

## **INFRASTRUCTURE**

### **« Consortium d'exploitation des données administratives québécoises » (CEDAL)**

#### **Plan d'action 2009-2013**

Deux regroupements stratégiques (RS) du Réseau de recherche en santé des populations du Québec, celui sur les bases de données et celui sur le longitudinal et les parcours de vie, ont joint leurs efforts depuis trois ans pour tenter de mettre en place au Québec un « Consortium d'exploitation des données administratives longitudinales » (CEDAL). Il s'agit là d'une infrastructure de recherche qui a fait ses preuves ailleurs. On le constate en particulier à voir l'abondante production d'une infrastructure semblable, établie au Manitoba depuis deux décennies et qui a inspiré la création de centres similaires dans plusieurs provinces canadiennes. Ces efforts donnent d'ailleurs lieu en ce moment à la création, à l'initiative de Statistique Canada, d'un réseau de collaboration entre plusieurs provinces pour faire des études comparatives.

Les données administratives longitudinales sont souvent le seul moyen de faire de telles recherches, extrêmement révélatrices d'une part des tendances lourdes au plan de la santé et de l'efficacité des soins, et d'autre part des transformations dans les déterminants de la santé des populations. En effet, les données d'enquêtes longitudinales, si elles sont plus fines, fournissent souvent des séries assez courtes, dont il est difficile de dégager des tendances longues pour apprécier, en particulier, les effets de différentes politiques et de diverses interventions en santé. Quand, en plus, de telles données peuvent être jumelées d'un domaine à l'autre et appariées avec des données d'enquête, elles peuvent devenir de formidables instruments de recherche et fournir une aide puissante à la décision éclairée.

Les infrastructures de rassemblement et d'accès aux données administratives longitudinales visent à lever trois obstacles majeurs à la poursuite de telles recherches. En premier lieu, et c'est l'obstacle le plus évident, les données de telles séries administratives posent évidemment des problèmes de confidentialité considérables. En l'absence de telles infrastructures, l'exploitation de telles données pour fins de recherche fondamentale ou appliquée suppose des démarches à la pièce auprès de la Commission d'accès à l'information qui sont certes indispensables, mais également extrêmement longues et frustrantes. On ne s'étonnera donc pas que la recherche faisant usage de ces riches bases de données soit très rare au Québec. Dans d'autres provinces, on est parvenu à mettre en place des protocoles et des technologies qui simplifient radicalement cet accès, tout en maintenant les normes de protection des informations individuelles les plus hautes. C'est là une première fonction que pourrait remplir le CEDAL.

En deuxième lieu, la recherche faite à la pièce avance péniblement parce que les efforts consentis ne sont pas cumulatifs : chacun doit repartir à zéro. Or les fichiers administratifs sont notoirement difficiles à manipuler. Il faut valider les informations, souvent en travaillant de concert avec ceux qui en sont les dépositaires, qui doivent parfois intervenir eux-mêmes auprès de ceux qui les ont recueillies pour connaître les circonstances de cette cueillette et les biais qui s'y sont glissés. Une fois ce patient travail complété, les chercheurs profitent de ce lourd investissement pour faire avancer leurs propres travaux, puis sont obligés de détruire le précieux nettoyage des données ainsi réalisé, faute d'un lieu qui pourrait les accumuler dans des conditions qui en garantissent la validité et qui puissent les mettre à disposition des chercheurs voulant utiliser les mêmes données par la suite. Une infrastructure comme le CEDAL servirait donc à stocker l'« intelligence » qui a été mise dans les données par ses usagers successifs. Le deuxième chercheur profiterait de l'investissement du premier, quitte à laisser lui-même le fruit de son travail comme base du travail du troisième, et ainsi de suite. C'est là une clé importante du développement cumulatif de telles recherches.

