

Le cadre qui sous-tend l'Outil de classification de la confidentialité et son rôle dans la modernisation

Joseph Duggan, Jack Gambino, Claude Girard et Michelle Marquis¹

Résumé

Avec l'Outil de classification de la confidentialité, Statistique Canada met en œuvre une composante, certes petite mais clé, de soutien à sa récente initiative de modernisation. Les renseignements statistiques de nature sensible, aussi appelés renseignements confidentiels, sont maintenant classifiés selon un continuum de risque qui remplace la classification binaire traditionnelle qui sous-tendait deux environnements de travail distincts : un réseau pour l'utilisation et le traitement internes de l'information protégée, et un autre réseau pour la communication externe et la diffusion de nos produits statistiques. La combinaison de ce changement avec d'autres initiatives - en lien avec l'utilisation de sources de données variées et combinées, à la modernisation de l'accès aux microdonnées et au regroupement d'un plus grand nombre de partenaires dans le cadre d'efforts de collaboration - permettra d'harmoniser nos pratiques de contrôle de la divulgation avec les infrastructures de TI actuelles et futures. L'outil vise à faciliter tout cela en sensibilisant davantage les gens aux questions et aux pratiques en matière de confidentialité, tout en aidant les administrateurs des données à déterminer le niveau de confidentialité de toutes données sélectionnées détenues à Statistique Canada. Le présent document décrit la méthodologie qui sous-tend l'Outil de classification de la confidentialité de conception simple, ainsi que les leçons apprises au cours de son élaboration.

Mots Clés : Renseignements statistiques de nature délicate; contrôle de la divulgation; gestion de l'information; accès aux microdonnées.

1. Introduction

La partie principale de ce travail décrit les détails techniques de l'outil, mais inclut aussi de l'information reliée au développement conceptuel du cadre de l'outil. À la suite de cette introduction, et après quelques mots traitant de la modernisation dans section 2, l'outil est décrit à la section 3 et puis décrit de façon plus technique à la section 4. L'article se conclut par quelques remarques sur le développement de l'outil et son avenir.

1.1 Une remarque sur la vie privée et la confidentialité

Même s'ils sont reliés, les deux concepts distincts de vie privée et de confidentialité méritent d'abord d'être clarifiés. Comme cela est décrit dans la Politique sur la protection des renseignements personnels et la confidentialité de Statistique Canada, les deux notions ne sont pas équivalentes. Le concept de vie privée réfère au droit de vivre en paix ainsi que celui d'être protégé de toute interférence, surveillance ou intrusion. Les gouvernements ont des obligations envers la collecte, l'usage, la divulgation, et la rétention des renseignements personnels. La confidentialité fait référence à la protection contre la diffusion d'information qui permettrait d'identifier une personne, une entreprise ou un établissement. Cette relation de confiance entre le fournisseur de l'information et l'organisation qui la recueille est fondée sur l'assurance que l'information fournie ne sera pas divulguée sans le consentement de la personne concernée ou d'une l'autorité législative appropriée.

Jusqu'à très récemment, la confidentialité a été perçue à Statistique Canada comme un sujet noir ou blanc, sa nature binaire étant due au caractère confidentiel ou non de l'information détenue. Ce principe a dicté la façon dont l'organisation fonctionnait, incluant la structure qu'elle a adoptée pour elle-même ainsi que pour ses avoirs en matière de données.

¹Joseph Duggan, Direction de la méthodologie, Statistique Canada, Ottawa, Ontario, Canada, K1A 0T6; joseph.duggan@canada.ca; Jack Gambino, Statistique Canada; jack.gambino@canada.ca, Claude Girard, Statistique Canada; claudio.girard@canada.ca et Michelle Marquis, Bureau de gestion de la protection de la vie privée et de coordination de l'information, Statistique Canada; michelle.marquis@canada.ca

2. Modernisation à Statistique Canada

La récente initiative de modernisation de Statistique Canada est basée sur cinq piliers : Prestation de services axés sur l'utilisateur; Méthodes et intégration des données à la fine pointe de la technologie; Renforcement des capacités statistiques et leadership; Partage et collaboration; ainsi qu'un Effectif moderne et un milieu de travail flexible.

