Confidentialité des données au sein des projets de recueil et d'analyse de données en soins primaires : une revue systématique de la littérature

Thomas Le Berre

N° d'ordre : **ANNÉE 2018**

UNIVERSITÉ DE RENNES 1

UNIVERSITE BRETAGNE LOIRE

THÈSE D'EXERCICE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Bretagne Loire

Thèse en vue du

DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE

présentée par

Thomas LE BERRE

Né(e) le 17 juin 1981 à Marseille

Confidentialité des données au sein des projets de recueil et d'analyse de données en soins primaires : une revue systématique de la littérature.

Thèse soutenue à Rennes

le 10 octobre 2018

devant le jury composé de :

Marc CUGGIA
PU-PH − CHU de Rennes / *Président*

Bruno LAVIOLLE
PU-PH − CHU de Rennes / *Juge*

Ronan GARLANTEZEC
MCU-PH − CHU de Rennes / *Juge*

Marie-Line GENTIL
CCU-MG − Faculté de Médecine de Rennes / *Directrice de thèse*

Professeurs des Universités - Praticiens Hospitaliers

Nom Prénom	Sous-section CNU
ANNE-GALIBERT Marie-Dominique	Biochimie et biologie moléculaire
BARDOU-JACQUET Edouard	Gastroentérologie; hépatologie; addictologie
BELAUD-ROTUREAU Marc-Antoine	Histologie; embryologie et cytogénétique
BELLISSANT Eric	Pharmacologie fondamentale; pharmacologie clinique; addictologie
BELOEIL Hélène	Anesthésiologie-réanimation; médecine d'urgence
BENDAVID Claude	Biochimie et biologie moléculaire
BENSALAH Karim	Urologie
BEUCHEE Alain	Pédiatrie
BONAN Isabelle	Médecine physique et de réadaptation
BONNET Fabrice	Endocrinologie, diabète et maladies métaboliques; gynécologie médicale
BOUDJEMA Karim	Chirurgie générale
BOUGET Jacques Professeur des Universités en surnombre	Thérapeutique; médecine d'urgence; addictologie
BOUGUEN Guillaume	Gastroentérologie; hépatologie; addictologie
BOURGUET Patrick Professeur des Universités Emérite	Biophysique et médecine nucléaire
BRASSIER Gilles	Neurochirurgie
BRETAGNE Jean-François Professeur des Universités Emérite	Gastroentérologie; hépatologie; addictologie

BRISSOT Pierre Professeur des Universités Emérite	Gastroentérologie; hépatologie; addictologie
CARRE François	Physiologie
CATROS Véronique	Biologie cellulaire
CATTOIR Vincent	Bactériologie-virologie; hygiène hospitalière
CHALES Gérard Professeur des Universités Emérite	Rhumatologie
CORBINEAU Hervé	Chirurgie thoracique et cardiovasculaire
CUGGIA Marc	Biostatistiques, informatique médicale et technologies de communication
DARNAULT Pierre	Anatomie
DAUBERT Jean-Claude Professeur des Universités Emérite	Cardiologie
DAVID Véronique	Biochimie et biologie moléculaire
DAYAN Jacques (Professeur associé)	Pédopsychiatrie; addictologie
DE CREVOISIER Renaud	Cancérologie; radiothérapie
DECAUX Olivier	Médecine interne; gériatrie et biologie du vieillissement; addictologie
DESRUES Benoît	Pneumologie; addictologie
DEUGNIER Yves Professeur des Universités en surnombre + Consultanat	Gastroentérologie; hépatologie; addictologie
DONAL Erwan	Cardiologie
DRAPIER Dominique	Psychiatrie d'adultes; addictologie
DUPUY Alain	Dermato-vénéréologie

ECOFFEY Claude	Anesthésiologie-réanimation; médecine d'urgence
EDAN Gilles	Neurologie
FERRE Jean Christophe	Radiologie et imagerie Médecine
FEST Thierry	Hématologie; transfusion
FLECHER Erwan	Chirurgie thoracique et cardiovasculaire
FREMOND Benjamin	Chirurgie infantile
GANDEMER Virginie	Pédiatrie
GANDON Yves	Radiologie et imagerie Médecine
GANGNEUX Jean-Pierre	Parasitologie et mycologie
GARIN Etienne	Biophysique et médecine nucléaire
GAUVRIT Jean-Yves	Radiologie et imagerie Médecine
GODEY Benoit	Oto-rhino-laryngologie
GUGGENBUHL Pascal	Rhumatologie
GUIGUEN Claude Professeur des Universités Emérite	Parasitologie et mycologie
GUILLÉ François	Urologie
GUYADER Dominique	Gastroentérologie; hépatologie; addictologie
HAEGELEN Claire	Anatomie

HOUOT Roch	Hématologie; transfusion
HUSSON Jean-Louis Professeur des Universités Emérite	Chirurgie orthopédique et traumatologique
HUTEN Denis Professeur des Universités Emérite	Chirurgie orthopédique et traumatologique
JEGO Patrick	Médecine interne; gériatrie et biologie du vieillissement; addictologie
JEGOUX Franck	Oto-rhino-laryngologie
JOUNEAU Stéphane	Pneumologie; addictologie
KAYAL Samer	Bactériologie-virologie; hygiène hospitalière
KERBRAT Pierre, RETRAITE	Cancérologie; radiothérapie
LAMY DE LA CHAPELLE Thierry	Hématologie; transfusion
LAVIOLLE Bruno	Pharmacologie fondamentale; pharmacologie clinique; addictologie
LAVOUE Vincent	Gynécologie-obstétrique; gynécologie médicale
LE BRETON Hervé	Cardiologie
LE GUEUT Mariannick Professeur des Universités en surnombre + consultanat	Médecine légale et droit de la santé
LE TULZO Yves	Réanimation; médecine d'urgence
LECLERCQ Christophe	Cardiologie
LEDERLIN Mathieu	Radiologie et imagerie Médecine

