Pratiques de gouvernance d'une entité donnant accès à des données intégrées pour la recherche, l'évaluation et la prise de décision en santé et services sociaux

Rapport d'ETMI classique

Unité d'évaluation des technologies et des modes d'intervention en santé et services sociaux (UETMISSS)

Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale

Mars 2018

Québec

Dépôt légal : 2018
Bibliothèque et Archives nationales du Québec

ISBN : 978-2-550-81274-6 (imprimé)
ISBN : 978-2-550-81275-3 (PDF)

Pratiques de gouvernance d'une entité donnant accès à des données intégrées pour la recherche, l'évaluation et la prise de décision en santé et services sociaux

Rapport d'ETMI classique

Auteurs

Julie Dussault, Ph. D., professionnelle en ETMISSS
Sylvie St-Jacques, Ph. D., responsable scientifique
Unité d'ETMISSS du CIUSSS de la Capitale-Nationale

Léa Langlois, B. Sc., professionnelle de recherche
Hervé Tchala Vignon Zomahoun, Ph. D., coordonnateur
Composante « Recherche en système de santé, l'application des connaissances et mise en œuvre », Unité de soutien-SRAP du Québec

Mars 2018

Demandeur et initiateur du projet
Serge Dumont, Ph. D., directeur scientifique de l'Institut universitaire de première ligne en santé et services sociaux, et du Centre de recherche sur les soins et les services de première ligne de l'Université Laval, CIUSSS de la Capitale-Nationale. Représentant de l'Alliance santé Québec (jusqu'en décembre 2017)

Gestionnaires du mandat
Francine Blackburn, directrice adjointe de l'enseignement et des affaires universitaires, CIUSSS de la Capitale-Nationale (jusqu'en octobre 2017)

Sandra Lavigne, adjointe au directeur de l'enseignement et des affaires universitaires, CIUSSS de la Capitale-Nationale

Julie Villeneuve, directrice adjointe de l'enseignement et des affaires universitaires, CIUSSS de la Capitale-Nationale

Collaborateurs
Alain Vanasse, M.D., Ph. D., directeur de la composante « Accès aux données » de l'Unité de soutien-SRAP du Québec, centre de recherche du CHU de Sherbrooke, Université de Sherbrooke

France Légaré, M.D., Ph. D., directrice de la composante « Recherche sur les systèmes de santé et services sociaux, l'application des connaissances et la mise en œuvre » de l'Unité de soutien-SRAP du Québec, Centre de recherche sur les soins et les services de première ligne de l'Université Laval, l'Institut universitaire de première ligne en santé et services sociaux; Centre de recherche du CHU de Québec, Université Laval

Conseil de validation scientifique
Bernard Candas, Ph. D., professionnel scientifique principal - analyses populationnelles, Institut national d'excellence en santé et en services sociaux (INESSS)

Pierre Dagenais, M.D, Ph. D., Service de rhumatologie, CIUSSS de l'Estrie-CHUS-Hôpital Hôtel-Dieu

Comité de suivi
Michel Boivin, professeur titulaire, École de psychologie de l'Université Laval

Christian Chabot, citoyen partenaire, membre de la table de travail de la composante AC, CHU de Québec-Université Laval

Serge Dumont, directeur scientifique de l'Institut universitaire de première ligne en santé et services sociaux du CIUSSS de la Capitale-Nationale

Jean-Claude Forest, médecin-chercheur, CHU de Québec-Université Laval

Jean-Paul Fortin, médecin-chercheur, CIUSSS de la Capitale-Nationale

Annie LeBlanc, chercheuse, CHU de Québec-Université Laval

Guy Thibodeau, président-directeur général adjoint, CIUSSS de la Capitale-Nationale

Alain Vanasse, médecin-chercheur, CIUSSS de l'Estrie

Hervé Tchala Vignon Zomahoun, Ph. D., coordonnateur de la composante « Recherche en système de santé et services sociaux, l'application des connaissances et la mise en œuvre », Unité de soutien-SRAP du Québec

Bibliothécaires
Roxanne Lépine, bibliothécaire, composante « Recherche en système de santé et services sociaux, l'application des connaissances et la mise en œuvre », Unité de soutien-SRAP du Québec (jusqu'en mars 2017)

Nathalie Mousseau, bibliothécaire, CIUSSS de la Capitale-Nationale (jusqu'en février 2017)

Conception graphique et mise en page
Sylvie Bélanger, technicienne en administration, CIUSSS de la Capitale-Nationale

Correspondance
Sandra Lavigne, adjointe au directeur de l'enseignement et des affaires universitaires, CIUSSS de la Capitale-Nationale sandra.lavigne.ciussscn@ssss.gouv.qc.ca

Soutien financier
Cette évaluation a bénéficié d'un financement de la composante « Recherche sur les systèmes de santé et services sociaux, l'application des connaissances et la mise en œuvre » de l'Unité de soutien-SRAP du Québec

Pour citer ce document : Dussault, J., St-Jacques, S., Langlois, L., Zomahoun, H.T.V. (2018), Pratiques de gouvernance d'une entité donnant accès à des données intégrées pour la recherche, l'évaluation et la prise de décision en santé et services sociaux. Rapport d'ETMI. UETMISSS, CIUSSS de la Capitale-Nationale, 41 p.

RÉSUMÉ

Objectif

Cette ETMI a pour but d'alimenter les réflexions locales et nationales sur l'accès et la valorisation de l'information pour la recherche, l'évaluation et la prise de décision en santé et services sociaux. Plus spécifiquement, elle vise à évaluer l'efficacité des pratiques de gouvernance des entités intendantes d'accès aux données de santé et de services sociaux qui pourraient être applicables dans le contexte québécois.

Méthode

Une revue systématique de la littérature a été réalisée à partir d'une recherche dans des bases de données bibliographiques et sur internet. La sélection des documents et l'extraction des informations pertinentes ont a été réalisées par deux évaluateurs de façon indépendante, puis par consensus. La crédibilité des documents a été évaluée à partir de critères spécifiques. À partir d'un cadre d'analyse, quatre domaines de la gouvernance ont été documentés : la protection de la vie privée, la recherche, l'information et les réseaux collaboratifs intersectoriels. Des informations sur la performance des entités intendantes ont aussi été recherchées concernant les délais et les coûts reliés aux demandes d'accès aux données intégrées, le nombre de demandes traitées, le nombre de publications ainsi que le nombre d'incidents liés à la violation de la confidentialité.

Résultats

Vingt-six documents présentant des informations sur 17 entités intendantes ont été repérés. Plusieurs éléments clés de leur gouvernance permettant d'assurer l'accès à des données intégrées tout en préservant le caractère confidentiel des données personnelles ont été mis en perspectives. Bien qu'aucun des documents répertoriés n'a pour objectif de mesurer l'efficacité des entités, plusieurs présentaient des données sur les délais d'accès, le nombre de demandes traitées et le nombre de publications.

Conclusion

Plusieurs éléments du contexte québécois semblent favorables à l'implantation réussie d'une entité intendante de données intégrées. Le modèle basé sur la centralisation de données, appliqué au Royaume-Uni et ailleurs au Canada, permettrait la fluidification du processus d'accès à des données intégrées et assurerait une utilisation innovante des données québécoises de santé et des services sociaux. Selon ce modèle, l'entité intendante agit à titre de fiduciaire des données et assure la transmission sécuritaire et efficace des données intégrées vers les utilisateurs qui n'ont jamais accès aux données d'identification personnelle.

ABSTRACT

Objective

This Health Technology Assessment aims to stimulate the local and national discussions regarding the access and promotion of information intended for research, assessment and decision-making in the Health and Social Services field. More specifically, it aims to measure the effectiveness of the governance practices of the health and social services data stewards that may be suitable in the Quebec context.

Methodology

We have conducted a systematic literature review based on a search in four bibliographic databases and 85 websites using the concepts of “Data Access”, “Issues”, “Outcomes” and “Governance”. Two evaluators carried out the selection of documents and the extraction of relevant information, first independently and then by consensus. The reliability of the documents was assessed based on specific criteria. Findings were analyzed based on an analytical framework comprising four areas of governance: privacy, research, information and cross-sectoral collaborative networks. Furthermore, information on the performance of the stewardship entities was gathered.

Findings

The systematic review comprised of twenty-six documents. Information was found on 17 entities, including eight in Australia, five in the United Kingdom, three in Canada, and one in Denmark. Several key elements of the governance of stewardship entities to ensure access to integrated data while preserving confidentiality of personal data were put into perspective. Information on the performance from some stewardship entities was available for the access delays, the number of requests processed and the number of publications. However, no document was intended to measure the effectiveness of the entities. However information gathered on the performance of entities were not sufficient to ensure consistent assessment of their effectiveness.

Conclusion

Several elements of the Quebec context appear to be right for the successful implementation of an integrated data steward. The centralized data model, used in the United Kingdom and elsewhere in Canada, would streamline the process of accessing the integrated data and ensure the innovative use of Quebec’s health and social services data. Under this model, the steward acts as the trustee of the data and ensures the secure and efficient transmission of the integrated data to users who never have access to personally identifiable information.

SOMMAIRE

Pratiques de gouvernance d'une entité donnant accès à des données intégrées pour la recherche, l'évaluation et la prise de décision en santé et services sociaux

Julie Dussault, Sylvie St-Jacques, Léa Langlois et Hervé Tchala Vignon Zomahoun

Soucieuse de la disponibilité de données de qualité, l'Alliance santé Québec a pour mission d'accroître la performance en recherche et en innovation de la grande région de Québec dans le domaine de la santé et des services sociaux. Elle vise à en maximiser les retombées positives sur la santé et le mieux-être de la population, sur l'écosystème des soins de santé et des services sociaux et sur l'économie régionale[1]. Dans ce contexte, une évaluation des technologies et des modes d'intervention (ETMI) sur la gouvernance des entités intendantes donnant accès à des données intégrées a été demandée à l'unité d'ETMI en santé et services sociaux (ETMISSS) du Centre intégré universitaire de santé et de services sociaux (CIUSSS) de la Capitale-Nationale en collaboration avec la Composante « Recherche sur les systèmes de santé et services sociaux, l'application des connaissances et la mise en œuvre » de l'Unité de soutien – SRAP du Québec.

Cette ETMI a pour but d'alimenter les réflexions locales et nationales sur l'accès et la valorisation de l'information et sur les modes de gouvernance d'une entité intendante d'accès aux données de santé et de services sociaux adaptés au contexte québécois.

Éléments clés de la gouvernance des entités selon le domaine

Protection de la vie privée (p. 6)

- Principe de séparation des données d'identification et des autres données;
- Clés de couplage cryptées;
- Recours à une tierce partie pour créer les clés de couplage;
- Application de mesures d'atténuation des risques de réidentification.

Recherche (p. 12)

- Approbation éthique;
- Processus interne d'autorisation des projets;
- Autres modalités d'engagement des utilisateurs pour le respect de la vie privée (ex. : signature de formulaires de confidentialité);
- Autorisation des projets par les détenteurs de données.

Information (p. 15)

- Mesures de sécurité informatique;
- Mesures de sécurité physique;
- Examen des requêtes et sorties;
- Évaluation des processus de sécurité;
- Soutien pour la présentation des projets et des demandes d'accès;
- Coordination de la rencontre entre les détenteurs de données et les utilisateurs;
- Formation et soutien pour l'analyse des données.

Réseaux (p. 17)

- Standardisation de la collecte de données;
- Partage des données;
- Structures nationales.

GLOSSAIRE

Accès en temps opportun
Accès accordé et procuré à l'intérieur d'une période de temps raisonnable. Dans le document publié en 2015 par le Conseil des académies canadiennes, un accès accordé à l'intérieur des quatre mois suivant la soumission d'une requête est considéré comme étant en temps opportun.

Anonymisation (synonyme de dépersonnalisation)
L'anonymat implique une démarche supplémentaire qui rend absolument impossible l'accès à l'identité du participant parce que la liste associant le nom des participants et des codes a été détruite (Comité d'éthique de la recherche avec des êtres humains de l'Université Laval https://www.cerul.ulaval.ca/cms/site/cerul/page84320.html).

Base de données
Collection de données structurées pour permettre des opérations, parfois très complexes, de lecture, de suppression, de déplacement, de tri, de comparaison ou autres opérations. Lorsque plusieurs bases de données sont constituées sous forme de collection, on parle alors d'une banque de données.

Centre de données
Endroit physique où sont regroupés des équipements informatiques (ex. serveurs) et où des bases de données peuvent être hébergées. Il permet de stocker des données, de les traiter et de les protéger.

Clé de couplage
Champ de liaison créé à partir des données d'identification.

Confidentialité
Fonctions et pratiques des personnes et des organisations pour assurer que l'information n'est accessible qu'à ceux dont l'accès est autorisé.

Couplage de données (data linkage)
Processus de fusion de bases de données identifiables provenant de deux ou plusieurs sources permettant ainsi de consolider des données reliées à un individu ou à un événement lorsque ces données ne sont disponibles dans aucune autre base de données.

Détenteur de données
Organisation qui collecte et/ou détient les données et prend les décisions initiales sur leur utilisation, leur divulgation, leur conservation et leur élimination. Le dépositaire joue un rôle central en permettant ou en freinant l'accès aux données par l'implantation de politiques sur la collecte, l'utilisation et l'élimination des données. Il doit s'assurer que ses employées suivent les pratiques appropriées pour le traitement des données.

Données cliniques
Données reliées à la santé.

Enregistrements
La représentation du stockage d'une ligne de données.

Entité intendante de données
Entité responsable de l'administration des données dont l'objectif est de permettre l'accès aux données tout en assurant un équilibre avec la protection de la vie privée.

Entrepôt de données (*Data repository, Safe haven*)
Endroit physique ou électronique où l'accès aux données est sécurisé et auquel les utilisateurs autorisés peuvent accéder soit par une visite du site physique, soit par l'intermédiaire d'une connexion Internet sécurisée.

Gouvernance
Ensemble des politiques et procédures pour l'acquisition et la gestion des données. Elle permet d'assurer aux partenaires, aux utilisateurs et aux patients que des mesures appropriées ont été prises pour assurer la confidentialité et la sécurité des informations.

Jeu de données
Un jeu de données est un sous-ensemble de données associées entre elles. Il s'agit d'une agrégation d'enregistrements de données organisés pour former un ensemble cohérent.

