

Sous la direction scientifique de
Nathalie de Marcellis-Warin – Benoit Dostie
Sous la coordination de
Genevieve Dufour

Le Québec **9** économique

**Perspectives et défis
de la transformation
numérique**

Chapitre 10

**POINT DE VUE SUR L'ACCESSIBILITÉ
AUX DONNÉES DES ADMINISTRATIONS
PUBLIQUES**

CATHERINE HAECK, MARIE CONNOLLY

Chapitre 10

POINT DE VUE

L'ACCESSIBILITÉ AUX DONNÉES DES ADMINISTRATIONS PUBLIQUES

Catherine Haeck

Professeure agrégée à l'Université
du Québec à Montréal, chercheuse
et fellow au CIRANO

Marie Connolly

Professeure titulaire à l'Université
du Québec à Montréal, chercheuse
et fellow au CIRANO

Résumé

Ce point de vue présente l'accessibilité des données des administrations publiques en portant une attention particulière aux données fiscales ainsi qu'aux données des deux plus grands postes de dépenses du gouvernement du Québec, soit la santé et l'éducation. Les données fiscales ont permis des avancées importantes dans l'étude de la mobilité intergénérationnelle du revenu au Canada et au Québec. Les données administratives de la santé en cours de développement à travers le pays vont révolutionner la capacité des chercheurs en santé. Finalement, les données en éducation, bien qu'existantes, tardent à être mises à la disposition des chercheurs malgré l'importance des retombées que pourrait procurer l'accès à ces données. Ces trois thèmes sont abordés tour à tour et montrent la position du Québec relativement aux développements qui ont lieu dans d'autres provinces canadiennes, notamment l'Ontario et la Colombie-Britannique.

Ce chapitre dresse un portrait sommaire¹ de l'accessibilité des données des administrations publiques, en portant une attention particulière aux données fiscales ainsi qu'à celles qui se rapportent aux domaines de la santé et de l'éducation, les deux plus grands postes de dépenses du gouvernement du Québec. Nous ne sommes certainement pas les premiers à parler du potentiel des données de source administrative : nommons, parmi d'autres, les écrits de Grusky, Hout, Smeeding et Snipp (2019), Card, Chetty, Feldstein et Saez (2010), Einav et Levin (2014), Statistique Canada (2009) et Connelly, Playford, Gayle et Dibben (2016). Les vertus de l'analyse quantitative pour outiller les décideurs avaient déjà été mises en avant par Amos Tversky et Daniel Kahneman (prix Nobel d'économie) il y a de cela 40 ans (Lewis, 2016).

Notre contribution consiste ici à présenter le point de vue de chercheurs québécois² et de discuter de l'accès aux données des administrations publiques canadiennes et québécoises. Dans ce texte, nous mettons l'accent sur les microdonnées administratives anonymisées sur les individus. Les données agrégées sont plus facilement accessibles, mais ne permettent pas de répondre à l'ensemble des questions que l'on peut se poser afin de mieux comprendre le fonctionnement de notre société. Ce point de vue dresse l'état de nos connaissances sur le sujet à l'heure d'écrire ces lignes ; il est évident que l'accès aux données évolue continuellement à travers le Canada, et que nous ne sommes pas en mesure de couvrir l'ensemble des initiatives dans chaque province.

Au Canada, l'accès aux microdonnées fiscales a permis des avancées importantes dans l'étude de la distribution des revenus et de la mobilité intergénérationnelle du revenu. Dans le domaine de la santé, les données sur les individus provenant des systèmes d'information sont utilisées depuis plusieurs années, mais la démocratisation de l'accès à ces données pour fins de recherche pourrait engendrer d'importantes retombées en ce qui a trait à l'amélioration de la santé des populations. Quant aux données en éducation, bien qu'elles existent, elles tardent à être mises à la disposition des chercheurs malgré l'importance des progrès qui pourraient être réalisés grâce à un accès à ces données. Au Québec, plus de 25 % des jeunes n'obtiennent pas leur diplôme d'études secondaires en sept ans. Ce ratio est inacceptable. Il faut exploiter le pouvoir des microdonnées pour repérer plus rapidement les futurs décrocheurs ainsi que les meilleures pratiques en enseignement, et ce travail doit être réalisé par des individus qui ne sont

pas impliqués dans la création des programmes. L'accès et la disponibilité des données fiscales, de santé et d'éducation seront abordés tour à tour et nous dresserons un portrait de la position du Québec relativement aux développements qui ont eu lieu dans d'autres provinces canadiennes, notamment le Nouveau-Brunswick, l'Ontario et la Colombie-Britannique.