LEGUERRIER Alain Professeur des Universités Emérite	Chirurgie thoracique et cardiovasculaire
LEJEUNE Florence	Biophysique et médecine nucléaire
LEVEQUE Jean	Gynécologie-obstétrique; gynécologie médicale
LIEVRE Astrid	Gastroentérologie; hépatologie; addictologie
MABO Philippe	Cardiologie
MAHE Guillaume	Chirurgie vasculaire ; médecine vasculaire
MALLEDANT Yannick Professeur des Universités Emérite	Anesthésiologie-réanimation; médecine d'urgence
MENER Eric (Professeur associé)	Médecine générale
MEUNIER Bernard	Chirurgie digestive
MICHELET Christian Professeur des Universités en surnombre	Maladies infectieuses; maladies tropicales
MOIRAND Romain	Gastroentérologie; hépatologie; addictologie
MORANDI Xavier	Anatomie
MOREL Vincent (Professeur associé)	Epistémologie clinique
MOSSER Jean	Biochimie et biologie moléculaire
MOURIAUX Frédéric	Ophtalmologie
MYHIE Didier (Professeur associé)	Médecine générale
ODENT Sylvie	Génétique

OGER Emmanuel	Pharmacologie fondamentale; pharmacologie clinique; addictologie
PARIS Christophe	Médecine et santé au travail
PERDRIGER Aleth	Rhumatologie
PLADYS Patrick	Pédiatrie
RAVEL Célia	Histologie; embryologie et cytogénétique
REVEST Matthieu	Maladies infectieuses; maladies tropicales
RICHARD de LATOUR Bertrand (Professeur associé)	Chirurgie thoracique et cardiovasculaire
RIFFAUD Laurent	Neurochirurgie
RIOUX-LECLERCQ Nathalie	Anatomie et cytologie pathologiques
ROBERT-GANGNEUX Florence	Parasitologie et mycologie
ROPARS Mickaël	Chirurgie orthopédique et traumatologique
SAINT-JALMES Hervé	Biophysique et médecine nucléaire
SAULEAU Paul	Physiologie
SEGUIN Philippe	Anesthésiologie-réanimation; médecine d'urgence
SEMANA Gilbert	Immunologie
SIPROUDHIS Laurent	Gastroentérologie; hépatologie; addictologie
SOMME Dominique	Médecine interne; gériatrie et biologie du vieillisement; addictologie

SOULAT Louis (Professeur associé)	Thérapeutique; médecine d'urgence; addictologie
SULPICE Laurent	Chirurgie générale
TADIÉ Jean Marc	Réanimation; médecine d'urgence
TARTE Karin	Immunologie
TATTEVIN Pierre	Maladies infectieuses; maladies tropicales
TATTEVIN-FABLET Françoise (Professeur associé)	Médecine générale
THIBAULT Ronan	Nutrition
THIBAULT Vincent	Bactériologie-virologie; hygiène hospitalière
THOMAZEAU Hervé	Chirurgie orthopédique et traumatologique
TORDJMAN Sylvie	Pédopsychiatrie; addictologie
VERHOYE Jean-Philippe	Chirurgie thoracique et cardiovasculaire
VERIN Marc	Neurologie
VIEL Jean-François	Epidémiologie, économie de la santé et prévention
VIGNEAU Cécile	Néphrologie
VIOLAS Philippe	Chirurgie infantile
WATIER Eric	Chirurgie plastique, reconstructrice et esthétique; brûlologie
WODEY Eric	Anesthésiologie-réanimation; médecine d'urgence

Maîtres de Conférences des Universités - Praticiens Hospitaliers

Nom Prénom	Sous-section CNU
ALLORY Emmanuel (Maître de conférence associé des universités de MG)	Médecine générale
AME-THOMAS Patricia	Immunologie
AMIOT Laurence (Baruch)	Hématologie; transfusion
ANSELMI Amédéo	Chirurgie thoracique et cardiovasculaire
BEGUE Jean-Marc	Physiologie
BERTHEUIL Nicolas	Chirurgie plastique, reconstructrice et esthétique ; brûlologie
BOUSSEMART Lise	Dermato-vénéréologie
CABILLIC Florian	Biologie cellulaire
CAUBET Alain	Médecine et santé au travail
CHHOR-QUENIART Sidonie (Maître de conférence associé des universités de MG)	Médecine générale
DAMERON Olivier	Informatique
DE TAYRAC Marie	Biochimie et biologie moléculaire
DEGEILH Brigitte	Parasitologie et mycologie
DROITCOURT Catherine	Dermato-vénéréologie
DUBOURG Christèle	Biochimie et biologie moléculaire
DUGAY Frédéric	Histologie; embryologie et cytogénétique
EDELINE Julien	Cancérologie; radiothérapie

FIQUET Laure (Maître de conférence associé des universités de MG)	Médecine générale
GARLANTEZEC Ronan	Epidémiologie, économie de la santé et prévention
GOUIN Isabelle épouse THIBAULT	Hématologie; transfusion
GUILLET Benoit	Hématologie; transfusion
JAILLARD Sylvie	Histologie; embryologie et cytogénétique
KALADJI Adrien	Chirurgie vasculaire; médecine vasculaire
LAVENU Audrey	Sciences physico-chimiques et technologies pharmaceutiques
LE GALL François	Anatomie et cytologie pathologiques
LEMAITRE Florian	Pharmacologie fondamentale; pharmacologie clinique; addictologie
MARTINS Pédro Raphaël	Cardiologie
MATHIEU-SANQUER Romain	Urologie
MENARD Cédric	Immunologie
MOREAU Caroline	Biochimie et biologie moléculaire
MOUSSOUNI Fouzia	Informatique
NAUDET Florian	Thérapeutique ; médecine d'urgence ; addictologie
PANGAULT Céline	Hématologie; transfusion
RENAUT Pierric (maître de conférence associé des universités de MG)	Médecine générale
ROBERT Gabriel	Psychiatrie d'adultes; addictologie
SCHNELL Frédéric	Physiologie

THEAUDIN Marie épouse SALIOU	Neurologie
TURLIN Bruno	Anatomie et cytologie pathologiques
VERDIER Marie-Clémence (Lorne)	Pharmacologie fondamentale; pharmacologie clinique; addictologie
ZIELINSKI Agata	

Remerciements

Je remercie **Monsieur le Professeur Marc CUGGIA**

pour avoir accepté de présider le jury et d'évaluer cette thèse.