Ensemble, ces piliers soutiennent Statistique Canada dans la réalisation de son énoncé de mission : « Au service du Canada avec des renseignements statistiques de grande qualité qui comptent. » En étant davantage axé sur l'utilisateur, Statistique Canada améliorera l'accès, pour des besoins statistiques, à des données plus pertinentes et plus actuelles. Ceci sera accompli en étendant les modes d'accès et en améliorant la mise en disponibilité des données, pour que celles-ci soient plus accessibles et que les sources dont nous disposons soient plus visibles aux chercheurs ainsi qu'autres utilisateurs. Il y aura une plus grande dépendance envers des sources de données multiples et variées, ainsi que sur les nouvelles technologies et les nouvelles approches méthodologiques liées à l'utilisation de ces sources et la protection de la confidentialité. Ces choses seront requises pour être en conformité avec l'engagement envers l'utilisation de méthodes de pointe et l'intégration des données. L'expansion des partenariats stratégiques avec des chercheurs académiques et des analystes en politique ainsi qu'avec les fournisseurs des données, et le renforcement de la capacité statistique sera clé dans ces avancements. Tous peuvent bénéficier d'une meilleure compréhension de la confidentialité, du risque, de la qualité, et des avantages et limitations des renseignements statistiques que Statistique Canada diffuse sous des formes variées.

2.1 Augmentation de l'accès

L'objectif de l'accroissement de l'accès aux renseignements doit toujours être tempéré par l'obligation de protéger la confidentialité des données. D'une part, l'accès non-restreint pour qui que ce soit à toutes les données n'accomplira rien en matière de confidentialité. De l'autre, la confidentialité absolue ne peut seulement être garantie qu'en empêchant tout accès aux données. Dans le cadre de sa modernisation, Statistique Canada a reconnu la tension entre ces deux états potentiels; Anil Arora, le statisticien en chef du Canada, a déclaré que « *Nous constatons déjà que, même si nous continuons de changer et de moderniser notre travail, nos valeurs fondamentales liées à la confidentialité et à la grande qualité des données restent les mêmes.* » (Statistique Canada, 2018) Statistique Canada cherche à accroître l'accès aux renseignements tout en conservant la confidentialité comme valeur fondamentale en mettant en place l'Outil de classification de la confidentialité.

2.2 La classification de la confidentialité

Le principe qui sous-tend l'Outil de classification de la confidentialité est de définir la confidentialité selon un continuum. Fondamentalement, tous les renseignements confidentiels demeurent confidentiels. Cependant, il y aura des degrés de confidentialité qui seront mis en correspondance avec les degrés de protection et de contrôle des données. En entrée au processus, il était admis que Statistique Canada devrait faire le bilan des renseignements sous sa responsabilité. Un outil auto-administré, à la fois sobre et simple a été conçu pour l'usage des praticiens des programmes statistiques, sous les conseils d'experts en matière d'accès et de confidentialité afin de compléter la classification.

3. L'Outil de classification de la confidentialité

Au fond, l'Outil de classification de la confidentialité (ci-après, l'outil) quantifie le niveau de risque de divulgation pour un produit statistique et la sensibilité des renseignements qu'il contient. Le risque est évalué en considérant quatre types de divulgation généralement étudiés (Duncan et al, 2011), qui sont légèrement modifiés pour les besoins de l'outil. Les pointages de sensibilité ont été adaptés à partir des niveaux proposés par le Bureau du commissaire à l'information et à la protection de la vie privée de l'Ontario. Ces évaluations du risque et de la sensibilité sont combinées pour obtenir un pointage final qui détermine le niveau de confidentialité du produit.