Principe de séparation
Action de scinder la base de données originale pour créer des bases séparées pour les données propres aux participants (exemple : cliniques) et pour les données d'identification, faisant en sorte que seuls les détenteurs de données ont accès à la base originale.

Protection de la vie privée (*privacy*)
Nécessité de préserver et de protéger l'accès à une troisième partie de toute information personnelle collectée par une organisation.

Standardisation des données
Processus de développer et d'implanter des méthodes et des outils similaires pour la collecte de données afin que celles-ci soient harmonisées et comparables.

Système d'information
Ensemble des éléments participant à la gestion, au traitement et à la diffusion de l'information au sein d'une organisation.

Utilisation appropriée des données
Utilisation des données selon des standards et des exigences de légalité, de sécurité, d'efficience et d'efficacité de façon à maintenir la confiance du public.

Utilisateurs
Personnes qui utilisent les données fournies par les entités pour des fins de recherche ou d'administration.

TABLE DES MATIÈRES

LISTE DES ANNEXES

LISTE DES FIGURES

LISTE DES TABLEAUX

LE PROJET D'ETMI

1 PROBLÉMATIQUE

La problématique en lien avec le projet d'ETMI et la méthodologie utilisée pour réaliser le projet ont été publiées précédemment[2].

Les entités intendantes, plus que de simples dépositaires de données, sont des lieux d'expertises où des banques de données intégrées à valeur ajoutée sont produites en jumelant des données (de mêmes individus) provenant de sources différentes[3]. Ces données peuvent être cliniques, médico-administratives, populationnelles ou d'informations géographiques et sociales[4]. Elles offrent de nombreuses possibilités d'analyses secondaires par exemple l'étude étiologique d'une maladie, l'étude de plusieurs phénomènes à l'intérieur d'une même cohorte ou la réalisation d'études portant sur l'utilisation des services ou sur l'évolution de pathologies[5]. Dans plusieurs endroits à travers le monde, de nombreux effets positifs sur la santé des populations ont été démontrés grâce à l'accès facilité à des données couplées par les entités[3]. La gouvernance de ces entités intendantes concerne la pertinence des questions de recherche, l'identification des bases de données appropriées, le couplage des données provenant de ces bases et l'assurance de la confidentialité et du respect de la vie privée[6].

La question initiale retenue pour la réalisation de la présente ETMI est la suivante :

« *Quelles sont les meilleures pratiques de gouvernance d'une entité intendante permettant l'accès à des données intégrées pour la recherche, l'évaluation et la prise de décision en santé et services sociaux?* »

En plus des domaines de gouvernance (la protection de la vie privée, la recherche, l'information et les réseaux), la revue systématique de la littérature rapporte, dans un deuxième temps, les résultats concernant la performance des entités intendantes identifiées incluant les délais d'accès aux données, les coûts relatifs aux demandes d'accès, le volume de demandes traitées, les incidents liés à la violation de la confidentialité des données, et le nombre de publications utilisant des données intégrées.

2 MÉTHODOLOGIE

L'approche méthodologique utilisée est une revue systématique de la littérature. La stratégie de recherche documentaire a été élaborée par une bibliothécaire et validée par une seconde. Quatre concepts ont été définis et une liste de termes de vocabulaire libre et de vocabulaire contrôlé a été établie pour chacun d'eux (encadré 1). Quatre bases de données bibliographiques (Medline (Ovid), Embase (Elsevier), CINAHL (EBSCO) et *Web of Science* (Thomson Reuters)) ont été interrogées. De plus, 85 sites Internet pertinents à la problématique ont été visités, des bases de données de littérature grise (ex. : Pro Quest Dissertations & Thesis Global) et les moteurs de recherche Google et *Google Scholar* ont été utilisés pour repérer les documents de la littérature grise. Les stratégies de recherche détaillées sont disponibles sur demande. Aussi, les bibliographies des publications retenues ont été consultées afin d'identifier des études pertinentes qui n'auraient pas été repérées initialement par la stratégie de recherche documentaire.

Les critères d'inclusion et d'exclusion qui ont servi à la sélection des documents sont rappelés dans l'**annexe 1**. La sélection des documents et l'extraction des données ont été effectuées de façon indépendante par deux

personnes et les désaccords ont été résolus par consensus. Des sites Internet, des guides d'utilisateurs et politiques d'entreprises peuvent avoir été consultés pour compléter l'information extraite des documents retenus.

Encadré 1 : Concepts utilisés pour la recherche documentaire

Concept 1 : **Accès aux données** (données, bases de données, centre de données, accès aux données, données intégrées);

Concept 2 : **Enjeux** (éthiques et légaux, confidentialité, anonymisation);

Concept 3 : **Retombées** (recherche, intérêt public, qualité des soins à la population).

Concept 4 : **Gouvernance**

La crédibilité des documents retenus a été évaluée à l'aide de critères inspirés de Couture (2015)[7] concernant 1) la validation du document par une organisation reconnue ou par la compétence ou la réputation de l'auteur; 2) l'insertion de documents dans la littérature spécialisée; et 3) certains aspects en lien avec la forme du document.

2.1 Cadre théorique

Les informations portant sur la gouvernance des entités intendantes d'accès aux données intégrées ont été traitées en fonction du cadre théorique choisi préalablement[2]. Ce cadre, illustré dans la **figure 1**, est inspiré des principales constatations issues du rapport réalisé par le Conseil des académies canadiennes sur l'accès aux données sur la santé et aux données connexes au Canada[8]. Dans cette figure, l'image du balancier exprime l'idée selon laquelle l'accès aux données intégrées comporte à la fois des avantages d'intérêt public et des risques quant au respect de la vie privée. Les modalités de gouvernance doivent viser la maximisation des retombées positives de la mise à disposition des données tout en minimisant les risques d'atteinte à la vie privée.

La gouvernance permet de déployer le plein potentiel de l'exploitation des données tout en respectant la vie privée des citoyens et tout en préservant le caractère confidentiel des données personnelles[8]. Quatre domaines de gouvernance, soit la protection de la vie privée, la recherche, l'information et les réseaux collaboratifs intersectoriels sont concernés. La gouvernance doit également tenir compte, d'une part, des considérations juridiques et éthiques et, d'autre part, des défis technologiques et méthodologiques associés à la création de banques de données intégrées.

3 RÉSULTATS

3.1 Sélection des documents et identification des entités intendantes

Le processus de sélection des documents inclus dans la revue systématique est illustré à l'aide d'un diagramme de flux *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) qui est présenté dans l'**annexe 2**. Au final, 26 documents ont été retenus. Dix ont été publiés entre 1998 et 2012, et quinze documents ont été publiés entre 2013 et 2016. La date de publication n'est pas disponible pour un document[9]. Le niveau de crédibilité est élevé pour 14 documents[10-23], modéré pour 11 documents[24-34] et faible pour un seul document[9].

Les documents retenus présentent des informations sur 17 entités intendantes qui

sont présentées dans le **tableau 1**. Parmi celles-ci, l'entité *Farr North England* n'était pas encore opérationnelle au moment de la rédaction du document[20]. Huit entités sont situées en Australie, cinq au Royaume-Uni, trois au Canada et une au Danemark. À titre informatif, les données peuvent provenir de bases de données médico-administratives (ex. : données sur les hospitalisations, paiements aux médecins, données d'imagerie médicale), d'institutions (ex. : hôpitaux, écoles), de registres (ex. : décès, naissances, cancers), d'études épidémiologiques (ex. : cohortes) et de recensements. Deux des entités étudiées incluent des biobanques (BioGrid[21] et SSI[12]).

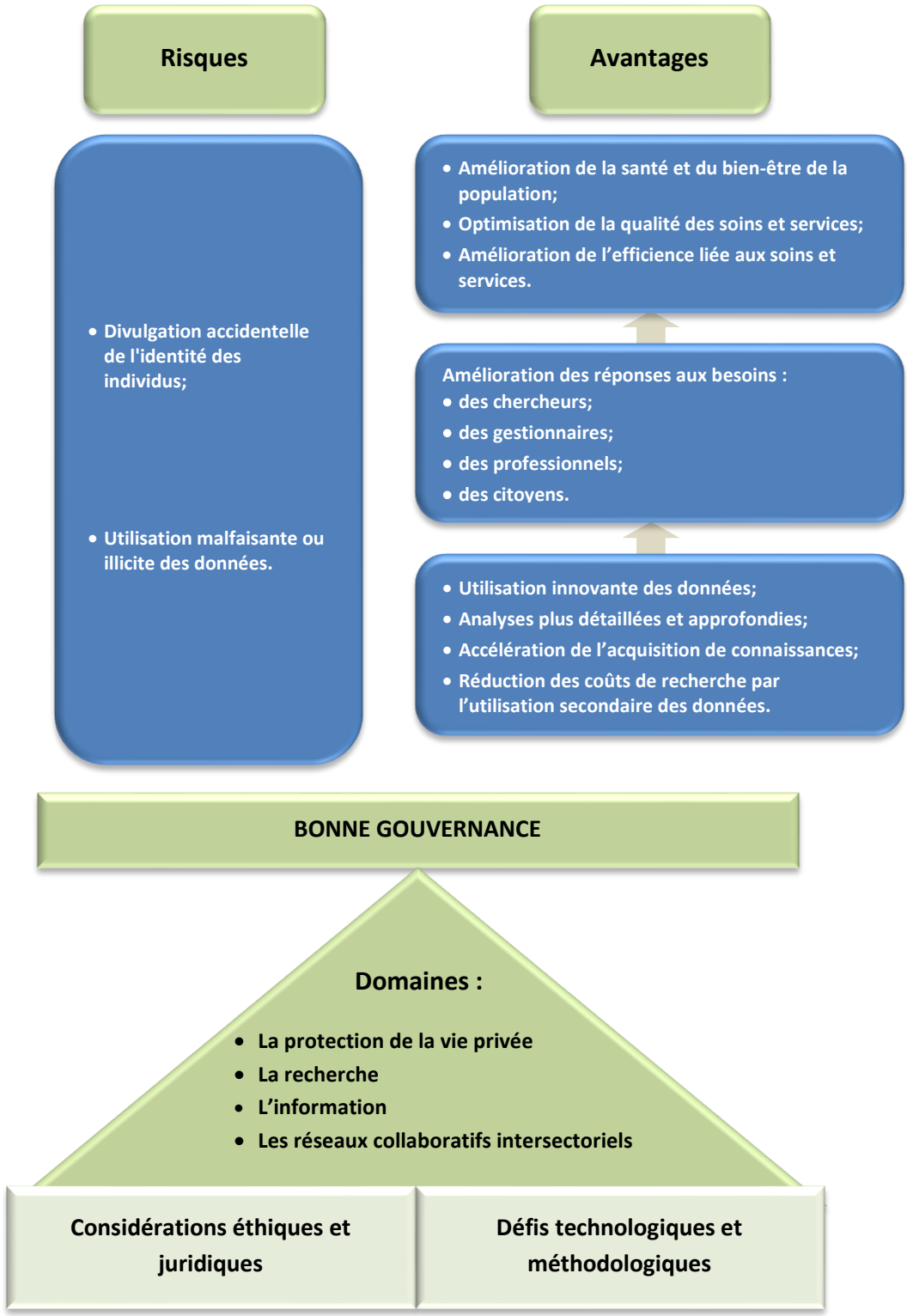
3.2 GOUVERNANCE DES ENTITÉS INTENDANTES

Les données extraites des documents retenus et portant sur la gouvernance des entités intendantes d'accès aux données intégrées sont présentées selon les domaines identifiés dans le cadre d'analyse.

Tableau 1 : Liste des entités identifiées par la recherche documentaire

Sigle ou acronyme	Entité	Portée	Juridiction territoriale
BioGrid	**Bio Grid Australia**[21]	Nationale	Australie
CDL	**Center for Data Linkage**[10]	Supra-régionale	Australie
CHeReL	**Center for health record Linkage**[24, 33]	Régionale	Nouvelle-Galles du Sud, Australie
CPRD	**The Clinical Practice Research Datalink**[16]	Nationale	Royaume-Uni
CVDL	**Victorian data linkages unit**[30]	Régionale	Victoria, Australie
FARR London	**The Farr Institute of Health Informatic Research London**[20]	Régionale	Sud-Est Angleterre, Royaume-Uni
Farr North England	**The Farr Institute of Health Informatic Research North England**[20]	Régionale	Nord de l'Angleterre, Royaume-Uni
ICES	**Institute for Clinical Evaluative Sciences**[18]	Régionale	Ontario, Canada
MCHP	**Manitoba Population Research data Repository**[23]	Régionale	Manitoba, Canada
PopData	**PopData BC**[13, 22]	Régionale	Colombie-Britannique, Canada
RLG	**Data linkage in Queensland**[27]	Régionale	Queensland, Australie
SAIL	**SAIL Databank**[15, 19, 26]	Régionale	Pays de Galles, Royaume-Uni
SA NT Datalink	**South Australia and North Territory datalink**[29, 32]	Régionale	Australie du Sud et Territoire du Nord, Australie
SILC	**The Scottish Informatic and Linkage Collaboration**[20, 28, 31]	Régionale	Écosse, Royaume-Uni
SSI	**Staten Serum Institute**[12]	Nationale	Danemark
TDLU	**Tasmanian data linkage unit**[9, 34]	Régionale	Tasmanie, Australie
WALDS	**Data linkage Western Australia**[11, 14, 17, 25]	Régionale	Australie occidentale, Australie

Figure 1 : Cadre théorique sur l'accès en temps opportun à des banques de données intégrées en santé et services sociaux



(Basé sur *The expert panel on timely access to health and social data for health research and health system innovation*, 2015)[8]

3.2.1 DOMAINE DE LA PROTECTION DE LA VIE PRIVÉE

La protection de la vie privée implique la surveillance et la réduction au minimum des risques touchant les intérêts personnels des individus et le caractère confidentiel des données. Elle concerne les processus spécifiques de protection de la vie privée au moment de permettre l'accès à des données. Ces processus visent à protéger le nom et la réputation des individus ainsi que leur droit à l'anonymat.

