pousser les gens vers certains résultats. 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 ». Calo Ryan. *Digital Market Manipulation*, University of Washington School of Law, rapport de recherche no 2013-27, 2013.

La gestion de la responsabilité à l'égard de l'impartition est un autre problème soulevé par le traitement de données permettant d'identifier une personne. Il revêt une importance particulière dans le contexte des données massives, car les entreprises possédant de grandes quantités de données n'ont peut-être pas les capacités analytiques nécessaires pour les traiter elles-mêmes et impartissent leurs activités d'analyse et de présentation de l'information.

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 (« but principal ») avec le consentement de cette personne, à moins que cela ne soit permis par la loi. L'utilisation des renseignements personnels aux fins d'analytique des données massives pourrait être interdite en vertu du consentement initial qui a été accordé, car cela pourrait constituer une utilisation secondaire de ces renseignements, sauf si la personne consent à cette utilisation secondaire.

Les demandes des consommateurs exercent également des pressions

Le rapport de 2013 de Deloitte intitulé *Croissance axée sur le client : des attentes grandissantes et des occasions émergentes* explorait diverses tendances qui transforment les interactions entre les entreprises et les consommateurs. Par exemple, nous verrons probablement un jour le lancement de Facebook en 2004 comme un point tournant dans la vie privée des consommateurs. Soudainement, les utilisateurs ont vu un nouvel avantage sur le plan social à partager volontairement toutes sortes de renseignements personnels en ligne. Aujourd'hui, l'information qui était jadis privée – nos amis, nos projets de vacances, nos restaurants et nos marques préférées – est très publique. Parallèlement, la capacité à repérer, à stocker et à analyser ces données massives et d'autres activités en ligne s'est accru de façon exponentielle et a transformé les médias et le marketing.

Les demandes croissantes des consommateurs pour des produits et des expériences personnalisés remettent en question les politiques et les pratiques des entreprises. Les clients veulent maintenant recevoir des publicités ciblées et d'autres avantages en fonction des renseignements qu'ils ont fournis (ou pensent l'avoir fait), même si les règles de protection de la vie privée l'interdisent.

Les entreprises elles-mêmes vont encore plus loin et demandent aux utilisateurs la permission d'accéder au mur et aux photos de leurs amis Facebook ou offrent des avantages additionnels à ceux qui fournissent plus de renseignements personnels, par exemple leur revenu. Nous prévoyons que les entreprises réviseront et modifieront leurs politiques de protection de la vie privée et les consentements de leurs clients en fonction des changements dans les comportements et les attentes de ces derniers et des nouvelles pratiques relatives à la collecte, à la divulgation et à l'utilisation des données.

Dans ce rapport, Deloitte conseille aux entreprises qui veulent exploiter le potentiel des données massives d'être transparentes en ce qui concerne leurs intentions et leurs pratiques, et de s'assurer que la valeur associée à cette transparence encouragera les clients à fournir les consentements demandés. Nous leur recommandons aussi d'utiliser avec circonspection les consentements additionnels accordés par leurs clients : ces derniers doivent savoir qu'ils sont ciblés et que leur choix d'accorder ou non leur consentement sera respecté.

Ne laissez pas les risques vous empêcher d'innover!

Les risques d'accès non autorisé aux données, si un tel accès mène à la divulgation ou à l'usage abusif de renseignements personnels, peuvent avoir de graves conséquences pour une entreprise. Ces conséquences comprennent une atteinte à la réputation, une poursuite, une atteinte à la marque, 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. Selon l'indice de confiance des consommateurs dans la protection de la vie privée de TRUSTe, 93 pour cent des gens se préoccupent de la protection de leur vie privée en ligne, 45 pour cent ne confient aucun renseignement personnel à des entreprises et 89 pour cent évitent de faire affaire avec des entreprises qu'ils soupçonnent de ne pas protéger la vie privée.

Or, même si ces risques existent, les entreprises ne devraient pas craindre d'innover en utilisant l'analytique des données. L'application d contrôles de protection de la vie privée et l'utilisation d'outils connexes peuvent réduire adéquatement les risques et permettre aux entreprises d'exploiter le potentiel transformateur des données massives tout en respectant la confidentialité des renseignements personnels[15].

15. TRUSTe, Consumer Privacy Confidence Index, 2014, http://www.truste.com/us-consumer-confidence-index-2014/.

Protection proactive de la vie privée/protection des renseignements personnels et facilitation de l'innovation

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. Par conséquent, elles ne peuvent pas répondre aux demandes de l'ère des données massives. Nous recommandons donc aux entreprises d'intégrer de façon consciente et proactive des stratégies de protection de la vie privée à leurs opérations en les incorporant directement à leurs technologies, leurs stratégies d'affaires et leurs processus opérationnels.