Comme il est mentionné dans l'introduction du guide de l'utilisateur de l'outil, « *l'Outil de classification de la confidentialité a été conçu afin d'appuyer le programme de modernisation de Statistique Canada[...] les renseignements statistiques de nature délicate (RSD) doivent maintenant être classés selon un continuum de risques. Le fait de remplacer la classification binaire classique des renseignements statistiques [...] facilitera la mise en place*

d'un milieu de travail mobile. Une telle approche modernisera l'accès aux microdonnées[...] ». Par ailleurs, l'un des dividendes prévus de rendre l'outil accessible aux personnes qui participent à la création, au stockage et à l'utilisation des données détenues au sein de Statistique Canada consiste à les sensibiliser davantage quant aux questions de confidentialité au sein de l'organisme.

L'objectif ultime était d'élaborer un outil qui aiderait à déterminer le niveau de confidentialité de tout ensemble de données que possède Statistique Canada. Un « ensemble de données » a été défini d'une manière très large : il peut s'agir tout autant d'un fichier de microdonnées, d'un tableau de statistiques, d'un registre statistique ou de tout autre élément qui pourraient contenir des renseignements statistiques de nature délicate, comme un graphique de points de données ou un dictionnaire de données provisoire renfermant des fréquences qui accompagne un fichier d'enquête.

3.1 Gouvernance

Les directeurs des divisions responsables des renseignements contenus dans le produit doivent apposer leur signature sur l'évaluation et affirmer que les évaluateurs avaient une connaissance suffisante des risques associés au produit. Un comité interne pour la revue et diffusion des microdonnées examinera les évaluations effectuées en utilisant les métadonnées saisies par l'outil, et comparera les résultats à l'intérieur des programmes et à travers l'organisation. Les pointages qui proviennent de l'utilisation de l'outil pourront être présentés au comité à des fins d'appel ou de dérogation, selon les circonstances et à l'initiative de l'administrateur des données.

3.2 Application

L'outil a été conçu pour l'usage par des personnes qui connaissent les données sous évaluation, mais qui ne sont pas nécessairement experts dans les méthodes de contrôle de la divulgation. Les programmes peuvent mettre en œuvre l'outil pour tous leurs produits existants (fichiers diffusés et de production, fonds de données) ou peuvent l'appliquer uniquement lorsque cela s'avère nécessaire. Par exemple, le recours à l'outil peut se faire en ciblant des nouveaux programmes et projets, ou ceux qui sont remaniés, ou l'outil être appliqué que lors de diffusions spéciales, de nouvelles initiatives de collaboration/partage, ou qu'à des fichiers uniques figurant dans les avoirs de données. Pour faciliter l'emploi de l'outil, les utilisateurs sont informés qu'il peut être appliqué à un seul produit ou à un groupe de produits à la fois – suivant la nomenclature du Modèle générique du processus de production statistique, développé pour le Groupe de haut niveau sur la modernisation de la production et des services statistiques en 2013.

3.3 Contenu et forme actuelle

En ce moment, la première version de l'outil existe sous la forme d'un classeur Excel avec quatre feuilles interdépendantes. La première feuille contient les instructions de base et de l'information choisie tirée du guide d'utilisateur distinct qui s'accompagne l'outil. La deuxième feuille présente quatre questions pour guider l'évaluation du risque de divulgation et la troisième sert à l'évaluation du niveau de sensibilité. La quatrième et dernière feuille résume l'information recueillie dans les deuxième et troisième feuilles, et présente le pointage global qui en résulte. L'information sommaire qui y est présentée inclut des détails qui identifient le produit évalué et l'évaluateur principal. Jointe à cette information est l'attestation par le directeur tel que décrit dans la section 3.1 ci-haut.

3.3.1 Questions sur le risque de divulgation

La feuille portant sur l'évaluation du risque de divulgation comporte quatre questions suivant le même modèle, avec une question pour chacun des quatre grands types de divulgation : divulgation d'identité, divulgation par attributs, divulgation inférentielle, et la divulgation résiduelle. L'évaluation du risque engendré par chaque type de divulgation doit se faire indépendamment d'un type à un autre. Les évaluateurs sont amenés à traiter chaque type de divulgation comme un « point d'entrée » aux données, et ce sans considération quant aux mesures de protection de confidentialité déjà en vigueur.