La synthèse des informations concernant le domaine de la protection de la vie privée est présentée dans le **tableau 2.** Selon les informations recueillies dans les documents retenus, le domaine de la protection de la vie privée inclut a) les modèles de développement des centres d'accès aux données (définis selon l'application de principes de séparation et le processus d'anonymisation) et b) les mesures d'atténuation des risques de réidentification.

a) Modèles de développement des entités donnant accès à des banques de données intégrées en santé et services sociaux

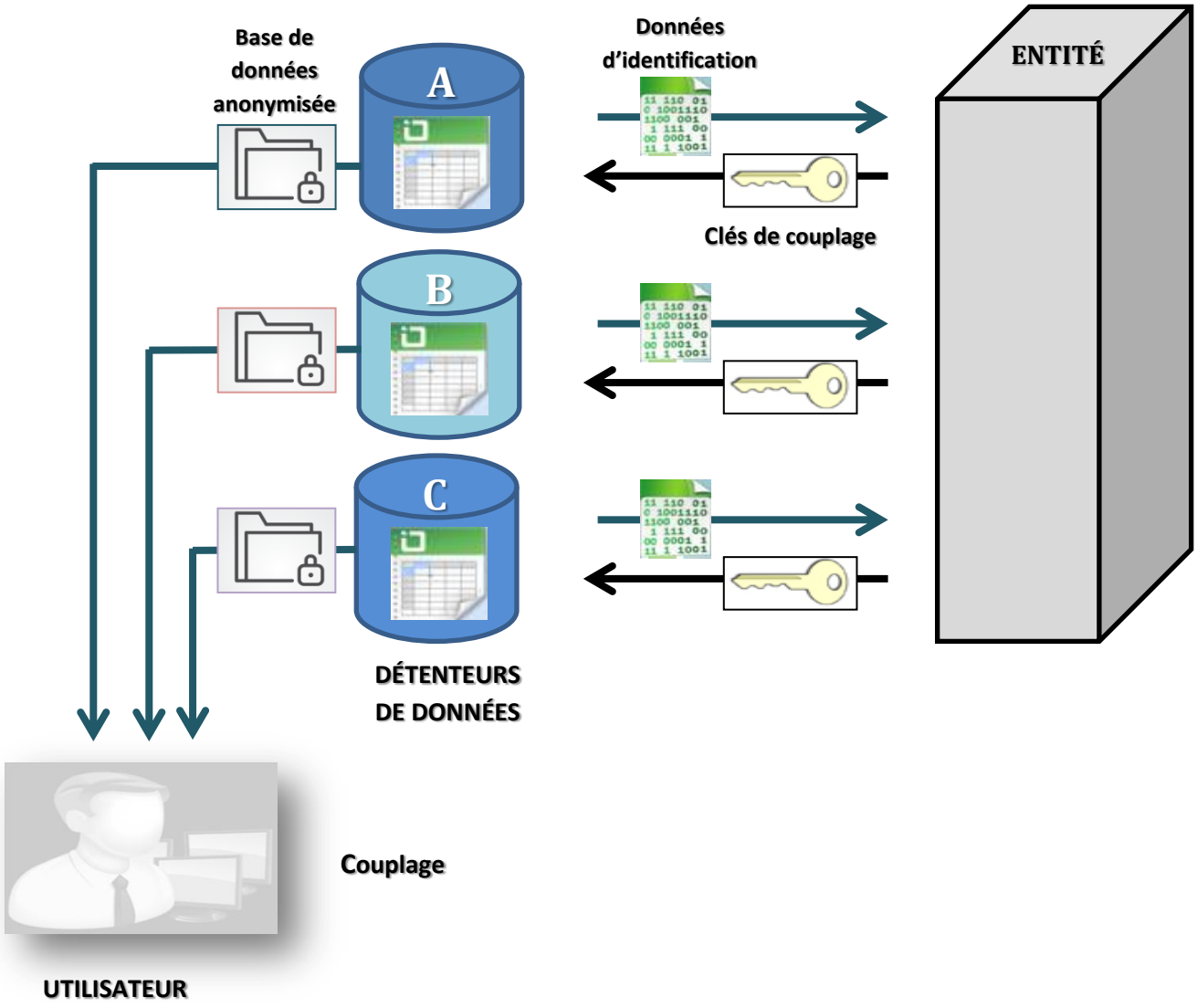
Le premier élément de la gouvernance de la protection de la vie privée concerne le modèle adopté par les différentes entités donnant accès à des banques de données intégrées en santé et en services sociaux. Bien qu'un modèle ne constitue pas en soi un élément de gouvernance, il est néanmoins déterminant pour l'établissement des règles, des mesures et des processus décisionnels et de surveillance d'une organisation. Deux principaux modèles ont été identifiés pour 15 des 17 entités incluses : la cartographie des liens de couplage et la centralisation des données. Outre ces modèles, deux entités sont structurées différemment (BioGrid[21] et SSI[12]).

Premier modèle : la cartographie des liens de couplage

Le modèle basé sur la cartographie des liens de couplage a été adopté par sept entités australiennes. Six de ces entités (CHeReL[24, 33], CDL[10], CVDL[30], RLG[27], SA NT Datalink[29] et TDLU[9, 34]) utilisent le modèle standard et une autre présente une variation de ce modèle (WALDS[14, 17, 25]). Les informations spécifiques à chacune des entités intendantes de données sont présentées dans l'**annexe 3**.

Dans le modèle standard (**figure 2**), l'entité ne reçoit des détenteurs que les données d'identification. Les bases de données originales (incluant les données cliniques) restent chez les détenteurs qui demeurent les seuls dépositaires des données. L'entité utilise les données d'identification pour créer une clé de couplage qui servira à relier entre elles les enregistrements concernant un même individu.

Figure 2 : Modèle de cartographie des liens de couplage



Afin de permettre l'utilisation de données couplées dans le cadre d'un projet, l'entité retourne à chaque détenteur de données les clés de couplage spécifiques au projet. Ce sont ces détenteurs de données qui procèdent à l'anonymisation et à l'extraction des données requises par le projet. Ils se chargent d'envoyer les données anonymisées aux utilisateurs qui procèdent au couplage. Les données de contenu et les données d'identification demeurent donc séparées en tout temps conformément au principe de séparation.

L'entité WALDS[14, 17, 25] a adopté le même modèle de fonctionnement, mais en plus, cette entité héberge une copie partielle de certaines bases de données sur le serveur CARES (*Custodian Administered Research Extract Server*)[14]. Ces données sont séparées des clés de couplage et elles sont mises à jour régulièrement.

Deuxième modèle : la centralisation des données

Le deuxième modèle est basé sur la centralisation des données (**figure 3**). Ce modèle a été adopté par les cinq entités situées au Royaume-Uni (SAIL[15, 19], CPRD[16], Farr London[20], Farr North England[20] et SILC[31]), dont une fonctionne selon une variante du modèle (SILC[31]), et par les trois entités situées au Canada (PopData[22], MCHP[23] et ICES[18]). Les informations spécifiques à chacune des entités sont présentées dans l'**annexe 4**.

Selon le modèle standard, les entités agissent au nom des détenteurs de données de qui elles reçoivent les jeux entiers de données. Elles sont gestionnaires de l'intégration des données (entrepôts dynamiques de données). Les données d'identification sont séparées des autres données soit à la source, par le détenteur, ou à leur réception, par l'entité intendante, appliquant ainsi le principe de séparation. Les données d'identification sont confiées à une tierce partie interne ou externe qui crée les clés de couplage. L'entité utilise ces clés pour procéder à la liaison des données provenant de différentes bases, créant ainsi une nouvelle base de données anonymisée.

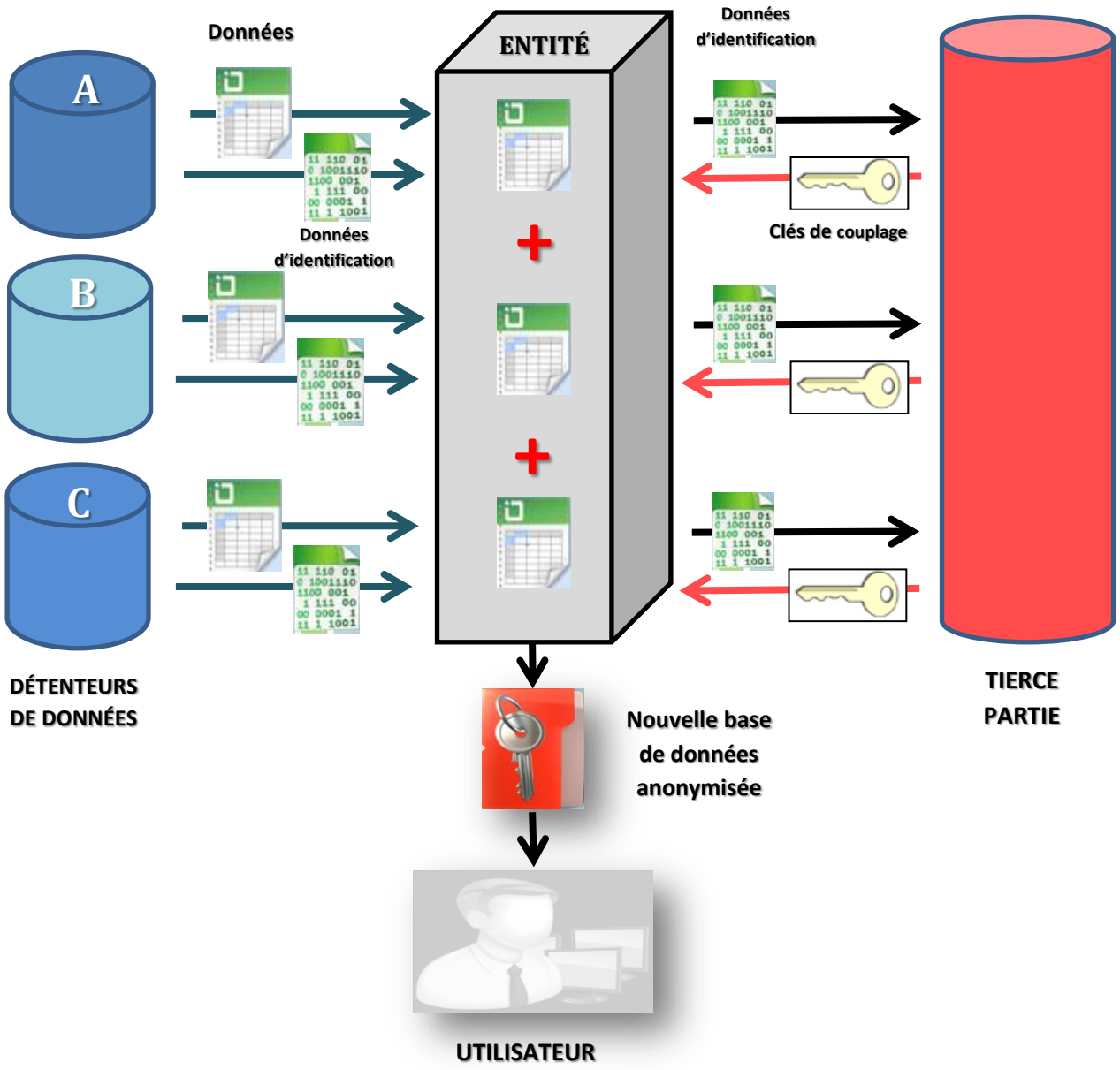
L'entité SILC fonctionne selon le même modèle, mais implique deux tierces parties[31]. La première agit au nom des détenteurs de données de qui elle reçoit les données séparées des tables contenant les identifiants personnels. Cette tierce partie reconstitue des bases de données anonymisées et crée des clés de couplage. Les bases de données et les clés sont ensuite envoyées à une autre tierce partie qui procède à la liaison des bases de données. Les données couplées et anonymisées sont conservées seulement pour la durée du projet et sont ensuite détruites.

Autres modèles

L'entité BioGrid, située en Australie, constitue un entrepôt virtuel organisé selon un modèle de stockage fédéré[21]. Cette entité est constituée d'un réseau de membres provenant de différentes organisations (ex. : hôpitaux, centres d'expertises). Ces détenteurs versent leurs données cliniques sur des serveurs locaux et transmettent les données d'identification à BioGrid qui crée des clés de couplage. Un numéro identifiant est créé pour chaque patient et est logé dans un index encrypté sur un serveur central. BioGrid réunit ces données dans un dépôt virtuel. Un intégrateur de données fédérées installé sur un serveur séparé contient des vues de chacun des dépôts locaux. Une technique probabiliste (ou un algorithme de hachage) est appliquée pour coupler les données des patients à partir des données démographiques. Les nouvelles entrées de patients sont scannées chaque nuit.

L'entité SSI est une infrastructure nationale danoise qui gère l'accès aux grandes bases de données de santé nationales[12]. De plus, elle est constituée d'une banque d'échantillons biologiques de couverture nationale[12]. Le numéro d'identification national unique attribué aux Danois à la naissance sert au couplage des fichiers. Selon les besoins des utilisateurs, le SSI réalise des appariements entre les bases de données.

Figure 3 : **Modèle de centralisation des données**



b) Mesures d'atténuation des risques de réidentification

Des informations concernant les mesures d'atténuation des risques de réidentification étaient disponibles pour cinq entités (CVDL, RLG, WALDS, SAIL et ICES). Deux de ces entités utilisent des outils spécifiques (SAIL et ICES). L'entité SAIL a développé un algorithme de contrôle du risque de divulgation (NEMO) qui procède par évaluation du nombre d'entrées uniques ou à faible occurrence[15]. Ceci permet de prévoir des agrégations de données ou encore la suppression de données ou de variables particulières. De plus, avec l'ajout de la passerelle d'accès à distance *SAIL Gateway*, un analyste sénior peut procéder à l'agrégation et à la suppression de données ou à la limitation du nombre de variables fournies aux utilisateurs qui accèdent aux données à distance[19]. Pour sa part, ICES utilise un outil d'évaluation des risques de réidentification, le *Privacy Analytics Risk Assessment Tool*[18]. Cet outil permet, par exemple, de grouper, de brouiller ou de supprimer les données qui pourraient conduire à une réidentification.