Je remercie **Monsieur le Professeur Bruno LAVIOLLE**

pour avoir accepté de participer au jury et d'évaluer cette thèse.

Je remercie **Monsieur le Docteur Ronan GARLANTEZEC**

pour avoir accepté de participer au jury et d'évaluer cette thèse.

Je remercie **Madame le Docteur Marie-Line GENTIL** pour avoir accepté que nous travaillions ensemble sur ce sujet, pour m'avoir soutenu et aidé tout du long et pour sa patience indéfectible.

Merci à toutes celles et ceux qui ont su m'apprendre au cours de mes stages hospitaliers, toutes ces personnes dont j'ai croisé le chemin à un moment, ASH, AS, IDE, Internes, PH… trop nombreuses pour que je les cite, je m'efforce de mettre en pratique vos enseignements.

Merci à Alexandra ELBAKYAN.

Merci aux Docteurs Pascale ROBLIN et Jean-Paul LAPIERRE qui chacun à leur manière m'ont donné l'envie de faire ce métier.

Merci au Docteur Marina OGIER pour sa rigueur et sa patience.

Merci au Docteur Faustine SAIGOT pour son accompagnement, son écoute et ses conseils.

Merci à ma famille et à mes amis, pour leur soutien et pour les bon moments passés et à venir malgré les kilomètres.

Merci à Hanan qui me supporte quotidiennement !

À Leïla.

Table des matières

Table des illustrations

Table des abréviations

ACHIL :Ambulatory Care Health Research Laboratory

AMSTAR : A MeaSurement Tool to Assess systematic Reviews

ASIP Santé : Agence des systèmes d''information partagée de santé

CAPriCORN : Chicago Area Patient Centered Outcomes Research Network

CEREES : Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé

CNIL : Commission nationale de l'informatique et des libertés

CPCSSN : Canadian Primary Care Sentinel Surveillance Network (Réseau canadien de surveillance sentinelle en soins primaires)

CPRD : Clinical Practice Research Datalink

DHHS : Department of Health & Human Services (Etats-Unis)

DMSP : Dossiers médicaux de soins primaires

EMRALD : Electronic Medical Record Administrative data Linked Database

FDA : Food and Drugs Administration (Etats-Unis)

GRAPHC : National centre for geographic & resource analysis in primary health care

ICES : Institute for Clinical Evaluative Sciences (Canada)

INDS : Institut national des données de santé

INSERM : Institut national de la santé et de la recherche médicale

INTEGO : Integrated computerized network

IRB : Institutional Review Board (terme américain)

ISAC : Independent Scientific Advisory Committee

MeSH : Medical Subject Headings

MHRA : Medicines and Healthcare products Regulatory Agency (Royaume-Uni)

NHS : National Health Service (Royaume-Uni)

NIVEL : Netherlands institute for health services research

PRISMA : Preferred Reporting Items for Systematic Reviews and Meta-Analyses

REB : Research Ethic Board (terme canadien)

REC : Research Ethic Committee (terme anglais)

RGPD : Règlement Général sur la Protection des Données (Union Européenne)

1 Introduction

L'exploitation automatisée des données de santé issues des bases de données des logiciels de gestion de dossiers médicaux en soins primaires est un enjeu important afin de développer la recherche en soins primaires. De nombreuses bases de données de ce type existent déjà (1) :

- dans les pays anglo-saxons : CPRD, THIN et Qresearch au Royaume-Uni ; the Veterans Health Administration Datawarehouse aux Etats-Unis ; EMRALD, CPCSSN au Canada.
- dans les pays Européens : le Fire Project en Suisse, NIVEL-PCD ou IPCI aux Pays-Bas, INTEGO en Belgique, SIDIAP en Espagne.

En France, de rares initiatives locales existent comme Primege PACA (2) mais restent confinées à de petites échelles.

Les données issues des dossiers médicaux de soins primaires (DMSP) permettent d'obtenir une synthèse de l'histoire médicale des patients et une vue globale de la santé de la population. Si elles sont particulièrement intéressantes dans plusieurs domaines de recherche, elles restent des données sensibles et nécessitent une protection adaptée.

Notre revue systématique de la littérature a donc recherché les méthodes de protection de la confidentialité des données au sein de projets de recueil de données existant à l'international dans la perspective de la création d'un projet de recueil de données de routine de soins primaires en France. Cette étude a exploré cinq axes de la protection des données : les acteurs intervenant sur le projet, les modalités de recueil de consentement des patients, les algorithmes de dé-identification des données, les droits d'accès aux données, les cadres réglementaire et éthique autour de ces projets (3).

La responsabilité de la protection de ces entrepôts de données est une composante fondamentale et est nécessaire à la confiance accordée aux projets par les patients et les médecins. Quelles sont les différents acteurs impliqués dans les projets de recueil de données de soins primaires en routine ?

La confidentialité et la protection de la vie privée sont des préoccupations importantes des patients (4,5). Dans quelle mesure leur consentement est-il pris en compte ?

Quels algorithmes de dé-identification ou d'anonymisation sont mis en place ?

Comment sont gérées les autorisations d'accès aux données issues des dossiers médicaux informatisés ?

Enfin, depuis le mois de mai 2018, les états membres de l'Union Européenne appliquent le Règlement Général sur la Protection des Données (RGPD) (6). Aux Etats-Unis, un règlement de ce type existe déjà depuis 1996. Dans quelle mesure les principales bases de données se conforment-t-elles à ces contraintes ?

2 Matériel et méthodes

Une revue systématique de la littérature internationale a été réalisée de janvier à juin 2018 sur la base des critères de Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA (7)).

Les critères 1, 8, 9, 10, 11, 12 et 15 de la grille AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews) (8), n'étaient pas pertinents dans le cadre de notre revue, car notre travail ne portait pas sur des essais cliniques randomisés et les méta-analyses comme précisé ensuite dans nos critères d'inclusion. Tous les autres critères de la grille AMSTAR2 ont été respectés.

2.1 Sources d'information

Tout d'abord, une recherche automatisée de la littérature dans la base de données PubMed ainsi que dans Google Scholar a été réalisée. Les requêtes ont été formulées par deux auteurs.