La question pour la *Divulgation d'identité* est précédée par la mention qu'elle est à risque de se produire lorsque des renseignements confidentiels peuvent être divulgués par l'entremise des identificateurs directs que le produit contient. La question se lit : Des renseignements confidentiels peuvent-ils être révélés directement à partir des renseignements affichés par le produit? L'évaluation du risque est menée en fonction du niveau d'effort requis pour arriver à une

divulgarion. Les catégories de réponse sont : Risque FAIBLE car cela nécessite un effort IMPORTANT; Risque MOYEN car cela nécessite un CERTAIN effort; et Risque ÉLEVÉ car cela nécessite PEU ou pas d'effort.

L'outil décrit la *Divulgarion par attributs* comme l'association des données avec une entité précise. Cela survient habituellement lorsqu'on regroupe les éléments décrits comme étant des identificateurs indirects ou des renseignements « visibles ». La question qui est posée par l'outil est : Des renseignements confidentiels peuvent-ils être révélés en regroupant les attributs disponibles dans le produit? Ici aussi, le risque est évalué en termes du niveau d'effort requis pour qu'une divulgation ne survienne en utilisant les mêmes catégories que précédemment.

On dit qu'il y a *Divulgarion inférentielle* lorsqu'une affirmation avec un haut niveau de confiance quant à son exactitude peut être émise à propos d'un individu en particulier. Des renseignements confidentiels peuvent-ils être révélés par une affirmation probabiliste tirée des renseignements contenus dans le produit ? Habituellement, les produits diffusés par des agences statistiques sont créés pour soutenir l'inférence statistique et, en conséquence, ce type de divulgation n'est pas typiquement aussi sévère que les autres. La manière dont ceci est reflété dans le pointage calculé par l'outil sera décrite plus tard dans la section 4.3.

La *Divulgarion résiduelle* prend en compte l'environnement complet de diffusion; les évaluateurs sont tenus d'évaluer le risque de divulgation résiduelle au meilleur de leurs connaissances en tant qu'experts du sujet et en ce qui a trait aux autres sources d'information pertinentes, n'attendant pas d'eux qu'ils sachent parfaitement quelles informations sont déjà connues du public. Des renseignements confidentiels peuvent-ils être révélés en combinant les renseignements contenus dans ce produit et ceux d'autres produits? Il est à noter que cette définition de la divulgation résiduelle porte sur les produits entre eux et ne traite pas d'enjeu de divulgation résiduelle qui puisse exister au sein d'un même produit. Toute manifestation de risque de divulgation résiduelle interne doit être traitée selon le type de divulgation approprié considéré par l'outil.

3.3.2 Question sur la sensibilité

Dans la section de l'outil où la sensibilité est évaluée, il n'y a qu'une seule question à répondre. Est-ce que la divulgation de renseignements dans un format identifiable :

- ...causerait de graves dommages à une personne ou à une entreprise?
- ...causerait des dommages considérables à une personne ou à une entreprise?
- ...ternirait la réputation d'une personne ou d'une entreprise ou la gênerait-elle?
- ...causerait des dommages minimes à une personne ou à une entreprise?
- ...ne causerait aucun dommage puisque l'on considère ces renseignements accessibles au public?

La catégorie de réponse choisie devrait refléter un scénario plausible et non un scénario extrême. Notez aussi que l'incidence sur Statistique Canada d'une divulgation n'a pas été prise en compte ici, ni ailleurs dans l'outil. Il ne s'agit pas d'une omission, mais de la reconnaissance que cette incidence est jugée être significative quel que soit le produit évalué au moyen de l'outil.

4. Description technique de l'outil

4.1 Aspects de la conception

Les différents paramètres utilisés dans les calculs pourraient être modifiés dans les futures versions de l'outil. La structure sous-jacente de l'outil et le modèle sur lesquels les calculs s'appuient ont également été conçus d'une manière assez simple qui permette de les adapter aux besoins futurs. Il est par ailleurs également possible de modifier le fonctionnement de ce modèle, puisque la façon décrite dans le présent document ne constitue qu'une seule des nombreuses possibilités.