Pour les trois autres entités (CVDL, RLG et WALDS), aucun outil particulier n'est décrit pour la réduction des risques de réidentification. Au CVDL, la confidentialité est traitée au cas par cas. Différentes approches peuvent être appliquées selon la demande d'accès aux données[30]. Au RLG, les données fournies peuvent être « confidentialisées » en fonction des objectifs de la recherche (ex. : agrégation, données partielles, intervalles)[27]. Chez WALDS, les utilisateurs ont accès à des données dénominalisées ou qui ont été préalablement agrégées[21].

Tableau 2 : Gouvernance de la protection de la vie privée selon le modèle de développement des entités intendantes de données

ENTITÉS	PROTECTION DE LA VIE PRIVÉE		
	Principe de séparation	**Anonymisation**	**Mesures d'atténuation des risques de réidentification**
Cartographie des liens de couplage			
CDL	L'entité ne reçoit que les données d'identification. Les bases de données demeurent chez leurs détenteurs.	L'entité crée des clés de couplage à partir des données d'identification et les envoie aux détenteurs de données. Les détenteurs procèdent à la dénominalisation et à l'extraction des données requises pour le projet et les envoient à l'utilisateur. L'utilisateur procède au couplage.	nd
CHeReL			nd
CVDL			Confidentialité traitée au cas par cas.
RLG			Agrégation ou autres approches pour conserver la confidentialité des données en fonction des objectifs du projet.
SA NT Datalink			nd
TDLU			nd
WALDS			Les utilisateurs ont accès uniquement à des données anonymisées ou agrégées.
Centralisation des données			
CPRD	L'entité reçoit les bases de données des détenteurs. Les données d'identification sont séparées des autres données (par le détenteur ou par l'entité).	Une tierce partie de confiance à l'interne ou à l'externe crée des clés de couplage. L'entité reçoit les clés et procède au couplage des données, créant ainsi une nouvelle base de données anonymisées.	nd
Farr London			nd
ICES			Outil d'évaluation des risques de réidentification qui permet aussi de le réduire au niveau jugé acceptable en groupant, brouillant ou supprimant des données.
MCHP			nd
PopData			nd
SAIL			Algorithme de contrôle du risque de divulgation permettant de prévoir des agrégations de données ou la suppression de données/variables particulières.
SILC			nd
Dépôt virtuel. Modèle de stockage fédéré			
BioGrid	Les détenteurs de données (ex. : hôpitaux) versent leurs données sur des serveurs locaux. Les données d'identification sont transmises à l'entité et gérées sur un serveur distinct.	Les détenteurs transmettent les données d'identification à l'entité. L'entité crée les clés de couplage. Chaque patient a un identifiant qui est conservé dans un index crypté sur un serveur central. L'entité réunit les données dans un dépôt virtuel. Un intégrateur de données fédérées contient des vues de chacun des dépôts locaux. Une technique probabiliste est utilisée pour coupler les données.	nd
Infrastructure nationale. Accès aux grandes bases de données nationales			
SSI	**Ne s'applique pas.** L'entité est une infrastructure nationale qui gère l'accès aux grandes bases de données de santé nationales.	Les individus reçoivent un numéro d'identification national unique à la naissance. Le numéro d'identification national permet le couplage des données par l'entité.	nd

nd : information non disponible

3.2.2 DOMAINE DE LA RECHERCHE

Le domaine de la recherche concerne les procédures et les processus mis en place pour maximiser l'utilisation des données tout en préservant la vie privée des individus. La qualité scientifique de la recherche et les retombées pour la société doivent donc dépasser les risques encourus tant sur le plan éthique que légal. Sont incluses les mesures d'approbation préalables des projets de recherche (éthique et autres), le consentement des individus pour l'utilisation des données qui les concernent, les autorisations d'accès et autres engagements éthiques.

a) Approbation éthique

Pour 15 des 17 entités (**tableau 3**), l'obtention d'une approbation émise par un comité d'éthique de la recherche est obligatoire pour obtenir l'accès aux données couplées. D'autres autorisations éthiques sont nécessaires dans certains cas spécifiques (CHeReL et CPRD). Au CHeReL, l'approbation du *Population and Health Services Research Ethics Committee* (PHSREC) du *New South Wales* (NSW) est requise pour les données détenues par le *NSW Department of Health* ou pour l'accès à la clé maître[24, 33]. Au CPRD, un comité d'éthique national de la recherche (NRE) approuve les projets qui nécessitent le couplage de données issues des milieux cliniques de soins primaires avec des données provenant d'autres bases existantes[16].

Les deux autres entités intendantes (BioGrid et SSI) appliquent des processus distincts. Chez BioGrid, les sites d'accès doivent préalablement obtenir un certificat d'un comité d'éthique de la recherche qui approuve leur processus de gestion des données, de dé-identification, de couplage, d'accès et de collecte de données[21]. Ainsi, aucune approbation éthique n'est nécessaire pour la plupart des projets de recherche, à l'exception des projets incluant des données gouvernementales identifiables. Au SSI, les demandes d'accès à des données identifiables nécessitent la permission de l'Agence danoise pour la protection des données[35]. Pour obtenir l'autorisation de lier le matériel biologique d'un individu à des données provenant de registres nationaux, une approbation supplémentaire du Comité d'éthique scientifique doit être obtenue[35].

Tableau 3 : Entités exigeant un certificat émis par un comité d'éthique de la recherche

Entités	Références
CDL	[10]
CHeReL*	[24, 33]
CPRD*	[16]
CVDL	[30]
Farr London	[20]
Farr North England	[20]
ICES	[36]
MCHP	[23]
PopData	[22]
RLG	[27]
SA NT Datalink	[29]
SAIL	[15]
SILC	[31]
TDLU	[34]
WALDS	[17]

*D'autres autorisations éthiques sont nécessaires dans certains cas spécifiques.

b) Autres modalités d'autorisation

Les 17 entités appliquent d'autres modalités d'approbation des projets de recherche lorsqu'une demande d'accès à des données couplées est formulée (**annexe 5**). Ces modalités (demande d'accès ou formulaire d'application) varient d'une entité à l'autre et portent notamment sur la pertinence et la qualité scientifique du projet[10, 16, 21, 28, 37], sur sa faisabilité[15, 20, 28, 38-40], sur la disponibilité et la qualité des données[15, 21, 37, 38], sur le respect des mesures de gouvernance de

l'information[15, 22, 37, 41] et sur l'intérêt public[22, 27, 28, 37].

c) Consentement des individus pour l'utilisation des données qui les concernent

Le consentement des individus pour l'utilisation des données personnelles qui les concernent peut être explicite, implicite ou non requis. Des informations sur ces procédures sont disponibles pour cinq des 17 entités identifiées (BioGrid, CHeReI, CPRD, RLG et SSI) et sont présentées dans l'**annexe 6**.

Les participants qui émettent un consentement explicite (*opt-in*) expriment leur accord à ce que leurs données personnelles soient utilisées pour des fins de recherche ou d'évaluation. Chez BioGrid, ce type de consentement est nécessaire pour la réalisation de recherches-interventions et pour l'accès aux données génétiques[21]. Au RLG, le consentement signé des participants doit être obtenu pour l'utilisation de leurs informations personnelles ou d'identification[27].

Le consentement implicite (*opt-out*) consiste à informer les personnes que leurs données codées et dénominalisées pourraient être utilisées à des fins de recherche ou d'évaluation tout en leur offrant nettement l'option de signifier leur refus. Sans ce refus signifié, les personnes sont considérées comme ayant accepté. Ce type de consentement est utilisé chez BioGrid pour l'accès aux données cliniques[21] et au CPRD pour les données collectées à partir des dossiers électroniques des patients[16]. Au SSI, en ce qui concerne spécifiquement les biobanques, les personnes peuvent s'opposer, de façon éclairée, au prélèvement et à la conservation de leurs échantillons biologiques[12].

Dans des cas spécifiques, aucun consentement n'est nécessaire pour obtenir l'accès aux données. C'est le cas des données administratives chez BioGrid[21] et des données identifiables du département de la santé au RLG[27]. Au CPRD, l'accès à des données identifiables ou potentiellement identifiables peut également être octroyé sans consentement lorsque les données concernent un individu décédé ou introuvable[42].

Au CHeReL, aucun consentement n'est nécessaire sauf pour les données collectées pour des études spécifiques[43]. Au SSI, toutes les données sont collectées sans que le consentement des individus ne soit requis[12].

d) Autorisation d'accès par les détenteurs de données

Onze entités intendantes (**tableau 4**) ont établi une procédure par laquelle tous les détenteurs doivent autoriser l'accès à leurs données pour chaque projet. L'entité BioGrid fonctionne de façon similaire en nommant un « détenteur » des données sur chaque site d'accès. Ce dernier doit donner son autorisation d'accès pour chaque projet[21].

Tableau 4 : Entités nécessitant des autorisations d'accès des détenteurs de données

Entités	Références
CDL	[10]
CHeReL	[33]
CVDL	[30]
Farr London	[20]
PopData	[22]
RLG	[27]
SA NT Datalink	[29, 32]
SAIL	[26]
SILC	[28]
TDLU	[9, 34]
WALDS	[14, 25]

Les entités CPRD, ICES, MCHP et SSI fonctionnent différemment puisqu'elles ont la responsabilité de contrôler l'accès aux données qu'elles hébergent. Leurs modalités d'accès se résument à ce qui a été présenté aux sections **3.2.2 a et b**.

Le CPRD rassemble des données anonymisées collectées de façon continue à partir des dossiers de santé électroniques[16]. Parmi les cliniques impliquées, 75 % d'entre elles ont consenti à ce que leurs données soient couplées avec celles d'autres bases de données (statistiques et registres nationaux, données issues d'études et d'essais cliniques).

L'entité ICES héberge et contrôle l'accès aux données sur l'utilisation des services de santé, à des données provenant des fournisseurs de soins de santé, de différents registres et d'enquêtes populationnelles ontariennes[18]. Les détenteurs de l'information (ex. : les médecins, les hôpitaux) sont autorisés à divulguer à l'entité des renseignements personnels concernant la santé de leurs patients. Ces informations sont utilisées pour des analyses statistiques qui permettent l'évaluation et la surveillance du système de santé. Leur accès peut aussi être consenti pour des projets de recherche qui ont été approuvés par l'équipe du service de données et d'analyse[39].

Le MCHP est l'entité intendante hébergeant en permanence les données de santé et du registre anonyme sur la population du Manitoba[18]. Les utilisateurs désirant accéder à des données autres que celles de santé doivent toutefois obtenir l'approbation des détenteurs de données. Le SSI est un entrepôt de données national qui a aussi son propre processus d'autorisation d'accès, tel que décrit précédemment (sections **3.2.2 a et b**).

e) Engagements des utilisateurs de données

Indépendamment des exigences des comités d'éthique et des détenteurs de données, certaines entités exigent que les utilisateurs s'engagent formellement à respecter la vie privée des personnes de qui proviennent les données couplées (**annexe 7**). Pour les dix entités qui présentent des informations à ce sujet (BioGrid, Farr London, Farr North England, ICES, MCHP, PopData, RLG, SAIL, SILC et WALDS), ces engagements prennent la forme de signatures d'ententes[15, 17, 21, 28], de formations[20, 22, 28, 39-41], de formulaires d'autorisation approuvés en vertu de la législation en vigueur[27] ou appuyés par une personne de référence[28].

Les informations relatives au domaine de la recherche sont récapitulées dans le **tableau 5**.

Tableau 5 : Synthèse des informations sur la gouvernance de la recherche

	Comité d'éthique de la recherche	Autres modalités d'autorisation	Consentements éthiques			Autorisation d'accès des détenteurs de données	Autres engagements formels
			Opt-in	Opt-out	aucun		
BioGrid	−	+	+	+	+	+	+
CDL	+	+	nd	nd	nd	+	−
CheReL	+	+	nd	nd	+	+	−
CPRD	+	+	nd	nd	nd	−	−
CVDL	+	+	nd	+	+	+	−
Farr London	+	+	nd	nd	nd	+	+
Farr North England	+	+	nd	nd	nd	nd	+
ICES	+	+	nd	nd	nd	−	+
MCHP	+	+	nd	nd	nd	−	+
PopData	+	+	nd	nd	nd	+	+
RLG	+	+	+	nd	+	+	+
SAIL	+	+	nd	nd	nd	+	+
SA NT datalink	+	+	nd	nd	nd	+	−
SILC	+	+	nd	nd	nd	+	+
SSI	−	+	nd	+	+	−	−
TDLU	+	+	nd	nd	nd	+	−
WALDS	+	+	nd	nd	nd	+	+

− : élément non présent
+ : élément présent
nd : information non disponible

3.2.3 DOMAINE DE L'INFORMATION

Ce domaine de gouvernance concerne le traitement de l'information afin de permettre l'accès aux données dans des délais raisonnables tout en maintenant leur confidentialité. Les mesures de protection de l'information couvrent la sécurité informatique, la sécurité des requêtes et des sorties, la sécurité physique et l'évaluation des processus de sécurité. Les entités possèdent également les compétences et offrent le soutien nécessaire pour jumeler, associer, agréger, nettoyer et transformer des données en provenance de systèmes multiples et complexes.

a) Mesures de sécurité informatique

Les mesures de sécurité informatique mises en place par les entités pour assurer la protection des données sont détaillées dans l'**annexe 8**. Des informations sont disponibles pour 15 des 17 entités identifiées (toutes sauf CVDL et MCHP). Les mesures de sécurité informatique incluent les modes de transferts sécurisés de données ou d'analyses[15, 27, 34, 37, 44], les modalités d'accès aux dépôts sécurisés (*safe haven*), aux fichiers ou aux services (ex. : l'authentification à double facteur, les mots de passes et identifiants cryptés)[12, 14, 18-20, 22, 24, 27-29, 31, 37, 44, 45]

et les moyens de sécurisation des réseaux et des serveurs (ex. : pare-feu, VPN, antivirus, cryptage du réseau)[10, 15, 16, 18-20, 22, 27, 29, 37, 44].

b) Mesures de sécurité des requêtes et sorties

Des informations sur les mesures de sécurité relatives aux requêtes et sorties (**annexe 8**) sont disponibles pour 15 des 17 entités (toutes sauf CHeReL et CPRD). Ces mesures visent à contrôler l'usage et la diffusion des données. Elles prennent la forme de systèmes de monitorage[20-22, 29], de procédures de conduite à observer[21, 27], de déclaration et d'examens des analyses et des publications[15, 19, 20, 22, 29, 32, 34, 37, 38, 40]. De plus, la désactivation de certaines fonctionnalités des ordinateurs (ex. : impression, transfert sur une clé USB)[18, 20] et d'autres mesures peuvent être prises de manière à ce que les données et les analyses y soient confinées[12, 22, 29, 31].