En troisième lieu, enfin, les obstacles que nous venons d'évoquer ont eu pour conséquence que nous avons très peu de chercheurs, dans le réseau public ou dans les universités, qui ont une expérience de travail avec de telles données administratives longitudinales. Il faudra assurer que ceux qui ont déjà cette expertise contribuent, par des programmes de formation et des activités de promotion de ce type de recherche, à élargir considérablement le bassin de chercheurs compétents. C'est ce qu'a fait le Centre interuniversitaire québécois de statistiques sociales (CIQSS), tout comme le Réseau canadien de Centres de données de recherche, qui compte maintenant 24 laboratoires au Canada, pour soutenir l'exploitation des enquêtes longitudinales et les données complexes issues de Statistique Canada (et de l'Institut de la statistique du Québec, dans le cas du CIQSS).

## **Annexe résumée : éléments essentiels d'un projet de « Consortium d'exploitation des données administratives longitudinales (CEDAL) »**

### **Objectifs et principes de base du CEDAL**

Le CEDAL vise essentiellement quatre objectifs. Leur atteinte est indispensable si le Québec veut développer une capacité sérieuse d'exploitation des données administratives longitudinales pour en tirer des informations et des connaissances requises pour soutenir les décisions publiques.

1. La validation des données, pour transformer l'information brute en variables sur lesquelles on pourra fonder des interprétations valables. Pour cela, il faudra créer un noyau de chercheurs et professionnels dont la mission sera d'assurer le développement de l'expertise d'exploitation des bases de données et d'offrir le soutien aux équipes qui désireront réaliser des analyses des données. De plus, il faudra coordonner les efforts et s'assurer que l'intelligence incorporée dans les données par leurs utilisateurs successifs est conservée et mise à disposition de l'ensemble des chercheurs qui veulent les utiliser subséquemment.

2. Un accès facilité aux données pour fins de recherche, dans le respect intégral de leur caractère confidentiel, mais aussi dans un effort délibéré pour lever les obstacles aux travaux de production d'informations et de connaissance.
3. La formation de nouveaux chercheurs et la mobilisation de chercheurs établis autour de ce projet d'exploitation de ces données, afin que la capacité de recherche québécoise en la matière soit radicalement renforcée.
4. L'élaboration d'un plan de développement des bases de données et d'un système d'échanges d'information entre chercheurs, et avec les décideurs et cliniciens, afin de répondre aux besoins des trois groupes et qui s'inspire de la perspective des parcours de vie (trajectoires le moyen et de long terme, prise en compte des diverses facettes de la vie (santé, mais aussi éducation, emploi et revenu, famille). Ceci permettra de situer les travaux entrepris dans le contexte large des politiques favorables à la santé.

Pour atteindre ces objectifs, le CEDAL devra s'inspirer des principes suivants, qui ont fait leurs preuves dans des centres semblables au Canada (Manitoba, Colombie-Britannique, Ontario Nouvelle-Écosse et Terre-Neuve) de même que dans le « Réseau canadien des Centres de données de recherche », dont le Centre interuniversitaire québécois de statistiques sociales est l'un des plus beaux fleurons.

1. Le CEDAL devra bien sûr jouir de l'appui de son partenaire principal, le MSSS, mais il devra aussi disposer d'une bonne marge d'autonomie, afin de rallier l'ensemble des joueurs dont la collaboration est indispensable à son succès : Agences et CSS, INSPQ, ISQ, RAMQ, Commissaire à la santé et aux services sociaux et bien sûr les chercheurs universitaires. Ceux-ci devront pouvoir y mener des recherches indépendantes et publiables, en même temps qu'orientées vers les besoins de la prise de décision dans le système de la santé et des services sociaux. C'est avec cette marge d'autonomie que le CEDAL pourra devenir un véritable lieu d'initiative stratégique et méthodologique, de même qu'un centre de production de connaissance respecté et légitime.
2. Il faudra assurer un financement généreux des activités d'exploitation des données et de la recherche, au moins durant une assez longue phase de développement. Le défi principal du CEDAL n'est pas de fournir un meilleur accès aux données à des chercheurs qui seraient déjà nombreux. Au contraire, la stimulation de la demande des chercheurs est encore plus importante que la satisfaction de l'offre. Le CEDAL vise principalement une croissance importante de la capacité québécoise d'exploitation des données administratives longitudinales. Cela suppose d'intéresser de nouveaux chercheurs, de faire les difficiles investissements initiaux dans la validation des données, d'accumuler un fonds « d'intelligence des données » qui accélérera la cadence des projets futurs. Cela requiert une mise de fonds de démarrage significative, qui pourra par ailleurs, par l'entremise de l'équipe du CEDAL, favoriser les projets pouvant apporter à la décision publique un éclairage sur des questions dont la pertinence est prioritaire.

