

Revue de littérature pour le plan d'action CEDAL

2006

Le développement des infrastructures permettant l'appariement et l'utilisation des données administratives aux fins de recherches a permis de traiter une multitude de questions en rapport avec le secteur de la santé. Les perspectives offertes par de ces infrastructures ne se limitent pas, en référence à l'approche des parcours de vie (Groupe de travail/BDAL 2006) à la « recherche universitaire, mais aussi à celle des missions de surveillance et de gestion des organismes publics »¹.

Chez les chercheurs, en santé des populations, les bases de données sont de plus en plus utilisées. En effet, les informations regroupées dans les bases de données informatisées permettent d'étudier une panoplie de sujets et de relations liés à la santé. Plusieurs études relatent des faits intéressants reliés à l'utilisation des bases de données. Dans le domaine de la surveillance, par exemple, l'appariement des données permet d'évaluer l'impact et l'efficacité d'un nouveau vaccin (Andrews 2002).

L'utilisation grandissante du jumelage de bases de données dans le domaine de la recherche en santé s'explique fondamentalement par son potentiel d'augmenter la capacité et les possibilités de recherche, (Shanahan et al. 2002) mais aussi par le rôle important que l'appariement de données peut jouer pour éclairer (développement et analyse) les politiques publiques en santé et le processus décisionnel (Roos 2002).

Les projets de développement de telles infrastructures, au Canada et en Australie pour ne citer qu'eux, traduit cette communauté d'intérêts entre les besoins des chercheurs et les décideurs d'utiliser l'appariement des données et trouve sa concrétisation dans un partenariat réel. Ces projets ont pour caractéristiques d'être hébergés par des centres de recherche qui ont pour mission de contribuer à améliorer la santé des populations en réalisant des travaux qui s'inscrivent dans les préoccupations en matière de politique de la santé et qui travaillent

¹ Groupe de travail/BDAL, 2006.

« étroitement avec les décideurs pour transformer activement les résultats de recherche en options stratégiques. »²

Ainsi, le Manitoba Centre for Health Policy (MCHP) qui est le pionnier dans le domaine au Canada, unité de recherche universitaire du Department of Community Health Sciences de l'Université du Manitoba, en plus des travaux effectués par ses chercheurs, soumet dans le cadre d'un contrat avec le ministère de la Santé, chaque année, cinq rapports dont les sujets sont choisis au cours de discussions avec les décideurs (Verena). Le partenariat avec les décideurs ne se limite pas au ministère de la Santé. En effet, dans cet univers du développement d'infrastructure d'appariement de données aux fins de recherche, l'expérience du Manitoba constitue un cas significatif et innovateur (Shanahan et al. 2002). Sa création a pris la forme d'un projet pilote, dont les parties prenantes étaient Statistique Canada et le Manitoba Centre for Health Policy and Evaluation, qui visait à évaluer la faisabilité de jumeler les trois sources de données ce qui a permis de créer une nouvelle base de données (Houle et al. 1996).

Cette logique de partenariat se retrouve jusqu'à dans les conditions de succès (Roos et al. 1996) d'un projet de jumelage de données et qui sont : « 1) les données sont appariées au niveau individuel ; 2) l'appariement offre la capacité de réaliser des études de cohorte ou territoriale ; 3) de permettre de répondre aux préoccupations de politiques publiques à différents niveaux ; 4) inclus des données autant sur l'état de santé de la population, socioéconomiques et sur l'organisation et l'utilisation des services de santé »³.

Pour confirmer, si besoin est, la capacité de « rapprocher tous les joueurs du milieu » qu'a un centre de jumelage de données administratives, rappelons la tenue récente d'un symposium⁴ sur les avantages de l'appariement qui a réuni plus d'une centaine de chercheurs, de décideurs en matière de politiques publiques et de praticiens.

² Black et al., 2005.

³ Roos et al., 1996.

⁴ Data Symposium- Link and Multiply: the Benefits of Data Linkage, 27 juillet 2006, Sydney.

