

**ALLOCUTION D'OUVERTURE PAR LE  
STATISTICIEN EN CHEF DU CANADA  
ANIL ARORA**

— — — — —  
**INFONEX — LES MÉGADONNÉES ET  
L'ANALYSE POUR LE SECTEUR PUBLIC**

# **Allocution d'ouverture du statisticien en chef du Canada Infonex — Les mégadonnées et l'analyse pour le secteur public Le 1<sup>er</sup> octobre 2019**

---

## **Introduction**

Bonjour. Je m'appelle Anil Arora et je suis le statisticien en chef du Canada. Je suis ravi d'être ici aujourd'hui pour parler des mégadonnées et de l'analyse.

Comme vous le savez, nous faisons tous partie d'une société axée sur les données et nous y contribuons tous. Les chiffres, les mesures et les innombrables façons que nous avons adoptés pour quantifier les nombreux aspects de notre société font maintenant partie de notre quotidien. Nous utilisons nos téléphones, ou d'autres appareils, pour suivre le nombre de pas que nous faisons, le nombre de calories que nous consommons et l'argent que nous dépensons. De nos jours, les stimulateurs cardiaques sont dotés d'une intelligence artificielle qui permet de surveiller et d'ajuster le rythme cardiaque de manière optimale, ce qui contribue à sauver des vies. Cet interminable flux de données nous entoure et, à titre de société, nous reconnaissons maintenant à quel point les données représentent un produit précieux.

Aujourd'hui, j'aimerais aborder cinq sujets :

- le fait que le volume n'est pas synonyme de qualité;
- l'utilisation des données comme actif stratégique;
- l'exploration responsable de nouvelles technologies habilitantes;
- la collaboration à l'appui de l'innovation;
- la reconnaissance du rôle essentiel que jouent les statistiques officielles pour notre économie et notre société.

## **Volume et qualité**

Parlons d'abord de volume. La quantité de données produites ne cesse d'augmenter. Selon Forbes, 2,5 TRILLIONS d'octets de données sont créés chaque jour. Il est difficile de se faire une idée de ce que cela représente réellement. De plus, les sciences et les technologies de l'information ne peuvent que faire augmenter ce nombre. Par ailleurs, plus de 3,7 milliards de personnes utilisent Internet.

Cela nous amène à nous interroger sur la façon dont nous pouvons tirer parti de ces données pour produire des renseignements utiles.

Lorsqu'il est question de mégadonnées, nous devons reconnaître que le volume n'est pas synonyme de qualité. Même si des tonnes de renseignements sont disponibles, ils ne sont utiles à personne tant qu'ils n'ont pas été analysés. Ce n'est que lorsque les données sont analysées qu'elles peuvent être utilisées pour prendre des décisions.

Il est facile de se laisser emporter par les données et les mesures, surtout pour les entreprises. En fait, un article paru récemment dans le *Harvard Business Review* explique comment un *trop grand nombre de dirigeants confondent les chiffres et la stratégie*. Il avertit les dirigeants de ne pas trop se fier aux chiffres parce qu'ils peuvent les faire dévier de leur stratégie. C'est un point vraiment intéressant, parce que cela démontre comment l'absence d'analyse peut avoir un impact négatif sur les entreprises ou sur tout autre type d'organisation. Trop de données sans les moyens ou les connaissances nécessaires pour les utiliser de façon périodique pourraient être préjudiciables.

## **Les données comme actif stratégique**

Une bonne analyse permet de réellement comprendre les données. Cette compréhension appuie la prise de décisions éclairées, ainsi que l'élaboration de politiques concernant les enjeux réels qui nous touchent au quotidien. N'importe qui, et en fait même une machine, peut extraire des chiffres, mais il faut inclure un élément d'expertise humaine pour dégager la véritable valeur des données, afin de pouvoir les utiliser comme un actif stratégique.

Alors, quel est le rôle de Statistique Canada, l'organisme national de statistique, dans tout cela? Eh bien, c'est exactement cette expertise qui nous caractérise. Nous nous considérons comme l'élément humain, ou les « maniaques des données » pour certains, derrière les chiffres.

Le besoin de données sur l'épidémie d'opioïdes est un exemple clair de données constituant un atout stratégique. Comme vous le savez peut-être, c'est un problème répandu au Canada, qui a de graves répercussions physiques, psychologiques, sociales et économiques. Nous avons collaboré avec la Ville de Surrey pour obtenir des données des premiers intervenants au sujet des personnes qui ont malheureusement perdu la vie à cause de cette crise. Nous avons également pu obtenir des données d'organismes provinciaux qui montrent ce qui est arrivé à ces personnes six mois ou un an avant leur décès, comme leurs interactions avec les services de santé ou le système judiciaire. Cela a pour résultat que les données peuvent être utilisées pour éliminer la stigmatisation et comprendre vraiment ce qu'est la crise, ainsi que les politiques et les programmes qui nous permettraient le mieux de nous y attaquer, de nous informer sur les interventions opportunes et efficaces.