c) Mesures de sécurité physique

L'accès physique aux entités est limité par des mesures de sécurité, notamment pour les zones où les données sont hébergées (**annexe 9**). Neuf entités présentent des informations à ce sujet (CHeReL, Farr North England, MCHP, PopData, SA NT Datalink, SILC, TDLU, WALDS). Les mesures de contrôle d'accès peuvent comprendre des systèmes de verrouillage des entrées ou d'autres contrôles électroniques ainsi que des caméras de surveillance[20, 22, 28, 29, 37, 44]. Pour l'accès à distance, des mesures peuvent être exigées pour sécuriser l'endroit où les données sont consultées [20, 27, 28, 40, 46].

d) Évaluation des processus de sécurité

Des informations concernant l'évaluation des processus de sécurité sont disponibles pour neuf entités (BioGrid, CDL, CHeReL, CPRD, Farr London, PopData, SAIL, SA NT Datalink et WALDS) (**annexe 9**). Pour sept d'entre elles, les processus de sécurité ou de conformité aux pratiques de gouvernance de l'information font l'objet d'audits ou d'examens indépendants [10, 15, 20-22, 24, 37]. Une entité réalise des audits auprès des organisations qui reçoivent les données couplées[46] et une autre a intégré des procédures d'évaluation des risques à sa gestion opérationnelle[29].

e) Compétences et soutien

Les entités détiennent les expertises scientifique et technique permettant l'anonymisation et le couplage des données. Elles mettent ces savoirs en pratique ou à profit en offrant leur soutien aux utilisateurs (**annexe 10**). Cinq entités (CHeReL, CVDL, RLG, SA NT Datalink, TDLU) offrent de l'appui pour la présentation des projets et la formulation des demandes d'accès [24, 27, 34, 38, 47]. Deux entités (CVDL, SA NT Datalink) coordonnent la rencontre entre les utilisateurs et les détenteurs de données [38, 47]. Les entités WALDS et CHeReL offrent de l'aide et des formations pour l'analyse des données couplées[24, 34].

Les informations relatives au domaine de l'information sont résumées au **tableau 6.**

Tableau 6 : Synthèse des informations sur la gouvernance de l'information

	Sécurité			Évaluation des processus de sécurité	Compétences et soutien
	informatique	physique	Requêtes et sorties		
BioGrid	+	nd	+	+	nd
CDL	+	nd	+	+	nd
CheReL	+	+	nd	+	+
CPRD	+	nd	nd	+	nd
CVDL	nd	nd	+	nd	+
Farr London	+	nd	+	+	nd
Farr North England	+	+	+	nd	nd
ICES	+	nd	+	nd	nd
MCHP	nd	+	+	nd	nd
PopData	+	+	+	+	nd
RLG	+	nd	+	nd	+
SAIL	+	nd	+	+	nd
SA NT Datalink	+	+	+	+	+
SILC	+	+	+	nd	nd
SSI	+	nd	+	nd	nd
TDLU	+	+	+	nd	+
WALDS	+	+	+	+	+

- : élément non présent
+ : élément présent
nd : information non disponible

3.2.4 DOMAINE DES RÉSEAUX

Le réseautage en recherche peut impliquer non seulement les chercheurs, mais aussi les parties prenantes par exemple les détenteurs de données et les organismes subventionnaires. La création et le maintien de réseaux peuvent faciliter l'initiation de projets, le partage de données et la mise en place de mécanismes de valorisation des données pour un usage secondaire. Les informations recueillies dans les documents inclus concernent la standardisation des collectes de données ainsi que le partage des données (**annexe 11**).

a) Standardisation des collectes des données

Des informations relatives à la standardisation de la collecte de données sont disponibles seulement pour BioGrid. Dans cette entité, des groupes composés de chercheurs, de cliniciens, de gestionnaires et de spécialistes en connaissances sur des maladies spécifiques se sont d'abord mobilisés pour déterminer les données à recueillir par les sites participants[21]. Ensuite, les administrateurs de BioGrid et les détenteurs de données se sont concertés afin d'optimiser les champs et pour formater les données à collecter[21].

b) Partage des données

En ce qui concerne le partage des données, des informations sur les mécanismes mis en place pour faciliter cette étape sont disponibles pour deux entités, soit BioGrid et CDL. L'entité BioGrid a élaboré des lignes directrices détaillées de reconnaissance de la propriété intellectuelle afin d'encourager les chercheurs à partager leurs données[21]. La propriété intellectuelle est conservée par la partie qui fournit les données originales. Une licence non exclusive pour l'utilisation secondaire des données peut être accordée à d'autres groupes de recherche.

Au CDL, un groupe de travail a été créé spécifiquement sur la question du transfert de données[10]. Ce sous-comité est chargé de conseiller et de donner des orientations aux membres du Conseil de gestion qui est chargé de superviser la mise en œuvre du Programme national australien de couplage de données.

L'Australie et le Royaume-Uni se sont dotés de structures nationales facilitant le partage de données sur leurs territoires. En Australie, le *Population Research Network* réunit des collaborateurs dont l'objectif est de développer les capacités de liaisons de données et de positionner l'Australie comme chef de file mondial[48]. Les unités qui desservent chaque état et territoire et l'Unité nationale de liaison des données (CDL) composent ce réseau. Il comprend un laboratoire sécurisé d'accès à distance, un système de transfert de fichiers sécurisé et un bureau national de coordination[48].

Au Royaume-Uni, le *Farr Institute of Health Informatics Research* comprend quatre nœuds répartis sur l'ensemble du territoire (London, North England, SAIL et SILC). Ce réseau soutient la recherche de pointe qui relie les données de santé électroniques à d'autres types de données. Cette infrastructure physique et électronique facilite la collaboration et l'établissement de partenariats. Elle offre aussi un soutien pour l'utilisation sécuritaire des données[49].

3.3 RÉSULTATS DE PERFORMANCE

Des informations concernant la performance des entités intendantes ont été colligées pour huit d'entre elles (BioGrid, CHeReL, CPRD, CVDL, ICES, PopData, SAIL et WALDS). Ces résultats, présentés à l'**annexe 12**, concernent les délais d'accès aux données[14, 18], le nombre de demandes traitées[14, 18, 19, 21, 22, 24, 30, 37] et le nombre de publications[11, 16-18, 21]. Aucune donnée n'a été repérée sur les coûts reliés aux demandes d'accès ni sur les incidents liés à la violation de la confidentialité des données.

3.3.1 DÉLAIS D'ACCÈS AUX DONNÉES COUPLÉES

Pour l'entité ICES, les délais d'accès aux données varient de deux à quatre semaines après qu'un accord ait été convenu avec les utilisateurs de données[18]. Chez WALDS, l'extraction de données à coupler en provenance de certaines bases de données est simplifiée par l'utilisation du serveur CARES (*Custodian Administered Research Extract Server*) qui héberge des copies partielles de bases de données, mises à jour régulièrement. Des mesures de délai d'accès, défini par le nombre de jours écoulés entre la demande formelle d'extraction des données et la réception par le service à la clientèle du fichier de données couplées, ont été prises pour les projets complétés entre janvier 2012 et juin 2014. Les résultats ont montré des délais moyens de 39,1 jours, de 61,3 jours et de 69,7 jours selon que le serveur CARES ait été utilisé, respectivement, pour toutes les extractions, pour certaines extractions ou pas du tout[14].

3.3.2 NOMBRE DE DEMANDES TRAITÉES

Le nombre de demandes traitées est présenté de différentes façons selon les entités pour lesquelles cette information est disponible. Par

conséquent, la comparaison des entités entre elles n'est pas possible.

Chez BioGrid, 4000 requêtes de données de recherche sont effectuées chaque mois, à partir d'Internet, en fonction des autorisations préalablement obtenues[21]. En 2011, pour les recherches collaboratives impliquant l'industrie pharmaceutique, 25 projets avaient été complétés, 9 étaient en cours et plusieurs autres en discussion[21].

Pour les trois premières années d'opérations de CHeReL (2006-2009), 57 projets complétés ont été recensés[24]. En juin 2009, 30 projets impliquant le couplage de données étaient en cours de réalisation[24].

Au cours de la période de 2012 à 2016, le CVDL a fourni des données couplées pour répondre aux besoins d'environ 150 projets de recherche ou de planification de politiques[30]. En 2016, au moment où cette information a été publiée, 26 autres projets étaient en cours et 46 étaient en phase de développement[30].

Entre 2008, au moment où ICES a lancé son Programme de diffusion de données sur le cancer, et l'année 2016, plus de 42 ensembles de données administratives de santé couplées ont été distribués aux chercheurs[18]. En 2016 ICES avait reçu 34 demandes d'accès admissibles depuis la mise en service de sa plateforme de données et de services d'analyse en mars 2014[18]. Ces demandes ont nécessité le couplage de données externes avec des données conservées par l'entité.

Entre 1998 et 2012, PopData et son prédécesseur, le *Columbia Linked Health Database*, ont couplé des données administratives de santé qui ont contribué à la réalisation de plus de 350 projets de recherche[22].

Depuis sa création en 2006 et jusqu'en janvier 2014, SAIL a approuvé l'utilisation de données couplées pour plus de 100 projets[19]. Enfin, lors de sa première année d'opération, en 1995, WALDS a permis la réalisation de 30 projets impliquant des données couplées[14]. En 2016, plus de 800 projets avaient été réalisés[37].

3.3.3 NOMBRE DE PUBLICATIONS

Le nombre de publications découlant du couplage de données est disponible pour quatre des entités incluses. Des données couplées chez BioGrid ont mené, en 2010, à la publication de 87 articles scientifiques et à la réalisation de 22 affiches[21]. Depuis sa fondation en 1987, les données en provenance du CPRD (ou anciennement le *Value Added Medical Products dataset*, puis le *General Practice Research Database*) ont permis de produire près de 2000 rapports de recherches dont plus de 1000 ont été publiés dans des revues évaluées par des pairs[16].

Le Programme de diffusion de données sur le cancer de ICES a conduit, de 2008 à 2016, à 34 publications scientifiques portant sur des projets réalisés à partir de données couplées[18]. Chez WALDS, plus de 250 publications dans des journaux scientifiques ont été répertoriées de 1995 à 2005[17]. Pour la période de 1995 à 2003, on comptait 177 présentations orales, 159 articles multimédias, 96 rapports et 104 autres produits (résumés de conférence, affiches, thèses, chapitres de livre, lettres à l'éditeur, articles soumis ou sous-presse)[11].

4 ANALYSE DES RÉSULTATS

Les informations issues des publications concernant les 17 entités intendantes d'accès aux données intégrées ont été analysées selon les quatre domaines de gouvernances présentés dans le cadre théorique choisi (**Figure 1, section 2.1**), soit la protection de la vie privée, la recherche, l'information et les réseaux.

À cet effet, la gouvernance de la **protection de la vie privée** circonscrit les conditions d'accès et d'usage de données permettant de conserver leur confidentialité. Parmi les pratiques de gouvernance identifiées, le principe de séparation entre les données d'identification et les autres données représente un processus qui contribue à préserver la confidentialité. Le cryptage des identifiants par la création de clés permet de coupler des banques de données et de les anonymiser. Le fait de confier cette étape à une tierce partie ajoute un élément de sécurité.

Les mesures d'atténuation des risques de réidentification constituent une pratique de protection supplémentaire. Ces mesures impliquent l'agrégation ou la suppression de données ou de variables particulières. Des outils spécifiques sont utilisés dans certaines entités, par exemple un algorithme de contrôle du risque de divulgation.

La **gouvernance de la recherche** implique la mise en place de divers processus ou mécanismes assurant une conduite responsable quant à l'utilisation des données couplées. L'approbation des projets par un comité d'éthique de la recherche est incontournable. De plus, l'évolution du rôle des entités intendantes a mené à l'émergence d'une nouvelle pratique éthique de certification de l'entité.

Par ailleurs, puisqu'il n'y a pas de ligne directrice claire sur la conduite à adopter pour l'usage secondaire des informations confidentielles en santé[50], toutes les entités se sont dotées de leurs propres structures internes d'évaluation des projets nécessitant le couplage de données. Des demandes d'accès doivent être formulées auprès des détenteurs de données, sauf dans le cas où l'entité est autorisée à les représenter. D'autres engagements formels à respecter la vie privée peuvent aussi être exigés (ex. : signature d'ententes).

La **gouvernance de l'information** concerne les mesures physiques et informatiques permettant d'assurer la sécurité des données. Ces mesures sont applicables à la fois au niveau de l'entité et pour les accès à distance. Des mesures physiques telles que le verrouillage des locaux, l'utilisation de caméras de surveillance ou d'autres contrôles électroniques permettent de sécuriser l'endroit où les données sont hébergées ou les sites d'accès à distance.

En ce qui concerne la sécurité informatique, des mesures incluant des modalités d'accès aux données et de transferts sécurisés ainsi que des moyens permettant de sécuriser les réseaux et les serveurs sont documentées. Les analyses des données couplées sont également scrutées avant leur sortie et leur publication. Toutes ces mesures peuvent aussi faire l'objet de vérifications afin d'identifier les failles dans la sécurité.

Enfin, **la gouvernance des réseaux** concerne la standardisation des collectes de données ainsi que des mécanismes de partage des données, facilitant la réalisation de projets de recherche impliquant des données couplées. La mise sur pied de réseaux nationaux structurés facilite le partage de données et offre un soutien pour l'utilisation sécuritaire des données.

En ce qui concerne la performance des entités, aucune étude n'avait pour objectif de l'évaluer. Cependant, certaines informations étaient disponibles pour huit entités dans les documents identifiés par la recherche. Ceci a permis de documenter, selon les entités, les délais d'accès aux données, le nombre de demandes traitées et le nombre de publications. Toutefois, la quantité et la disparité des informations recueillies empêchent la comparaison des entités entre elles.