Accès aux données et environnement sécurisé

Garantir l'anonymisation des microdonnées dans un environnement hautement sécurisé est une condition préalable à l'ouverture de leur accès. Sur ce point, le Canada est certainement un chef de file dans le monde. Créé en 2001, le Réseau canadien des Centres de données de recherche (RCCDR) héberge les microdonnées confidentielles de Statistique Canada dans des laboratoires informatiques hautement sécurisés. Cette infrastructure de recherche assure la protection des renseignements personnels. Les chercheurs n'ont pas les données en leur possession, celles-ci n'étant accessibles qu'à l'intérieur des laboratoires, sous surveillance constante et uniquement aux chercheurs autorisés. Les ordinateurs qui les contiennent ne sont pas connectés à Internet, et il est impossible d'utiliser des clés USB pour sauvegarder des informations. Les résultats tirés de l'analyse des données doivent toujours être validés par un analyste de Statistique Canada qui s'assure que l'information divulguée ne permet pas d'identifier un individu. De plus, la structure des ensembles de données ne permet pas d'obtenir en quelques clics seulement des renseignements sur un individu en particulier. Les ensembles de données se décomposent en centaines de sous-ensembles qui ne sont pas tous accessibles aux chercheurs. Seules les données pertinentes au projet en cours le sont.

En pratique, le RCCDR réunit plus de 32 laboratoires à travers le Canada situés en milieu universitaire³, dont 5 se trouvent au Québec, dans le réseau du Centre interuniversitaire québécois de statistiques sociales (CIQSS). Ces laboratoires permettent l'accès à plusieurs types de données, incluant les données d'enquêtes, les données de recensement et certaines données des administrations publiques. Il est aussi possible d'effectuer des couplages de données grâce au soutien de Statistique Canada, ce qui permet d'enrichir des bases de données existantes. Ces couplages sont réalisés par Statistique Canada, car les clés d'appariement ne sont jamais

rendues disponibles aux chercheurs. Durant la dernière décennie, un effort soutenu de la part de Statistique Canada et du RCCDR a permis la création d'une banque offerte pour la recherche comprenant un large éventail de données des administrations publiques. Bien que les progrès faits soient importants, certains types de données sont plus accessibles que d'autres, notamment en fonction de leur province d'origine.

Au Québec, l'ISQ offre l'accès à des données d'enquête et à certains ensembles de données administratives dans ses centres d'accès aux données de recherche (appelés CADRISQ). Le fonctionnement des laboratoires de l'ISQ est extrêmement similaire à celui des laboratoires de Statistique Canada. Au Nouveau-Brunswick, le New Brunswick Institute for Research, Data and Training (NB-IRDT) est un chef de file au pays en ce qui concerne le développement de l'utilisation de données de nature administrative, particulièrement dans le domaine de la santé. Plusieurs modifications de lois entourant la protection des renseignements personnels pour des fins de recherche et l'adoption de deux lois (connues d'abord comme le projet de loi n° 57, adopté en mai 2017, et le projet de loi n° 29, adopté en juin 2019) ont permis à NB-IRDT de recevoir des données anonymisées, mais munies d'un identifiant unique rendant possible le couplage entre les diverses sources de données.

Il est évident que les infrastructures permettant un accès sécurisé existent déjà. Les avancées récentes dans l'accès aux données du Québec suggèrent une certaine ouverture du gouvernement, mais pour offrir un réel accès aux chercheurs, celui-ci devra en faire plus, en commençant notamment par une réforme de la Loi sur l'accès à l'information pour des fins de recherche et de la Loi sur l'Institut de la statistique du Québec.