Dans de tels cas, la contribution des organismes nationaux de statistique, ce que nous appelons les ONS, est très évidente. Pour répondre aux besoins d'information de la société, les ONS ne peuvent pas fonctionner uniquement comme des fournisseurs de données. Nous devons fournir également l'expertise en analyse nécessaire pour résoudre les problèmes complexes qui touchent la société aujourd'hui.

J'aimerais prendre quelques instants pour parler de certaines des façons dont nous, à Statistique Canada, nous adaptons et nous réorientons, à l'ère des mégadonnées, pour continuer de remplir ce rôle.

## **Évolution des attentes et des besoins des utilisateurs / nouvelles technologies**

Notre organisme diffuse des faits depuis plus de 100 ans. Nous reconnaissons que les temps ont changé et que la relation des Canadiens avec les données a changé. Les attentes continuent d'augmenter. Les Canadiens veulent plus d'information, ils veulent qu'elle soit adaptée à leurs besoins personnels, tout en ayant l'assurance que leurs renseignements seront dûment protégés.

Nous avons dépassé le stade où nous parlions de la nécessité de nous moderniser et nous avons pris des mesures pour nous adapter à ce nouvel environnement. Nous avons fermement ancré la modernisation dans notre façon de travailler, dans notre façon de penser... dans notre culture. Et c'est ce qui nous propulse vers l'avant, nous met en meilleure position pour réagir aux changements qui nous entourent.

Parce qu'il ne s'agit pas seulement de moderniser nos processus. Il faut aussi changer notre point de vue sur la façon dont nous servons les gens comme vous et les Canadiens qui participent au processus de politique publique, dans les entreprises, les universités, les associations et d'autres entités qui contribuent à améliorer nos collectivités.

L'amélioration continue, pour un organisme qui existe depuis 100 ans, est une nécessité, car nous reconnaissons que nous ne sommes pas les seuls intervenants qui existent. Beaucoup d'entreprises, de gouvernements et d'organisations approfondissent le monde des données et produisent ou fournissent leurs propres données. À Statistique Canada, nous voyons cela comme une occasion de nous transformer pour mieux servir le secteur public et celui des affaires.

Nous avons amélioré l'actualité des données accessibles au public en trouvant de nouvelles sources d'information, en élaborant de nouvelles méthodes et de nouveaux processus et en collaborant plus étroitement avec les partenaires. Nous continuons de mettre l'accent sur la production de renseignements de base sur lesquels les utilisateurs des données s'appuient, comme l'Indice des prix à la consommation ou le produit intérieur brut. De plus, nous offrons aux Canadiens de nouveaux outils qui rendent l'utilisation des statistiques plus pertinente, comme notre Explorateur du commerce international, une série de quatre présentations visuelles de données interactives qui permettent de découvrir les relations commerciales du Canada et leur évolution au fil du temps.

Bien que nous soyons déterminés à produire des données de grande qualité, nous maintenons notre engagement de protéger la vie privée et la confidentialité des Canadiens. Le maintien de la confiance de la population canadienne est au cœur de nos activités et de tout ce que nous faisons. C'est notre devise depuis plus de 100 ans. Nous prenons la vie privée des Canadiens très au sérieux et nous protégeons les données que nous recueillons. Les données obtenues à partir de nouvelles sources sont utilisées uniquement à des fins statistiques et sont traitées avec le même degré de

vigilance, de confidentialité et de sécurité que les données recueillies au moyen de méthodes traditionnelles.

### ***Données administratives***

L'une des façons d'améliorer l'actualité des données consiste à utiliser de façon responsable de nouvelles sources de données, comme les données administratives. En fait, Statistique Canada utilise et protège les données administratives depuis sa création.

Ces types de données offrent une occasion de collaboration entre les producteurs de ces données et nous, à Statistique Canada. Les données administratives peuvent mener à une situation « gagnant-gagnant » pour tout le monde, car nous pouvons fournir des renseignements plus rapidement et plus fréquemment. Nous pouvons aussi réduire le fardeau de réponse et les coûts en nous éloignant des enquêtes traditionnelles. Ces occasions de collaboration nous permettent de profiter de l'expertise de chacun. De plus, elles nous permettent d'être plus centrés sur l'utilisateur, tout en maintenant notre rôle de fiduciaire de confiance de données clés pour les Canadiens.

### ***Intelligence artificielle et apprentissage automatique***

Nous avons également recours à l'intelligence artificielle et à l'apprentissage automatique :

- L'intelligence artificielle a servi à analyser 1,1 million de commentaires sur le Recensement de 2016, afin d'éclairer les recensements futurs.
- Nous entrevoyons l'automatisation complète de fonctions comme le codage des réponses d'enquête, avec une plus grande exactitude et efficacité à l'avenir.
- Nous pouvons également enrichir les données grâce aux couplages d'enregistrements.