Requête PubMed

("Electronic Health Records"[Mesh] OR "Medical records systems, computerized"[Mesh] OR "Electronic Medical Record"[All Fields])
AND
("Data Anonymization"[Mesh] OR Pseudonymisation[All Fields] OR "de-identification"[All Fields] OR "Privacy"[Mesh] OR "Confidentiality"[Mesh] OR "Informed Consent"[Mesh] OR "Patient Rights"[Mesh] OR "Access to Information"[Mesh] OR "Ethics Committees, Research"[Mesh] OR "Computer Security"[Mesh])
AND
("General Practice"[Mesh] OR "General Practitioners"[Mesh] OR "Primary Health Care"[Mesh] OR "Physicians, Primary Care"[Mesh] OR "Ambulatory care"[Mesh] OR "Ambulatory Care Facilities"[Mesh] OR "Primary care"[All Fields]) AND ("2010/01/01"[PDAT] : "3000/12/31"[PDAT])

Requête Google Scholar

Since 2010

privacy consent OR anonymization OR pseudonymization OR "de identification" OR Confidentiality OR "Access to Information" OR "Ethics Committees" OR "data protection" "primary care database"

Figure 1. ***Formulation des équations de recherche***

2.1.1 PubMed

Afin d'identifier tous les articles correspondant à notre recherche, notre requête a été structurée en 3 parties (Figure 1) :

1. le terme « electronic health records » (Dossiers Médicaux Electroniques) et les termes synonymes du MeSH ;

2. le terme « data anonymization » (Anonymisation de données) et leurs synonymes, les termes « privacy » et « confidentiality », « informed consent » et « patients rights », « access to information », « ethics committees », « computer security » ;

3. Les termes « general practice » et les synonymes en lien avec les soins primaires.

2.1.2 Google Scholar

Nous avons formulé notre requête à l'aide des termes « privacy », « consent », « anonymization », « pseudonymization », « de identification », « confidentiality », « access to information », « ethics committees », « data protection » et « primary care database » (Figure 1).

L'équation de recherche Google Scholar apparait plus simple que celle utilisée pour la recherche PubMed, cela est dû au fait que Google Scholar prend en compte automatiquement les synonymes des termes utilisés.

2.1.3 Recherches complémentaires

Secondairement à nos deux requêtes, nous avons identifié sept projets de recueil de données issues des dossiers médicaux de soins primaires. Les données des sites internet officiels de chaque projet extrait ont été intégrées à notre étude. Lorsque le site web n'était pas cité dans les références de l'article, nous avons utilisé le moteur de recherche Google en associant le nom de la base au mot-clé « database » ou « primary care database ».

Nous avons aussi analysés les sites web de réglementation de la protection des données Européen et des Etats-Unis.

2.2 Sélection des articles

2.2.1 Critères d'inclusion

Les articles ont été retenus s'ils concernaient des projets de recueil de données en soins primaires, et s'ils comportaient des informations sur la protection et la confidentialité des données personnelles et de la vie privée. Des études expérimentales ont été incluses car elles étaient destinées à être appliquées à des projets en lien avec les soins primaires (preuve de concept).

Les projets de recueil de données issues des dossiers médicaux informatisés de soins primaires sont définis par ces critères:

- des réseaux qui collectent les dossiers médicaux informatisés (et non des dossiers papiers)

- des réseaux qui effectuent une collecte automatisée et non manuelle des données

- sur des dossiers médicaux informatisés de médecins généralistes

- à l'échelle régionale ou nationale

Les projets contenant des données à la fois de soins primaires et secondaires étaient inclus si l'étude portait sur les données de soins primaires.

Nous avons inclus des revues de littérature, des articles scientifiques et des éléments de la littérature grise, via les sites web des projets extraits et les sites web de réglementation.

2.2.2 Critères d'exclusion

Les articles dont la publication est antérieure à 2010 ont été exclus en raison de l'introduction de nouveaux termes dans le MeSH, notamment « electronic medical record » en 2010. Cette limite permet donc de recueillir des articles plus pertinents et de centrer la recherche sur des projets de recueil de données actifs.

Nous avons également exclu les articles rédigés dans une langue autre que l'anglais ou le français, ceux qui ne relevaient pas des critères d'inclusion, ceux dont le texte intégral n'a pas pu être obtenu en raison de l'absence d'abonnement à la revue.

2.3 Collecte des données

Les deux requêtes ont été lancées indépendamment par deux des auteurs, qui ont ensuite lu les résumés afin de sélectionner les articles pertinents sur la base des critères d'inclusion et d'exclusion. Une liste des articles retenus à partir des résumés a été élaborée par consensus. Ensuite, ces deux auteurs ont lu de façon indépendante le texte intégral de chacun de ces articles afin de confirmer qu'ils correspondaient aux critères de sélection. La liste finale a été obtenue par consensus.

Par la suite, chaque article a été étudié par les deux auteurs afin d'en extraire les données au sein d'un formulaire standardisé. Ce formulaire était constitué de 5 parties : les acteurs des projets, les modalités de consentement des patients, les méthodes de dé-identification des données, les conditions d'accès aux données pour les chercheurs, et les considérations éthiques et réglementaires.

Le formulaire a également été complété à l'aide des sites web officiels des projets et des textes réglementaires de l'Union Européenne et des Etats-Unis concernant la protection des données.

3 Résultats

251 articles ont été identifiés via PubMed et 926 via Google Scholar.

Après lecture du titre et du résumé de chaque article, 26 références de PubMed et 11 références de Google Scholar ont été retenues. Un article a été identifié dans les deux bases de données.

Après lecture du texte intégral, 23 articles ont été conservés.

Sept sites web officiels des projets de recueil extraits, ainsi que le site web de la plateforme PopMedNet, ont été étudiés et leur contenu a été ajouté au formulaire.

Les textes officiels Européens et des Etats-Unis concernant la protection des données ont été intégrés à notre analyse.