Données fiscales pancanadiennes

Des microdonnées fiscales longitudinales sont disponibles à l'échelle canadienne, notamment par la Banque de données administratives longitudinales (DAL), la Base de données sur la mobilité intergénérationnelle du revenu (BDMIR) ainsi que l'Étude longitudinale et internationale des adultes (ELIA). Ces ensembles de données sont disponibles dans tous les centres de données de recherche du RCCDR. Les données fiscales reposent sur les « fichiers T1 des familles », lesquels proviennent du fichier de données fiscales recueillies chaque année auprès de tous les

contribuables canadiens dans le cadre de leur déclaration de revenus, et les archives sont disponibles à partir de 1982. Ainsi, il est possible de suivre les trajectoires de revenus de milliers de Canadiens et d'étudier les facteurs influençant les inégalités de revenus ou bien la mobilité économique d'une génération à l'autre.

Nos propres travaux basés sur la BDMIR ont permis de démontrer que, parallèlement à une hausse des inégalités de revenus depuis le début des années 1980, le Canada a connu une hausse du degré auquel ces inégalités se perpétuent d'une génération à l'autre (Connolly, Haeck et Lapierre, 2019). Que ce soit pour l'ensemble du pays ou pour chacune des provinces, la transmission intergénérationnelle du revenu, soit le pouvoir explicatif du revenu parental sur le revenu d'un jeune à l'âge adulte, a augmenté entre les cohortes nées au début des années 1960 et celles du milieu des années 1980. Autrement dit, la mobilité socioéconomique intergénérationnelle est en baisse : si l'on visualise la distribution des revenus comme une échelle, non seulement les échelons sont devenus plus distants l'un de l'autre, mais il est plus difficile de gravir ces échelons. Un tel travail de recherche, basé sur des millions de jeunes et leurs parents, n'est réalisable que grâce à l'accès aux données fiscales. Celles-ci fournissent également une source d'information fiable quant aux revenus, et leur couplage avec d'autres données est très utilisé par Statistique Canada. Un bon exemple est celui du Recensement canadien de la population : plutôt que de demander aux répondants « quel est votre revenu » comme c'est le cas pour une enquête, Statistique Canada couple les données du Recensement avec des données de l'Agence du revenu du Canada afin d'avoir des informations beaucoup plus exactes sur le revenu. Ceci réduit le fardeau de réponse et augmente la précision des résultats.

Santé, un grand chantier en cours

À notre connaissance, la première province à avoir créé un centre de recherche pour étudier des questions de santé à partir de données médico-administratives (les réclamations des médecins et des hôpitaux) est le Manitoba. Marchessault (2011) documente que, dès les années 1970, les docteurs Noralou Roos et Paul Henteleff se rencontrent pour discuter et commencer le développement de ce qui va devenir en 1991 le Centre des politiques de santé du Manitoba (*Manitoba Centre for Health Policy*). Dès le départ, leur modèle utilise les données du monde réel (« *real-world*

data ») pour étudier des questions de santé. On parle aujourd'hui de *pragmatic clinical trial*, mais cette idée ne date pas d'hier. Plusieurs facteurs ont contribué au développement du modèle manitobain. Les données étant colligées à un seul endroit par une seule organisation, une relation de collaboration et de partage d'information entre les chercheurs et les administrateurs publics a pu être établie plus facilement. De plus, la petite taille démographique de la province, à une époque où les capacités d'analyse des données administratives n'étaient pas celles que l'on connaît aujourd'hui, a aussi facilité ce développement. La compétence des personnes engagées dans ce projet novateur, ainsi qu'un financement stable, ont grandement contribué au succès du projet.

Depuis plus de 25 ans, l'Ontario a développé un centre de recherche, aujourd'hui connu sous le nom de Institute for Clinical Evaluative Sciences (ICES), pour étudier des questions de santé à partir de données administratives. L'ICES est aujourd'hui considéré comme un chef de file non seulement au Canada, mais aussi à l'international. Ce regroupement de plus de 250 chercheurs cliniciens de haut calibre mène des recherches sur des thèmes variés en santé des populations avec les données médico-administratives de la province. L'ICES est un dépositaire légal des données, lesquelles sont obtenues à la source. L'institut reçoit un financement de base du ministère de la Santé de l'Ontario qui lui assure une stabilité financière et lui permet de préparer les données pour des fins de recherche. L'institut travaille en partenariat avec ce ministère afin de répondre à certaines questions. Depuis près de cinq ans, les données médico-administratives sont complémentées par des données sociales, provenant pour certaines des autres ministères. L'ICES gère aussi les données de santé de certains groupes autochtones avec qui il travaille en étroite collaboration. De manière générale, les données de l'ICES sont accessibles à tout chercheur qui en fait la demande. En pratique, les coûts d'accès sont très élevés et peuvent constituer un frein au développement de la recherche pour les chercheurs ne faisant pas directement partie du regroupement.