4.2 Déterminer le niveau de classification des renseignements statistiques

Le niveau de classification d'un ensemble de données (ou d'un produit), p , est une fonction du pointage associé au risque de divulgation multiplié par le pointage de sensibilité du produit. La première étape présente simplement les

deux facteurs (voir tableau 4.3.1-1 ci-dessous) et montre le résultat de leur multiplication. Lors de la deuxième étape, le tableau 4.3.1-2 est utilisé comme tableau de correspondance servant à donner le pointage global (ou niveau de classification) du produit. Un tableau de correspondance comportant une seule étape serait plus simple, mais comme l'outil fait tous les calculs lui-même, ces tableaux montrent bien aux évaluateurs la relation entre les deux facteurs.

Le pointage associé au risque de divulgation est la somme des quatre composantes de risques de divulgation qui sont évaluées de façon *indépendante* : la divulgation de l'identité, la divulgation par attributs, la divulgation inférentielle et la divulgation résiduelle. Les évaluations, soit « Faible », « Moyen » ou « Élevé », sont associées au risque que pose chaque type en tant qu'avenue potentielle menant à une divulgation en supposant que les seules mesures prises pour la prévenir sont celles déjà en vigueur dans les données. Ces évaluations sont transformées en valeurs qui dépendent du type de divulgation en question. Pour nos besoins, leur somme est plafonnée à une valeur de 10.

Le pointage associé à la sensibilité correspond à la valeur directe du niveau d'incidence qu'une violation aurait sur un particulier, une entreprise ou une institution à la suite d'une divulgation de renseignements. Elle peut être évaluée comme étant « Grave », « Élevé », « Moyen », « Faible » ou « Négligeable »; ces pointages prennent respectivement les valeurs de cinq à un.

4.3 Calcul du niveau de classification des renseignements statistiques

Le niveau de classification C d'un ensemble de données p est calculé à partir de la fonction en escalier $F(RD_p \times S_p)$, où le pointage associé au risque de divulgation, $RD_p = \min [10, \Sigma (\text{Ident}_p + \text{Attr}_p + \text{Inf}_p + \text{Rés}_p)]$ est une valeur entière entre 0 et 10 et le pointage pour la sensibilité, S_p , est une valeur entière entre 1 et 5.

Les quatre éléments du vecteur de risque de divulgation (Ident_p , Attr_p , Inf_p , Rés_p) sont, respectivement, les valeurs d'évaluation du risque découlant de la divulgation de l'identité, de la divulgation par attributs, de la divulgation inférentielle et de la divulgation résiduelle pour p .

$$\text{Ident}_p = \begin{cases} 10, & \text{si la cote est } \textit{Élevé}, \\ 2, & \text{si la cote est } \textit{Moyen}, \\ 0, & \text{si la cote est } \textit{Faible}. \end{cases}$$

$$\text{Attr}_p = \begin{cases} 3, & \text{si la cote est } \textit{Élevé}, \\ 2, & \text{si la cote est } \textit{Moyen}, \\ 0, & \text{si la cote est } \textit{Faible}. \end{cases}$$

$$\text{Inf}_p = \begin{cases} 2, & \text{si la cote est } \textit{Élevé}, \\ 1, & \text{si la cote est } \textit{Moyen}, \\ 0, & \text{si la cote est } \textit{Faible}. \end{cases}$$

$$\text{Rés}_p = \begin{cases} 3, & \text{si la cote est } \textit{Élevé}, \\ 1, & \text{si la cote est } \textit{Moyen}, \\ 0, & \text{si la cote est } \textit{Faible}. \end{cases}$$

Les valeurs attribuées à chacune des trois mêmes cotes « Faible », « Moyen » ou « Élevé » varient afin de refléter l'incidence relative qu'ont les divers types sur le niveau global de risque de divulgation. Ces valeurs peuvent également être interprétées comme étant des poids ou facteurs d'importance relative.