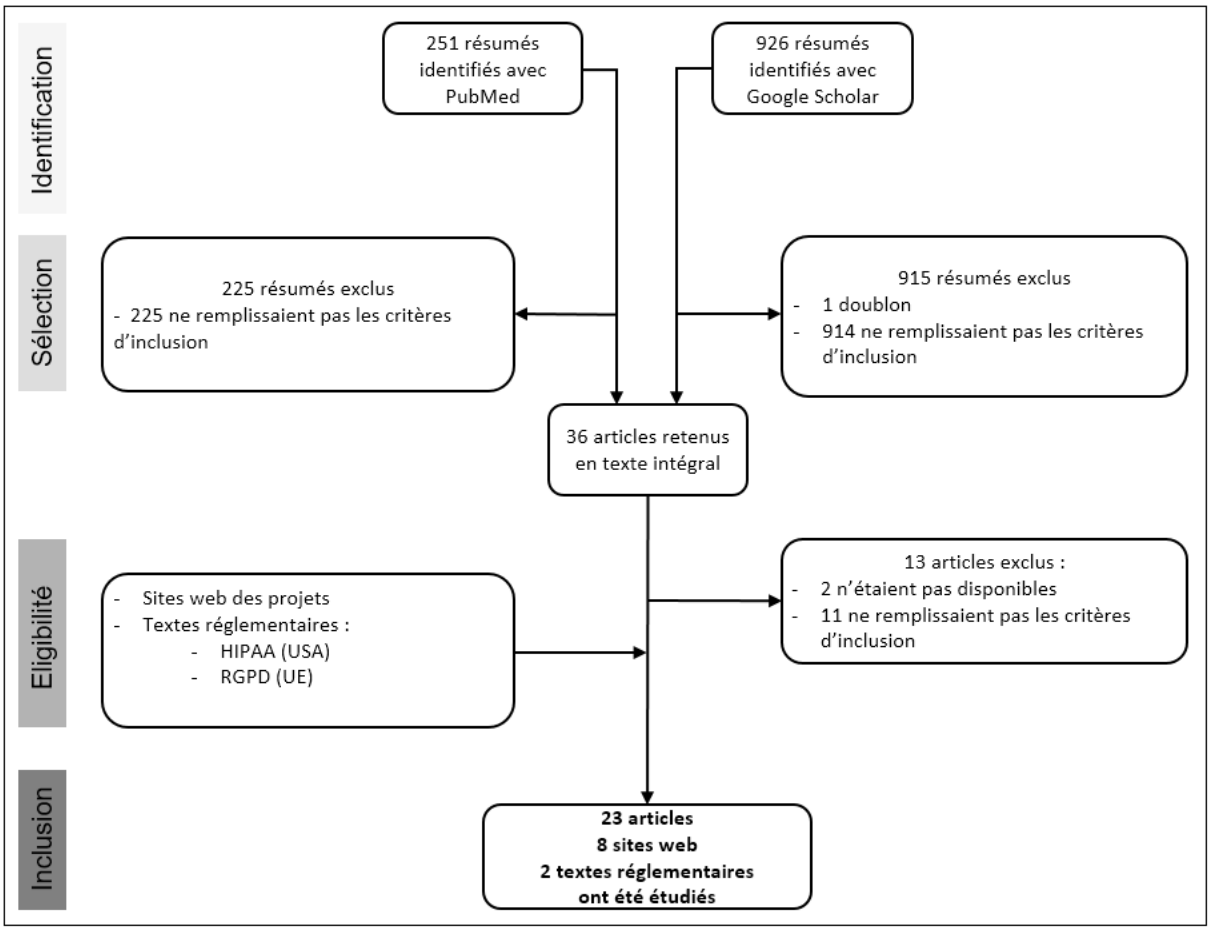

Figure 2. ***Diagramme de flux***

3.1 Description des projets de recueil de données

3.1.1 CAPriCORN

Le Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) (9) est un réseau de recueil de données à Chicago aux Etats-Unis (USA) qui a pour objectif de développer, tester et mettre en œuvre des politiques et des programmes pour améliorer la qualité des soins, l'état de santé et l'équité en santé des populations de la région de Chicago.

3.1.2 EMRALD

L'Electronic Medical Record Administrative data Linked Database (EMRALD) (10) est une base de données hébergée à l'Institute for Clinical Evaluative Sciences (ICES). Cette base est constituée d'informations cliniques issues des dossiers médicaux électroniques des médecins généralistes exerçant en Ontario au Canada. Les données peuvent être chainées aux bases de données administratives auxquelles l'ICES a accès. EMRALD a pour objectif d'évaluer le système de soins et d'améliorer l'état de santé de la population en Ontario.

3.1.3 CPCSSN

Le Réseau canadien de surveillance sentinelle en soins primaires CPCSSN (ou RCSSSP, acronyme francophone) (11) est un réseau de surveillance de maladies chroniques utilisant les dossiers informatisés de soins primaires du Canada. Son objectif est de mieux comprendre ces maladies chroniques afin d'en améliorer la prise en charge en soins primaires. Le programme CPCSSN a reçu le Prix d'innovation en protection de la vie privée 2013 de l'Association internationale des professionnels de la protection de la vie privée (IAPP). Ce réseau est un pionnier du concept de « privacy by design » qui sera développé dans la partie 3.4.1.1.

3.1.4 GRAPHC

GRAPHC (12) est un projet du National Centre for Geographic Resources & Analysis in Primary Health Care au sein de l'Australian National University, qui a pour but de renforcer la recherche en soins primaires en fournissant des services permettant des analyses géographiques.

3.1.5 CPRD

Au Royaume-Uni, le Clinical Practice Research Datalink (13) est un projet de recueil de données observationnel et interventionnel dont les données les plus anciennes datent de 1987. Il a pour but de fournir aux chercheurs des données anonymisées issues de soins primaires.

3.1.6 INTEGO

INTEGO (14) est un projet Belge de collecte de données issues de soins primaires de la région Flamande regroupant des données cliniques, biologiques et thérapeutiques. Il est rattaché au département de Médecine Générale de l'Université Catholique de Louvain. Son objectif est de faciliter la recherche en soins primaires en fournissant aux chercheurs des données issues des cabinets de médecine générale. ACHIL(15) est un projet associé à INTEGO.

3.1.7 NIVEL

NIVEL Primary Care Database (16) est un projet de recueil de données issues des soins primaires aux Pays-Bas destiné à observer la santé et l'usage des services de santé d'un échantillon représentatif de la population du pays.

