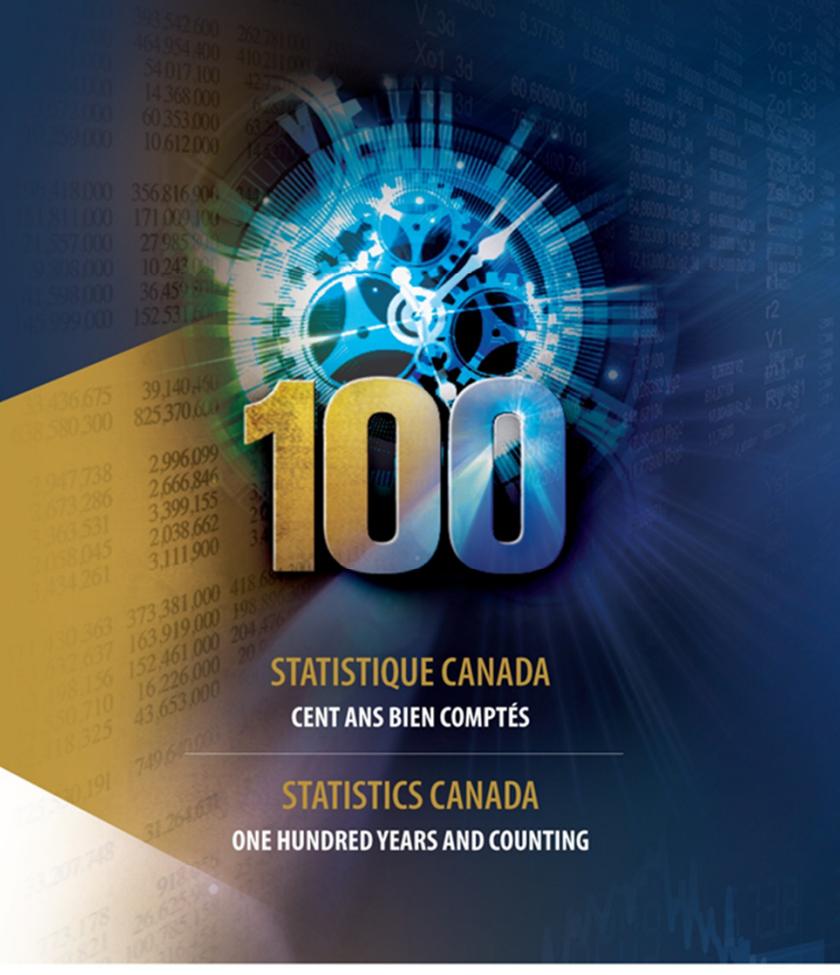


La Protection des Données: Le point de vue canadien

Wesley Yung, Statistique Canada



100

STATISTIQUE CANADA
CENT ANS BIEN COMPTÉS

STATISTICS CANADA
ONE HUNDRED YEARS AND COUNTING



Statistique
Canada

Statistics
Canada

Canada

Plan de la présentation

- Le contexte
- L'outil de classification de confidentialité
 - Le risque de divulgation
 - L'impact d'une divulgation
- Niveau de sensibilité
- Utilisation potentielle
- Sommaire





Le contexte

- Modernisation de Statistique Canada
- Cinq piliers
 - Prestation de services axés sur l'utilisateur
 - Méthodes à la fine pointe
 - Renforcement des capacités statistiques
 - Partage et collaboration
 - Effectif moderne et milieu de travail flexible





Le contexte

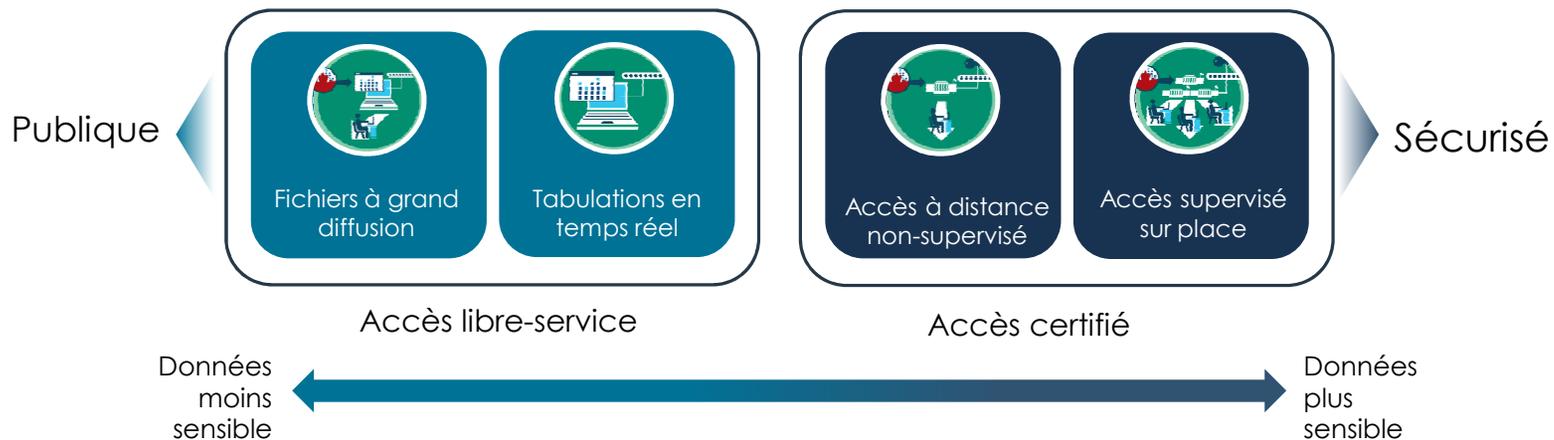
- Prestation de services axés sur l'utilisateur
 - Les utilisateurs disposent des renseignements et des données dont ils ont besoin
- Les utilisateurs demandent de plus en plus de données
 - De plus en plus détaillées (micro-données)
- Mais comment protéger la confidentialité?
 - En plus des trucs typiques (anonymisation des données, etc.)