### ***Nécessité et proportionnalité***

Une autre façon de relever ce défi consiste à travailler en consultation avec des spécialistes de la statistique et de la protection de la vie privée du monde entier pour élaborer un nouveau cadre méthodologique fondé sur les principes de la nécessité et de la proportionnalité.

Tout d'abord, nous devons déterminer la nécessité. Statistique Canada a pour mandat de produire des données pour le Canada, des données qui sont essentielles pour les gouvernements, les municipalités, les entreprises et les particuliers. Nous mesurons notre société, notre économie et notre environnement.

Puis, nous avons recours à la proportionnalité, ce qui signifie que lorsque nous planifions nos enquêtes, nos experts s'efforcent d'établir un équilibre entre le volume et

la sensibilité des sources de données et la nécessité de réduire le fardeau de réponse pour les personnes, tout en maintenant la protection de votre vie privée.

Le cadre s'appuie sur les principes qui ont toujours guidé Statistique Canada. Nous modernisons maintenant notre approche en élargissant ces principes pour élaborer un cadre scientifique. Ce cadre permettra d'évaluer explicitement la proportionnalité et la nature délicate des données et fera en sorte que les valeurs statistiques, comme la protection des renseignements personnels et la confidentialité, sont respectées.

## **Collaboration / innovation**

Maintenant que nous en savons un peu plus sur la façon dont nous nous modernisons, j'aimerais vous démontrer que cette valeur s'est véritablement ancrée dans la culture de notre organisme et se manifeste de diverses façons dans l'ensemble de l'organisme. On pourrait dire que la modernisation a ouvert la voie de l'innovation et de la collaboration.

## ***Écosystème de l'innovation***

Pour favoriser une solide culture de l'innovation et de l'amélioration continue, Statistique Canada a créé un écosystème de l'innovation.

Cet écosystème a été conçu pour :

- donner aux employés les moyens d'expérimenter, d'établir des liens entre des idées et des personnes;
- enrichir et partager leur expertise;
- fournir une infrastructure, des outils, de l'encadrement et du soutien;
- tirer parti de l'innovation afin d'atteindre nos objectifs.

L'écosystème comprend une équipe d'innovation numérique, un centre et un laboratoire d'innovation, un radar de l'innovation, des marathons de programmation mensuels, des centres d'expertise et un Conseil de recherche et développement.

## ***Culture LEAN, bac à sable infonuagique, marathons de programmation***

Nous avons également lancé une initiative de culture LEAN. L'objectif est de favoriser des communications ouvertes et transparentes avec les partenaires et les intervenants, de promouvoir une culture d'innovation et d'encourager les façons de penser et les comportements innovateurs grâce à la reconnaissance. Pour ce faire, nous avons un entraîneur LEAN spécialisé et une formation LEAN à l'interne. Pour ceux qui se demandent ce que signifie LEAN, il s'agit de l'état d'une organisation qui s'efforce toujours de créer de la valeur pour ses clients.

Nous reconnaissons que l'inclusion est le principal critère de réussite pour l'innovation. L'expérimentation réussie mobilise et met à profit divers talents, connaissances et expériences multidisciplinaires. Qu'il s'agisse de tenir régulièrement des marathons de

programmation ou de créer un bac à sable infonuagique de l'innovation, nous continuons de bâtir notre écosystème.

Nous le faisons d'une manière qui met l'accent sur la mobilisation des Canadiens. Par exemple, dans le cadre de notre marathon de programmation reposant sur la ludification, nous avons mis au défi les participants de proposer des façons uniques de « ludifier » nos données pour les rendre amusantes et accessibles à un plus grand nombre de personnes.

## **Gouvernance**

Alors, vous vous demandez peut-être comment une grande organisation gère tous ces changements et toutes ces initiatives de modernisation avec des délais serrés et des ressources fixes.

Une autre mesure que nous prenons pour appuyer la modernisation consiste à repenser notre infrastructure de technologie de l'information. Statistique Canada a organisé son effectif de TI de façon à se concentrer sur une stratégie axée sur l'infonuagique pour former, en étroite collaboration avec Services partagés Canada, la base d'une infrastructure de TI moderne et sécurisée, afin de migrer des anciens centres de données vers l'informatique en nuage.

## **Composantes de la stratégie de données (au pays et à l'étranger)**

La Stratégie de données de Statistique Canada est axée sur cinq capacités :

1. Gouvernance des données
2. Intendance des données, y compris :
  - Recherche de données
  - Numérisation des données
  - Interopérabilité des données
  - Gestion des données
3. Ressources de données
4. Cadre de confiance en matière de données
5. Leadership en matière de données

Les objectifs de haut niveau de la stratégie sont de veiller à ce que les données soient disponibles, accessibles, interopérables, réutilisables, reproductibles et ouvertes.