## Questions pratiques à résoudre

**1. La validation des données.** La première mission sera d'assurer la validation des données pour fins de recherche utiles tant aux missions de surveillance et de gestion des organismes que de la recherche universitaire. Cette validation pourra se faire en collaboration avec le nouvel infocentre.

- En assurant que les données sont mises à jour
- En assurant qu'une version standard des données validées est identifiée pour chaque année et, de ce fait, que cette version passe dans l'usage courant
- en assurant que, dans toute la mesure du possible, les données d'une même source sont comparables d'une année à l'autre ;
- Et surtout **en assurant que l'intelligence des données** (les améliorations apportées aux données au cours de leur utilisation en gestion/surveillance et en recherche) soit incorporées dans les bases de données pour qu'elles puissent profiter aux usagers ultérieurs (variables dérivées, sous échantillons spécifiques, etc.). Ces améliorations se matérialiseront dans des routines informatiques applicables aux données, accompagnées d'une documentation qui devrait se conformer aux normes internationales les plus avancées et **en évitant les duplications** c'est-à-dire que la même validation ou le même jumelage ne soient réalisés à deux endroits différents
- S'assurer du développement et du maintien de « l'intelligence » de ces données, ce qui implique :
  - la vérification et la validation des variables dans chacune des banques;
  - la standardisation (ex. résultats de laboratoire);
  - l'assurance de la compatibilité ou complémentarité des informations se retrouvant dans plus d'une banque;
  - le développement d'indices (la combinaison de plusieurs variables disponibles dans les fichiers), améliorant le potentiel informatif et la définition de sous populations;
  - la participation à des groupes de travail pour établir ou redéfinir des standards;
  - l'élaboration de programmes informatiques permettant la création d'indicateurs ou de trajectoires et le stockage des programmes et routines utilisés; *Il doit y avoir des ententes avec les fournisseurs et les experts pour que ces opérations soient réalisées.*
- Assurer le jumelage<sup>1</sup> des banques de données et leur conservation;

---

<sup>1</sup> *Jumelage des banques de données* : la possibilité de jumeler les différentes banques de données est primordiale pour atteindre les objectifs mais il est important de considérer qu'il est possible de le faire sans qu'il y ait création d'une méga banque de données.

- Donner accès à de l'aide technique – sur les données et les analyses; ce qui suppose *une équipe minimale et que des ententes soient convenues avec des partenaires pour que cette aide soit toujours disponible aux chercheurs et aux professionnels du réseau de la santé et des universités.*
- Mettre à la disposition des décideurs ou de leurs représentants, des indicateurs identifiés comme pertinents pour leur planification et la reddition de compte selon différents découpages (national, régional, local) par le biais d'ententes avec les organismes qui ont l'infrastructure de diffusion; on doit *assurer la pérennité d'un ensemble d'indicateurs dont la valeur est reconnue par les chercheurs et la nécessité avalisée par les décideurs.*
- Faire la promotion de la recherche à partir de ces banques de données administratives; ce qui suppose que des *chercheurs soient associées au projet et intéressés par les possibilités offertes de répondre à des questions qui pourraient être retenues par des décideurs.*
- Tenir à jour un dictionnaire de données;
- Colliger la documentation et les articles et rapports de recherche basés sur ces ensembles de données;
- Développer et maintenir une bibliothèque de tous les développements mentionnés pour que l'apprentissage et le potentiel informatif soient progressifs et cumulatifs;
- Participer à la formation des étudiants et des chercheurs
- S'associer à des organismes pour partager les expertises.
- Assurer le transfert de connaissances vers le réseau de la santé et ceux d'autres ministères associés;
- Réaliser des activités de recherches commanditées par les décideurs ou qui correspondent aux intérêts propres des chercheurs associés au CEDAL
- Constituer un noyau de chercheurs intéressés à s'impliquer dans les projets de démarrage.