5 LIMITES

Les résultats de la revue systématique de la littérature doivent être interprétés avec prudence en raison de certaines limites. D'abord, les législations locales auxquelles doivent se conformer les entités intendantes étudiées ne sont pas prises en considération dans l'analyse des résultats. Ensuite, ces entités sont des organisations complexes dont les pratiques de gouvernance ne peuvent être repérées et décrites uniquement à partir d'une revue de la littérature. Des informations plus exhaustives seraient nécessaires pour se prononcer quant aux meilleures pratiques de gouvernance.

De plus, les modalités de fonctionnement particulières et les pratiques de gouvernance des entités intendantes n'ont pas toutes fait l'objet de publications. Aussi, les informations rapportées dans les documents retenus sont parfois incomplètes ou inaccessibles. Par ailleurs, la nature de l'information disponible varie d'une entité à l'autre. Enfin, les entités n'ont pas toutes atteint le même degré de maturité organisationnelle. Alors que certaines sont opérationnelles depuis plusieurs années, d'autres sont sur le point de l'être, ce qui influence potentiellement leurs pratiques de gouvernance et l'accessibilité des informations qui les concernent.

6 APPLICABILITÉ DES RÉSULTATS DE L'ETMI DANS LE CONTEXTE QUÉBÉCOIS

Cette ETMI apporte plusieurs éléments de réflexion concernant la mise en place, au Québec, d'une entité intendante permettant l'accès à des données couplées.

D'abord, la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* (RLRQ, c. A-2.1) assure déjà la protection de la vie privée dans un contexte de création et d'utilisation de banques de données. Les niveaux de rigueur et de protection de la vie privée que procure cette loi pourraient se transposer aux pratiques de gouvernance d'une entité donnant accès à des données couplées.

La responsabilité d'évaluer si les conditions de la Loi sont remplies incombe à la Commission d'accès à l'information (CAI) qui traite les demandes d'accès aux données au cas par cas. Ce processus comporte plusieurs barrières opérationnelles qu'une entité intendante permettrait de surmonter. Le modèle de centralisation des données, adopté au Royaume-Uni et ailleurs au Canada, pourrait par exemple améliorer la fluidité du processus. Dans ce modèle, l'entité agit comme fiduciaire des données et assure la transmission sécuritaire et efficace des données intégrées vers les utilisateurs qui n'ont jamais accès aux données d'identification personnelle.

La CAI, ou une autre instance pourrait se porter garante de la protection de la vie privée en s'assurant que le risque de ré-identification soit abaissé à un seuil minimal. Cela pourrait, par exemple, impliquer un processus de certification éthique de l'entité. C'est le cas pour BioGrid et SSI qui doivent obtenir un certificat d'un comité d'éthique de la recherche qui approuve leur processus de gestion des données, de dé-identification, de couplage, d'accès et de collecte de données.

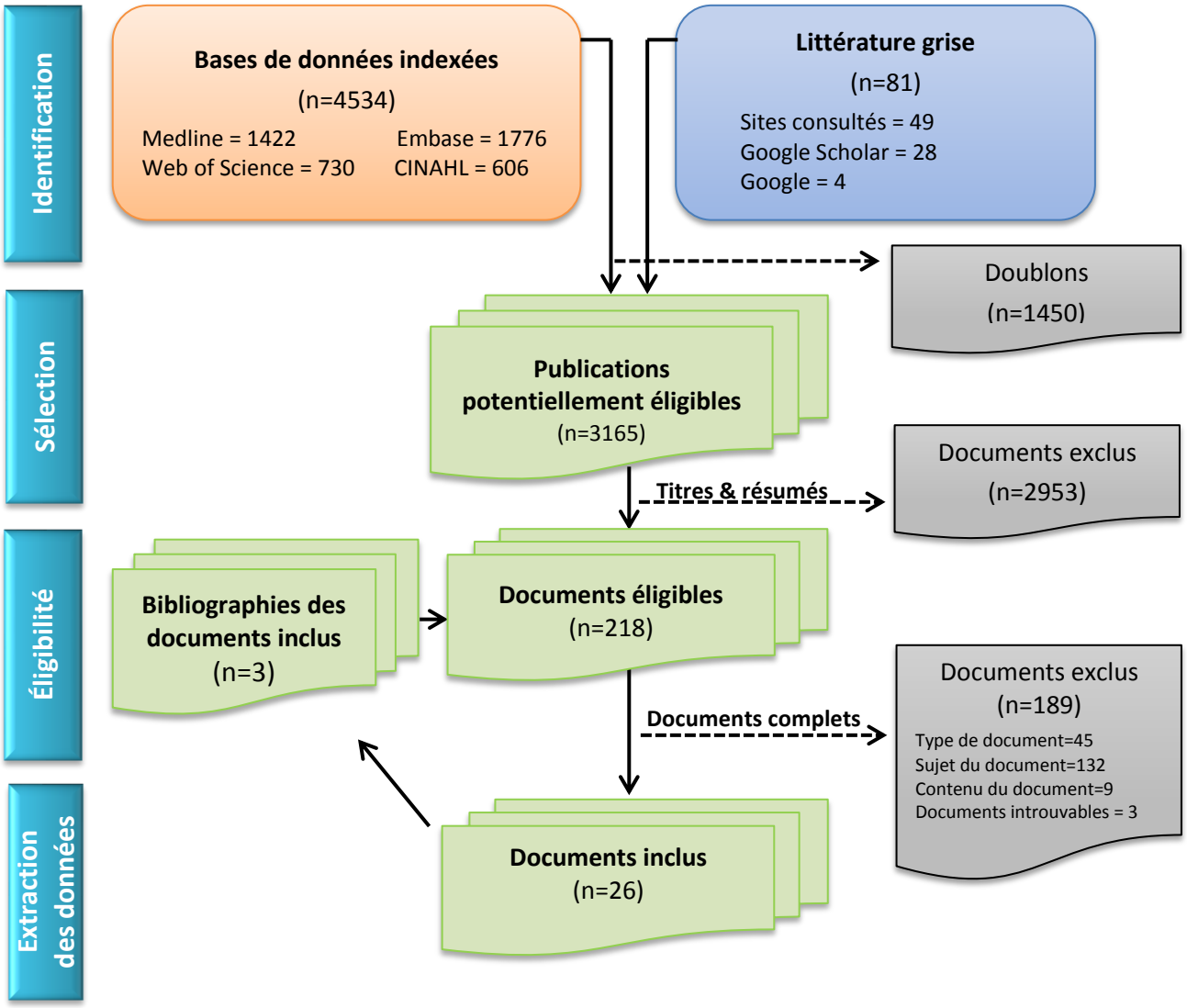
Enfin, le Québec n'a pas d'enjeux aussi complexes qu'un état fédéral, tels que le Royaume-Uni ou l'Australie, où des partenariats entre différentes juridictions sont nécessaires pour accéder aux données. Au Québec, une centralisation des services de couplage pourrait répondre aux différents besoins des utilisateurs de données couplées.

ANNEXES

Annexe 1 : Critères de sélection des documents

Critères	Inclusion	Exclusion
1er critère : Type de document	Type de document ✓ Études originales (tous les types de devis) ✓ Revues systématiques ou utilisant une approche systématique ✓ Rapports d'ETMISSS ✓ Rapports d'évaluation ✓ Rapports gouvernementaux ou d'organisations savantes (pratiques exemplaires, guides)	✓ Éditoriaux ✓ Commentaires ✓ Résumés de conférence ✓ Lettres d'opinion ✓ Publicités ✓ Articles de journaux ou de magazines de vulgarisation ✓ Revues n'utilisant pas une approche systématique ✓ Rapports annuels d'organisations
2e critère : Langue de publication	En français ou en anglais	
3e critère : Pays concernés	Pays de l'Organisation de coopération et de développement économiques (OCDE)	
Critères	**Inclusion**	**Exclusion**
4e critère : Sujet du document	Concerne une ou plusieurs entités dont le **mandat principal** est de coupler des banques de données de santé et de services sociaux avec des banques de données connexes, **et de rendre accessibles** les banques de données intégrées aux chercheurs, aux décideurs et aux citoyens.	✓ Système d'information exclusivement à l'intérieur d'un établissement (ex : dossiers médicaux électroniques, dossiers cliniques informatisés) ✓ Des banques de données originales (ex. : recensements, études épidémiologiques et cliniques) ✓ Des entités qui effectuent le couplage de bases de données orientées autour d'une seule pathologie ✓ Cohortes
5e critère : Contenu du document	**Présente des processus de gouvernance** d'une entité d'accès à des données intégrées	

Annexe 2 : Diagramme de sélection des documents PRISMA



n= nombre de documents

Annexe 3 : Entités ayant adopté le modèle de cartographie des données

Entités	Références	Application du principe de séparation et processus d'anonymisation
CHeReL	Lawrence, 2008[33]; *Cancer Institute* NSW, 2009[24]	- Cryptage et transfert sécuritaire des données entre les détenteurs de données et CHeReL. - Création de clés de couplage spécifiques à chaque projet et maintien d'une clé maître de couplage. - Seul le personnel de CheReL assigné au couplage a accès aux informations personnelles d'identification des participants.
CDL	Boyd, 2012[10]	- Les variables d'identité sont utilisées seulement à l'étape du couplage des données et au niveau de l'unité juridictionnelle. - L'unité juridictionnelle fournit au CDL des données démographiques ou des identifiants cryptés pour le couplage avec les données nationales. - Pour les couplages de données nationales, le CDL attribue une clé de couplage unique pour toutes les entrées concernant une même personne et crée une cartographie à partir d'un identifiant unique crypté. - Transferts de clés uniques masquées et cryptées aux détenteurs de données.
CVDL	Chen, 2016[30]	- Création d'une cartographie de couplages spécifique à chaque projet avec des numéros d'identification et des pointeurs cryptés et liés aux bases de données pertinentes.
RLG	*Department of health, Queensland Government*, 2016[27]	- Les enregistrements contenant seulement les données démographiques et l'identifiant individuel de la personne ou de l'enregistrement sont transmis au RLG. - Le RLG crée les clés de couplage spécifiques aux projets. - Chez les détenteurs de données, une clé de couplage spécifique au projet est jointe aux données cliniques pertinentes et les variables requises sont extraites.
SA NT	SA NT Datalink, 2016[29]	- Le personnel de l'entité reçoit les informations d'identification et démographiques, et les identifiants des détenteurs de données. - L'entité réalise le couplage, crée des clés de couplage anonymisées spécifiques à chaque projet et stocke l'information démographique identifiée dans le fichier maître de couplage (*Master Linkage File*). - Seules des clés de couplage dépersonnalisées sont fournies aux chercheurs et aux analystes par l'entité. - Les détenteurs de données fournissent eux-mêmes les données aux chercheurs et aux analystes. - Séparation des rôles et des responsabilités (seuls les détenteurs de données ont accès aux données personnelles).

Entités	Références	Application du principe de séparation et processus d'anonymisation
TDLU	Wiggins, 2016[34]; Strokes[9]	- Le TDLU obtient des détenteurs uniquement les données pour les variables administratives. - Le TDLU conserve le mappage maître de couplage et crée des clés de couplage. - Les détenteurs de données reçoivent un fichier des clés de couplage comprenant un identifiant spécifique au projet et leurs identifiants de sources. - Les détenteurs de données y attachent les données cliniques et enlèvent les identifiants. - Les chercheurs obtiennent les ensembles de données dépersonnalisées et anonymisées de chacun des détenteurs (incluant le contenu clinique). - Les chercheurs couplent les données à l'aide de la clé unique.
WALDS	Eitelhuber, 2014[14]; Holman, 2008[17]; EuroREACH, 2011[25]	- L'entité possède une clé maîtresse de couplage. - Les employés de WALDS utilisent les champs démographiques pour créer les liens (clés de couplage) entre les données. - Les clés de couplage spécifiques aux projets sont retournées aux détenteurs de données qui font l'extraction. - Ces clés de couplage et les données qui n'identifient pas les participants sont transmises et utilisées par les chercheurs.

Annexe 4 : Entités ayant adopté le modèle de centralisation des données

Entités	Références	Application du principe de séparation et processus d'anonymisation
SAIL	Ford, 2009[15]; Jones, 2014[19]	- Transport des données sécurisé par le *Wales National Switching Service*. - Anonymisation et cryptage des données par *Health Solutions Wales*. - Séparation des données démographiques anonymisées et des données cliniques. - Création d'un champ de liaison anonyme pour chaque individu. - Réunification des données cliniques et démographiques anonymes par l'Unité de recherche sur l'information en santé (HIRU) + 2e niveau de cryptage.
CPRD	Herrett, 2015[16]	- L'entité rassemble des données anonymes collectées une fois par mois (données électroniques des pratiques de soins primaires). - Les données des patients sont couplées par le *Health and Social Care Information Centre*. - Principe de séparation physique entre les chercheurs ayant accès à des renseignements identifiables d'études primaires et ceux utilisant les données du CPRD.
PopData	Pencarrick Hertzman, 2013[22]	- Transfert annuel des données cryptées pour couplage proactif. - Résultats stockés à l'aide d'identifiants personnels uniques sans signification. - Les identifiants sont utilisés pour rassembler des données provenant de plusieurs ensembles pour chaque projet de recherche approuvé. - Les données sont cryptées en tout temps : lors du transfert, du stockage, de l'utilisation, de la sauvegarde et de la destruction.
MCHP	Roos, 2008[23]	- Les données sont hébergées au *Manitoba Centre for Health Policy* (registre populationnel anonymisé). - Les fiduciaires (entités responsables des données) préparent des dossiers contenant un numéro d'enregistrement brouillé pour chaque individu ainsi que des renseignements pour faciliter l'identification de ces personnes (le nom, l'adresse, la date d'entrée en vigueur et la date de naissance). - Santé Manitoba effectue le couplage des enregistrements pour trouver le numéro d'identification personnel de santé (PHIN) approprié. - Un PHIN crypté est attaché à chaque numéro d'enregistrement brouillé reçu du fiduciaire. - Un fichier de croisement contenant seulement un fichier crypté et le numéro d'enregistrement brouillé pour chaque individu est fourni à MCHP. - Le fiduciaire envoie le fichier de données administratives approprié à MCHP, toutes les informations d'identification étant retirées et le numéro d'enregistrement brouillé joint.

