FUTURS OBSTACLES ET OPPORTUNITÉS POUR STATISTIQUE CANADA

DOUG NORRIS

Doug Norris, Ph.D., est vice-président principal et démographe en chef chez Environics Analytics et l'un des principaux experts du recensement au Canada.

Statistique Canada a parcouru beaucoup de chemin depuis ses débuts en tant qu'organisme sur papier qui faisait les totalisations manuellement. Aujourd'hui, elle recueille et diffuse des données et des analyses de plus en plus complexes par voie électronique.

Au cours des 100 dernières années, l'organisme de statistique a relevé de nombreux défis, mais à chaque test, il a évolué et s'est renforcé. Cette expérience sera cruciale dans les années à venir, car il est sur le point d'être testé une fois de plus. Les progrès rapides de la technologie et la montée en puissance des « grandes données » qui en découle ont ouvert une nouvelle ère dans laquelle la prise de décision fondée sur des données probantes est le nouveau mantra.

Pour de nombreuses organisations, on s'attend de plus en plus à ce que les données et les analyses façonnent les politiques et les décisions d'affaires. Il ne sera pas facile de s'adapter à un large éventail de demandes croissantes, mais cela pourrait signifier une agence encore plus forte.

NOUVEAUX ENJEUX

L'un des principaux rôles de Statistique Canada est d'aider tous les ordres de gouvernement à élaborer de meilleures politiques en leur fournissant des renseignements essentiels sur les Canadiens. Bien que ce mandat n'ait pas changé, les questions auxquelles nos dirigeants sont appelés à répondre ont changé — et elles deviennent de plus en plus complexes à mesure que la société canadienne subit un changement important. Par exemple, la population du Canada vieillit, tandis que les régions rurales et les petites villes du pays diminuent. Dans l'ensemble, le pays est de plus en plus diversifié et nos dirigeants essaient de relever les défis liés à l'inclusion.

Pour compliquer les choses, l'économie subit un changement en raison de la croissance rapide de l'économie numérique et de l'adoption de l'automatisation, qui obligera la future main-d'œuvre à adopter de nouvelles compétences. Ce changement économique pourrait également exacerber les inégalités au sein de la population.

Les gouvernements ont besoin de données de meilleure qualité, fournies en temps opportun et avec plus de précision, pour s'attaquer à ces questions multiformes. Les gouvernements ne sont pas les seuls à avoir besoin de ces données ; les organismes sans but lucratif et les entreprises du secteur privé recherchent également des données pour les aider à évaluer les besoins changeants de leurs produits et services, ainsi qu'à améliorer la façon dont ils servent leurs clients. Cette prise de conscience accrue de la puissance des données s'est traduite par un regain d'intérêt pour les « lacunes dans les données », ce qui a récemment été souligné dans une vaste série du *Globe and Mail*.

Les pressions exercées pour fournir des données en temps opportun constitueront un défi particulier pour Statistique Canada. Historiquement, le cycle d'élaboration des données était un long processus, souvent pluriannuel. Parfois, les données ne sont devenues disponibles qu'après qu'un événement inattendu ait poussé une question à l'ordre du jour public. La technologie a amélioré la rapidité d'exécution, mais l'organisation a encore beaucoup de chemin à parcourir avant de pouvoir fournir des données en temps quasi réel.

Heureusement, l'agence va dans la bonne direction. Il est reconnu que les sources de données non traditionnelles peuvent également jouer un rôle. Un exemple récent a conduit à la légalisation du cannabis. Pour aider les gouvernements et les entreprises à se préparer au changement de politique, la demande de données sur le cannabis a bondi. En réponse, Statistique Canada a mené une nouvelle enquête et a utilisé des techniques de « crowdsourcing » pour estimer le prix payé pour le cannabis.

Un deuxième exemple concernait la demande de données sur la propriété étrangère de condominiums, que l'on croyait être un facteur important contribuant à l'inflation des coûts du logement. Statistique Canada collabore avec la Société canadienne d'hypothèques et de logement (SCHL) et les municipalités pour recueillir rapidement de nouvelles données sur cette question.

COLLECTION DE DONNÉES

La demande croissante de données survient à un moment où les méthodes traditionnelles de collecte de données sont remises en question. Une grande partie du programme de statistiques sociales de Statistique Canada s'articule autour d'enquêtes auprès des ménages. Cependant, les taux de réponse aux sondages ont diminué et il semble clair qu'il est temps d'envisager une nouvelle approche.

La solution consiste à se tourner vers des sources de données administratives ou autres, collectées à des fins autres que statistiques. Par exemple, d'importants ensembles de données administratives du gouvernement fédéral, comme les dossiers fiscaux et les dossiers d'immigration, sont utilisés depuis un certain temps dans le cadre de certains programmes statistiques, mais ils ne représentent que la pointe de l'iceberg.

Les gouvernements, le secteur privé et d'autres organisations recueillent une énorme quantité de données, notamment sur les médias sociaux. Ces données présentent des difficultés statistiques, notamment en ce qui concerne la couverture de l'ensemble de données et les concepts et définitions utilisés. L'expansion des sources de données peut également poser de nouveaux problèmes de comparabilité entre les ensembles de données. Il s'agit là d'un défi particulier pour les données relatives à la santé, à l'éducation, à la justice et au travail. Bon nombre de ces données relèvent de la compétence provinciale, où les programmes et les exigences en matière de protection de la vie privée diffèrent d'une province à l'autre. Les municipalités recueillent également des données qui seraient utiles, bien que les problèmes de comparabilité soient beaucoup plus importants.