Comme certains d'entre vous le savent peut-être, tout cela s'harmonise étroitement avec la stratégie de données du gouvernement du Canada : *Feuille de route de la Stratégie de données pour la fonction publique fédérale*. Statistique Canada est à l'avant-plan de cette stratégie et est là pour l'appuyer.

Cette stratégie représente une occasion majeure de créer des ressources de données unifiées et ouvertes dans l'ensemble de l'administration fédérale. Notre travail de modernisation s'harmonise bien avec les principaux défis et éléments de la stratégie

relative aux données et il constituera un facteur important de bon nombre de ces éléments clés, dans l'ensemble du secteur public.

Grâce aux processus et aux discussions que nous avons réalisés pour déterminer la meilleure façon de collaborer à cette stratégie fédérale, nous avons remarqué trois leçons essentielles :

1. être un organisme proactif et flexible
2. mobiliser les Canadiens et les organisations canadiennes
3. favoriser des partenariats stratégiques, penser au système statistique national

### **Priorités pour la gouvernance des données au Canada (collaboration nationale et internationale)**

Au fur et à mesure que nous progressons, nos priorités en matière de gouvernance des données sont les suivantes :

- une structure de gouvernance mieux adaptée à l'environnement en constante évolution, y compris en ce qui a trait à la révolution des données et aux changements législatifs comme la légalisation du cannabis;
- la collaboration avec des partenaires, à l'échelle nationale et internationale, dans le but de promouvoir l'interopérabilité;
- l'harmonisation avec la Stratégie de données pour la fonction publique fédérale.

L'une des priorités à l'échelle nationale consiste à établir une structure de gouvernance, afin d'élaborer et de réviser des normes qui correspondent aux besoins du gouvernement pour faciliter la prise de décisions fondées sur des données probantes.

Par le passé, Statistique Canada suivait un cycle de vie de cinq ans pour les principaux systèmes de classification. Nous avons découvert que lorsque nos séries de données approchaient de l'échéance de cinq ans, les données proprement dites ne correspondaient plus au paysage socioéconomique. Afin de s'adapter à cet environnement qui évolue rapidement, Statistique Canada a décidé d'être proactif et a adopté une approche évolutive.

### ***Cannabis***

Par exemple, cette façon de faire les choses a été utilisée pour répondre aux besoins en données afin de mesurer l'incidence de la légalisation du cannabis au Canada. Une révision des neuf principaux systèmes de classification a été diffusée pour s'harmoniser avec la légalisation. Il existe maintenant des catégories particulières à l'intérieur de ces classifications qui aideront à en mesurer l'incidence sur les industries, les produits, les professions, les importations et exportations, ainsi que les programmes d'éducation, pour n'en nommer que quelques-uns.



## ***Classification nationale des professions***

L'évolution du paysage professionnel est un autre exemple où l'approche évolutive sera utile. À mesure que le Canada devient plus axé sur l'information et que la technologie s'oriente vers l'IA et l'apprentissage automatique, nos utilisateurs de données nous disent que la Classification nationale des professions doit être mise à jour plus fréquemment pour tenir compte des nouvelles professions. Statistique Canada a collaboré avec Emploi et Développement social Canada à l'élaboration et à la gestion de la Classification nationale des professions, une structure de classification systématique qui catégorise toute la gamme des activités professionnelles au Canada. L'information sur les professions revêt une importance cruciale pour la production de renseignements sur le marché du travail et les carrières pour de nombreux programmes et services, et nous espérons maintenir cette relation et travailler en étroite collaboration pour répondre aux besoins de nos utilisateurs de données.

## ***Groupe de haut niveau sur la modernisation des statistiques officielles***

Ce n'est pas seulement la normalisation des systèmes de classification qui est importante, mais aussi les modèles d'information qui contribuent à définir les processus, de la collecte à la diffusion des données, ainsi que la structure de l'information qui définit les activités. Statistique Canada joue un rôle actif et a une forte présence à l'échelle internationale en matière de normalisation. Je préside actuellement le Groupe de haut niveau sur la modernisation des statistiques officielles, qui est géré par la Commission économique pour l'Europe des Nations Unies. Le groupe de haut niveau est chargé de déterminer les projets de collaboration internationale annuels qui s'alignent sur le programme de modernisation statistique des Nations Unies.

L'un des quatre groupes de modernisation qui relèvent du groupe de haut niveau est le Groupe des normes de soutien, qui est responsable de l'élaboration et de la mise à jour des modèles d'information nécessaires à la modernisation statistique. Ces modèles comprennent le Modèle générique du processus de production statistique, le Modèle générique d'informations statistiques, le Modèle générique d'activité des organismes statistiques et l'Architecture commune de la production statistique. Statistique Canada est actif au sein du Groupe des normes de soutien et de trois des quatre groupes de travail. Ces modèles d'information sont utilisés dans l'élaboration de systèmes communs aux processus qui permettent l'interopérabilité.

