

Constitution et utilisation des banques de données intégrées en santé : tour d'horizon international des modèles et des pratiques exemplaires

SYNTHÈSE

BiESP BUREAU D'INFORMATION
ET D'ÉTUDES EN SANTÉ
DES POPULATIONS

Sommaire

Introduction : les banques de données et la santé	2
Le cadre légal et réglementaire : tension permanente ou question d'approche?	3
Les banques de données intégrées en santé : plus qu'un système d'information	6
Conclusion : un développement basé sur le leadership et la stabilité	9

Résumé

L'utilisation secondaire des données de santé médico-administratives jumelées a permis des avancées importantes dans les domaines des sciences sociales, biomédicales et translationnelles. Elle joue également un rôle essentiel pour le soutien à la décision dans la transformation des services. Les difficultés d'accès aux données personnelles de santé pour une utilisation secondaire sont fréquemment citées comme étant le résultat des contraintes légales visant le respect de la vie privée. Pourtant, certaines juridictions ont grandement facilité l'utilisation secondaire, sans compromis sur la protection de la vie privée, par la création de centres de données intégrées en santé dont la mission est d'en maximiser l'exploitation au profit du système de santé et du bien-être de la population.

Le développement de tels centres soulève forcément des enjeux de respect de la confidentialité, de gouvernance et de structure opérationnelle. Heureusement, plusieurs modèles canadiens et internationaux sont des exemples de réussite dont on peut tirer les enseignements.

Le modèle de référence est constitué d'un centre de données indépendant commandité par les autorités publiques dont la gouvernance est assurée par un partenariat de l'ensemble des acteurs. Ce centre est aussi un lieu d'expertise capable de générer des données intégrées à valeur ajoutée – au-delà du jumelage, de soutenir les utilisateurs et de contribuer activement au transfert des connaissances générées grâce à l'utilisation des données. Le respect des lois et des règlements (y compris en amont et en aval du centre) restent sous la supervision de l'organisme chargé de veiller à la protection des renseignements personnels qui au Québec est la Commission d'accès à l'information.

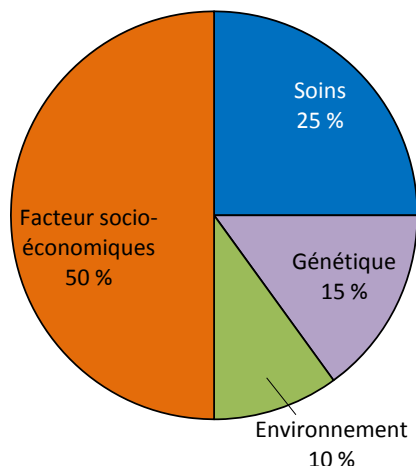
Le Québec possède déjà une banque de données intégrées en santé entourée de l'expertise nécessaire¹ pour mettre en place rapidement un projet pilote sur le modèle des centres de données exemplaires canadiens et internationaux. Au-delà de la fluidification de l'accès aux données, ce projet permettra à tous les acteurs de convenir des conditions requises pour constituer un véritable centre de données intégrées en santé, y compris les ajustements légaux et réglementaires qu'il y aurait lieu de faire.

Il est maintenant démontré que la création d'un centre de données intégrées en santé est entièrement compatible avec les principes de confidentialité et de protection de la vie privée. La véritable question est donc de savoir si le Québec a réellement les moyens d'en faire l'économie compte tenu des bénéfices potentiels pour la gouvernance du système de santé, le milieu de la recherche et la santé de la population.

Introduction : les banques de données et la santé

La santé des individus résulte de l'interaction entre de nombreux facteurs (1,2) qui constituent un écosystème complexe (caractéristiques sociales, économiques, organisationnelles, environnementales et individuelles).

Estimations de la contribution des principaux déterminants à l'écosystème de la santé



(Canadian Institutes of Advanced Research, 2012).

Le système de santé, partie intégrante de cet écosystème, intervient sur les facteurs modifiables par des politiques, des programmes et des infrastructures dont la gestion nécessite la collecte et le traitement d'un volume gigantesque de renseignements personnels² de nature médico-administrative. Au-delà de leur utilité primaire de gestion, l'utilisation secondaire de ces données permet d'effectuer des études, des analyses et des évaluations. Mais le plein potentiel ne se révèle que lorsque les différentes sources sont jumelées et bonifiées par des données organisationnelles et sur l'environnement des personnes. C'est alors que la recherche peut contribuer à accroître nos connaissances sur la santé dans la globalité de son écosystème afin d'informer les décideurs, d'améliorer les services et contribuer à la santé des individus. Dans son rapport « *Best care at lower cost* » paru en 2013, la première recommandation de l'*Institute of Medicine* était de créer de grandes banques de données intégrées pour permettre au système de santé d'apprendre de lui-même, en permanence et en continu (3).