Malgré ces difficultés, l'utilisation des données administratives présente plusieurs avantages. Dans de nombreux cas, les données seront plus précises. Par exemple, le passage à l'utilisation des données fiscales pour tous les répondants au Recensement de 2016 a considérablement amélioré l'exactitude des données sur le revenu, surtout dans les petites régions. L'utilisation de codes postaux ou de codes d'adresse permet également de produire des données au niveau des petites régions — ce qui était très rare en dehors du recensement — ce qui est important pour les entreprises et les administrations locales. De plus, l'utilisation de données administratives permet de combiner des données provenant de différentes sources ou au fil du temps, ce qui permet de développer des séries de données beaucoup plus riches.

Bien que les données administratives puissent offrir de grandes possibilités, toutes les initiatives d'élaboration de données doivent être soigneusement mises en balance avec les questions de protection de la vie privée, qui deviennent de plus en plus importantes au fil des jours. Nous avons la chance que Statistique Canada jouisse d'une excellente réputation en matière de protection de la vie privée et de la confidentialité des données personnelles tout en maintenant ou même en élargissant l'accès aux données pour éclairer les politiques publiques. Les attentes en matière de protection de la vie privée et de sécurité continueront d'augmenter, d'autant plus que nos vies sont stockées électroniquement. Statistique Canada doit être plus proactive pour informer le public de l'importance des données qu'elle recueille et de ses applications, mais aussi de tous ses mécanismes de protection. Elle doit gagner la confiance du public.

ACCÈS ET DIFFUSION

En plus des nombreux défis et possibilités en matière d'élaboration de données, Statistique Canada devra aussi continuer d'améliorer l'accès aux données et leur diffusion. Les analystes ont déjà constaté d'importantes améliorations dans ce domaine au cours de la période qui a suivi le recouvrement des coûts. Aujourd'hui, toutes les données non confidentielles produites par Statistique Canada sont disponibles gratuitement en ligne.

L'introduction et la croissance des centres de données de recherche, qui permettent aux chercheurs d'accéder à des données détaillées dans un environnement contrôlé, ont constitué un progrès important pour rendre les données plus accessibles. Une étape future, sur laquelle des travaux sont déjà en cours, consiste à fournir aux chercheurs l'accès à

des microdonnées détaillées en temps réel, avec, bien entendu, les garanties appropriées en matière de confidentialité et de sécurité des ensembles de données.

Bien que la technologie ait grandement facilité la diffusion des données, elle présente aussi des défis plus importants en matière de protection de la vie privée et de confidentialité. Par exemple, si l'utilisation de la technologie pour relier des ensembles de données peut rendre les données plus utiles, elle rend la divulgation plus complexe. Il est intéressant de noter que le *U.S. Census Bureau* a récemment décidé de renforcer son approche en matière de divulgation des données en adoptant une technique appelée protection différentielle de la vie privée, qui exige que toutes les totalisations de données comprennent un certain bruit aléatoire. De plus, au moment de la rédaction du présent rapport, le *Census Bureau* a déclaré qu'en raison de ce changement, il pourrait ne pas être en mesure de produire certains de ses tableaux détaillés et fichiers de données à grande diffusion à partir du recensement de 2020.

PARTIES PRENANTES

Un sous-produit précieux de l'intérêt généralisé pour les données est qu'il y a un nombre croissant d'intervenants actifs qui créent des données. Avec beaucoup plus de données disponibles, les possibilités d'élaboration de données sont beaucoup plus nombreuses, mais elles nécessiteront de nouvelles techniques et approches méthodologiques.

Il faut qu'il y ait beaucoup plus de coopération entre les nombreux intervenants dans le domaine des données. Par le passé, Statistique Canada s'est surtout efforcée de répondre aux besoins du gouvernement fédéral à l'échelle nationale et, dans une moindre mesure, à l'échelle provinciale, en accordant moins d'attention aux petites régions géographiques. La participation d'un plus grand nombre d'intervenants signifie qu'il faut établir des mécanismes pour mieux comprendre les besoins en matière de données, y compris le partage d'une partie du travail d'élaboration des données ainsi que les résultats de ces données.

Un autre défi consistera à décider comment classer par ordre de priorité les besoins en données. Dans certains cas, Statistique Canada devra peut-être cesser de produire certaines de ses données de moindre priorité, ce qui créera des lacunes dans les données pour certains, afin de libérer des ressources pour des besoins plus prioritaires.

Statistique Canada et, plus généralement, tous les intervenants dans le domaine des données, ont de nombreuses possibilités et de nombreux défis à relever. Espérons que le Conseil consultatif canadien de la statistique, qui vient d'être annoncé, pourra fournir des conseils à Statistique Canada. De façon plus générale, le gouvernement et Statistique Canada se sont lancés dans un programme de modernisation des données qui vise non seulement à trouver des moyens de remplacer les données existantes, mais aussi à élaborer des données nouvelles et plus détaillées pour résoudre les problèmes du moment.

Chez Environics Analytics, nous sommes ravis de l'engagement de ce gouvernement à l'égard des grandes données et de l'innovation. Nous croyons qu'une plus grande collaboration entre les geeks des données des secteurs public et privé accélérera le processus. De concert avec nos clients, certains

des meilleurs et des plus brillants utilisateurs de données dans les entreprises canadiennes, nous nous sommes engagés à aider à mettre les données en pratique et à faire du Canada une économie et une société axées sur les données.