D'intérêt public, ces infrastructures se trouvent à la convergence de deux des points majeurs de la recherche en santé ainsi que sur les soins de santé dispensés et qui sont l'utilisation de données administratives longitudinales et l'utilisation des techniques de jumelage (Roos 2002). L'importance de ces deux points, implique la nécessité d'un processus continu de collecte de données longitudinales afin qu'elles puissent être jumelées à d'autres données et ainsi supporter de nouveaux types de recherche (Roos 2002). En mettant sur pied un dictionnaire conceptuel (Concept Dictionary), consulté par un nombre grandissant de centres de recherche pour alimenter et faciliter leurs travaux, afin de rendre le plus de données possibles disponibles, le Manitoba Centre for Health Policy a favorisé l'émergence de nouveaux projets (Roos et al. 2001). L'expérience du Manitoba constitue la principale source d'enseignement pour la mise en œuvre de projets de jumelage des bases de données. Que ce soit à travers la consultation de son dictionnaire conceptuel, les résultats obtenus⁵ ou le leadership de ces chercheurs. Pour preuve, en Colombie-Britannique en Australie, particulièrement en ce qui concerne les problématiques d'accès aux ressources de santé, d'équité, d'utilisation et d'amélioration du système de soins de santé, et ce, à l'aide des bases de données (Shanahan et al. 2002).

Élaborer un projet d'appariement de données pour l'ensemble de la population est une activité complexe qui nécessite une vision et un investissement à long terme (Shanahan et al. 2002) et un environnement favorable à la collaboration des parties prenantes (Holman 1999). Si peu d'information est disponible sur les conditions d'émergence et de réalisation de ce type de projet, on retiendra néanmoins que les principes qui président à la réalisation d'un projet de cette nature peuvent se résumer comme suit : planification, pragmatisme et progressivité. L'efficacité de ce type d'infrastructure est aussi en lien avec la disponibilité d'un certain nombre d'outils comme le dictionnaire conceptuel (Concept Dictionary) et les logiciels d'appariement (Roos et al. 2003), ou les inventaires des bases de données (Black et al. 2005) et de manière générale l'importance des outils dédiés aux chercheurs dans un environnement de recherche.

Ces outils, lorsque supportés par des procédures transparentes (définition de la mesure des variables et description des étapes de la recherche), favorisent la réutilisation des données par les chercheurs, leur évaluation et leur réplique (Roos et al. 2003). De plus, le développement de ces

⁵ Houle et al., 1996.

outils permet aux chercheurs de poser de nouvelles questions ou d'approcher de vieilles questions plus efficacement (Dyson 1998).

Reste que le potentiel de ce type de projet est conditionné par la disponibilité de « données valides et fidèles à la réalité tout en ayant la possibilité de jumeler des données entre elles »⁶ ces conditions renvoient au contexte et aux finalités de la collecte et de l'entreposage des données qui ne sont pas de manière « explicite d'appuyer le milieu de la recherche »⁷. S'assurer de disposer des données qui permettent aux chercheurs de comprendre les effets des politiques publiques sur la population soulève des enjeux, règles du jeu dans l'univers des bases de données administratives et politiques de conservations, de divulgations et d'utilisations des données, qui concernent autant les chercheurs que les collecteurs/dépositaires de ces données (Black et al. 2005).

La principale limite recensée à ce potentiel est le respect de la confidentialité et la protection des renseignements personnels (Shanahan et al. 2002; Roos et al. 2003) sachant que les questions d'éthique qui entourent l'accès et l'utilisation des données aux fins de recherches sont beaucoup plus complexes dans le cas de données jumelées que ceux qui concernent l'accès à une source unique (Chamberlayne et al. 1998). Le degré de limitation varie aussi selon les normes et les principes légaux et éthiques d'un pays à l'autre ou d'une province à l'autre quant à l'utilisation des données (Anderson 1998). Si l'anonymisation des données est un processus généralement utilisé pour respecter cette limite, le respect de la confidentialité et la protection des renseignements personnels pose la question, parmi tant d'autres questions, de la nature et de la complexité de l'identifiant utilisé pour réaliser le jumelage (Rumeau-Rouquette 1997; Chamberlayne et al. 1998; Weil et al. 1999; Shanahan et al. 2002; IRIS-Q 2003). Notons que cette question de l'identifiant a aussi un impact important sur les taux de jumelage obtenus. Les paramètres qui influent également ce taux sont, notamment, la méthode employée, généralement la méthode probabiliste et le degré de cohérence entre les sources de données (Roos et al. 1986).