Entités	Références	Application du principe de séparation et processus d'anonymisation
ICES	Ishiguro, 2016[18]	- Une personne garante désignée couple les données importées de l'externe avec celles de ICES à partir d'un identifiant unique crypté. - Les variables d'identification sont retirées de l'ensemble de données avant l'analyse. (Méthode « *Safe Harbour* » qui consiste à supprimer 18 variables incluant le nom, l'emplacement géographique plus petit que le pays, tous les éléments datés reliés à un individu (sauf l'année) et les numéros de téléphone).
Farr London	Lea, 2016[20]	- Offre une solution technique pour stocker de façon sécurisée des données d'identification ou pseudonymisées. - La solution technique comprend un service d'indexation des patients qui repose sur la dépersonnalisation et sur un logiciel de couplage de données. - Ce service permet de classer les tables de données anonymisées ou pseudonymisées, et de les partager en toute sécurité aux chercheurs des institutions autorisées. - Tous les projets sont séparés les uns des autres dans un dépôt de données sécurisé (*Safe Haven*).
Farr North England		nd
SILC	Donnelly, 2015[31]	- Principe de séparation des fonctions (aucune des personnes qui participent au processus de couplage n'a accès à l'ensemble des données). - Le service de couplage est assuré par le *National Records of Scotland* (NRS) qui agit au nom des détenteurs de données (lesquels conservent leurs données). - Des clés de couplage spécifiques à chaque projet sont utilisées par le NRS pour remettre ensemble les données et les identifiants de chaque base de données avant de les transmettre pour le couplage. - Le *National Services Scotland* procède au couplage des bases de données anonymisées. - Les jeux de données couplés sont détruits à la fin des projets.

nd=information non disponible

Annexe 5 : Autres modalités d'autorisation des projets de recherche

Entités	Références	Modalités	Éléments évalués
BioGrid	Merriel, 2011[21]	✓ Demande d'accès révisée par deux scientifiques ✓ Révision finale par le *Management Committee*	➢ Importance de la question ➢ Qualité scientifique du projet ➢ Disponibilité des données ➢ Taille de l'échantillon
CDL	Boyd, 2012[10]	✓ Processus d'approbation du *PHRN*	➢ Le projet de recherche
	www.phrn.org.au[48]	✓ Approbation des unités régionales de couplage de données concernées	nd
CHeReL	www.cherel.org.au[43]	✓ Formulaire d'application	nd
CPRD	Herrett, 2015[16]	✓ Approbation d'accès par ISAC	➢ Le protocole de recherche
CVDL	Chen, 2016[30]	✓ Application	nd
	www2.health.vic.gov.au[38]		➢ La disponibilité des données ➢ La qualité des données et les limites ➢ La faisabilité du couplage des données
Farr London	www.ucl.ac.uk[41]	✓ Formulaire d'application	➢ La gouvernance de l'information
Farr North England	Lea, 2016[20]	✓ Comité indépendant de gouvernance	➢ Projet de recherche
ICES	Ishiguro, 2016[18]	✓ Formulaire	nd
	www.ices.on.ca[39]		➢ La faisabilité du projet ➢ Le calendrier de réalisation du projet
MCHP	Roos, 2008[23]	✓ Proposition examinée par le *Health Information Provincial committee* (HIPC) et les autres ministères impliqués	nd
	www://umanitoba.ca[40]		➢ La faisabilité du projet (incluant les coûts)
PopData	Pencarrick Hertzman, 2013[22]	✓ Demande d'accès	➢ La nécessité de l'accès aux données identifiables ➢ L'intérêt public ➢ La sécurité, l'utilisation et la destruction des données
RLG	Queensland Government, 2016[27]	✓ Application approuvée par le directeur général	➢ L'intérêt public
SA NT Datalink	Harrison, 2016[32]	✓ Formulaire d'application	nd
SAIL	Ford, 2009[15]; Jones, 2014[19]; Wellcome Trust, 2015[26]	✓ Demande d'accès évaluée par le *Collaboration Review System* ✓ Révision par HIRU et IGRP	➢ La pertinence ➢ La disponibilité des données ➢ Les exigences en termes de ressources ➢ La faisabilité du projet ➢ Les mesures de gouvernance de l'information
SILC	Lea, 2016[20]; Data Linkage Scotland, 2016[28]	✓ Processus d'application ✓ Approbation des chercheurs par l'*Electronic Data Research and Innovation Service* ✓ Formation accréditée ✓ Jury d'experts indépendants	➢ Les bénéfices pour la société ➢ La faisabilité du projet ➢ La pertinence scientifique
SSI	www.ssi.dk[35]	✓ Approbation des projets par le *Danish data Protection Agency*	nd
TDLU	Wiggins, 2016[34]	✓ Processus d'autorisation	nd
WALDS	Euro REACH, 2011[25]	✓ Formulaire d'application approuvé par le *Department of Health of Western Australia* et WALDS	nd
	www.datalinkage-wa.org[37]		➢ L'expérience et la formation des chercheurs ➢ L'intérêt public ➢ La qualité scientifique ➢ Les objectifs de santé et de bien-être ➢ Disponibilité des ressources et respect des lois et normes ➢ Le calendrier de réalisation du projet

nd= information non disponible

Annexe 6 : Consentement des participants selon les entités

Entités	Références	Type de consentement		
		Explicite	implicite	aucun
BioGrid	Merriel, 2011[21]	Recherche Intervention Données génétiques	Données cliniques	Données administratives
CHeReL	www.cherel.org.au[43]	na	na	Toutes les données
CPRD	Herrett, 2015[16]	na	Données des dossiers électroniques	na
	www.health.qld.gov.au[42]	na	na	Données identifiables ou potentiellement identifiables; si consentement impossible à obtenir
RLG	*Queensland Government, Department of Health*, 2016[27]	Informations personnelles ou d'identification	na	Données identifiables du département de la santé
SSI	Burgun, 2016[12]	na	Prélèvements systématiques	Données personnelles

na = non applicable

Annexe 7 : Engagements à respecter la vie privée des personnes

Entités	Références	Accords signés	Formations	Autres
BioGrid	Merriel, 2011[21]	Signature d'un accord concernant les exigences éthiques et la protection de la vie privée	nd	nd
Farr London	www.ucl.ac.uk[41]	nd	Formation sur la gouvernance des données + tests en ligne	nd
Farr North England	Lea, 2016[20]	nd	Formation sur la gouvernance des données	nd
ICES	www.ices.on.ca[39]	nd	Formation sur les politiques et procédures de confidentialité	nd
MCHP	www.umanitoba.ca[40]	nd	Formation sur la protection de la vie privée	nd
PopData	Pencarrick Hertzman, 2013[22]	nd	Formation sur la protection de la vie privée	nd
RLG	*Queensland Government, Department of Health*, 2016[27]	nd	nd	Demande d'autorisation approuvée en vertu de la Loi sur la santé publique
SAIL	Ford, 2009[15]	Signature d'ententes de comportement responsable et de conformité avec les mesures et politiques de sécurité	nd	nd
SILC	Data Linkage Scotland, 2016[28]	Signature des termes d'utilisation de la politique de violation de la vie privée	Formation sur la protection de la vie privée	Signature par une personne qui se porte garante
WALDS	Holman, 2008[17]	Signature d'une entente sur la sécurité des données	nd	nd

nd = information non disponible

Annexe 8 : Mesures de sécurité informatique et relatives aux requêtes et sorties

Entités	Références	Sécurité informatique	Sécurité des requêtes et des sorties
BioGrid	www.biogrid.org.au[45]	- Accès à l'aide d'un identifiant personnel	nd
	Merriel, 2011[21]	nd	- Monitorage des requêtes (audits)
CDL	Boyd, 2012[10]	- Réseau informatique autonome sécurisé basé sur des normes et des codes de pratique	- Régime strict de déclaration - Cadre de confidentialité - Cadre de gouvernance de l'information - Ententes contraignantes de diffusion
CHeReL	Cancer Institute NSW, 2009[24]	- Procédures d'accès aux services - Procédures pour assurer la sécurité électronique des données	nd
CPRD	Herrett, 2015[16]	- Serveurs sécurisés	nd
CVDL	www2.health.vic.gov.au[38]	nd	- Révision des publications par les détenteurs de données
	Lea, 2016[20]	nd	- Données détruites à la fin du projet - Révision de tout document avant publication
Farr London	Lea, 2016[20]	- Sessions virtuelles sécurisées - Infrastructure sécurisée - Authentification à double facteur	- Blocage des fonctions de téléchargement (copies, captures d'écran et collages)
Farr North England	Lea, 2016[20]	- Infrastructure sécurisée - Accès à distance sécurisé et à des outils et logiciels - Politiques et standards de sécurité du refuge sécurisé (*safe haven)* - Authentification à double facteur	- Infrastructure de monitorage
ICES	Ishiguro, 2016[18]	- Infrastructure de bureau virtuel sécurisé - Accès via connexion internet cryptée - Authentification multifactorielle - Mot de passe robuste - Limite d'accès aux répertoires et aux ressources autorisées	- Évaluation des risques avant le transfert ou la copie de données - Désactivation des périphériques de l'ordinateur
MCHP	http://umanitoba.ca[40]	nd	- Autorisation de diffusion nécessaire - Autorisation de certains détenteurs de données qui ne sont pas liées à la santé
PopData	Pencarrick Hertzman, 2013[22]	- Environnement de recherche sécurisé (SRE) - Pare-feu - Logiciel de détection des intrusions - Serveur virtuel avec outils et logiciels - Accès à distance avec authentification à double facteur et réseau virtuel privé (VPN)	- SRE empêche le transfert de données individuelles - SRE enregistre les accès et actions qui peuvent être audités - Surveillance des fichiers transférés
RLG	*Queensland Government, Department of Health*, 2016[27]	- Pare-feu - Fichier crypté et transport protégé par mot de passe - Mot de passe transmis par téléphone - Stockage sécuritaire des données obligatoire	- Application de la loi en vigueur (pénalités possibles)

Entités	Références	Sécurité informatique	Sécurité des requêtes et des sorties
SA NT DataLink	SA NT Datalink, 2016[29]	- Processus de détection d'intrusions - Contrôles d'accès et de l'environnement informatique - Authentification des utilisateurs - Cryptage - Pare-feu - Redondance et sauvegardes - Durcissement du système et de l'application - Accès avec le serveur sécurisé de l'Université	- Les données identifiées ne peuvent sortir de l'environnement le plus sécurisé - Stockage et monitorage des journaux de fichiers dans un emplacement sécurisé avec accès restreint - Monitorage des transferts de données - Contrôle des contenus et des utilisateurs
	Harrison, 2016[32]	nd	- Révision des publications par les détenteurs de données
SAIL	Ford, 2009[15]	- Terminal sécurisé - Produits d'analyses transférés dans un dépôt sécurisé	- Examen des sorties statistiques
	Jones, 2014[19]	- Passerelle SAIL Gateway : accès sécurisé à distance : - Pare-feu - Authentification à deux facteurs - Connexion réseau cryptée - Serveurs de sécurité	- Enregistrement des activités de l'utilisateur, suivi et analyse des risques de divulgation
SILC	Lea, 2016[20]; Data Linkage Scotland, 2016[28]	- Points d'accès sécurisés (safe haven) dans des institutions autorisées - Accès à distance via un réseau virtuel sécurisé	nd
	Donnelly, 2015[31]	- Dépôt de données sécurisé	- Aucune donnée individuelle ne peut sortir du dépôt
SSI	Burgun, 2016[12]	- Accès sécurisé - Système d'authentification basé sur le numéro unique d'identification	- Sorties contrôlées des résultats; données agrégées seulement
TDLU	Wiggins, 2016[34]	- Transport sécuritaire des données	- Révision des résultats par les détenteurs de données - L'entité doit être informée des publications ou communications de résultats
	www.menzies.utas.edu.au[44]	- Réseau autonome dans un environnement contrôlé - Protection par mot de passe stricte et multiniveaux - Logiciel antivirus - Cryptage pour le transfert de données	nd
WALDS	Eitelhuber, 2014[14]	- Accès électroniques restreints	nd
	www.datalinkage-wa.org.au [37]	- Transfert de données via des portails cryptés sécurisés ou par la livraison en personne - Cryptage des clés de couplage - Sauvegarde sécurisée (chiffrée) régulière des données avec stockage sécurisé hors site - Pare-feu avec protocoles de verrouillage automatique - Modification régulière des mots de passe de connexion	- Contrôle d'assurance qualité des données avant leur publication - Examen des sorties

nd = information non disponible

Annexe 9 : Mesures de sécurité physique est d'évaluation des processus de sécurité