## ***Stratégie de données pour la fonction publique fédérale***

L'harmonisation avec la Stratégie de données pour la fonction publique fédérale est une autre priorité majeure. Les ministères et agences du gouvernement du Canada sont confrontés à de nombreux défis semblables en matière de gouvernance des données. Il faut assurer la gouvernance des données d'entreprise pour appuyer la mise en œuvre de la Stratégie de données pour la fonction publique fédérale et pour susciter un changement de culture et l'utilisation stratégique des données.

La stratégie relative aux données résume certains des principaux défis du gouvernement : l'absence d'une gouvernance horizontale pour donner une orientation stratégique relativement aux problèmes liés aux données, l'absence de littératie en matière de données et la réticence culturelle à éliminer le travail en vase clos, l'absence d'une infrastructure numérique suffisante et d'un cadre de règles complexes, ainsi que les défis liés à l'acquisition, à la gouvernance et à la gestion de grands volumes de données.

Il faut commencer à considérer le gouvernement comme une entité unique, à collaborer et à établir un équilibre entre la nécessité de partager les renseignements entre les organisations en temps opportun, tout en respectant la protection des renseignements personnels et les responsabilités éthiques envers les Canadiens. Nous devons chercher d'autres sources fiables qui peuvent améliorer nos travaux en cours et combler les lacunes statistiques actuelles. Cela nécessitera un cadre d'évaluation de la qualité, ainsi que la normalisation des concepts pour permettre l'intégration de l'information, tout en protégeant la vie privée et en maintenant la confiance du public.

Une fonction publique proactive plutôt que réactive, qui répond à l'écosystème de données en évolution et aux besoins des Canadiens, aura besoin de systèmes de gouvernance solides qui peuvent communiquer facilement entre eux, ainsi que de concepts normalisés et de modèles d'information qui serviront de fondement à l'intégration de renseignements fiables.

### ***Collectif canadien de normalisation en matière de gouvernance des données***

Comme vous en avez peut-être entendu parler, 90 % des données dans le monde ont été générées au cours des deux dernières années. Comme l'offre de données augmente de façon exponentielle, il est simplement logique d'adopter des stratégies de normalisation à ce stade.

Le printemps dernier, des experts en la matière du Canada ont commencé à planifier un Collectif canadien de normalisation en matière de gouvernance des données afin d'accélérer l'élaboration de normes de gouvernance pour l'ensemble de l'industrie. Le but est de favoriser la coordination et la collaboration en ce qui concerne les questions touchant la normalisation de la gouvernance des données, notamment des définitions et des classifications. Statistique Canada est membre de ce collectif, qui comprendra des représentants de l'industrie, des gouvernements et des organismes qui exercent une fonction réglementaire reconnue par la loi, des organismes universitaires, de recherche et d'intérêt public, ainsi que des organismes d'élaboration de normes.

Notre organisme collabore avec des intervenants pour déterminer le type de normes nécessaires et le processus d'élaboration de ces normes. Les méthodes de gouvernance des données doivent être modernisées pour profiter aux organisations et aux citoyens canadiens.

De toute évidence, les classifications et les autres métadonnées doivent être révisées plus fréquemment pour s'adapter aux changements qui touchent l'économie et la société canadiennes.

Alors que nous commençons à travailler dans ce domaine, en vue d'améliorer la circulation de l'information et l'interopérabilité, ainsi que d'accroître la disponibilité et l'actualité de l'information, nous devons nous assurer que la protection des renseignements personnels est intégrée au processus et n'est pas considérée après coup ou perçue comme un compromis!

### ***Classification canadienne de la recherche et développement***

Statistique Canada collabore avec la Fondation canadienne pour l'innovation, les Instituts de recherche en santé du Canada, le Conseil de recherches en sciences naturelles et en génie du Canada et le Conseil de recherches en sciences humaines du Canada à l'élaboration d'une nouvelle classification canadienne de la recherche et développement pour 2019. Cette norme commune sera notamment utilisée par les organismes subventionnaires fédéraux pour la collecte et la gestion de données administratives sur les programmes ou les subventions de recherche et développement, ainsi qu'à des fins statistiques par Statistique Canada. Elle permettra également d'harmoniser les normes du Canada avec les normes internationales de classification de la recherche et développement.

### ***SCIAN***

Les classifications normalisées sont importantes à l'échelle internationale. Statistique Canada, de concert avec les organismes statistiques du Mexique et des États-Unis, a élaboré le Système de classification des industries de l'Amérique du Nord. Ce système a été conçu pour offrir des définitions communes de la structure industrielle des trois pays, ainsi qu'un cadre statistique commun, afin de faciliter l'analyse des trois économies, par suite de la mise en œuvre de l'Accord de libre-échange nord-américain. Nous utilisons ce système lorsque la classification des industries est requise dans le cadre de nos programmes. Le système est également utilisé par l'Agence du revenu du Canada, y compris dans les logiciels de l'impôt.

Nous continuerons de travailler avec nos partenaires internationaux pour veiller à ce que les données soient comparables d'un pays à l'autre.