⁶ Shanahan et al., 2002.

⁷ Black et al., 2005.

Notons que le jumelage peut aussi être considéré comme un moyen pour évaluer la qualité des données obtenues pour un même individu selon le niveau d'agrégation des données jumelées (Roos 2002). Pour accroître constamment le niveau de jumelage, la connaissance des facteurs d'erreurs dans l'appariement de données de sources multiples, et à contrario l'amélioration des techniques, sont l'objet de préoccupation constante à l'occasion de colloques (Symposium on Health Data Linkage 2002; Data Linkage Symposium 2006) ou de publications de chercheurs (Holman 1999; Christen 2002; Rosman 2002).

De ce qui précède et des connaissances acquises par l'équipe de l'axe BD, il ressort que l'objectif qui vise à améliorer l'accès aux données administratives et leur utilisation par les chercheurs travaillant dans le milieu universitaire ou dans le secteur public, ce tant à des fins cliniques, administratives, de santé publique que pour le développement des connaissances théoriques sur les déterminants de la santé est conditionné dans une large mesure par la dynamique des acteurs et ce à plusieurs niveaux.

De l'expérience canadienne et australienne, le leadership des chercheurs apparaît le véritable moteur du développement des projets d'infrastructures. Comme il a été déjà signalé, la terminologie de chercheurs utilisée recouvre un univers complexe, aux besoins variés, intéressé à un degré ou un autre par l'accès aux données administratives et par le potentiel de leurs appariements pour une utilisation « tant à des fins cliniques, administratives, de santé publique que pour le développement des connaissances théoriques sur les déterminants de la santé. »⁸ Dans cette perspective le partage des connaissances est déterminant sur le plan méthodologique et technique avec les chercheurs d'ici et d'ailleurs. La logique du partage des connaissances se retrouve dans la composition même des structures des centres d'appariement canadiens puisqu'on y retrouve des chercheurs et des centres associés et affiliés d'autres provinces. Nous remarquons également que cette dynamique de partage des connaissances est très active entre les chercheurs des provinces de l'ouest canadien, des pays scandinaves, de la Grande-Bretagne, de la Nouvelle-Zélande et de l'Australie. Ce dernier abritant régulièrement des rencontres scientifiques dont l'objectif est de faire le point sur diverses problématiques touchant l'appariement des données et les résultats de la recherche en santé.

⁸ Groupe de travail/BDAL, 2006.

Ce leadership s'avère productif dans le cadre d'un partenariat stratégique avec les décideurs en matière de politiques publiques, généralement le ministère de la Santé et les gestionnaires du réseau de la santé, mais aussi avec les responsables des aspects légaux et éthiques. Nous supposons qu'un des facteurs facilitant le partenariat avec les décideurs, est que les centres universitaires qui abritent les infrastructures d'appariement des données sont spécialisés dans la recherche en politique publique et contribuent donc au parcours de la décision éclairée. À la base de ce partenariat dans le cas du BCLHD et du MCHP une convergence d'intérêt en lien avec la finalité d'améliorer la santé des populations. À cet égard, les centres d'appariement des données sont dédiés aux projets de recherche « d'intérêt public d'intérêt public et respectant les politiques de protection de la vie privée ». Si nous connaissons encore mal les enjeux et les défis de la construction de ces partenariats et des conditions de leur fonctionnement, nous apprenons cependant que le ministère est non seulement parti prenante du processus de décision concernant les requêtes des chercheurs, il contrôle également les résultats de la recherche pour veiller au respect de la condition « d'intérêt public ». De la même manière, les responsables des aspects légaux et éthiques veillent au respect de la confidentialité et de la protection des renseignements personnels.

Le pré requis au fonctionnement de ces centres d'appariement est la qualité des données qui se trouvent être au cœur d'une autre dynamique d'acteur, celle qui lie les chercheurs aux collecteurs et aux dépositaires de ces données. La finalité de leur collecte, de leur entreposage et de leur utilisation n'étant pas forcément identique le rapprochement entre ces acteurs devient déterminant. À la préoccupation des chercheurs d'un accès facilité au moindre coût à des données de qualité, collecteurs et dépositaires font valoir les limites de leur mandat et des ressources nécessaires pour satisfaire les demandes de chercheurs.