Le pointage pour la sensibilité peut prendre les valeurs suivantes :

$$S_p = \begin{cases} 5, & \text{si la cote est } \textit{Grave}, \\ 4, & \text{si la cote est } \textit{Élevé}, \\ 3, & \text{si la cote est } \textit{Moyen}, \\ 2, & \text{si la cote est } \textit{Faible}, \\ 1, & \text{si la cote est } \textit{Négligeable}. \end{cases}$$

Et enfin, la fonction en escalier F peut être évaluée à l'aide du produit ($RD_p \times S_p$) afin d'obtenir le niveau de confidentialité résultant; où $C_p = F(RD_p \times S_p)$, comme suit :

$$C_p = \begin{cases} 9, & \text{si } RD_p \times S_p \in \{45, 50\}, \\ 8, & \text{si } RD_p \times S_p \in \{36, 40\}, \\ 7, & \text{si } RD_p \times S_p \in \{30, 32\}, \\ 6, & \text{si } RD_p \times S_p \in \{21, 24, 25, 27\}, \\ 5, & \text{si } RD_p \times S_p \in \{15, 16, 18, 20\}, \\ 4, & \text{si } RD_p \times S_p \in \{10, 12, 14\}, \\ 3, & \text{si } RD_p \times S_p \in \{7, 8, 9\}, \\ 2, & \text{si } RD_p \times S_p \in \{4, 5, 6\}, \\ 1, & \text{si } RD_p \times S_p \in \{1, 2, 3\}, \text{ et} \\ 0, & \text{si } RD_p \times S_p = 0. \end{cases}$$

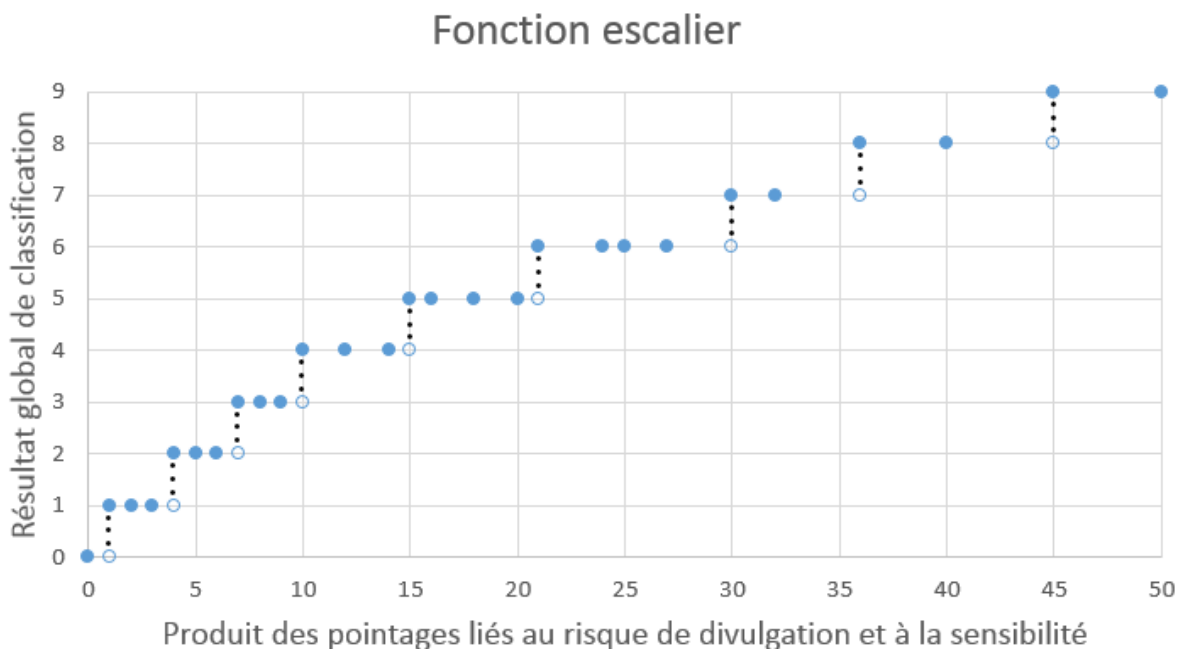
4.3.1 Exemple de calcul

Cette sous-section de l'article présente un exemple fictif de calcul d'un pointage global.