Pourquoi faisons-nous tout cela? Pourquoi apportons-nous des changements importants pour nous recentrer à l'interne tout en allant de l'avant avec des approches innovatrices en matière de collaboration? Nous faisons tout cela parce que si nous trouvons des façons de libérer le potentiel que comportent ces données, cela profitera énormément à notre société et à notre économie.

Cet aspect théorique est-il donc entièrement valable?

## **Le pouvoir des mégadonnées : Utilité des statistiques pour la société / l'économie**

Permettez-moi de donner cinq exemples concrets d'initiatives de Statistique Canada dans l'ensemble du secteur public qui utilise les mégadonnées pour l'analyse des données. Cela comprend presque toujours une approche de collaboration dans le cadre de laquelle nous échangeons de l'information avec nos partenaires fédéraux pour appuyer leurs travaux stratégiques.

**Exemple 1 (Immigration) :** Intégrer des mégadonnées structurées, comme des données administratives, pour mieux comprendre les résultats des immigrants.

- Saviez-vous que, depuis 2016, le Canada accueille en moyenne 300 000 nouveaux arrivants chaque année, ce qui fait que les décideurs ont besoin de solides analyses pour s'assurer qu'ils élaborent de bonnes politiques qui favorisent une intégration réussie?
- Statistique Canada utilise des données administratives intégrées pour appuyer l'analyse dans ce contexte.
- Par exemple, l'accession à la propriété est une étape clé de l'intégration sociale et économique des immigrants. Les nouvelles données sur le logement du Programme de la statistique du logement canadien ont été utilisées pour suivre l'accession à la propriété chez les immigrants. Selon les résultats :
  - À Toronto et à Vancouver, les immigrants étaient moins susceptibles de posséder un logement individuel non attenant que les propriétaires nés au Canada.
  - L'accession à la propriété chez les familles de réfugiés réinstallés est généralement plus faible que chez les familles nées au Canada. À Vancouver, on parle de 50 % des réfugiés réinstallés, comparativement à 61 % des familles nées au Canada. Les taux plus faibles sont en grande partie attribuables à des revenus plus faibles.
- Les immigrants agissent aussi comme moteur économique au pays. À partir de données intégrées de la Base de données canadienne sur la dynamique employeurs-employés, les chercheurs de Statistique Canada ont déterminé que les entreprises appartenant à des immigrants sont plus jeunes et plus susceptibles d'être des entreprises à forte croissance et qu'elles contribuent davantage à la croissance nette de l'emploi.

**Exemple 2 :** Analyse prédictive pour orienter les options de dépistage du cancer

- À une époque où les coûts des soins de santé augmentent, les fournisseurs cherchent à utiliser les ressources de la façon la plus efficace possible. Les modèles prédictifs peuvent aider à évaluer les répercussions des différentes options de politiques et de programmes.
- En collaboration avec le Partenariat canadien contre le cancer, Statistique Canada a créé OncoSim, un modèle prédictif utilisant une gamme de sources de

données pour estimer les répercussions de différentes options pour le dépistage et le traitement du cancer.

- Le modèle du cancer du poumon a récemment été utilisé pour estimer les répercussions de l'introduction d'un protocole de dépistage du cancer du poumon chez les personnes de 55 à 74 ans qui ont fumé l'équivalent d'un paquet par jour pendant 30 ans ou plus.
- Les résultats du modèle montrent qu'au fil du temps, l'approche structurée permettrait au système de santé de diminuer les coûts en réduisant de plus de 80 % le besoin de dépistage par tomodensitométrie et d'interventions diagnostiques. Cela pourra se faire sans influencer sur les effets sur la santé, à savoir le taux de cas de cancer.
- Le modèle est utilisé par les programmes de cancérologie partout au pays pour planifier de meilleurs services pour les patients.

**Exemple 3 (Éducation) :** Comprendre les résultats sur le marché du travail des diplômés postsecondaires : pouvons-nous déterminer qui a un emploi après l'obtention de son diplôme?

- Puisque les investissements dans les études postsecondaires sont importants pour les étudiants et les familles, il est primordial de comprendre les résultats sur le marché du travail après l'obtention d'un diplôme.
- Statistique Canada a recours à des données fiscales et sur l'éducation intégrées pour suivre les résultats sur le marché du travail des finissants.
- Selon les résultats, parmi les étudiants qui ont obtenu leur diplôme en 2014, 66 % sont entrés sur le marché du travail et 20 % sont retournés aux études à plein temps.
- Les domaines d'études où le revenu d'emploi médian est le plus élevé après l'obtention d'un diplôme comprennent l'architecture et le génie, la santé et les domaines connexes, ainsi que les affaires, la gestion et l'administration publique.