Enfin, les centres offrent généralement des services de soutien aux requêtes des chercheurs et de préparation des données grâce à la présence d'une variété de ressources (analystes, gestionnaires, informaticiens, documentalistes, professionnels de recherche) d'une part, et en assurant l'entreposage des données traitées dans des conditions qui en garantissent la sécurité et la

protection des renseignements personnels il évite de répéter les coûts primaires de collecte de données et facilite le développement continu de l'expertise (jumelage et extraction des données).

Projet BDAL

- Anderson, R. (1998). "The DeCODE proposal for an Icelandic Health Database." 1-12.
- Andrews, N. (2002). The value of linked data for research into surveillance and adverse events. Symposium on Health Data Linkage, University of Adelaide, Public Health Information Development Unit (PHIDU).
- Black, C. et al.(2005). Data, Data, Everywhere...: Improving acces to population health and health services research data in Canada. Centre for Health Services and Policy Research. Vancouver. 148 p.
- Chamberlayne, R. et al. (1998). "Creating a Population-based Linked Health Database: A new Ressource for Health Services research." 89, 4: Health Module: 270-273.
- Christen, P. (2002). Parallel computing techniques for high-performance probabilistic record linkage. Symposium on Health Data Linkage, University of Adelaide, Public Health Information Development Unit (PHIDU).
- Data Linkage Symposium (2006). Link and Multiply : the Benefits of Data Linkage, Sydney, À paraître.
- Dyson, F. (1998). Imagined Worlds. Cambridge, Harvard University Press.
- Groupe de travail/BDAL.(2006). Les bases de données administratives longitudinales en santé des populations : vers un accès et une utilisation facilités. Réseau de recherche en santé des population du Québec- Axe longitudinal et axe bases de données. 36 p.
- Holman, A. (1999). "Population-based linkage of health records in Western Australia : development of a health services reseach linked database." 23, 5: 453-459.
- Houle, C. et al. (1996). "Le projet d'appariement du recensement et des fichiers de soins de santé au Manitoba : Composante des ménages privés." No. 91.
- IRIS-Q.(2003). Groupe de travail sur l'accès aux données : Proposition d'un cadre global pour l'accès aux données rendues accessibles par le biais du projet IRIS-Québec. Version finale. 23 p.
- Roos, L. L. (2002). Record linkage: Data quality, tool development, and substantive research. Symposium on Health Data Linkage, University of Adelaide, Public Health Information Development Unit (PHIDU).
- Roos, L. L. et al. (1986). "The art and science of record linkage: methods that work with few identifiers." 16(1): 45-57.
- Roos, L. L. et al. (2003). "Working More Productively : Tools for Administrative Data." 38:5: 1339-1358.
- Roos, L. L. et al. (2001). An information-rich environment: linked-record systems and data quality in Canada. Proceedings of statistics Canada Symposium 2001 : Achieving data quality in a statistical agency : a methodological perspective.
- Roos, L. L. et al. (1996). "A reserch registry : use, developpement, and accuracy." 34 (6): 39-47.

- Rosman, D. (2002). Measuring data and link quality in a dynamic multi-set linkage system. Symposium on Health Data Linkage.
- Rumeau-Rouquette, C. (1997). "Croisement des données individuelles. Méthodes de chaînage." 45: 248-256.
- Shanahan, M. et al. (2002). Linking the prairies to the outback: What can New South Wales learn from the Manitoba Data Linkage Project? Symposium on Health Data Linkage, University of Adelaide, Public Health Information Development Unit (PHIDU).
- Symposium on Health Data Linkage (2002). Its value for Australian health policy development and policy relevant research, SYDNEY, Public Health Information Development Unit (PHIDU).
- Verena, H. M. "La Revue canadienne du vieillissement." 24 (Suppl. 1): 1 - 2.
- Weil, G. et al. (1999). "Les identifiants permanents des patients en Europe: une situation contrastée. Des noms aux nombres, de la désignation à l'identification." 11(Springer-Verlag France): 95-103.