Si le vecteur RD des éléments de risques (Identité, Attribut, Inférentiel, Résiduelle) se voyait attribuer les niveaux (*faible, moyen, élevé, moyen*), cela se traduirait en un vecteur de valeurs (0, 2, 2, 1) qui, lorsqu'additionnées, donnerait un total de 5 comme pointage de RD.

Pour poursuivre cet exemple, si le niveau de sensibilité du produit est évalué à « Moyen », il serait converti en une valeur de 3. Le produit mathématique du pointage de RD et le pointage de sensibilité serait alors 15 ($=5 \times 3$), et le niveau de classification résultant serait de 5 selon la fonction en escalier montrée en Figure 4.3.1-1.

Figure 4.3.1-1
Fonction escalier donnant la relation entre le pointage global et le niveau de classification final



*Les cercles ouverts et les lignes pointillées qui sont utilisés habituellement pour les fonctions en escalier à valeurs réelles sont présentés ci-dessus simplement afin de mettre en évidence les échelons.

Comme un aparté, une interprétation intéressante est que le pointage pour le risque de divulgation peut être vu comme la longueur maximale dans un espace multidimensionnel engendré par les risques potentiels. Notez que si les valeurs utilisées pour le pointage assigné aux quatre composantes de vecteur de Risque de divulgation ($Ident_p$, $Attr_p$, Inf_p , $Rés_p$) sont plutôt remplacées par leurs racines carrées, le pointage pour le risque de divulgation sera équivalent au carré de la longueur euclidienne du vecteur de risque de divulgation – avec un plafond à 10. Dans ce cas, pour l'exemple en haut ($Ident_p$, $Attr_p$, Inf_p , $Rés_p$) sera égal à $(0, \sqrt{2}, \sqrt{2}, 1)$. La longueur de ce vecteur, ou la racine carrée de son pointage RD, est alors $(\sqrt{0^2} + \sqrt{2^2} + \sqrt{2^2} + \sqrt{1^2})^{\frac{1}{2}}$, ou $\sqrt{5}$. Multipliant cette distance par le pointage pour la sensibilité définie un volume pour le niveau de confidentialité global.

Tableau 4.3.1-1
Détermination du pointage global via le risque de la divulgation et le pointage de sensibilité

Sensibilité	5	0	5	10	15	20	25	30	35	40	45	50
	4	0	4	8	12	16	20	24	28	32	36	40
	3	0	3	6	9	12	15	18	21	24	27	30
	2	0	2	4	6	8	10	12	14	16	18	20
	1	0	1	2	3	4	5	6	7	8	9	10
		0	1	2	3	4	5	6	7	8	9	10
		Risque potentiel de divulgation										

Tableau 4.3.1-2
Détermination tabulaire d'un niveau de classification basé sur le pointage global de risque-sensibilité

Niveaux de classification finaux fondés sur le résultat combiné du risque de divulgation et de la sensibilité :		
Limite inférieure	Limite Supérieure	Niveau
45	50	9
36	44	8
30	35	7
21	29	6
15	20	5
10	14	4
7	9	3
4	6	2
1	3	1
0	0	0

Au lieu d'une fonction escalier, l'outil présente le pointage global sous forme d'un tableau qui est plus facile à comprendre. Ce format utilise un gradient de couleurs s'étendant du vert qui correspond à un risque faible jusqu'au rouge pour la zone de risque élevé diagonalement opposée. Les valeurs qui se trouvent à l'intérieur du tableau donnent lieu, par leur nature, à un coloris comportant plus de vert que de rouge. Les associations pré-conditionnées au choix de couleurs sont faites pour renforcer la confiance à l'égard des évaluations de risque faibles et les préoccupations en ce qui concerne les évaluations de risque plus élevées. Les tailles relatives de ces régions sont faites pour refléter l'objectif de la modernisation envers l'augmentation de l'accès, tout en reconnaissant que la confidentialité a été évaluée correctement pour tous les produits.