Un autre exemple d'initiative provinciale est donné par le Nouveau-Brunswick. Le NB-IRDT permet l'accès et la fusion d'un remarquable éventail de données longitudinales : Citizen Database, Physician Billing, Vital Statistics, NB Suicide Registry, NB Cancer Registry, Healthy Toddler Assessment, NB Prescription Drug Program, et la liste continue⁴. La base de données Citizen Database inclut des données démographiques

et géographiques (par exemple le code postal) sur tous les résidents du Nouveau-Brunswick. La base de données Physician Billing inclut tous les paiements faits aux médecins par patient, les paiements réciproques et les salaires versés aux professionnels de la santé. Cette base de données est similaire à celle de la Régie de l'assurance maladie du Québec, à la différence qu'elle est entièrement accessible dans le laboratoire du NB-IRDT et peut-être facilement jumelée avec d'autres sources de données (jumelage réalisé par l'analyste sur place). La base de données Vital Statistics est similaire à celle des fichiers de naissances accessibles dans les laboratoires de l'ISQ. Mais, encore une fois, au Nouveau-Brunswick, il est possible de lier la base de données avec d'autres sources, ce qui n'est pas le cas à l'ISQ, sauf à grands frais et de manière moins automatisée. Enfin, le Healthy Toddler Assessment inclut les résultats de l'évaluation volontaire des enfants de 18 mois en ce qui a trait à la santé visuelle, dentaire et auditive, au développement langagier, à la croissance, aux habiletés physiques et aux relations sociales. Ces données rappellent celles collectées par les CLSC durant la petite enfance, mais qui ne sont pas rendues accessibles à des fins de recherche. L'accès et le couplage de données du NB-IRDT sont gratuits pour les chercheurs universitaires, et ce, même pour les données médico-administratives.

À l'échelle pancanadienne, il existe une plateforme pour les données sur les médicaments, C-NODES. Le futur réside dans le développement d'une plateforme nationale des données médico-administratives. Les Instituts de recherche en santé du Canada (IRSC) financent C-NODES ainsi que le développement d'une plateforme pancanadienne pour les microdonnées de santé au moyen de la Stratégie de recherche axée sur le patient (SRAP). Il existe plusieurs unités de SRAP, dont une se trouve au Québec. Par contre, la fusion de ces données à l'échelle canadienne est loin d'être réalisée, et il semble bien que ce projet de couplage ne se concrétisera pas avant plusieurs années.

Au Québec, certains groupes de chercheurs ont des accès privilégiés aux données administratives de la santé, quoique rien n'ait l'ampleur de l'ICES. L'ISQ a récemment développé une entente avec le ministère de la Santé pour gérer l'accès aux microdonnées, et le processus vient d'être mis en application. De nouvelles modalités ont été mises en œuvre à la suite du lancement, en juin 2019, du nouveau Guichet d'accès aux données de recherche. Celles-ci sont maintenant accessibles dans les CADRISQ, dont un est situé en milieu universitaire, et à distance de manière sécurisée⁵. Ce

modèle va cependant devoir évoluer pour permettre un accès plus simple aux chercheurs en délimitant le rôle de la Commission d'accès à l'information dans l'approbation des demandes d'accès. Certains doutent de la qualité des données administratives, puisque celles-ci ne sont pas collectées pour des fins de recherche, mais pour des fins administratives, par exemple pour le paiement des médecins. Lorsque l'objectif est purement administratif, le montant facturé doit être juste, l'intervention facturée doit aussi l'être puisqu'elle est associée au montant facturé. En revanche, le diagnostic associé à l'intervention, par exemple, n'entre pas dans la facturation et peut donc être considéré comme une donnée de moindre qualité. Bien entendu, si ces données commencent à être utilisées par les chercheurs cliniciens pour suivre l'évolution de certaines conditions et faire des analyses sur des données réelles (*real-world data*), alors de là naîtra l'incitatif pour améliorer la qualité des données. On sait que les expériences contrôlées ont par ailleurs aussi leurs limites. Obtenir des résultats dans l'ensemble de la population est la voie de l'avenir. Le Québec doit miser sur ses expertises en recherche et utiliser de manière démocratique le fort potentiel des données administratives réparties dans ses différents ministères pour répondre à des questions qui touchent directement le bien-être de sa population.