Une des approches de protection proactive de la vie privée la plus largement reconnue est la *protection intégrée de la vie privée* (PIVP), un cadre 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 et coauteure du présent document. Ce concept consiste à intégrer la protection de la vie privée directement aux spécifications de conception des technologies, des pratiques d'affaires et de l'infrastructure en réseau. M^me^ Cavoukian a créé ce cadre en réponse aux effets toujours croissants des technologies de l'information et des communications et des grands systèmes de données en réseau. Il offre une solution pratique à toute entreprise souhaitant trouver un juste équilibre entre la volonté d'innover et la nécessité de protéger la vie privée en lui fournissant un compromis qui satisfait ces deux critères.

Le concept de protection intégrée de la vie privée presse les entreprises à adopter une approche proactive en la matière. Il dicte par défaut la préservation de la confidentialité des renseignements et intègre des mesures protectrices directement dans les systèmes informatiques, les pratiques d'affaires et l'infrastructure en réseau. Il garantit la protection de la vie privée et le contrôle que chaque personne exerce sur ses renseignements tout en offrant aux entreprises un avantage concurrentiel à long terme.

La mise en œuvre du concept de protection intégrée de la vie privée peut avoir de vastes répercussions dans l'ensemble de l'organisation. Cela peut amener 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 a été adopté par de nombreuses organisations des secteurs public et privé aux États-Unis, dans l'Union européenne et ailleurs dans le monde[16]. En 2010, il a été adopté à l'unanimité comme cadre de protection de la vie privée par l'International Assembly of Privacy Commissioners and Data Protection Authorities[17]. Il a également été intégré aux suggestions concernant la création d'un conseil chargé d'examiner les aspects éthiques des projets utilisant des données massives[18].

16. Ces instances comprennent 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 d'ailleurs dans le monde qui ont adopté des lois de protection de la vie privée fondées sur les principes équitables de traitement de l'information (FIPP). 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.

17. *Ibid.*, IPC/Ontario, Résolution.

18. Calo Ryan. *Consumer Subject Review Boards: A Thought Experiment*, Stanford Law Review Online 66 (2013): 97-102.

Principes de la *protection intégrée de la vie privée*

Le concept de protection intégrée de la vie privée repose sur sept principes fondamentaux dont le but est 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 :

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 (des mesures ***proactives*** et non réactives; des mesures ***préventives*** et non correctives).
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 (assurer la protection ***implicite*** de la vie privée).
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. Elle fait partie intégrante du système, sans porter atteinte à ses fonctions (***intégrer*** la protection de la vie privée dans la conception des systèmes et des pratiques).
4. La protection de la vie privée tient compte de tous les intérêts et objectifs légitimes en cause selon un paradigme à somme positive (assurer une fonctionnalité intégrale selon un paradigme à ***somme positive*** [gagnante pour tous] et non à somme nulle [gagnant/perdant]).
5. La protection de la vie privée persiste pendant toute la période de conservation des données – elle assure la conservation sécurisée des données, puis leur destruction sécurisée à la fin de leur période de conservation (assurer la sécurité de bout en bout, pendant ***toute la période de conservation des renseignements***).
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; la transparence est la clé (assurer la ***visibilité*** et la ***transparence***).
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 (***respecter*** la vie privée des utilisateurs).

Protéger la vie privée et favoriser l'innovation : stratégies à déployer

Le fait de penser que la protection de la vie privée nuit à l'innovation est une approche périmée à somme nulle. L'idée qu'elle doive être sacrifiée au profit de l'innovation est une fausse croyance qui oblige à faire des compromis inutiles. 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.

Les organisations ont besoin d'un nouveau guide. Elles doivent délaisser les approches à somme nulle et adopter un paradigme à somme positive dans lequel l'innovation *et* la protection de la vie privée cohabitent. La protection intégrée de la vie privée est une façon particulièrement efficace d'intégrer la confidentialité à 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. Nous croyons qu'elle est un objectif très valable pour toute entreprise.

Les entreprises disposent de nombreuses stratégies pour intégrer la protection de la vie privée à l'analytique des données. La minimisation des données, l'anonymisation et les contrôles d'accès des utilisateurs sont trois stratégies qu'elles peuvent mettre en œuvre dès aujourd'hui pour acquérir des perspectives d'affaires utiles et mettre l'accent sur les innovations fondées sur les données tout en préservant les renseignements personnels.