Le contexte

- Dans le passé, les données étaient sensibles ou non-sensibles
 - Très restrictif car la plupart des données étaient classifiés comme sensibles
- Il est clair maintenant que la sensibilité est subjective
 - Pour un même produit une personne dirait qu'il est sensible, alors qu'une autre dirait qu'il ne l'est pas
- Nous avons besoin d'un continuum de sensibilité

Le contexte



- Mais comment classifier les données?



L'outil de classification de confidentialité

- Utilise d'une approche à l'échelle pour permettre
 - Une classification cohérente
 - Réduire la subjectivité
- L'outil de classification de confidentialité (OCC) a été conçue pour soutenir un pilier du modernisation de Statistique Canada
 - Développé pour aider à classer la sensibilité des données



L'outil de classification de confidentialité

- L'OCC classifiera l'information statistique le long d'un continuum de risque
- C'est un questionnaire court, simple et auto-administré
- Il attribue un score basé sur les réponses à diverses questions
- Il évalue deux aspects d'un fichier
 - Risque de divulgation
 - L'impact d'une divulgation



Le risque de divulgation

- C'est le risque de ré-identification
 - Quatre types de divulgation : identité, attributs, déductif et résiduel
 - Evalué sur une base de trois risques : faible, moyen ou élevé
 - Si vous nécessitez des efforts importants: risque faible
 - Par exemple, connaître les règles de Statistique Canada
 - D'autre côté, s'il y a peu d'efforts nécessaires: risque élevé
 - Par exemple, connaître l'industrie d'une entreprise dominante



Le risque de divulgation

- Pour chaque type de divulgation, il faut évaluer le risque et assigner un score pour chaque type et classe de risque

Évaluation de risque	Description
Risque faible	Effort significatif requis pour divulguer l'information confidentielle
Risque moyen	Effort requis
Risque élevé	Peu ou aucun effort requis

- Les scores peuvent varier selon le type de divulgation



Le risque de divulgation

- Par exemple, pour la divulgation d'identité, les valeurs pour faible, moyen et élevé peuvent être 1, 3 et 8, respectivement, alors que pour divulgation par déduction, elles pourraient être 1, 3 et 4
- Nous obtenons ensuite la somme des quatre scores (plafonnés à 10) pour obtenir le score de risque.



L'impact d'une divulgation

- Basé sur la nature et contenu de fichier, assigne un score entre 1 et 5

Évaluation d'impact	Description
Négligeable	La divulgation ne causerait aucun dommage étant donné que l'information est considérée comme accessible au public
Faible	La divulgation causerait un dommage minime
Moyen	La divulgation causerait un dommage à la réputation ou à la gêne d'un individu ou d'une entreprise
Élevé	La divulgation causerait un dommage considérable à un individu ou à une entreprise
Sévère	La violation de l'information porterait gravement dommage à un individu ou à une entreprise



Niveau de sensibilité

- Le score de risque est multiplié par l'impact pour obtenir un valeur entre 0 et 50
- En utilisant une fonction échelonnée, le score est traduit en niveaux. Les niveaux de risque sont incrémentiels

Impact	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										



Niveau de sensibilité

- 10 niveaux sont identifiés
- Les bornes pourraient être différentes étant donné la situation

Borne inférieure	Borne 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



Utilisation potentielle

- Milieu de travail flexible
 - Niveau 1 et 2 : Travail dans les endroits publiques
 - Niveau 3 et 4 : Travail dans les endroits avec accès au publique restreint (par exemple, les chambres d'hôtel)
 - Niveau 5 : Travail chez vous
 - Niveau 9 : Travail seulement dans le siège social de Statistique Canada



Sommaire

- La confidentialité des données reste une pierre angulaire de Statistique Canada
- Mais, il faut changer avec le temps
- Aujourd'hui, la classification de sensible/non-sensible ne fonctionne plus
- En utilisant l'OCC, Statistique Canada essaie de protéger la confidentialité des données tout en augmentant l'accès

MERCI!

Pour de plus amples
renseignements, visitez le
www.statcan.gc.ca



#StatCan100