Éducation et petite enfance, les enfants pauvres des données

Pour ce qui est des études postsecondaires, la Plateforme de liens longitudinaux entre l'éducation et le marché du travail (PLEMT) est un excellent exemple de ce qui peut être fait à l'échelle canadienne. Cette source de données est une plateforme qui permet le couplage de plusieurs sources d'information. Au cœur de cette plateforme se trouvent le Système d'information sur les étudiants postsecondaires, qui contient des données sur chaque étudiant dans tous les établissements postsecondaires canadiens (principalement les universités), ainsi que le Système d'information sur les apprentis inscrits, lequel fournit des renseignements similaires, mais pour les programmes de formation de qualification professionnelle et d'apprentissage. Ces données, disponibles depuis 2009 pour l'ensemble des provinces canadiennes, permettent le suivi du nombre de diplômés par établissement et par programme. La plateforme contient également des clés de couplage permettant de faire le lien avec les données fiscales contenues dans les fichiers de déclarations de revenus. Il est donc possible de suivre

les jeunes diplômés sur le marché du travail et d'évaluer, par exemple, le rendement de l'éducation universitaire selon le profil socioéconomique de chacun (Frenette, 2019). Cette plateforme est remarquable en ce sens que toutes les institutions postsecondaires sans exception y participent. Grâce à un degré de coopération nationale peu observé dans d'autres domaines, les chercheurs et les décideurs ont accès à des informations d'une portée jusqu'ici inégalée.

En revanche, il existe un grand vide concernant les microdonnées de l'éducation primaire et secondaire, ou même de la petite enfance. Pourtant, le discours du budget fédéral prononcé le 22 mars 2016 par le ministre des Finances Bill Morneau faisait explicitement référence à l'importance des données sur les enfants : « Il est impossible de mettre en œuvre des politiques efficaces sans prendre appui sur des données rigoureuses. Si nous souhaitons sortir les enfants de la pauvreté, nous devons d'abord en comprendre la cause » (Morneau, 2016). Or, les données pancanadiennes sur les enfants et les jeunes sont rares. Celles provenant des systèmes d'information des ministères le sont encore plus.

Sur ce point, le Nouveau-Brunswick et la Colombie-Britannique semblent devancer les autres provinces canadiennes. Les microdonnées du système d'éducation primaire et secondaire sont récemment devenues accessibles au Nouveau-Brunswick, et leur documentation est en cours d'élaboration. À l'heure actuelle, ces données longitudinales individuelles peuvent facilement être couplées avec tous les ensembles de données administratives du NB-IRDT présentés précédemment, les clés de couplage existant déjà. L'accès aux données et leurs couplages sont gratuits pour les chercheurs universitaires. Ce modèle en faveur de la recherche est différent du modèle en vigueur au Québec. En admettant que les données deviennent disponibles, le modèle québécois n'a pas créé des environnements de couplage entre les sources administratives, et quand des couplages sont effectués, les chercheurs doivent assumer des coûts importants pour les financer. De plus, les fichiers de recherche ainsi créés doivent par la suite être détruits (bien que les clés soient conservées). Ce modèle constitue un frein pour la recherche et engendre une redondance de coûts pour les contribuables québécois.

En Colombie-Britannique, les politiciens ont décidé de mettre les enfants au cœur de leurs priorités et, pour ce faire, ils se sont donné des outils pour pouvoir évaluer leur parcours de vie durant le primaire et le secondaire.