Cette recommandation repose sur les contributions à l'amélioration de la santé qui ont été réalisées dans plusieurs provinces canadiennes et pays développés par la création de banques de données intégrées en santé (4). D'une part, l'utilisation secondaire d'informations jumelées provenant de différentes banques de données médico-administratives est une pratique de recherche et d'évaluation répandue (5), mais les banques de données intégrées en santé à l'échelle populationnelle ouvrent de nouvelles perspectives par :

- l'augmentation de la **puissance des études** multifactorielles, d'interactions complexes ou avec de nombreux facteurs confondants;
- l'étude de phénomènes qui se développent sur des **décennies ou des générations**;
- l'inclusion de tous les segments de la population, **sans discrimination**, aussi divers et petits soient-ils;
- l'étude des **faibles prévalences** (maladies, combinaisons de facteurs, etc.);
- le suivi de l'implantation d'innovations : **efficience** des programmes, **sécurité** des médicaments, **issues populationnelles** de santé;

¹ Le Système intégré de surveillance des maladies chroniques du Québec (SISMACQ) géré et exploité par l'Institut national de santé publique du Québec.

² Un renseignement est considéré personnel s'il s'agit d'informations nominatives ou qui permettent, seules ou en les combinant à d'autres, d'identifier une personne.

- la compréhension **des barrières et des facilitateurs** aux interventions et de leur valeur sociétale;
- la priorisation **des problématiques** (tenant compte des prévalences, facteurs associés et variantes locales).

Il existe une volumineuse littérature démontrant l'impact des banques de données intégrées populationnelles et particulièrement en santé (5–15). Les exemples en recherche sociale, biomédicale et translationnelle sont nombreux, mais il faut aussi noter la place essentielle que les banques de données intégrées en santé occupent dans la transformation des services (16). Dans ce domaine, l'accès aux données populationnelles intégrées est essentiel pour permettre à la recherche d'informer adéquatement la prise de décision stratégique.

Ces grandes banques de données intégrées en santé constituent une source d'information à très haute valeur ajoutée pour un coût relativement faible. Aux États-Unis, la création de valeur par l'utilisation de grandes banques de données intégrées en santé est estimée entre 12 et 17 % des dépenses annuelles de santé (17). Environ le tiers est associé à l'amélioration coût-efficacité des soins, à l'identification des meilleures approches de prestation de soins et d'innovations efficaces ainsi qu'à la stimulation de la recherche et de l'innovation dans les domaines prioritaires pour le système de santé. L'accès à une banque de données intégrées en santé est donc actuellement perçu comme une nécessité plutôt qu'une option. Il offre deux avantages majeurs: celui d'améliorer le système de santé et de permettre à la recherche de rester compétitive sur le plan international. Au Québec, les secteurs de la santé et de la recherche sont tous deux financés par le secteur public et il y a donc encore plus d'avantages à les faire travailler l'un au profit de l'autre.

S'il y a consensus sur les bénéfices apportés par une infrastructure informationnelle intégrée en santé, la réalisation d'un tel projet soulève plusieurs enjeux de confidentialité des données, de gouvernance et de structure opérationnelle permettant l'atteinte des objectifs de gains pour la santé. Ce document propose un tour d'horizon des modèles de développement et d'utilisation d'une banque de données intégrées en santé à l'échelle populationnelle qui constituent

actuellement des pratiques exemplaires applicables au contexte législatif et au système de santé québécois.

Le cadre légal et réglementaire : tension permanente ou question d'approche?

Le Québec a été un pionnier et un modèle pour la mise en place d'une législation de protection de la vie privée régissant la création et l'utilisation des banques de données, notamment administratives (18). À ce jour, cette législation n'a connu aucune faille notable concernant l'utilisation des données médico-administratives de santé par la recherche. Pour constituer une grande banque de données intégrées en santé, il faudra donc s'assurer que la même rigueur et au moins le même niveau de protection sont offerts aux citoyens.

La constitution et l'utilisation des banques de données médico-administratives sont régies par la Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels (ci-après la Loi) qui crée la Commission d'accès à l'information (CAI) pour veiller à son application. D'autres lois particulières prévoient la création de banques de données pour répondre à des besoins spécifiques, mais l'utilisation secondaire des données reste régie par la Loi.

En bref, la Loi permet la cueillette et l'utilisation de renseignements personnels par les organismes publics, sans le consentement explicite des personnes (art. 64), dans la mesure où cela est nécessaire à l'exercice des attributions de l'organisme ou pour la gestion d'un programme dont il a la responsabilité. Les usages à d'autres fins requièrent généralement le consentement de chaque personne³ (art. 65.1 de la Loi). La communication de renseignements personnels entre des organismes publics est aussi encadrée sur les mêmes bases de la nécessité et de l'exercice d'une responsabilité de gestion de politiques ou de programme (art. 67 ss.).

³ Toutefois le consentement ne sera pas requis si l'utilisation de l'information a un lien pertinent et direct avec les fins pour lesquelles il a été recueilli, ou si son utilisation est manifestement au bénéfice de la personne concernée ou encore si est nécessaire à l'application d'une (autre) loi au Québec.