3.2 Acteurs des projets de recueil de données

Les projets de recueil de données de soins primaires rassemblent de nombreux acteurs. Chaque réseau dispose d'une architecture propre mais leur analyse retrouve des composantes similaires :

- Un conseil d'administration
- Des partenaires publics et privés
- Un comité d'éthique ou IRB (Institutional Review Board)
- Un comité scientifique
- Une personne chargée de veiller au respect de la politique de confidentialité (Data Protection Officer)
- Un tiers de confiance

3.2.1 Le conseil d'administration

Le conseil d'administration du réseau CPCSSN(11) est composé de médecins participants au projet. Chaque structure locale est représentée au conseil d'administration. Au sein de CPCSSN, des comités permanents sont chargés de missions spécifiques : vie privée et éthique, recherche, fidélisation et recrutement des cabinets, gestion des données et technologies de l'information.

NIVEL est pilotée par un conseil d'administration comprenant des médecins, et des personnes compétentes dans le secteur social.

La composition des conseils d'administration des autres projets n'était pas détaillée.

3.2.2 Les partenaires

Notre étude a permis de mettre en évidence différents modes de financement, dans la majorité des cas publics. Des partenaires privés peuvent être associés aux projets notamment aux Etats-Unis.

CAPriCORN est l'un des 13 réseaux de données cliniques pour la recherche, financés par le Patient-Centered Outcomes Research Institute, une organisation non gouvernementale à but non lucratif, financée par l'Etat américain. On note que le site CAPriCORN affiche un partenariat avec un laboratoire pharmaceutique (9).

L'ICES (17), hébergeur du projet EMRALD, est une organisation indépendante à but non lucratif financée par le Ministère de la Santé et des soins de longue durée de l'Ontario. ICES a de nombreux partenariats avec des organismes gouvernementaux et locaux qu'ils soient académiques (université de Toronto) ou institutionnels (Health Canada) ou associatifs (Canadian Diabetes Association).

CPCSSN, qui dépend du Collège des médecins de famille du Canada et de la Queen's University, est financé par l'agence de santé publique du Canada.

GRAPHC est financé par des subventions de recherche.

Le Clinical Practice Research Datalink est un projet de recueil de données observationnel et interventionnel, gouvernemental à but non lucratif, financé conjointement par le National Health Institute (NHS), le National Institute of Health Research (NIHR), et le Medicine and Healthcare Products Regulatory Agency (MHRA).

Le projet INTEGO est financé par le ministère de la santé Belge.

NIVEL est l'institut national pour les services de recherche en santé aux Pays-Bas, c'est une organisation non gouvernementale à but non lucratif financée notamment par le Ministère de la Santé.

3.2.3 Les comités d'éthique de la recherche

Des comités sont chargés de s'assurer de la conformité des projets de recherche à l'éthique médicale et, plus largement, au cadre juridique de la recherche sur la personne humaine.

3.2.3.1 Deux types de comités d'éthique

On retrouve deux modèles : institutionnel, piloté par l'institution qui mène les recherches et non-institutionnel, sous le contrôle direct des autorités nationales.

3.2.3.1.1 Les comités d'éthique institutionnels

Les comités institutionnels sont constitués auprès de chaque institution de recherche (Université, hôpital,…). Ils sont la norme en Amérique du Nord et sont dénommés Institutional Review Board (IRB) aux USA et Research Ethic Board (REB) au Canada.

Aux USA, les IRB sont des comités d'éthique de la recherche scientifique. Ils étudient les projets de recherche afin de déterminer s'ils sont conformes à l'éthique et sont chargés d'approuver, désapprouver, surveiller ou demander des mises en conformité. Chaque institution de recherche doit se munir d'un IRB obéissant à la réglementation fédérale, sous le contrôle des autorités sanitaires fédérales (FDA et DHHS).

En dehors des USA, certaines institutions de recherche, par exemple l'INSERM en France, se dotent d'un IRB agréé par l'administration des USA afin de se conformer aux exigences des revues scientifiques américaines.

3.2.3.1.2 Les comités d'éthique non-institutionnels

Ces comités d'éthique sont contrôlés par les autorités administratives. Par exemple, en France, les Comités de Protection des Personnes sont agréés par le Ministère de la Santé et ses membres sont nommés par le représentant de l'Etat au niveau régional (Article L1123-1 du code de la Santé Publique). Au Royaume-Uni les Research Ethic Committees sont gérées par le NHS.

3.2.3.2 Fonction des comités d'éthique

On distingue deux types de fonction :

- les conditions d'autorisation de la création des projets de recueil
- les autorisations accordées par la suite pour l'accès aux données des projets constitués.

3.2.3.2.1 Des comités d'éthique pour valider la mise en place des projets de recueil

Dans un article sur les enjeux éthiques du CPCSSN (18), les auteurs rappellent que le projet a obtenu l'accord des REB de chaque province à l'initiation du projet, mais que du fait des variations d'interprétation des réglementations et des règles éthiques, l'obtention de ces accords ont retardé la mise en œuvre du projet.

Les principales difficultés dans l'obtention du feu vert des REB ont concerné les points suivants : l'explication du recueil de certaines données spécifiques, le respect de la vie privée, la confidentialité, la sécurité des données, le chainage des données, l'absence de recueil du consentement. Même au sein de la même juridiction, les auteurs ont constaté de grandes différences de traitement entre deux REB.

Afin de simplifier les procédures d'approbation, les auteurs proposent la création d'un comité d'éthique centralisé, spécialisé et responsable d'approuver ce type d'études, sur le modèle de ce qui existe déjà au Canada pour la recherche en oncologie et dans d'autres pays.

Le projet INTEGO(19) a été approuvé à son initiation par la Commission de la Protection de la Vie Privée (CPVP, devenue en 2018 l'Autorité de Protection des Données) ainsi que par la commission d'éthique de l'Université Catholique de Louvain. Par la suite, la CPVP a revu les différentes procédures du projet et demandé l'implication d'un tiers de confiance dans le recodage des données ainsi que la création d'un comité d'éthique et scientifique au sein du projet.

3.2.3.2.2 Des comités d'éthique pour valider les demandes d'accès des chercheurs

L'IRB du projet CAPriCORN appelé CHAIRb, est chargé d'évaluer les demandes d'accès aux données du réseau.