**Exemple 4 (Municipalités) :** Utiliser des données ouvertes pour aider les villes à atteindre leurs objectifs de développement durable conformément au Programme de développement durable à l'horizon 2030 des Nations Unies

- L'un de ces objectifs de développement durable vise à rendre les villes inclusives, sûres, résilientes et durables.
- Cet objectif est également de mesurer la proportion de la population qui a un accès facile au transport en commun, ce qui signifie que les arrêts d'autobus se trouvent à moins de 0,5 kilomètre des endroits où les citoyens vivent, travaillent, étudient et magasinent, et qu'il y a un service fréquent pendant les périodes de pointe.
- Pour aider les villes à atteindre ces objectifs, Statistique Canada utilise des données ouvertes, et plus particulièrement la *General Transit Feed Specification*, un protocole de gestion en temps réel qui communique la

position GPS des véhicules, ainsi que la latitude et la longitude des arrêts et les différents horaires.

- En utilisant ces données ouvertes avec nos propres données, nous sommes en mesure d'estimer la proportion de la population qui a accès facilement à des services de transport en commun.
- Ces renseignements aideront les villes à faire le suivi de leurs progrès et à déterminer les secteurs où des améliorations pourraient être apportées, ce qui se traduira par de meilleurs services de transport pour les citoyens.

**Exemple 5 (Agriculture) :** Utiliser des données de satellites pour fournir des renseignements à jour sur le rendement des cultures

- Ag-zéro est un autre exemple de la façon dont Statistique Canada utilise des méthodes et des technologies de collecte de données innovatrices, tout en les adaptant pour mieux répondre aux besoins des utilisateurs.
- Ag-zéro est une initiative emballante dans le cadre de notre Programme de la statistique agricole, qui vise à recueillir les données requises par le programme au moyen de sources de données administratives et d'imagerie par satellite d'ici 2026.
- Il s'agit d'un autre exemple où mes collègues de Statistique Canada ont montré qu'ils écoutent les commentaires des Canadiens qui veulent des données à jour, exactes et détaillées, tout en répondant à un moins grand nombre d'enquêtes.
- Cette initiative de modernisation tire parti d'autres sources de données, comme la technologie d'observation de la Terre et les données historiques sur l'assurance-récolte pour le Manitoba, ainsi que des progrès dans les techniques de traitement des données, afin de transmettre l'information aux agriculteurs tout en réduisant le temps qu'ils doivent consacrer à répondre aux questionnaires statistiques.
- Statistique Canada mettra à l'essai cette approche innovatrice, qui peut également profiter à d'autres organismes de la chaîne de valeur.
- Par exemple, imaginez comment les données d'Ag-zéro pourraient aider les responsables des politiques publiques et les entreprises à intervenir plus rapidement auprès des agriculteurs après un événement météorologique important ayant des répercussions sur les cultures, ainsi que les gouvernements et les assureurs qui doivent rapidement apporter un soutien aux agriculteurs.

Comme les autres organismes nationaux de statistique, nous tenons compte des changements découlant de la législation et nous nous adaptons dans une large mesure pour répondre aux besoins d'information en matière de politiques.

Prenons la légalisation du cannabis. Nous avons examiné des façons d'obtenir des données sur la production, la consommation et les prix, mais le cannabis était illégal à l'époque, et cette information dépassait la portée de nos systèmes statistiques.

Nous avons donc dû utiliser des méthodes assez originales.

- L'une était l'externalisation ouverte :
  - Nous avons lancé un site appelé **StatsCannabis** et nous avons demandé aux gens de nous dire à quelle fréquence ils consommaient du cannabis et combien ils payaient pour en obtenir.
  - Nous avons reçu plus de 20 000 réponses le premier mois! C'est l'une des raisons pour lesquelles la confiance est importante pour un organisme statistique. Les Canadiens faisaient suffisamment confiance à Statistique Canada pour fournir ces renseignements de nature délicate.
- Nous avons également commencé à utiliser l'analyse des eaux usées municipales pour en apprendre davantage sur la consommation dans différentes municipalités.
  - Nous avons utilisé une technique appelée épidémiologie des eaux usées.
  - Cette nouvelle technique nous a permis de relever le défi de la sous-déclaration. C'est un problème auquel nous sommes toujours confrontés lorsque nous essayons de produire des données au moyen d'outils comme des enquêtes.
- Nous commençons également à intégrer les ventes de cannabis aux données sur les ventes au détail.

### **Plateforme mondiale des Nations Unies sur l'utilisation des mégadonnées en statistique officielle (collaboration)**

À un niveau plus élevé, on reconnaît également la nécessité de normaliser et de collaborer. Pour avoir une collaboration mondiale permettant d'exploiter le pouvoir des mégadonnées afin d'améliorer la vie des gens, il faut nous harmoniser de manière à favoriser notre travail à l'échelle internationale.