1. Minimisation des données

L'analytique des données massives n'implique pas toujours l'utilisation de renseignements permettant d'identifier une personne. Cependant, lorsque c'est le cas, la minimisation des données est la méthode qui permet le mieux de gérer les risques d'entrave à la vie privée, puisqu'elle élimine efficacement ces risques dès le début du cycle de vie de l'information.

En vertu de cette stratégie, il faut concevoir dès le début les systèmes d'analytique des données massives afin qu'il n'y ait *aucune* collecte de renseignements permettant d'identifier une personne, à moins qu'un objectif spécifique et impérieux ne le justifie et jusqu'au moment où cet objectif est défini. Par exemple, l'utilisation de renseignements personnels doit être limitée aux fins principales prévues de leur collecte, et peut être élargie à d'autres fins uniquement si la personne y consent de façon explicite. Dans d'autres cas, les entreprises estiment que des données sommaires ou agrégées sont plus que suffisantes pour combler leurs besoins. Les stratégies de minimisation des données cadrent également avec les stratégies d'anonymisation (voir ci-dessous).

Astuce pour la minimisation des données : la première question que vous devez vous poser au sujet de votre processus d'analytique des données vise à déterminer si vous avez besoin de renseignements personnels pour faire cette analyse. Si la réponse est non – autrement dit, si aucun renseignement personnel n'est nécessaire –, aucun enjeu de protection de la vie privée n'entre en ligne de

compte. Cependant, nous recommandons toujours aux entreprises d'utiliser des contrôles appropriés pour s'assurer que les renseignements *confidentiels* sont traités et stockés de façon appropriée. L'application des grands principes de protection de la vie privée peut les aider à définir les types de contrôle qu'elles doivent utiliser.

Obtenir de l'information en temps réel sur la gestion du rendement d'une compagnie aérienne

Une grande compagnie aérienne a demandé à Deloitte de créer un outil de gestion du rendement en temps réel pour ses cadres et ses décideurs. Cet outil devait intégrer l'information de diverses sources, surtout celle de l'entreprise, mais aussi certains renseignements externes, afin d'offrir un aperçu du rendement global du réseau du transporteur aérien.

Pour réaliser ce mandat, nous avons conçu un outil qui recueille un large éventail de renseignements non personnels sur chacune des destinations du transporteur : horaires et statuts des vols, coordonnées des personnes-ressources, évaluations du service à la clientèle anonymisées, statistiques sur les employés et données financières de l'entreprise. Nous avons également inclus certains renseignements externes anonymes, notamment sur les activités sur les médias sociaux, afin de fournir au client un moyen de surveiller ce que ses clients disent en temps réel. Les données recueillies sont livrées en temps réel au moyen d'une application très visuelle de type « tableau de bord » (accessible par un écran tactile ou une tablette), ce qui permet aux dirigeants de l'entreprise de vérifier en un coup d'œil le rendement d'un site donné et de consulter des renseignements additionnels en touchant simplement un bouton.

Toutes les données visées par ce projet étaient des renseignements non personnels. Bien qu'il ait été important de s'assurer que les renseignements confidentiels étaient sécurisés, il n'était pas nécessaire de prendre des mesures additionnelles pour préserver des renseignements personnels, car il n'y en avait aucun dans l'ensemble de données. .

2. Anonymisation

L'anonymisation consiste à utiliser un ensemble d'outils ou de techniques pour éliminer 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. Ces techniques comprennent la suppression ou le masquage des « identifiants directs », par exemple le nom

ou le numéro d'assurance sociale, et la suppression ou la généralisation des « identifiants indirects », par exemple le code postal ou la date de naissance. Même si les identifiants indirects ne permettent pas d'identifier une personne, ils peuvent être liés à d'autres ensembles de données contenant des identifiants directs et pourraient donc servir à identifier une personne. Si l'anonymisation est faite correctement, les données ainsi traitées peuvent être utilisées aux fins de recherche et d'analyse afin de permettre l'acquisition de nouvelles perspectives et l'innovation tout en minimisant les risques de divulgation de l'identité des personnes à qui correspondent ces données.

Protection différentielle et données synthétiques

Tandis que les outils et techniques d'anonymisation ont gagné en popularité au fil des ans et sont devenus des produits commerciaux, 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. Deux de ces technologies sont la protection différentielle de la vie privée et les données synthétiques.