Depuis plus de 10 ans, les chercheurs ont accès aux microdonnées du ministère de l'Éducation de la province pour la période allant de 1995 à 2016. Ces renseignements concernent l'ensemble des enfants inscrits à l'école en Colombie-Britannique, et un identifiant unique est attribué à chacun afin de pouvoir le suivre durant tout son parcours primaire et secondaire. Chaque année, pour chaque élève, plusieurs informations sont recueillies, dont les suivantes : (1) des variables démographiques personnelles (sexe, langue parlée à la maison, identité autochtone, date de naissance, code postal), (2) des informations sur le programme d'études (année d'études, langue seconde, éducation spécialisée, immersion française, cours suivis au secondaire, nombre de reprises de cours), (3) des informations sur l'école (identifiant unique anonymisé de l'école au 30 septembre, école publique ou indépendante), (4) les résultats aux tests provinciaux et les notes par matière (participation, rang centile, etc.) en 4^e, 7^e, 10^e et 12^e années, le nombre d'écoles où l'enfant a été inscrit durant son parcours et l'état d'obtention du diplôme, et (5) des caractéristiques du quartier de résidence (valeur moyenne des maisons, pourcentage de personnes vivant sous le seuil de faible revenu, pourcentage de personnes par type de diplôme, etc.). Ces données sont accessibles aux chercheurs universitaires par l'entremise de PopData BC et éventuellement Statistique Canada. L'accès est conditionnel à ce que le projet soit d'intérêt public. Pour environ 20 % des enfants de 4^e, 7^e, 10^e et 12^e années, il est possible d'accéder aux informations collectées en 2016 et en 2017 lors d'une enquête de satisfaction effectuée auprès des jeunes et de leurs parents.

Ces données font partie d'un projet plus vaste qui vise à comprendre le parcours des enfants. En pratique, elles peuvent être couplées avec un large éventail de données administratives qui sont également sous la responsabilité de PopData BC, et peuvent ainsi être fusionnées avec celles du système d'éducation postsecondaire, des fichiers de naissance ainsi que du système d'assurance sociale. En théorie, elles peuvent aussi être couplées aux données du système de santé, mais il semblerait que cela ne soit pas possible, en pratique, à court terme. Le couplage des données n'est, pour l'instant, pas gratuit en Colombie-Britannique.

Au Québec, il est, à l'heure actuelle, quasiment impossible d'obtenir les microdonnées administratives de l'éducation primaire et secondaire. L'ISQ mène des enquêtes portant sur les jeunes, l'une des plus connues étant l'Étude longitudinale du développement des enfants du Québec

(ELDEQ). Cette enquête longitudinale se concentre sur une cohorte de personnes nées en 1998. D'autres, telles que l'Enquête québécoise sur le développement des enfants à la maternelle (EQDEM), offrent aussi des portraits statistiques transversaux de différentes cohortes d'enfants à un moment de leur vie. Bien que ces données soient riches et pertinentes, elles ne permettent pas de comparer le Québec aux autres provinces canadiennes ou d'étudier les trajectoires de plusieurs cohortes d'enfants.

Nous avons besoin de microdonnées comparables et représentatives des enfants de chaque province. Au-delà des données scolaires, il serait utile d'en avoir également sur leur parcours de vie, leur bien-être, leur réussite à l'école, leur développement comportemental et cognitif, leur milieu familial, leur milieu scolaire, etc. Ces données continuellement mises à jour permettraient aux chercheurs d'évaluer les programmes et les politiques qui touchent les enfants et de voir si ceux-ci en sortent gagnants ou perdants. Historiquement, le Canada avait l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) de Statistique Canada, une enquête d'une richesse peu commune, couvrant plusieurs cohortes d'enfants âgés de 0 à 11 ans et suivis sur plusieurs années de manière longitudinale. La variété des questions couvertes et d'informations recueillies permettait d'aborder plusieurs points concernant les enfants. Cette enquête a débuté en 1994 et a été discontinuée en 2008. Elle avait entre autres permis l'analyse du programme québécois des services de garde (par exemple mentionnons Baker, Gruber et Milligan (2008), Lefebvre et Merrigan (2008), Haeck, Lefebvre et Merrigan (2015) et Haeck, Lebihan et Merrigan (2018)). Aucune autre enquête pancanadienne n'offre la richesse de l'ELNEJ depuis. Il est évident que les provinces doivent collaborer pour parvenir à instaurer un système d'information comparable à travers le Canada.