Par exemple, le Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle de l'ONU a été créé en 2014. Il a pour mandat d'assurer l'orientation sur l'utilisation des mégadonnées en statistique officielle. La Plateforme mondiale des Nations Unies nous permet de collaborer et d'apprendre ensemble – nous devons relever les mêmes défis et nous pouvons collaborer en formation et en renforcement des capacités. Nous pouvons profiter de l'accès à de nouvelles technologies et sources de données, ainsi qu'à une expertise.

Le Canada fait partie de ce groupe de travail, avec 27 autres États membres et 16 organisations internationales. Le groupe de travail a constaté la nécessité d'une plateforme qui permettrait aux chercheurs de collaborer, d'échanger des idées et de travailler ensemble à des projets, tout en utilisant des données et des méthodes fiables. Cette collaboration stimule l'innovation, d'autant plus que le partage des connaissances est traditionnellement limité aux présentations et aux communications. Avec la numérisation de l'information et les technologies de communications, ainsi que l'émergence de communautés virtuelles à travers le monde, les mégadonnées ouvrent

des possibilités importantes pour l'élaboration de politiques fondées sur des données probantes, qui peuvent être particulièrement pertinentes à l'échelle internationale.

Ce groupe de travail sur les mégadonnées sert les gouvernements, les entreprises, le milieu universitaire, les médias et d'autres utilisateurs des données, dans le but de fournir des données améliorées pour une vie meilleure.

En fournissant aux membres des orientations concernant l'utilisation des nouvelles sources de données et des technologies, services et applications qui y correspondent pour les statistiques officielles à l'échelle mondiale, le groupe tente de maintenir la pertinence de ces statistiques dans un contexte de données en évolution rapide, en les rendant plus actuelles, plus granulaires et plus fréquentes, et il tente de trouver des solutions de calibre mondial à la compilation de statistiques et d'indicateurs des objectifs de développement durable, en établissant un lien entre les gens, les données mondiales, les nouvelles méthodes et les nouveaux algorithmes, ainsi que les services technologiques innovateurs et les logiciels les plus récents.

Le but ultime est de mieux soutenir la prise de décisions fondées sur des données probantes aux niveaux local, national et mondial pour le développement durable de l'économie, de l'environnement et de la société.

Le Canada participe à plusieurs équipes de travail, y compris l'Équipe de travail sur les données de scanners. Cette équipe cherche des moyens d'accroître l'utilisation de ces données dans les statistiques officielles. On pourrait, par exemple, utiliser les données de scanners des détaillants pour aider au calcul des indices de prix.

## **Conclusion**

J'aimerais profiter de l'occasion pour vous remercier tous de votre participation aujourd'hui à cette conférence stimulante. Nous aurons l'occasion d'approfondir les mégadonnées et ce qu'elles signifient pour vous.

Au cours des deux prochains jours, nos conférenciers parleront d'une vaste gamme de sujets, allant de l'IA à l'apprentissage automatique et à la robotique, en passant par la vérification des TI, l'analyse du mode de vie, la visualisation des données, les données ouvertes et bien plus encore.

J'attends avec impatience les exposés de deux de mes collègues. Éric Rancourt, directeur général de la Direction de la méthodologie à Statistique Canada, donnera un aperçu des différents types de données – leurs possibilités et leurs limites – ainsi que des approches que nous adoptons à Statistique Canada pour évaluer la qualité des données.

Nous entendrons Monica Pickard, de la Division des sciences des données, qui nous parlera de la façon d'utiliser en toute confiance des mégadonnées non structurées.

En conclusion, j'aimerais vous faire part de quelques leçons à retenir...



- Il faut reconnaître que le volume n'est pas synonyme de qualité. Il y a tout un processus pour produire de bonnes données qui peuvent appuyer des décisions fondées sur des données probantes.
- Il faut utiliser les données comme un actif stratégique. Pour bien les comprendre, il faut tirer parti des connaissances des experts de Statistique Canada.
- Il faut explorer de nouvelles technologies. À mesure que les attentes des utilisateurs continueront d'augmenter, ces nouvelles possibilités nous permettront de suivre le rythme.
- Il faut collaborer pour soutenir l'innovation. Il faut travailler ensemble pour trouver des solutions innovatrices aux défis, que ce soit en collaborant avec les intervenants ou en faisant de la littératie en matière de données une compétence.
- Il faut reconnaître l'utilité des statistiques pour notre économie et notre société. L'analyse des données par des experts peut mener à de meilleurs services de santé, à la création d'emplois, à des entreprises plus intelligentes... On ne connaît pas encore bien toutes les possibilités.

Comme nous le savons tous, le monde des données évolue si rapidement qu'il peut être difficile d'en suivre le rythme. Saisissons tous cette occasion d'apprendre les uns des autres à mesure que nous découvrons les dernières innovations, les dernières technologies et les derniers défis liés aux mégadonnées. Je suis impatient à l'idée d'en entendre plus sur ce sujet et j'espère que vous l'êtes tout autant! Les données nous permettent de nous ouvrir aux possibilités et j'ai hâte d'explorer ces possibilités avec vous.

Merci.