La Loi reconnaît toutefois le bien-fondé de l'utilisation secondaire de renseignements personnels à des fins d'étude, de recherche ou de statistique et, plutôt que de fixer des critères stricts, elle fournit des balises qui permettent à chaque projet de faire valoir ses contraintes, ses spécificités et ses mérites auprès de la CAI (art. 125) :

La Commission peut [...] accorder [...] l'autorisation de recevoir à des fins d'étude, de recherche ou de statistique [...] des renseignements personnels sans le consentement des personnes concernées, si elle est d'avis que:

- l'usage projeté n'est pas frivole;
- les fins recherchées ne peuvent être atteintes que si les renseignements sont communiqués sous une forme nominative;
- les renseignements personnels seront utilisés d'une manière qui en assure le caractère confidentiel.

C'est donc à la CAI que revient la responsabilité d'évaluer si les conditions de la Loi sont remplies.

La non-frivolité de l'usage sera évaluée à partir d'un protocole de recherche ou d'étude qui doit définir avec précision les objectifs spécifiques visés. Pour les chercheurs, il existe des mécanismes institutionnels indépendants qui valident cette exigence comme l'évaluation par les pairs opérée par les organismes subventionnaires et les évaluations par les comités d'éthiques.

Les deux exigences suivantes sont en quelque sorte reliées puisque la CAI tentera de limiter la communication de renseignements personnels au strict minimum ce qui réduit aussi le risque d'atteinte à la vie privée. Pour réaliser cet exercice, la CAI demandera au chercheur de justifier chacune des variables requises et le nombre d'individus pour lesquels ces renseignements sont requis. Elle évaluera le risque en faisant aussi appel à l'expertise des détenteurs de données et en exigeant que le chercheur obtienne leur accord. S'il doit y avoir un jumelage entre des renseignements personnels issus de plusieurs bases de données, il devra être réalisé par l'un des détenteurs et non par le requérant pour encore une fois réduire la communication de renseignements personnels aux chercheurs.

Cette loi reconnaît donc une spécificité aux besoins de la recherche en lui accordant un accès aux données sans fixer des normes administratives rigides, mais plutôt en guidant le jugement sur la base de trois conditions très légitimes.

Le processus comporte toutefois des barrières opérationnelles importantes qui réduisent, voire découragent, l'utilisation secondaire des données pour le développement des connaissances par la recherche académique de même que par les organismes et administrations publiques. Il présuppose d'abord que les requérants ont une connaissance pointue des banques de données administratives afin de traduire les questions de recherche concernant les algorithmes de sélection de cohortes et des variables à extraire qui sont entièrement conditionnés par l'architecture et la codification complexes des informations dans les banques médico-administratives. Les négociations qui doivent être entreprises entre la CAI, l'ensemble des détenteurs auxquels on s'adresse et le requérant impliquent des coûts et des délais de transactions importants. Tout au long du processus, il persistera donc une incertitude sur la faisabilité « réglementaire » du projet, peu compatible avec la compétitivité exigée des chercheurs par les organismes de financement et les universités. Dans les organisations et les administrations publiques, cette faisabilité « réglementaire » est régulièrement source de tensions et nuit à l'accomplissement des mandats et des obligations. Enfin, dans ce contexte, un chercheur ne peut envisager de faire de préétudes pour tester la pertinence de sa démarche.

Dans plusieurs juridictions où des législations similaires à celle du Québec ont été adoptées, il existe aussi une perception générale qu'elles ont freiné l'utilisation des données au détriment de la santé des populations et qu'il en sera de même pour une banque de données intégrées en santé (19). Les tensions entre les organismes chargés de veiller à la protection de la vie privée et les chercheurs en santé ont été traduites dans de nombreuses publications (19–25). L'Academy of Medical Sciences au Royaume-Uni a très bien résumé le regard que portent ces milieux sur les mécanismes de contrôle :

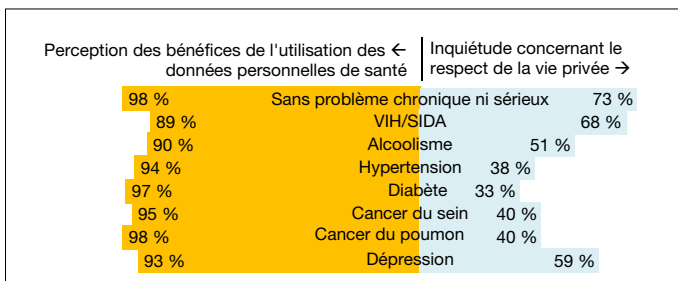
Il n'y a pas de données démontrant un risque systémique de méfaits par l'utilisation de renseignements personnels en recherche en santé...

... mais les progrès en santé publique et en santé des populations sont réduits par les contraintes issues de ces lois et règlements.

La première affirmation ne peut être réfutée par des exemples concrets d'utilisation abusive de renseignements personnels, mais cela tend plutôt à confirmer l'efficacité des mécanismes en place qu'à douter de leur pertinence. De même, l'affirmation voulant que le respect des règles de préservation de la confidentialité des données se fasse au détriment de l'amélioration de la santé et des systèmes de soins ne semble pas justifiée. En effet, il existe des structures efficaces de partage de données intégrées qui contribuent significativement à l'amélioration de la santé tout en renforçant l'application des lois et règlements dans des juridictions tout aussi attachées à la confidentialité des données que le Québec(4). Il existe donc des modèles de partage de données intégrées en santé qui réduisent les barrières opérationnelles et donc accroissent l'exploitation des données au profit de la collectivité sans compromettre la confidentialité des données et le respect de la vie privée.