Conclusion

Le gouvernement fédéral a fait des avancées importantes en matière d'accès aux données administratives au cours des cinq dernières années. L'environnement de couplage des données créé par Statistique Canada permet de limiter le coût des fusions de données et d'augmenter les possibilités de recherche. Un exemple remarquable est la possibilité de couplage avec des données fiscales (comme pour le recensement) et les

bases de données dérivées des fichiers fiscaux (comme la Banque de données administratives longitudinales). Par contre, les données portant sur la santé et l'éducation sont de propriété provinciale.

Les administrations publiques du Québec, incluant celles opérant dans les domaines de la santé et de l'éducation, collectent une très grande quantité d'informations. Le manque d'accessibilité à ces sources pour des fins de recherche menée par des chercheurs universitaires et non financée par le secteur privé freine les travaux et la capacité de ces chercheurs à informer nos dirigeants. La recherche se fait donc souvent à partir de données collectées au moyen d'enquêtes financées par des fonds publics, mais qui, la plupart du temps, demeurent la propriété des chercheurs ayant obtenu la subvention. Il y a, par conséquent, une multiplication des enquêtes. Cette situation n'est pas justifiée dans un environnement où les données sont anonymisées et disponibles uniquement dans des laboratoires sécurisés comme ceux du CIQSS.

De plus, si le Québec ne participe pas aux initiatives pancanadiennes pour la création d'ensembles de données à grande échelle, il devient impossible pour les chercheurs de faire des analyses comparatives en santé et en éducation. Pour rendre ces données accessibles et utiles, il faut une volonté politique à l'échelle provinciale et, surtout, il faut réformer la Loi sur l'accès à l'information ainsi que les lois constitutives des ministères, des organismes et de l'ISQ. Le Québec est la seule province au Canada qui exige des chercheurs de passer par deux paliers d'approbation pour accéder à des données : il faut premièrement une autorisation de la part de la Commission d'accès à l'information, et deuxièmement une autre de la part du ministère concerné et détenteur des données. Même l'ISQ doit se plier à cet exercice, ce dont est dispensé Statistique Canada. La Commission d'accès à l'information n'est pas outillée pour faire face à des demandes utilisant des méthodologies statistiques complexes impliquant plusieurs sources de données. Le processus d'accès est un labyrinthe de formulaires et de démarches qui semble sans fin lorsque les sources de données sont multiples. La Loi sur l'accès à l'information devrait être réformée pour permettre à l'ISQ de faire avancer le Québec en suivant le modèle du Nouveau-Brunswick. Le Plan économique du Québec de mars 2018 mentionnait explicitement la promotion de l'accès aux données de

L'accessibilité aux données des administrations publiques

recherche (Gouvernement du Québec, 2018, section 5.6.2), et des sommes ont été octroyées en ce sens à l'ISQ, ce qui témoigne d'une certaine ouverture sur cette question. Mais il reste encore du travail à faire.

Bien entendu, il faut absolument un environnement sécurisé où le couplage des différentes bases de données est contrôlé par divers processus pour en assurer la confidentialité. Statistique Canada et l'ISQ ont une vaste expertise dans ce domaine, acquise sur plusieurs décennies. L'accès à leurs données est contrôlé et exclut l'information identificatrice (nom, prénom, adresse de résidence, numéro d'assurance sociale). La réglementation et l'infrastructure physique pour accéder à des données confidentielles existent déjà. Il faut maintenant que le gouvernement réforme la Loi sur l'accès à l'information et donne le mandat aux différents ministères de rendre leurs microdonnées accessibles, et ce, uniquement dans les environnements sécurisés existants.

Le potentiel de retombées est vaste. L'accès à des données de qualité permet aux chercheurs québécois de tirer le maximum de leurs travaux, favorisant ainsi un rayonnement à l'échelle canadienne et internationale. Il permet également la formation de la relève en recherche au Québec dans les techniques de pointe. Les nouveaux diplômés ont ainsi de meilleures perspectives professionnelles et peuvent davantage contribuer à notre société et à la croissance économique. Finalement, sans données, nous ne pourrions jamais apprendre de nos erreurs et tirer profit de réussites, afin de mieux guider nos politiques et d'améliorer le bien-être de notre société.