Entités	Références	Sécurité physique	Évaluation des processus de sécurité
BioGrid	Merriel, 2011[21]	nd	- Deux audits indépendants
CDL	Boyd, 2009[10]	nd	- Processus d'audit indépendant en lien avec la sécurité de l'environnement informatique
CHeReL	Cancer Institute NSW, 2009[24]	- Mesures d'accès aux services	- Évaluation indépendante des risques associés aux technologies de l'information (3e année)
CPRD	www.cprd.com[146]	- Mesures mises en place par les utilisateurs	- Audits potentiels auprès des organisations qui reçoivent les données couplées
Farr London	Lea, 2016[20]	nd	- Audits annuels du fonctionnement de la solution technique
Farr North England	Lea, 2016[20]	- Entité située dans un environnement physique sécurisé	nd
MCHP	http://umanitoba.ca[40]	- Accès à l'établissement restreint et mesures supplémentaires pour les espaces hébergeant les données - Accès à distance limité à des emplacements sécurisés et désignés, et utilisation des équipements approuvés	nd
PopData	Pencarrick Hertzman, 2013[22]	- Établissement divisé en trois zones de sécurité - Serrures - Système d'alarme - Surveillance par caméra - Murs fortifiés	- Évaluation des pratiques de sécurité de l'information - Évaluation de la sécurité par une firme externe
RLG	*Queensland Government*, 2016[27]	- Coffre-fort verrouillé - Accès limité à la pièce où sont consultées les données	nd
SAIL	Ford, 2009[15]	nd	- Audits externes de conformités avec la gouvernance de l'information et formulation de recommandations
SA NT DataLink	SA NT Datalink, 2016[29]	- Toit sécurisé - Entrées sécurisées - Contrôle électronique et portes extérieures surveillées et contrôlées par électronique - Serrures - Sécurisation des conduits, ascenseurs et ouvertures - Conteneurs sécurisés dans une armoire à verrous pour les périphériques de stockage informatique	- Procédure d'évaluation des risques (internes, externes et incidents survenant naturellement) intégrée à la gestion opérationnelle
SILC	Lea, 2016[20]; Data Linkage Scotland, 2016[28]	- Points d'accès sécurisés d'institutions autorisées dont certains sont munis de surveillance vidéo	nd
TDLU	www.menzies.utas.edu.au[44]	- Contrôles physiques d'entrée - Accès limités	nd
WALDS	Eitelhuber, 2014[14]	- Accès via laboratoire sécurisé	nd
	www.datalinkage-wa.org.au[37]	- Carte de sécurité avec photo - Serveurs dans une pièce verrouillée avec accès restreint supplémentaire - Accès physique aux serveurs limité	- Examen indépendant de sécurité des protocoles de confidentialités, d'infrastructures techniques et de traitement des données suivis de recommandations

nd =information non disponible

Annexe 10 : Champs de compétences des entités et soutien offert aux utilisateurs

Entités	Références	**Compétences** Couplage et traitement des données	**Soutien** Accès aux données et couplage
BioGrid	Merriel, 2011[21]	- *Unique Subject Identifer* - Techniques de correspondance probabiliste - Optimisation des champs de données et formatage	nd
CHeReL	Cancer Institute NSW, 2009[24]	nd	- Site Internet avec guides pour les services offerts - Réunions avec les chercheurs pour assurer un soutien dans l'approbation des projets - Cours universitaires sur l'analyse des données couplées
CVDL	www2.health.vic.gov.au[38]	nd	- Soutien et conseils pour les demandes d'accès - Personnel coordonne et facilite le processus d'accès aux données
RLG	*Queensland Government, Department of Health*, 2016[27]	nd	- Soutien pour la soumission de projets
SA NT Datalink	www.santdatalink.org.au[47]	nd	- Soutien pour la demande d'accès aux données auprès des détenteurs - Facilite la rencontre entre les chercheurs et les détenteurs de données - Aide à la rédaction de la lettre de faisabilité
SSI	Burgun, 2016[12]	Expertises : - bases de données et variables - algorithmes de sélection de cas - méthodes de traitement des données - conditions d'interopérabilité sémantiques des bases de données et de l'appariement de données	nd
TDLU	Wiggins, 2016[34]	nd	- Soutien pour remplir l'application pour accéder aux données
WALDS	EuroREACH, 2011[25]	nd	- Soutien et formation pour l'utilisation des données

nd =information non disponible

Annexe 11 : Réseaux

Entités	Références	Standardisation de la collecte de données	Partage de données
BioGrid	Merriel, 2011[21]	- Au départ, des groupes composés de chercheurs, de cliniciens, de gestionnaires de données et de spécialistes ont déterminé les données qui seraient recueillies à chacun des sites participants - Disponibilité d'un glossaire des données qui est accessible librement - Travail entre les officiers de BioGrid et les détenteurs afin d'optimiser les champs de données et formater les données collectées	- Partage des données facilité par un processus de reconnaissance de la propriété intellectuelle des données
CDL	Boyd, 2012[10]	nd	- Groupe de travail sur le transfert de données - Membre du réseau *Population Health Research Network* (PHRN) national auquel participent tous les états et territoires, et qui vise à développer des infrastructures et des capacités pour valoriser les nombreuses collections de données de santé australiennes et faciliter la recherche
CHeReL	*Cancer Institute NSW*, 2009[24]	nd	- Participation au *National Collaborative Research Infrastructure Strategy Population Research Network* pour faciliter l'accès à des données provenant de différents états et territoires
CVDL	Chen, 2016[30]	nd	- Membre PHRN
Farr London	Lea, 2016[20]	nd	- Membre du *Farr Institute*
RLG	*Queensland Government*, 2016[27]	nd	- Membre du PHRN
SA NT DataLink	SA NT Datalink, 2016[47]	nd	- Membre du PHRN
SAIL	Lea, 2016[20]	nd	- Membre du *Farr Institute*
SILC	Lea, 2016[20]	nd	- Membre du *Farr Institute*
TDLU	Wiggins, 2016[34]	nd	- Membre du réseau PHRN
WALDS	Eitelhuber, 2014[14]	nd	- Membre du PHRN

nd =information non disponible

Annexe 12 : Mesures de performance

Entités	Références	Délais d'accès	Nombre de demandes traitées	Nombre de publications	Autres
BioGrid	Merriel, 2011[21]	nd	- 4000 requêtes de données/mois Pharmacologie : - 25 projets complétés - 9 en cours	87 publications scientifiques (2007-2009) 22 affiches	- Cinq chercheurs avec reconnaissances internationales - Développement de trois systèmes informatisés de gestion de maladies En pharmacologie : - 1,2 million $ en provenance des compagnies - 9 projets en cours - 1 brevet commercial
CHeReL	*Cancer Institute NSW*, 2009[24]	nd	- 57 projets complétés pendant la période de 3 ans - Juin 2009 : 30 projets en cours - Augmentation du nombre de requêtes répondues	nd	- Cinq bourses (3,9 millions $) Révision indépendante (2008) : - Mise en place rapide avec succès; - Utilisation importante et croissante des services de l'entité parmi les décideurs; - Appréciation des capacités de l'entité; - Reconnaissance croissante des services.
CPRD	Herrett, 2015[16]	nd	nd	(1987-2014) ≈2000 rapports de recherche ≈1000 articles scientifiques	nd
CVDL	Chen, 2016[30]	nd	- 150 projets de planification de politiques ou de recherches entre 2012 et 2016; - 26 projets en cours; - 46 projets en développement.	nd	nd

Entités	Références	Délais d'accès	Nombre de demandes traitées	Nombre de publications	Autres
ICES	Ishiguro, 2016[18]	- 2 à 4 sem. après l'accord; - Demandes d'analyses varient selon la complexité du projet.	*Cancer data linkage programme*: - 42 ensembles de données intégrées ont été produits *Data & Analytic services (DAS) platform* (mars 2014) : - 200 demandes (187 admissibles); - 136 demandes d'accès; - 24 requêtes de services d'analyse; - 34 requêtes d'importation de données couplées aux données d'ICES.	*Cancer data linkage programme* 34 publications	nd
PopData	Pencarrick Hertzman, 2013[22]	nd	- Plus de 350 projets de recherche (14 ans)	nd	nd
SAIL	Jones, 2014[19]	nd	- 100 projets approuvés depuis 2006 (2006-2013)	nd	75 points d'accès
WADLS	Brook, 2008[11]	nd	nd	- 177 présentations; - 172 publications scientifiques; - 159 articles multimédias; - 96 rapports; - 104 autres produits.	nd
	Eitelhuber, 2014[14]	- Moyenne de 39 jours	- Plus de 30 projets impliquant CARES (2011)	nd	nd
	www.datalinkage-wa.org.au[37]	nd	- Plus de 800 projets de 1995 à 2016	nd	nd
	Holman, 2008[17]	nd	- Plus de 400 projets sur 10 ans	- Plus de 250 publications scientifiques sur 10 ans	+ de 35 étudiants ont gradué ou sont en processus de l'être Fonds de recherche externes à WADLS : 58 millions$

nd = information non disponible

RÉFÉRENCES

Les références marquées d'un * sont incluses dans la revue systématique de la littérature.

1. Clair, M. *Mot du président, Alliance santé Québec*. 2015 [cited janvier 2016; Available from: https://www.alliancesantequebec.com/.
2. St-Jacques, S. and J. Dussault, *Meilleures pratiques de gouvernance d'une entité donnant accès à des données intégrées de recherche clinique, populationnelle et informationnelle. Plan de réalisation.*, UETMISSS-PL, Editor. 2017, CIUSSS de la Capitale-Nationale: Québec. p. 25 pages.
3. Candas, B., *Constitution et utilisation des banques de données intégrées en santé.* 2015, INSPQ: Québec.
4. Hovenga, E.J.S. and H. Grain, *Health Data and Data Governance.* Studies in healthtechnology and informatics, 2013(193): p. 67-92.
5. Jutte, D.P., L.L. Roos, and M.D. Brownell, *Administrative Record Linkage as a Tool for Public Health Research.* Annual Review of Public Health, 2011(32): p. 91–108.
6. Egelstaff, R. and M. Wells, *Data governance frameworks and change management.* Studies in health technology and informatics, 2013(193): p. 108-119.
7. Couture, M. *L'évaluation de la crédibilité des documents en ligne.* 2015 [cited 2017; Available from: http://benhur.teluq.ca/ST/sciences/sci1021/evalweb.htm.
8. *Accessing health and health-related data in Canana*, C.c.d. académies, Editor. 2015, The expert panel on timely access to health and social data for health research and health system innovation: Ottawa, Canada. p. 208.
9. *Stokes, B., *Data Linkage in Tasmania.* Tasmanian data Linkage Unit.
10. *Boyd, J.H., et al., *Data linkage infrastructure for cross-jurisdictional health-related research in Australia.* BMC Health Services Research, 2012. **12**(1): p. 480.
11. *Brook, E.L., D.L. Rosman, and C.D. Holman, *Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System.* Australian and New Zealand Journal of Public Health, 2008. **32**(1): p. 19-23.
12. *Burgun, A., et al., *Partage de données patients pour la recherche : aspects organisationnels et éthiques.* Ethics, Medicine and Public Health, 2016. **2**: p. 435-441.
13. *Chamberlayne, R., et al., *Creating a Population-based Linked Health Database: A New Resource for Health ServicesResearch.* Canadian Journal of Public Health / Revue Canadienne de Santé Publique, 1998. **89**(4): p. 270-273.
14. *Eitelhuber, T. and G. Davis, *The custodian administered research extract server: "improving the pipeline" in linked data delivery systems.* Health Information Science and Systems [Electronic Resource], 2014. **2**(6).
15. *Ford, D.V., et al., *The SAIL Databank: building a national architecture for e-health research and evaluation.* BMC Health Services Research, 2009. **9**(157).
16. *Herrett, E., et al., *Data Resource Profile: Clinical Practice Research Datalink. .* International Journal of Epidemiology, 2015. **44**(3): p. 827-836.
17. *Holman, C.D., et al., *A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system.* Australian health review : a publication of the Australian Hospital Association, 2008. **32**(4): p. 766-777.
18. *Ishiguro, L., et al., *Increasing Access to Health Administrative Data with ICES Data & Analytic Services. .* Healthcare Quarterly, 2016. **19**(1): p. 7-9.

19. *Jones, K.H., et al., *A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: a privacy-protecting remote access system for health-related research and evaluation.* Journal of biomedical informatics, 2014. **50**: p. 196-204.
20. *Lea, N.C., et al., *Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research.* JMIR medical informatics, 2016. **4**(22): p. e22.
21. *Merriel, R.B., et al., *BioGrid Australia facilitates collaborative medical and bioinformatics research across hospitals and medical research institutes by linking data from diverse disease and data types. .* Human Mutation, 2011. **32**(5): p. 517-525.
22. *Pencarrick Hertzman, C., N. Meagher, and K.M. McGrail, *Privacy by Design at Population Data BC: a case study describing the technical, administrative, and physical controls for privacy-sensitive secondary use of personal information for research in the public interest.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 25.
23. *Roos, L.L., et al., *From health research to social research: Privacy, methods, approaches.* Social Science & Medicine, 2008. **66**(1): p. 117-129.
24. **Center for Health Record Linkage. The first Three Years 2006-07 to 2008-09.* 2009, Cancer Institute NSW: Australia.
25. **Good practice on data linkages and performance measurement in relation to access to national health care data systems.* 2011, EuroREACH: EuroREACH WP3 Advisors meeting. Tel Aviv, Israel.
26. **Enabling Data Linkage to Maximise the Value of Public Health Research Data: full report.*, in *Public Health Research Data Forum* 2015, Wellcome Trust: England and Wales.
27. **Queensland Data Linkage Framework.*, Q. Health, Editor. 2016: Brisbane, Australia: State of Queensland.
28. **Data Linkage Scotland. Information and case studies about data linkage research and services.* 2016 8 septembre 2017].
29. **SA NT Datalink. Security Manual -Overview.* 2016, University of South Australia: South Australia.
30. *Chen, Y., *Data Linkage in Victoria. Unlocking the power of data linkage.* 2016.
31. *Donnelly, G., *Scottish Informatics and Linkage Collaboration*. 2015.
32. *Harrison, J., *Using Data Linkage.*, F.U.C. Research Center for injury Studies, Steering Committee SA NT Datalink, Editor. 2016.
33. *Lawrence, G., I. Dinh, and L. Taylor, *The Centre for Health Record Linkage: a new resource for health services research and evaluation.* Healt Information Management Journal, 2008. **37**(2): p. 60-62.
34. *Wiggins, N., *Applying for Linked data.* 2016, Tasmanian data Linkage Unit.
35. *www.ssi.dk*. [cited 2017 8 septembre].
36. L.W., D. and K.A. Ross, *ICES Report 2014 Prescribed Entity Review.* 2014: Institute for Clinical Evaluative Sciences.
37. *www.datalinkage-wa.org.au*. [cited 2017 septembre 2017].
38. *www2.health.vic.gov.au*. [cited 2017 12 juin].
39. *www.ices.on.ca*. [cited 2017 12 juin].
40. *www.umanitoba.ca*. [cited 2017 12 juin].
41. *www.ucl.ac.uk*. [cited 2017 12 juin].
42. *www.health.qld.gov.au*. [cited 2017 17 septembre].
43. *www.cherel.org.au*. [cited 2017 18 septembre].
44. *www.menzies.utas.edu.au*. [cited 2017 13 juin].
45. *www.biogrid.org.au*. [cited 2017 13 juin].

46. *www.cprd.com*. [cited 2017 16 septembre].
47. *www.santdatalink.org.au*. [cited 2017 16 septembre].
48. *www.phrn.org.au*. [cited 2017 16 septembre].
49. *www.farrinstitute.org*. [cited 2017 18 septembre].
50. Yiannakoulias, N., *Understanding Identifiabilitly in Secondary Health Data.* Canadian Journal of Public Health / Revue Canadienne de Santé Publique, 2011. **102**(4): p. 291-93.