De façon générale, le public est conscient, à plus de 90 % des bénéfices qu'il y a à exploiter les données individuelles de santé pour la recherche (26). Par contre, les risques à la vie privée restent une préoccupation importante.

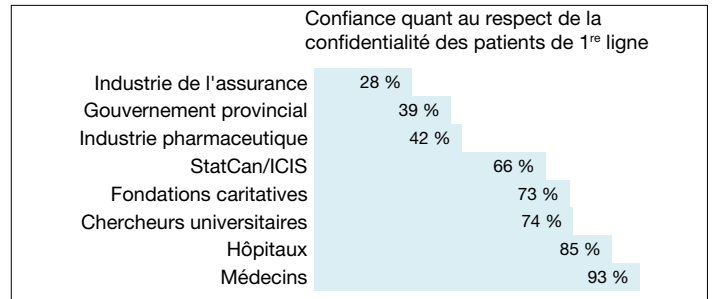
Pourcentage de répondants canadiens, selon l'état de santé



(Adapté de DJ Willison et coll (2009). BMC Medical Ethics.10:10).

Il est important de noter que la confiance du public quant au respect de la confidentialité dépend largement de l'organisation à qui les données sont confiées.

Confiance envers le respect de la confidentialité selon l'utilisateur des données



(Adapté de DJ Willison et coll (2009). BMC Medical Ethics.10:10).

La confiance la plus élevée est attribuée à ceux auxquels le public confie déjà ses données personnelles de santé (médecins et hôpitaux), mais qui ne possèdent généralement pas l'expertise d'exploiter les données populationnelles. Ensuite viennent les institutions du savoir reconnues pour leur expertise (universitaires et organismes publics) qui obtiennent la confiance d'au moins deux tiers de la population. En dernier, on retrouve les gouvernements et les industries (assurances et pharmaceutiques).

Cette étude montre bien que l'appui du public pour une utilisation des données de santé par la recherche existe, mais qu'il est conditionnel. La confiance du public n'est pas acquise et l'application d'une réglementation solide reste donc incontournable.

Les banques de données intégrées en santé : plus qu'un système d'information

Des modèles exemplaires

La création d'une banque de données intégrée en santé qui respecte la confidentialité des données n'en assure pas de facto une exploitation optimale. Il ne s'agit plus ici d'assurer la gestion de politiques ou de programmes, mais de réaliser des études d'évaluation et de recherche pour générer de nouvelles connaissances sur la santé qui évolue dans un écosystème complexe.

Il existe de nombreux centres de données jumelées populationnelles dont l'objectif est de maximiser l'utilisation secondaire au profit de la santé et des communautés. Les plus importants sont au Royaume-Uni, en Australie, au Canada (hors Québec), en Nouvelle-Zélande et en Allemagne et sont regroupés dans un réseau international (*International Population Data Linkage Network-IPDLn*)⁴ :

- Canada
 - Canadian Institute for Health Information (CIHI)
 - The Child and Youth Data Lab (CYDL), AB
 - Population Data BC
 - Health Services Analysis Unit, AB
 - Institute for Clinical Evaluative Sciences, ON
 - Manitoba Centre for Health Policy
- Royaume-Uni
 - Avon Longitudinal Study of Parents and Children (ALSPAC)
 - Oxford Record Linkage Group (ORLS)
 - Information Services Division, Scotland
 - SAIL Databank
 - UK Biobank
- Australie
 - Australian Institute of Health and Welfare
 - Australian Bureau of Statistics
 - Centre for Health Record Linkage
 - Health LinQ

- SA-NT DataLink
- Western Australia Data Linkage Branch
- Population Health Research Network
- The Centre for Data Linkage
- The Tasmanian Data Linkage Unit (TDLU)
- Nouvelle Zélande
 - Department of Population Health, University of Otago, Christchurch
- Allemagne
 - German Record Linkage Center

La communauté scientifique reconnaît parmi les centres mondiaux, un certain nombre de modèles exemplaires particulièrement productifs (5) :

- Oxford Record Linkage Study (ORLS), UK
- Scottish Record Linkage System, UK
- MigMed2 Database, Sweden
- Statistics Norway
- Rochester Epidemiology Project, Minnesota, USA
- Manitoba Center for Health Policy (MCHP), Canada
- Center for Health Services and Policy Research (CHSPR), Canada
- Institute for Clinical and Evaluative Sciences (ICES), Canada
- Public Health Research Network (PHRN), Australia
Western Australia Data Linkage System (WADLS), Australia

Le modèle de développement des banques de données intégrées en santé est déterminant pour le succès du projet en ce qui concerne la richesse des données et la maximisation de leur utilisation et donc de leur impact (27). Les projets qui sont sous la responsabilité soit de prestataires de services (p. ex. hôpitaux, centres de services, etc.) ou de celle des payeurs (p. ex. assurances publiques ou privées) font généralement face aux mêmes grandes difficultés. Ils tendent à favoriser les intérêts des promoteurs du projet dans le choix des données à intégrer et limitent le partage et l'utilisation des données, parfois sur une base discrétionnaire. De plus, les méthodes et les analyses sont susceptibles de ne pas être entièrement partagées comme elles le devraient selon les pratiques scientifiques reconnues. Ces caractéristiques

⁴ Source: <http://www.ipdln.org/>, consulté le 23 février 2015.

engendrent des réserves quant aux résultats et nuisent à leur adoption pour améliorer les systèmes de santé.