Références

Baker, M., Gruber, J. et Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, 116(4), 709-745. doi:10.1086/591908

Candas, B. (2016). *Constitution et utilisation des banques de données intégrées en santé : tour d'horizon international des modèles et des pratiques exemplaires*. Institut national de la santé publique du Québec. Repéré à : https://www.inspq.qc.ca/sites/default/files/publications/2188_banques_donnes_sante_modeles_pratiques.pdf

Card, D., Chetty, R., Feldstein, M. S. et Saez E. (2010). Expanding access to administrative data for research in the United States. *American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas*. doi:10.2139/ssrn.1888586

Connelly, R., Playford, C. J., Gayle, V. et Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1-12. doi:10.1016/j.ssresearch.2016.04.015

Connolly, M., Haeck, C. et Lapierre, D. (2019). Social mobility trends in Canada: Going up the Great Gatsby curve. Cahier de recherche n° 19-03, Groupe de recherche sur le capital humain, mai 2019 (version révisée). Repéré à : <https://econpapers.repec.org/paper/grcwpaper/19-03.htm>

Einav, L. et Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210). doi:10.1126/science.1243089

Frenette, M. (2019). *Do Youth from Lower-and Higher-income Families Benefit Equally from Postsecondary Education?* Statistique Canada, Analytical Studies Branch Research Paper Series. Repéré à : <https://www150.statcan.gc.ca/n1/en/pub/11f0019m/11f0019m2019012-eng.pdf?st=1ynUaiEo>

Gouvernement du Québec. (2018). *Le plan économique du Québec*. Budget 2018-2019. Bibliothèque et Archives nationales du Québec. Repéré à http://www.budget.finances.gouv.qc.ca/budget/2018-2019/fr/documents/PlanEconomique_18-19.pdf

Grusky, D. B., Hout, M., Smeeding, T. M. et Snipp, M. (2019). The american opportunity study: A new infrastructure for monitoring outcomes, evaluating policy, and advancing basic science. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(2), 20-39. doi:10.7758/rsf.2019.5.2.02

Haeck, C., Lebihan, L. et Merrigan, P. (2018). Universal child care and long-term effects on child well-being: Evidence from Canada. *Journal of Human Capital*, 12(1), 38-98. doi:10.1086/696702

Haeck, C., Lefebvre, P. et Merrigan, P. (2015). Canadian evidence on ten years of universal Preschool Policies: The good and the bad. *Labour Economics*, 36C, 137-157. doi:10.1016/j.labeco.2015.05.002

Lefebvre, P. et Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada, *Journal of Labor Economics*, 26(3), 519-548. doi:10.1086/587760

Lewis, M. (2016). *The undoing project: A friendship that changed our minds*. W.W.Norton and Company.

Marchessault, G. (2011, février). The Manitoba Centre for Health Policy: A case study. *Healthcare Policy*, 6(numéro spécial), 29-43. doi:10.12927/hcpol.2011.22117

L'accessibilité aux données des administrations publiques

Morneau, B. (2016, 22 mars). *Rétablir l'espoir pour la classe moyenne : le budget fédéral de 2016*. Notes de discours de l'honorable Bill Morneau, C.P., député, à l'occasion de la présentation du budget fédéral de 2016. Repéré à : <https://www.budget.gc.ca/2016/docs/speech-discours/2016-03-22-fr.html>

Statistique Canada. (2009). Lignes directrices concernant la qualité. Cinquième édition – octobre 2009. Repéré à : <https://www150.statcan.gc.ca/n1/fr/pub/12-539-x/12-539-x2009001-fra.pdf?st=Pv13etHV>

Notes

1. Une version plus détaillée de ce texte est disponible sur le site <https://grch.esg.uqam.ca/>, dans la rubrique Publications, en tant que Cahier de recherche n° 19-04 du Groupe de recherche sur le capital humain.
2. Notre point de vue, mais aussi celui de plusieurs : des opinions glanées au fil du temps, ainsi que d'autres ayant fait l'objet d'études (par exemple Candas, 2016).
3. Certains ensembles de données sont également disponibles à distance, entre autres par le Système d'accès à distance en temps réel (voir <https://www.statcan.gc.ca/fra/adtr/adtr>).
4. Voir la liste complète en visitant ce site Web : <https://www.unb.ca/nbirdt/data-holdings/holdings.html>.
5. Les fichiers accessibles à distance sont toutefois différents de ceux disponibles dans les CADRISQ, car certaines informations peuvent avoir été modifiées pour satisfaire les obligations de l'ISQ en matière de protection des renseignements personnels.