Le projet CPCSSN est doté d'un comité de recherche chargé d'évaluer les demandes d'accès aux données qui lui sont soumises par des chercheurs.

Le projet INTEGO dispose d'un comité d'éthique et scientifique.

Au Royaume-Uni, un comité nommé Independent Scientific Advisory Committee (ISAC) est chargé d'approuver les demandes d'accès au CPRD, en complément du comité d'éthique. Il a pour rôle de conseiller la Medicines & Healthcare products Regulatory Agency (MHRA) sur la qualité méthodologique, scientifique et éthique des projets utilisant les données du CPRD.

3.2.3.3 Composition des comités d'éthique

Aux Etats-Unis, la réglementation prévoit qu'un comité d'éthique (Institutional Review Board, IRB) soit composé d'au moins 5 membres issus d'horizons variés, parmi lesquels : au moins un membre reconnu pour son expérience auprès de populations vulnérables (enfants, détenus, femmes enceintes, personnes en situation de handicap), au moins un membre scientifique et un membre non-scientifique, au moins un membre sans lien d'intérêt direct avec l'institution, au moins un membre représentant les intérêts des chercheurs de l'institution. Si le projet évalué concerne la population carcérale, un représentant des détenus doit être présent. Les membres doivent refléter la diversité sur le plan ethnique, culturel, le genre, la profession, et ne doivent pas présenter de conflit d'intérêt avec le projet évalué.

Dans le cas de CHAIRb, il a été requis que chaque institution participant à CAPriCORN soit représentée. Les sessions des IRB se déroulent sous la surveillance de l'Office for Human Research Protections qui dépend du Department of Health & Human Services.

L'Autorité de protection des données (anciennement CPVP) en Belgique a imposé au projet INTEGO de se doter d'un comité éthique et scientifique, composé de chercheurs du Département de Médecine Générale, de chercheurs d'autres départements de l'université et d'autres universités, de médecins participant à INTEGO, d'un participant non médecin, d'un éthicien et d'un juriste.

3.2.4 Un comité scientifique

Des comités scientifiques évaluent la pertinence des questions de recherche posées.

Le projet CPRD (20) donne accès aux données des patients dans le cadre de projets de recherche en santé après approbation du protocole d'étude par le Independent Scientific Advisory Committee (ISAC) de la MHRA après validation du comité d'éthique.

Le projet INTEGO est doté d'un comité chargé de donner son avis sur la qualité scientifique des études.

3.2.5 Un responsable de la protection des données personnelles (Data Protection Officer)

3.2.5.1 Des responsables nationaux

Une personne ou une commission peut être nommée comme référent indépendant du respect de la protection des données personnelles. En France, la CNIL joue ce rôle.

Au Canada, il s'agit du Commissariat à la protection de la vie privée.

Au Royaume-Uni, Dame Fiona Caldicott est la National Data Guardian. L'exemple suivant illustre l'influence de ce garant de la protection des données.

Une thèse (21) soutenue en 2015 rappelle l'abandon du projet « Care.data » qui trouve ses causes en partie dans le manque de confiance du public dans la protection des données personnelles. Depuis 2012, le « Health and Social Care Act » prévoyait l'obligation légale pour les médecins généralistes de transmettre leurs données au Health and Social Care Information Center (HSCIC) afin d'alimenter Care.data à partir du printemps 2014. Le gouvernement avait prévu d'informer la population par l'envoi d'une plaquette d'information dans chaque foyer, mais beaucoup n'ont pas été reçues. Les médecins étaient chargés d'informer systématiquement les patients de leur droit d'opposition, et pouvaient marquer le dossier patient le cas échéant afin que les données ne soient pas transmises.

Devant l'opposition d'associations de médecins et de patients, le projet a été repoussé plusieurs fois puis abandonné en 2016 suite au rapport de Dame Fiona Caldicott (qui occupe le poste de National Data Guardian) qui formule 20 recommandations afin d'améliorer la protection de la vie privée et de retrouver la confiance du public.

3.2.5.2 Un responsable au sein des projets

Au sein des projets de recueil de données, une personne peut être responsable de la mise en œuvre de la politique de protection des données.

Au Canada, l'agente de la protection des renseignements personnels et d'éthique en matière de recherche, Madame Jannet Ann Leggett, est la garante du programme de protection des données du CPCSSN (22).

Le CPRD dispose d'un Data Protection Officer.

En Europe, le RGPD (6) rend obligatoire la nomination d'un délégué à la protection des données pour les autorités ou les organismes publics, les organismes réalisant un suivi régulier et systématique des personnes à grande échelle, les organismes traitant à grande échelle des informations sensibles.

3.2.6 Tiers de confiance

La notion de tiers de confiance a été retrouvée dans plusieurs projets. Il peut s'agir d'un outil logiciel, d'une personne ou d'un prestataire.

Cet intermédiaire neutre garantit la dé-identification des données à des niveaux différents selon les projets. Nous reviendrons dans la partie 3.4 sur la définition de la dé-identification. Soit il s'assure que les données sont bien dé-identifiées, soit il a en charge lui-même le procédé de dé-identification.

Il peut aussi avoir une autre fonction, celle de permettre la ré-identification des données à la demande des chercheurs. En effet, lorsqu'un chaînage des données est requis, les réseaux font appel à un tiers de confiance pour limiter le risque de ré-identification des patients. Dans CAPriCORN, un tiers de confiance garde une table de correspondance en cas de nécessité de ré-identification.

Aux Etats-Unis, les IRB proposent des procédures de certification des tiers de confiance (Honest Broker). La personne certifiée par l'IRB aura alors la responsabilité des opérations effectuées par son intermédiaire.

En Belgique, INTEGO utilise les services d'eHealth, une plateforme publique de services électroniques en santé, comme tiers de confiance.

3.3 Consentement des patients

3.3.1 Définitions

3.3.1.1 Consentement (ou consentement exprès ou explicite ou « opt-in »)

Selon le RGPD, est appelé consentement « *de la personne concernée, toute manifestation de volonté, libre, spécifique, éclairée et univoque par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement.* » (Article 4-11). Nous utiliserons par la suite le terme consentement comme équivalent des termes consentement exprès ou explicite ou opt-in.