Une troisième option consiste à créer un centre de données indépendant et commandité par les autorités publiques. Cette approche a permis de créer des banques de données intégrées populationnelles fiables, méthodologiquement rigoureuses et coûts efficaces (16,27,28). Faire intervenir une tierce partie (le centre de données indépendant) permet de positionner les discussions sur les éléments au cœur de la réussite du centre et de faciliter l'émergence de compromis sur l'accès aux données, leur utilisation et le contrôle de la confidentialité. Aborder la création d'une banque de données intégrée sous l'angle de la propriété des données et des privilèges qui s'y rattachent serait nuisible.

La gouvernance

Le centre de données indépendant a donc la mission de faciliter l'accès aux données et d'en promouvoir l'utilisation afin de maximiser la production de connaissances dans l'intérêt public (21,16,27,29-32). La vision capable de porter cette mission repose sur l'établissement et le maintien d'un partenariat entre les acteurs et sur des valeurs de qualité, d'efficacité, d'efficience et de satisfaction.

Une gouvernance partagée entre les acteurs - les détenteurs de données, les utilisateurs, le public, les prestataires institutionnels et professionnels du système de santé - est un élément central pour la réussite d'un tel projet, car elle permet de construire une confiance mutuelle entre toutes les parties. Ce partenariat est aussi essentiel à la définition des grandes orientations permettant d'assurer le maintien de l'efficience et de la pertinence sociale et politique du centre, régulièrement validées par des évaluations d'activités et d'impacts. L'autorité de gouvernance doit aussi veiller à l'encadrement des activités par des politiques et des pratiques explicitement documentées, approuvées et transparentes. Le respect des lois et des règlements d'une telle structure, y compris des processus reliés en amont et en aval, doivent rester sous la supervision de l'organisme indépendant chargé de veiller à la protection des renseignements personnels qui au Québec est la CAI.

L'objectif opérationnel du centre de données intégrées est la fluidification des échanges et la réduction des coûts de transaction au profit de l'utilisation des données pour l'amélioration du système de santé et du bien-être de la population. Cela dépasse néanmoins la gestion mécanique des échanges, basée sur les technologies de l'information. La maximisation des retombées repose sur la création d'une banque de données à valeur ajoutée requérant un haut niveau d'expertise concernant les processus de jumelage (déterministe et probabiliste), d'harmonisation des données de sources différentes, de maintien de la cohérence longitudinale malgré l'évolution des banques de données sources, d'évaluation de la qualité, de désidentification et d'évaluation des risques de réidentification ainsi que de bonification des données par l'intégration d'indicateurs et de scores composites. C'est ainsi que le centre acquiert un haut degré de crédibilité, car il peut soutenir les utilisateurs par sa connaissance étendue des caractéristiques, des forces et des limites des données.

La fonction de transfert des connaissances entre les recherches effectuées grâce au centre de données et les décideurs ou les services publics doit aussi faire partie des activités du centre pour en maximiser l'impact et demeurer pertinent. Dans le même esprit, le centre doit penser à intégrer dans les données accessibles, les résultats des travaux qui pourraient être bénéfiques aux autres chercheurs (p. ex. ajout d'indicateurs ou offre de calculs en prétraitements).

Un centre de données intégrées en santé doit être une structure dynamique qui intègre une culture du changement susceptible de capter la progression des besoins, la transformation du milieu de la santé et l'avancement des connaissances. Il doit régulièrement intégrer de nouvelles sources de données, des innovations méthodologiques et les évolutions technologiques.

La protection des renseignements personnels

Le respect de la confidentialité et la protection des renseignements personnels resteront les points les plus sensibles de l'utilisation secondaire des données. Or, la création d'un centre de données peut permettre de réduire les risques de la violation de confidentialité.

Le premier point consiste à dénominaliser les renseignements personnels (p. ex., nom et prénom, numéro d'assurance maladie, numéro de dossier, etc.) par un identifiant anonyme unique qui ne sera utilisé que par le centre de données pour jumeler les données concernant un même individu. Ce sont les organismes détenteurs des banques de données sources qui conservent la clé de correspondance entre les données nominatives et les identifiants anonymes. De plus, l'identifiant anonyme sera à nouveau crypté avant de le transmettre à l'utilisateur afin qu'il soit impossible de combiner les données obtenues pour des projets différents.