5. Le passé n'est qu'un prologue

5.1 Rétrospective

En résumé, il peut être instructif de passer en revue quelques-uns des événements majeurs qui ont marqué le développement de l'outil. Les premières tentatives de créer une « liste de vérification » pour guider l'évaluation des niveaux de confidentialité ont donné lieu à des arbres de classification. Ces arbres devenaient très grands et lourds lorsqu'est venu le moment de prendre correctement en compte les divers types de données (sociale / entreprise/ institutionnelle, enquête/données administratives, etc.).

Comme alternative, il a été proposé d'utiliser les quatre grands types de divulgation comme cadre, et de mettre de côté les considérations spécifiques aux diverses formes de données. En débattant de la manière de combiner les pointages obtenus à l'égard de ces types en un pointage global du risque de divulgation, l'application des poids s'est avérée la direction logique, et méthodologique, à prendre – mais cela a exigé du leadership et du travail de consultation afin d'arriver à la fonction et aux paramètres décrits dans cet article. Mettre de côté le concept de sensibilité jusqu'au moment qu'il y ait entente quant à la façon d'évaluer le risque de divulgation s'est révélé être presque providentiel, puisqu'une notion de sensibilité ou de valeur rattachée à une chose semble être enracinée dans la manière dont la nature humaine fait l'évaluation du risque. Pour les dimensions de la confidentialité, il semblait aussi naturel d'utiliser une grille à deux dimensions pour combiner les facteurs et pour présenter leurs relations. Le dernier événement marquant méritant une mention était une autre généralisation: l'usage du Modèle générique du processus de production statistique pour créer des classes des produits similaires provenant d'un secteur de programme et à réduire le nombre d'évaluations requises dans la pratique.

5.2 Développements futurs

Pour conclure, rappelons que la confidentialité est, comme toujours, une valeur fondamentale à Statistique Canada. L'Outil de classification de la confidentialité doit être, tout d'abord, un outil simple. Il existe des plans pour développer une version de l'outil qui fonctionne sur le Web et rendre disponible les résultats des évaluations disponibles sous forme de métadonnées rattachées aux produits statistiques. L'outil doit aussi contribuer à accroître davantage la sensibilisation quant à la manière dont la confidentialité est maintenue lorsque l'accès aux renseignements statistiques est accru. Il ne s'agit là que d'une petite partie du plus grand tableau qu'est la modernisation au sein de Statistique Canada, et l'outil sera - et l'a déjà été - partagé avec autres départements et agences. En définitive, et tel que prévu depuis ses tout débuts, l'outil sera modifié et amélioré au fil de temps.

Reconnaisances

Les auteurs souhaitent reconnaître la participation des autres membres du groupe de travail, et de ses gestionnaires, qui ont collaboré au développement de l'Outil de classification de la confidentialité. De plus, plusieurs secteurs de programme doivent être nos remerciés pour leurs travaux d'évaluation de produits sélectionnés dans les essais avec l'outil et le guide d'utilisateur. Finalement, ils remercient François Brisebois pour ses contributions perspicaces au projet et pour ses revues utiles de cet article.

Bibliographie

Duncan, G., Elliot, M. et Salazar-Gonzalez, J. (2011), *Statistical Confidentiality: Principles and Practice*, New York: Springer-Verlag.

Federal Committee on Statistical Methodology (2005). "Statistical Policy Working Paper 22 (Second version): Report on Statistical Disclosure Limitation Methodology." NTIS PB94-165305.

OECD Glossary of Statistical Terms. <https://stats.oecd.org/glossary/index.htm>

Statistique Canada (2018), Le Bulletin sur la modernisation: juin 2018, bulletin mensuel interne, Ottawa, Canada

Statistique Canada (2018b), "L'outil de classification de la confidentialité : Guide de l'utilisateur (version 1.0 external)", Ottawa, Canada

UNECE (2013). Version 5.0 of the Generic Statistical Business Process Model (GSBPM)
<https://statswiki.unece.org/display/GSBPM>