**2. La formation et promotion.** Le CEDAL pourra assumer une fonction de formation à l'analyse des bases de données administratives longitudinales, doublée d'une fonction de promotion de celle-ci. Le CEDAL devra faire appel à des chercheurs externes pour des collaborations de recherche pour réaliser ses mandats. Le CEDAL sera une porte d'entrée privilégié pour les chercheurs désirant réaliser des recherches indépendantes utilisant les données administratives.

**3. Une perspective de développement à long terme** Le CEDAL offrira le développement des bases de données administratives longitudinales et de l'analyse de ces données, couvrant préférentiellement la moyenne et la longue durée, avec des données « multi facettes », par le jumelage de bases de données de différents domaines qui intègrent des bases de données sociales à celles de la santé en replaçant les individus dans le contexte des ménages et des familles auxquels ils appartiennent sur des bases locales, régionales ou nationale ;

## **Financement**

Pour fonctionner adéquatement, on doit compter sur un financement de base, pour assurer la coordination entre les fournisseurs, les chercheurs et les décideurs, et une équipe technique de base.

Ce financement pourrait provenir de plusieurs sources. Un fond de démarrage pourrait être fourni par le MSSS, le FRSQ et le RRSPQ.

Des ententes avec des partenaires institutionnels doivent être établies pour qu'en contrepartie d'un financement, des études pertinentes à leurs besoins décisionnels soient produites par des chercheurs associés au CEDAL comme cela se fait au Manitoba, par exemple.

## **Organisation administrative**

Le CEDAL doit se doter d'une structure simple. Il doit être un facilitateur.

Il ne s'agit pas de créer une immense infrastructure mais un lieu de coordination pour permettre l'intégration fonctionnelle de l'ensemble des expertises requises pour assurer une recherche utile à la prise de décision. Une équipe minimale permanente est requise pour réaliser les fonctions décrites ci-haut. Celle-ci pourra être composée de chercheurs provenant du milieu universitaire et du secteur public et ses membres pourront se situer dans des lieux physiques distincts. Le mandat de cette équipe sera d'assurer la pérennité des opérations, de coordonner les besoins et les demandes et l'échange entre les partenaires tant fournisseurs que chercheurs ou décideurs.

Le CEDAL aura donc trois fonctions fondamentales pour le développement de la recherche utilisant les données administratives longitudinales comme source d'informations utiles à la recherche et à la décision éclairé. Il devra favoriser l'accès aux données dans le plus grand respect de la confidentialité; créer un pôle d'expertise en analyse des données, essentiellement nourri par le cumul de l'intelligence mise dans les données par les efforts des chercheurs; et mobiliser les chercheurs en assurant leur formation, en illustrant l'intérêt de telles recherches pour le milieu universitaire et pour le réseau de la santé et des services sociaux (de même que pour les autres ministères et agences gouvernementales, puisque par le biais de l'étude des déterminants sociaux de la santé, objet premier du CEDAL, seront également abordées des questions relatives à la pauvreté, à l'éducation, à la famille, à l'emploi et au revenu, etc.)

Le RRSPQ a mobilisé depuis trois ans les milieux concernés par un tel projet, aussi bien dans les universités que dans les ministères, les Agences de santé, l'INSPQ, l'ISQ, la RAMQ et ainsi de suite. Notre objectif est en effet de créer un consortium, où travailleraient en synergie diverses institutions qui pourraient fournir des ressources et des contributions correspondant à leur mission propre et en tirer des bénéfices quant à l'accomplissement de cette mission. Nous ne voulons pas créer un organisme qui se substituerait aux acteurs qui sont déjà présents et actifs. Le CEDAL sera plutôt un lieu d'expertise méthodologique et d'initiative au plan des politiques publiques, qui favorise

une collaboration, une mise en commun des ressources, une organisation efficace de l'accès, une continuité dans la création de bases de données valides et une mobilisation des chercheurs, comme utilisateurs de données, et des décideurs, comme demandeurs de recherche aidant la décision.