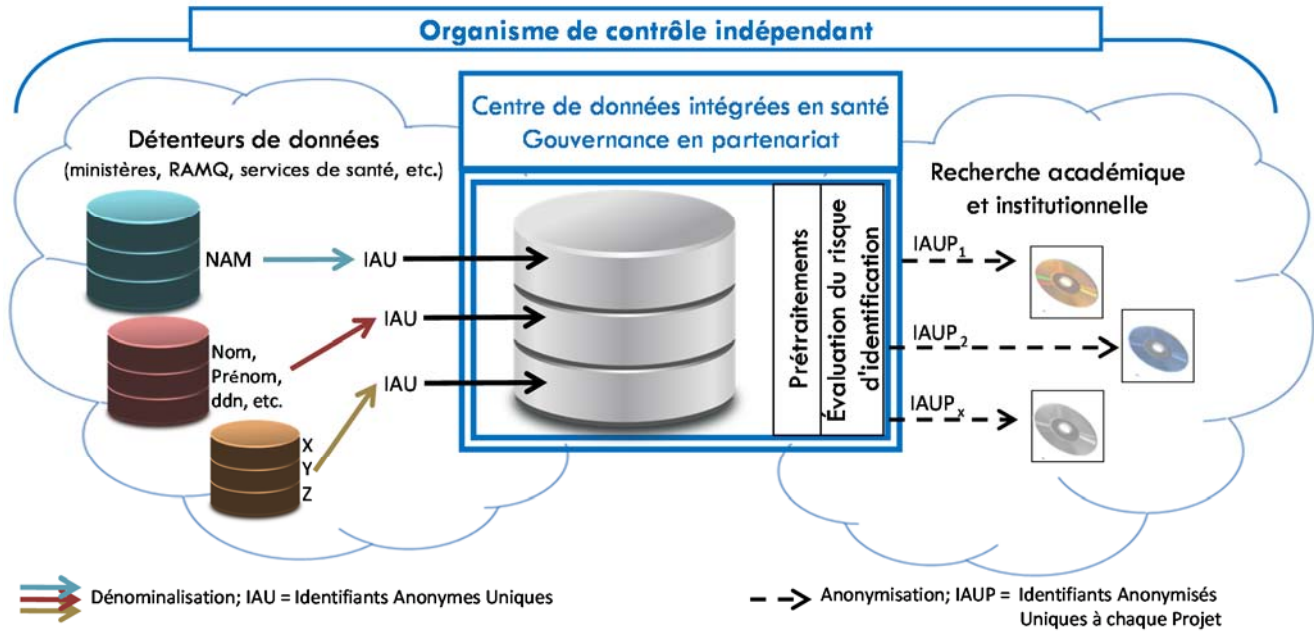
Le deuxième point est l'utilisation de l'expertise du centre de données pour minimiser la transmission de données qui, combinées ensemble, permettraient d'identifier des individus sans même avoir de données nominatives. Ces combinaisons sont aussi qualifiées de renseignements personnels. Il est très fréquent que des données (dates de naissances ou de décès, codes postaux, pathologies, etc.) soient requises par les chercheurs non pas pour leur valeur intrinsèque, mais pour en dériver une autre information comme l'âge à un certain événement, le délai entre deux interventions, l'indice de défavorisation du milieu de vie, l'état de santé général d'une personne, etc. Si le centre de données possède l'expertise pour réaliser toutes ces tâches, voire les automatiser, la plupart des renseignements personnels n'ont plus à être transmis. La capacité de prétraitement du centre offre donc un accès élargi avec une réduction des risques.

Les progrès récents apportés par les algorithmes de désidentification et d'évaluation des risques de réidentification sont toutefois l'élément le plus important. Ils permettent d'objectiver l'évaluation du risque et de standardiser les procédures, ce que les pratiques actuelles d'évaluation des demandes d'information par un analyste ne peuvent faire. Le projet cd-link

(<http://www.ices.on.ca/Research/Research-programs/Cancer/cd-link>) en Ontario a adopté avec succès cette technologie pour faciliter l'utilisation des données de santé de l'ICES (Institute for Clinical Evaluative Sciences) concernant les personnes atteintes de cancer (33). L'ex-commissaire à la vie privée de l'Ontario a étudié cette approche pour conclure que c'est un mécanisme efficace de protection des renseignements personnels (34).

Les données ne sont pas transmises tant que le risque de réidentification n'a pas atteint un seuil prédéterminé. Plus le seuil est bas, moins il y a de risque de réidentification, car plus il y a d'observations qui présentent des caractéristiques similaires dans le fichier fourni, par exemple le seuil est de 0,33 si trois personnes ont des caractéristiques identiques et de 0,2 s'il y en a cinq. Les techniques mentionnées plus haut sont alors utilisées par le centre de données pour atteindre le seuil visé: transformation, réduction de la précision, prétraitement, etc. Le seuil requis peut éventuellement être défini selon les mesures de sécurité appliquées dans l'environnement où les données seront utilisées. Les seuils de risque et les exigences de sécurité requises pour obtenir les données doivent être déterminés par l'organisme de supervision de l'application des lois et règlements qui au Québec est la CAI.