Protection différentielle

La protection différentielle de la vie privée[1] consiste à intégrer 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, ce qui protège son identité et sa vie privée à l'intérieur de cet ensemble.

Habituellement, on configure la protection différentielle en créant une interface de requête, ou « curateur », qui s'interpose entre les renseignements personnels contenus dans un ensemble de données et les personnes qui veulent y accéder. Un algorithme évalue les risques d'entrave à la vie privée des requêtes; en fonction de cette analyse, le logiciel détermine ensuite le niveau de « bruit » qu'il faut introduire dans le résultat avant de le présenter. Normalement, cette distorsion est assez faible pour ne pas affecter la qualité des réponses de façon importante, mais elle est suffisante pour protéger l'identité des personnes incluses dans l'ensemble de données.

Données synthétiques

La plupart des méthodes de protection différentielle de la vie privée ne permettent pas aux chercheurs d'accéder à un ensemble de données afin de s'autoanalyser. Cela limite donc les types de questions que ceux-ci peuvent poser. Pour contourner ce problème, certains envisagent la possibilité de créer des ensembles de données « synthétiques » à l'intention des chercheurs.

Tant que le nombre de personnes incluses dans un ensemble de données est suffisamment grand comparativement au nombre de champs ou de dimensions, on peut 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[2]. Bien qu'il soit déjà possible de générer de tels ensembles synthétiques, l'effort informatique que cela nécessite est habituellement très élevé. Toutefois, d'importants progrès ont été réalisés afin de rendre la génération d'ensembles de données synthétiques plus efficiente, et les avancées dans ce domaine se poursuivent[3].

1. Voir Dwork Cynthia. Differential Privacy, Proceedings of the *33rd International Colloquium on Automata, Languages and Programming*, ICALP partie 2, 2006, p. 1 à 12; Dwork Cynthia. *A firm foundation for private data analysis*, communications de l'ACM, vol. 54, 2011.

2. Voir Blum Avrim, Katrina Ligett et Aaron Roth. *A Learning Theory Approach to Non-Interactive Database Privacy, Proceedings of the 40th, ACM SIGACT Symposium on Theory of Computing,* 2008, p. 609 à 618.

3. Thaler Justin, Jonathan Ullman et Salil Vadhan. *Faster Algorithms for Privately Releasing Marginals*, arXiv : 1205.1758 [cs.DS]; Ullman Jonathan et Salil Vadhan. *PCPs and the Hardness of Generating Synthetic Data, Electronic Colloquium on Computational Complexity*, rapport technique TR10-07, [En ligne], février 2010 [http://people.seas.harvard.edu/~salil/research/synthetic-Feb2010.pdf]

Astuce pour l'anonymisation : l'anonymisation peut aussi aider les entreprises à utiliser des données sans enfreindre les exigences ou les restrictions relatives à l'utilisation secondaire. La suppression des renseignements permettant d'identifier une personne des ensembles de données leur permet d'exploiter toutes les données dont elles disposent tout en respectant leurs engagements à l'égard de l'objectif principal pour lequel ces données ont initialement été recueillies.

Or les pratiques d'anonymisation ne procurent pas toutes le même niveau de confidentialité et les outils utilisés ne fournissent pas tous la même qualité de résultats pour garantir un risque suffisamment faible de réidentification. Le choix des outils et techniques utilisés pour anonymiser un ensemble de données varie. Les entreprises doivent fonder leur choix sur des cadres solides d'anonymisation. Un excellent exemple de tels cadres est celui proposé par Khaled El Emam pour anonymiser les données sur la santé à des fins d'utilisation secondaire[19].

Faciliter la recherche en santé grâce aux données agrégées

Les professionnels de la santé, les chercheurs et les autorités en matière de santé publique doivent, par nécessité, recueillir et traiter une énorme quantité de renseignements personnels sur les patients, notamment leur numéro d'identification, leur numéro d'assurance maladie, leurs dossiers médicaux et leurs antécédents. La protection de la confidentialité de ces renseignements extrêmement personnels est de la plus haute importance, mais ceux-ci sont peut-être aussi la clé pour améliorer les diagnostics et les traitements, ce qui serait très bénéfique pour l'ensemble de la société.

Par exemple, Deloitte a été engagé par un organisme gouvernemental dont les données étaient réparties dans deux bases distinctes : celle de la santé et sécurité au travail, et celle des normes d'emploi. En regroupant ces deux bases de données, l'équipe de gestion pourrait déterminer s'il existe ou non une corrélation entre les demandes d'indemnisation des employés d'un certain employeur et la fréquence à laquelle ce même employeur n'a pas respecté ses obligations financières envers ses employés (p. ex., le versement rapide des salaires).