Pour arriver à créer un tel consortium, le RRSPQ a tenu au cours des trois dernières années un grand colloque et de nombreuses réunions avec tous les partenaires possibles d'un tel projet : réunions techniques avec des experts et réunions politiques avec des acteurs qui sont de plus en plus sensibilisés à l'impérative nécessité de tirer parti de telles données administratives longitudinales, à l'instar de ce qui se fait ailleurs au Canada et dans le monde. Il est évident qu'un tel projet suscite énormément d'enthousiasme et très peu de réticences. De plus, le RRSPQ a clairement acquis le statut d'honnête courtier, d'acteur indépendant et bien intentionné qui est le mieux placé pour qu'advienne l'effort de concertation requis pour un consortium.

Mais le chemin se révèle quand même parsemé d'embûches. L'accès à l'information est toujours une question très délicate. Les ressources requises sont loin d'être négligeables. De plus, les détenteurs de données ont des mandats stricts quant à l'usage qu'ils peuvent en faire et qui ne permet pas toujours de place à la recherche faite à l'extérieur, par des chercheurs indépendants. Le recours à ces forces extérieures nous paraît pourtant, à la lumière des expériences étrangères, à la fois indispensable pour tirer tout le parti de ces données et tout à fait gérable dans le respect des intérêts de chacun.

Nous en sommes pour le moment à l'étape où tous les acteurs pertinents nous font confiance et partagent notre point de vue. Mais il reste beaucoup de travail à faire pour assurer l'émergence du CEDAL. Travail technique, certes. Mais aussi travail de persuasion par rapport à des formules précises. Nous avons à cet égard produit un grand nombre de documents qui spécifient comment nous entendons nous y prendre (**annexe 9.6**). Mais le travail technique plus poussé ne pourra s'enclencher que si le feu vert est donné à cet égard par le MSSS, qui est un joueur incontournable. Le ministère en est justement à réévaluer son action en regard des ressources informationnelles qu'il détient ou contrôle. Nous avons de la part du sous-ministre un accord de principe pour établir un Groupe de travail conjoint MSSS/RRSPQ en vue de recommander des modes d'implantation d'un consortium (**annexe 10.6.1**). Il reste à ce que les choses démarrent vraiment : politiquement, par rapport à la formule que prendra le consortium, et techniquement, pour résoudre la myriade de problèmes d'accès aux données, d'aide experte à leur utilisation, de stockage, de l'intelligence des données, de formation, de diffusion des résultats de recherche, de financement. Soulignons à cet égard que l'une des formules les plus intéressantes consiste, comme au Manitoba, en un contrat cadre de recherche entre ministère(s) et chercheurs pour la production annuelle d'un certain nombre de rapports de recherche, convenus par mutuel accord, en contrepartie d'un financement régulier.

Il nous faut donc poursuivre, au cours des quatre années qui viennent, les efforts pour que le CEDAL devienne réalité. Basé sur notre expérience passée, nous estimons qu'il nous faudra 18-24 mois de rencontres formelles au MSSS pour négocier un plan opérationnel de mise en œuvre d'un CEDAL. Aussitôt que le CEDAL démarrera comme projet pilote, le gros de ses ressources devra venir du réseau public. Mais il nous semble important que le RRSPQ, en tant que partie prenante au consortium représentant les chercheurs universitaires et ceux du secteur public, ait voix au chapitre en fournissant lui-même une partie du financement. Soulignons par ailleurs que si les partenaires formant CEDAL exigent que le pôle universitaire du consortium s'engage aussi financièrement, le RRSPQ s'associerait sans doute à une demande de financement plus substantielle, et distincte de son financement de base, auprès du FRSQ. Ultimement un financement par la Fondation Canadienne pour l'innovation est envisagé et des discussions préliminaires en vue d'un concours dédié à l'exploitation des BDAL ont récemment eu lieu (Lyne Sauvageau, FQRSC, communication personnelle).