Le centre peut donc travailler directement avec les requérants pour se conformer au seuil de risque fixé par la CAI tout en comblant les besoins de l'utilisation secondaire. Par contre, les niveaux de risque acceptables et les exigences de sécurité qui y sont associées doivent être établis et contrôlés par la CAI. Les normes de sécurité étant préétablies, le requérant connaîtra a priori les normes à respecter pour être autorisé à recevoir et utiliser des données. Un système d'inspection et d'accréditation par la CAI devrait alors être mis en place.



Représentation schématique d'un centre de données intégrées en santé

Conclusion : un développement basé sur le leadership et la stabilité

Les modèles exemplaires de centres de données partagées en santé reposent sur un modèle opérationnel dédié à l'utilisation secondaire des données dans le but de maximiser les bénéfices pour la santé de la population. Permettre l'accès aux données par un cadre législatif et réglementaire est loin d'être suffisant.

Comme tout partenariat entre des acteurs provenant d'univers très différents, un centre de données intégrées devra les réunir autour de l'objectif commun: maximiser l'utilisation secondaire des données de santé pour le mieux-être de la population. Le deuxième facteur de réussite semble être de prendre le temps qu'il faut pour construire un centre par étape en commençant par un noyau central composé des données démographiques et de consommation des soins de santé, sans pour autant perdre de vue qu'il faut viser l'exhaustivité et la globalité des données reliées à la santé. Par étape signifie aussi de procéder par un projet pilote qui permet à toutes les parties de déployer, de tester et d'ajuster les pratiques entourant le fonctionnement d'une telle structure. Dans cette configuration initiale, l'accès aux données est aussi restreint à un nombre limité d'utilisateurs.

Prendre le temps comporte néanmoins un risque important. Celui d'échouer par perte d'intérêt qui peut éventuellement se traduire par une perte de financement. La création d'un centre de données doit aussi miser sur la mobilisation des acteurs et doit être confiée à des partenaires qui font preuve de leadership afin de faire émerger les solutions aux divers problèmes qui se manifesteront forcément et maintenir le dynamisme et l'intérêt dans la durée. Enfin, la mobilisation est aussi tributaire d'un engagement ferme avec un financement stable et prévisible.

Références

1. The King's Fund. Broader determinants of health [Internet]. [cited 2014 Feb 10]. Available from: <http://www.kingsfund.org.uk/time-to-think-differently/trends/broader-determinants-health>
2. Marmot M, Allen JJ. Social determinants of health equity. *Am J Public Health*. 2014 Sep;104(Suppl 4):S517–9.
3. Institute of Medicine. Best care at lower cost: the path to continuously learning health care in America. Smith M, Saunders R, Stuckhardt L, McGinnis JM, editors. Washington, DC, USA: National Academies Press (NAP); 2013.
4. International Population Data Linkage Network. Data linkage centres [Internet]. [cited 2015 Feb 10]. Available from: <http://www.ipdln.org/data-linkage-centres>
5. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011 Jan;32:91–108.
6. Canadian Academy of Health Sciences. Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research. In: Canadian Academy of Health Sciences (CAHS), editor. Return on Investment in Health Research. Ottawa, Canada; 2009. p. 331.
7. Holman CDJ, Bass AJ, Rouse IL, Hobbs MST. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health*. 1999 Oct;23(5):453–9.
8. Holman CDJ, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, et coll. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev*. 2008 Nov;32(4):766–77.
9. Goldacre MJ, Griffith M, Gill L, Mackintosh A. In-hospital deaths as fraction of all deaths within 30 days of hospital admission for surgery: analysis of routine statistics. *BMJ*. 2002 May 4;324(7345):1069–70.
10. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. *J Clin Epidemiol*. 2000 Dec;53(12):1258–67.