Pour aider l'équipe de gestion, Deloitte a créé un tableau de bord et a utilisé l'analytique avancée pour repérer dans chacune des bases de données des tendances au fil du temps, puis a comparé les résultats obtenus pour chaque base.

Nous avons également conçu un algorithme d'anonymisation pour masquer les renseignements protégés sur les employés grâce à des identifiants de substitution et permettre l'analyse des données agrégées sur les demandes d'indemnisation. L'algorithme masque les identifiants alphanumériques (p. ex., le numéro d'identification de l'employé, ses numéros de dossier médical, son salaire et les dates des événements médicaux ou des demandes d'indemnisation pertinents), tout en supprimant tous les autres renseignements protégés sur l'employé. Deloitte a ainsi pu démontrer à l'équipe de gestion qu'elle pouvait faire une analyse complète des données sur ses employés en mettant l'accent sur les corrélations à l'intérieur des données, sans pour autant compromettre la confidentialité des renseignements personnels des employés[1].

1. Deloitte Health Informatics LLC. Deloitte De-Identification Algorithm, 2012.

19. El Emam Khaled. *De-identifying Health Data for Secondary Use: A Framework*, [En ligne], octobre 2008 [http://www.ehealthinformation.ca/documents/SecondaryUseFW.pdf].

3. Contrôles d'accès par les utilisateurs

Il est toujours important de protéger les renseignements personnels contre tout accès non autorisé. Cependant, dans le domaine de l'analytique des données massives, la quantité et la variété des renseignements analysés rendent leur protection particulièrement vitale. Pour les ordinateurs en réseau, le contrôle de l'accès est le processus par lequel on accorde ou on refuse les demandes spécifiques d'obtention et d'utilisation de l'information et des services de traitement de l'information connexes20. Si on applique simultanément d'autres principes de sécurité intégrée tels que le droit d'accès minimal, le besoin de connaître, la confiance minimale et la séparation des tâches21, le contrôle d'accès est une façon efficace de protéger les renseignements personnels.

Par exemple, la base de données sur les clients d'une institution financière peut contenir une foule de renseignements sur ces personnes : le nom de leur employeur, leur revenu, l'identité de leur conjoint ou de leurs enfants, leur adresse et plus encore. Cependant, très peu de gens au sein de l'institution ont besoin d'accéder à cette information, ou du moins à toute cette information. Il est important de créer des niveaux d'accès appropriés aux renseignements personnels en suivant les principes du besoin de connaître et du droit d'accès minimal.

Astuce pour les contrôles d'accès par les utilisateurs : sécurité n'est pas synonyme de confidentialité. Bien qu'un niveau élevé de sécurité soit essentiel au maintien de la confidentialité, la protection de la vie privée englobe un ensemble beaucoup plus vaste de barrières autres que des mécanismes de sécurité. La protection de la vie privée régit non seulement la façon dont l'information est protégée et la façon d'y accéder, mais aussi la façon dont elle est recueillie et utilisée. Les contrôles d'accès par les utilisateurs protègent les renseignements personnels des menaces internes en éliminant même la possibilité de divulguer ou de faire un usage abusif des données, que ce soit de façon accidentelle ou intentionnelle. Cette protection est particulièrement nécessaire dans un monde où les ensembles de données sont de plus en plus volumineux. Les entreprises doivent revoir et évaluer régulièrement les protocoles de contrôle d'accès par les utilisateurs afin de s'assurer que leurs contrôles sont améliorés de façon continue à mesure que les systèmes évoluent et que les pratiques d'analytique changent.

20. Voir *Glossary of Key Information Security Terms*, NIST, [En ligne] [http://nvlpubs.nist.gov/nistpubs/ir/2013/NIST.IR.7298r2.pdf], p. 2.

21. Cavoukian Ann et Mark Dixon. *Privacy and Security by Design: An Enterprise Architecture Approach*, [En ligne], septembre 2013 [http://www.ipc.on.ca/site_documents/pbd-privacy-and-security-by-design-oracle.pdf].

Améliorer la sécurité au travail dans le secteur minier

Notre client, une multinationale du secteur minier qui avait beaucoup investi dans les processus, les structures, les contrôles et la culture de la sécurité, était aux prises avec un nombre inacceptable d'incidents graves, de blessures et de décès. Il a demandé à Deloitte de lui fournir des services d'analytique de la sécurité permettant de faire une évaluation objective et factuelle de son rendement actuel et de déterminer les principaux liens et les causes profondes qui auraient pu échapper aux gestionnaires de la sécurité.