11. Brook EL, Rosman DL, Holman CDJ. Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System. *Aust N Z J Public Health*. 2008 Feb;32(1):19–23.
12. Warren JL. Using SEER-Medicare Data For Cancer Survivorship Research. 2012. p. 33.
13. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care*. 2002 Aug;40(8 Suppl):IV – 3–18.
14. Mekel M, Shortt SED. Coming of age and taking stock: the state of academic health policy research centres in Canada. *Healthc Policy*. 2005 Sep 15;1(1):140–50.
15. Ji J, Hemminki K, Sundquist J, Sundquist K. Ethnic differences in incidence of type 1 diabetes among second-generation immigrants and adoptees from abroad. *J Clin Endocrinol Metab*. 2010 Feb;95(2):847–50.
16. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. *Health Serv Res*. 2010 Oct;45(5 Pt 2):1468–88.
17. Groves P, Kayyali B, Knott D, Van Kuiken S. The “ big data ” revolution in healthcare -- Accelerating value and innovation. Center for US Health System Reform, McKinsey&Company, editors. 2013.
18. Assemblée nationale du Québec. Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels. Quebec City, Canada: Province du Québec; 1982.
19. The Academy of Medical Science. Personal data for public good: using health information in medical research. London, United Kingdom; 2006 p. 77.
20. Davies C, Collins R. Balancing potential risks and benefits of using confidential data. *BMJ*. 2006 Aug 12;333(7563):349–51.
21. Souhami R. Governance of research that uses identifiable personal data. *BMJ*. 2006 Aug 12;333(7563):315–6.
22. Mathers N, Watt G, Perrin N. Towards Consensus on Best Practice: Use of patient records from general practice for research. London, UK; 2009 p. 35.
23. Stanley FJ, Meslin EM. Australia needs a better system for health care evaluation. *Med J Aust*. 2007 Mar 5;186(5):220–1.
24. Kalra D, Gertz R, Singleton P, Inskip HM. Confidentiality of personal health information used for research. *BMJ*. 2006 Jul 22;333(7560):196–8.
25. Adams C, Allen J. Government databases and public health research: facilitating access in the public interest. *J Law Med*. 2014 Jun;21(4):957–72.
26. Willison DJ, Steeves V, Charles C, Schwartz L, Ranford J, Agarwal G, et coll. Consent for use of personal information for health research: do people with potentially stigmatizing health conditions and the general public differ in their opinions? *BMC Med Ethics*. 2009 Jan;10:10.
27. Love DE, Paita LM, Custer WS. Data sharing and dissemination strategies for fostering competition in health care. *Health Serv Res*. 2001 Apr;36(1 Pt 2):277–90.
28. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et coll. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc*. 2007;14(1):1–9.
29. Meyer A-M, Olshan AF, Green L, Meyer A, Wheeler SB, Basch E, et coll. Big data for population-based cancer research: the integrated cancer information and surveillance system. *N C Med J*. 2014;75(4):265–9.
30. Boyd JH, Ferrante AM, O'Keefe CM, Bass AJ, Randall SM, Semmens JB. Data linkage infrastructure for cross-jurisdictional health-related research in Australia. *BMC Health Serv Res*. *BMC Health Services Research*; 2012 Jan;12(1):480.
31. Williams PAH. Information security governance: a risk assessment approach to health information systems protection. *Stud Health Technol Inform*. 2013 Jan;193:186–206.
32. Hovenga EJS, Grain H. Health data and data governance. *Stud Health Technol Inform*. 2013 Jan;193:67–92.
33. Earle CC. Ontario Cancer Data Linkage Project: “cd-link”. *Healthc Q*. 2014 Jan;17(1):11–3.
34. Cavoukian A, Castro D. Big Data and Innovation , Setting the Record Straight: De-identification Does Work. Toronto, Canada: Government of Ontario; Canada; 2014 p. 13.
35. Blais C, Jean S, Sirois C, Rochette L, Plante C, Larocque I, et coll. Quebec Integrated Chronic Disease Surveillance System (QICDSS), an innovative approach. *Chronic Dis Inj Can*. 2014 Nov;34(4):226–35.

Constitution et utilisation des banques de données intégrées en santé : tour d'horizon international des modèles et des pratiques exemplaires

AUTEUR

Bernard Candas
Bureau d'information et d'études en santé des populations

REMERCIEMENTS

L'auteur souhaite remercier le Dr Denis A. Roy, vice-président scientifique à l'Institut national de santé publique du Québec, pour le partage de ses réflexions sur les enjeux de l'utilisation secondaire des données de santé au Québec qui ont contribué à enrichir ce tour d'horizon international des modèles et des pratiques exemplaires.

L'auteur souhaite aussi remercier Me François Côté, conseiller juridique de l'Institut national de santé publique du Québec, pour sa relecture attentive de ce document qui a permis de préciser l'environnement juridique entourant l'utilisation secondaire des données de santé au Québec.

AVERTISSEMENT

Ce document est produit afin de contribuer aux discussions concernant l'utilisation secondaire des données de santé dans le cadre du développement des connaissances. Ce document ne représente pas nécessairement l'opinion ou la position de l'Institut national de santé publique du Québec sur cette question et ne peut être présenté comme tel.

Mise en page

Sylvie Muller
Bureau d'information et d'études en santé des populations



Ce document est disponible intégralement en format électronique (PDF) sur le site Web de l'Institut national de santé publique du Québec au : <http://www.inspq.qc.ca>.

Les reproductions à des fins d'étude privée ou de recherche sont autorisées en vertu de l'article 29 de la Loi sur le droit d'auteur. Toute autre utilisation doit faire l'objet d'une autorisation du gouvernement du Québec qui détient les droits exclusifs de propriété intellectuelle sur ce document. Cette autorisation peut être obtenue en formulant une demande au guichet central du Service de la gestion des droits d'auteur des Publications du Québec à l'aide d'un formulaire en ligne accessible à l'adresse suivante : <http://www.droitauteur.gouv.qc.ca/autorisation.php>, ou en écrivant un courriel à : droit.auteur@cspq.gouv.qc.ca.

Les données contenues dans le document peuvent être citées, à condition d'en mentionner la source.

Dépôt légal – 4^e trimestre 2016
Bibliothèque et Archives nationales du Québec
ISBN : 978-2-550-77001-5 (PDF)

© Gouvernement du Québec (2016)

N° de publication : 2188