Nous avons analysé un ensemble imposant et complexe de données liées et non liées à la sécurité, notamment les dossiers des employés et les données sur la formation, la production et le rendement des actifs. Puisque ces données contenaient des renseignements personnels, nous devions obligatoirement prendre des mesures pour nous assurer que la vie privée des personnes serait protégée dès le début. Certains renseignements, par exemple les protocoles de traitement et les dossiers médicaux, n'ont pas été inclus dans la collecte. Les numéros d'identification des employés ont été entièrement masqués afin d'éliminer le risque de réidentification. Nous avons même conseillé au client d'inclure l'équipe d'analytique dans l'évaluation rigoureuse de la sécurité et des privilèges d'accès, garantissant ainsi que personne n'aurait un accès plus grand que nécessaire aux données. Nous avons retiré tous les privilèges d'accès qui n'étaient pas nécessaires afin de protéger encore mieux la vie privée.

Ce projet a permis de recueillir de nouveaux renseignements sur les liens et les causes possibles de la performance du client en matière de sécurité. Grâce à ces résultats, la société minière a été en mesure d'améliorer ses pratiques de sécurité, ce qui a progressivement réduit le nombre d'incidents.

Le déploiement de ces stratégies – minimisation des données, anonymisation et contrôles d'accès par les utilisateurs – peut avoir un effet positif immédiat sur la capacité d'une entreprise à protéger la confidentialité des renseignements personnels qu'elle détient. En réduisant la quantité totale de données recueillies, en anonymisant rigoureusement les données et en limitant l'accès des utilisateurs, une entreprise est en mesure de certifier de façon proactive qu'elle respecte les règles de protection de la vie privée tout en préservant sa capacité à utiliser les données massives pour obtenir de nouvelles perspectives sur ses activités.

Innovation et protection de la vie privée : vous *pouvez* gagner sur tous les tableaux

Les entreprises continueront d'utiliser l'analytique des données massives pour progresser vers leurs objectifs stratégiques et mieux servir leurs clients. Cela ne signifie pas pour autant qu'elles doivent délaisser la protection de la vie privée, loin de là. Grâce à une planification soignée et à l'application de techniques et de principes tels que ceux qui font partie de la protection intégrée de la vie privée, elles peuvent utiliser les données pour obtenir l'effet désiré tout en protégeant les renseignements personnels.

Il faut faire preuve d'un solide leadership pour faire de la protection de la vie privée une priorité. Il faut aussi prendre des décisions avisées sur la conception et la mise en œuvre afin de véritablement intégrer cette pratique à l'ADN de l'entreprise. Une surveillance et une évaluation rigoureuses garantissent que les mesures adoptées aujourd'hui permettront de relever les défis de demain relatifs aux données.

Chez Deloitte, nous avons pris cet engagement en créant le Laboratoire national d'analytique et de recherche, qui est géré par notre équipe de spécialistes en protection de la vie privée. Cette installation de classe mondiale se consacre à l'exécution des meilleures pratiques en matière d'analytique des données, de protection de la vie privée et de sécurité des renseignements.

Nous sommes également déterminés à aider nos clients à trouver des façons novatrices d'adopter l'analytique et à mettre en place des pratiques judicieuses afin de protéger les renseignements personnels et les données confidentielles des entreprises.

Les données massives ne sont pas près de disparaître, mais cela ne veut pas dire de sacrifier la protection de la vie privée ou de 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 obtenir l'effet désiré tout en protégeant les renseignements personnels. Il est possible de gagner sur tous les tableaux.

Commissaire à l'information
et à la protection de la vie privée de l'Ontario
2, rue Bloor Est
Bureau 1400
Toronto (Ontario)
Canada M4W 1A8
Téléphone : 416-326-3333
Télécopieur : 416-325-9195
Courriel : info@ipc.on.ca
Site Web : www.ipc.on.ca

Deloitte
2, rue Queen Est, Bureau 1200
P.O. Box 8
Toronto (Ontario)
Canada M5C 3G7
Téléphone : 416-601-6150
Télécopieur : 416-601-6151
Courriel : deloitteanalytics@deloitte.ca
Site Web : www.Deloitte.ca/analytics

Privacy by Design: www.privacybydesign.ca

Le 10 juin 2014

Commissaire à l'information
et à la protection de la vie privée
Ontario, Canada

Deloitte.