3.3.1.2 Non-opposition (ou consentement implicite ou « opt-out »)

Dans certains cas, notamment celui de la recherche scientifique sur des données dé-identifiées, il est permis de ne pas rechercher le consentement d'une personne avant de recueillir ses données personnelles. La personne doit être informée du recueil de ses données et dispose tout de même d'un droit d'opposition à leur utilisation.

3.3.2 Modalités de recueil du consentement

Le recueil de données dé-identifiées est uniformément basé sur le principe de non-opposition des patients. En cas d'utilisation de données identifiantes, le consentement des patients est systématiquement recherché.

Au sein de CAPriCORN (23,24), le consentement des patients n'est recherché que dans le cas où l'accès à des informations de santé protégées est nécessaire, au sens de la règlementation des USA (HIPAA Privacy Rules). Le réseau dispose d'un centre de relation (« Communication center ») qui lui permet de communiquer avec les patients par courrier électronique ou par SMS, notamment afin de recueillir leur consentement si nécessaire. Aucune information sur le droit d'opposition au recueil de données n'a été trouvée sur le site officiel.

Dans le contexte de données préexistantes et dé-identifiées, CPCSSN applique également le principe de non-opposition. Chaque cabinet participant affiche une information validée par le comité d'éthique destinée aux patients, afin de leur expliquer l'usage qui sera fait de leurs données et leur droit d'opposition. Dans une étude sur les enjeux éthiques du CPCSSN (18), les auteurs rappellent que dans le cadre d'une étude clinique où sont collectées des données nouvelles, le consentement éclairé des participants est généralement requis.

En 2013, Dame Fiona Caldicott (National Data Guardian) a recommandé dans son rapport « Information: To share or not to share? The information governance review » que les données ayant un faible risque résiduel de ré-identification du CPRD soient partagées sur un mode de non-opposition, alors que les données identifiables devraient être partagées uniquement si nécessaire, avec le consentement des personnes.

Les projets de recueil de données en soins primaires NIVEL (25) et INTEGO reposent sur le principe de non-opposition. Selon un article de 2014 (19), le projet INTEGO prévoit d'ajouter aux données cliniques des informations génomiques. Ceci nécessite le consentement éclairé des patients participants.

3.3.2.1 Opinion des patients sur le recueil de leur consentement

L'opinion des patients a notamment été étudiée au Royaume-Uni dans le cadre du CPRD.

Une étude réalisée 2011 (5) auprès de 5331 patients de centres de soins primaires et secondaires de l'Ouest de Londres a obtenu les résultats suivants : 90,8% des répondants ont considéré que leur consentement était nécessaire avant de donner accès à leurs données identifiables, alors qu'ils n'étaient plus que 50,7% à considérer que leur consentement était nécessaire pour accéder à leurs données dé-identifiées. Seulement 58,6% des répondants avaient la notion que leurs données étaient enregistrées dans un dossier médical électronique.

Dans un autre article publié en 2015 sur les enjeux éthiques du CPRD (26), les employés et les patients de deux cabinets ont été contactés pour participer à un groupe de discussion.

Les auteurs rappelaient en préambule que choisir entre consentement et non-opposition est sujet à controverse : comment être certain que les patients qui ne se sont pas opposés à l'usage de leurs données sont vraiment en accord avec cet usage. L'étude a en effet relevé que certains patients ne savaient pas que leur participation était automatique sans action de leur part. Le choix de considérer la non-opposition comme un mode de consentement semble avoir été vécu comme problématique par les participants à cette étude.

3.4 Dé-identification

Le terme de dé-identification est préféré au terme d'anonymisation. La dé-identification consiste à retirer les informations identifiantes des données recueillies. Une ré-identification reste cependant souvent possible, notamment lors du chaînage des données avec d'autres bases de données.

Tous les projets de recueil de données dé-identifient les données avant de traiter l'information. Il reste cependant possible, à la demande et sous contrôle d'un comité d'éthique, d'accéder aux données identifiantes. Dans ce cas, le recueil du consentement du patient est requis, comme précisé précédemment.

3.4.1 Procédés techniques de dé-identification

3.4.1.1 « Privacy by design »

Ce concept peut se traduire en français par la mise en œuvre de la protection des données dès la conception d'un projet, c'est-à-dire la prise en compte des risques pour les données personnelles et l'intégration de procédés de protection tels que la pseudonymisation. Le projet CPCSSN en a été un des pionniers.

L'Union Européenne rend incontournable cette notion par l'article 25 du RGPD.

3.4.1.2 Pseudonymisation

Selon le RGPD : il s'agit du *« traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable* » (article 4-5).

Ce procédé est cité par l'article 25 du RGPD comme un des moyens de protection des données.

Il est utilisé par plusieurs projets de recueil tels que NIVEL, INTEGO, CPRD, CPCSSN.

3.4.1.3 HASH-ID

Ce procédé particulier de pseudonymisation est utilisé par le réseau CAPriCORN. Il est ainsi conçu : un tiers de confiance crée une clé qu'il partage avec toutes les institutions participantes, chacune de ces institutions utilise la clé pour la combiner aux informations identifiantes de chaque patient (numéro de sécurité sociale, nom, prénom, date de naissance, genre) et obtenir un identifiant unique. Pour chaque étude, chaque institution sélectionne les patients et envoie à un deuxième tiers de confiance la liste des Hash-ID avec les données cliniques associées. Celui-ci dé-doublonne les données reçues et les ré-adresse aux institutions pour vérification. Le deuxième tiers de confiance réalise ensuite un masquage des institutions sources et remplace le Hash-ID par un identifiant non lié aux données patients.

3.4.1.4 k-anonymisation

La k-anonymisation est un procédé utilisé afin d'éviter la ré-identification par croisement de données.

Il s'agit de diminuer la précision de certaines informations qui pourraient permettre une ré-identification de sorte qu'il existe dans le jeu de données au moins k individus qui possèdent la même information. Cela peut se faire en « généralisant » les informations à risque (par exemple en remplaçant l'âge par une tranche d'âge ou en supprimant les derniers chiffres d'un code postal).

Ce procédé peut être complété par la l-diversité en introduisant l'obligation d'avoir au moins l valeurs différentes par catégorie.

3.4.1.5 Expérimentations

Une équipe a étudié au Danemark (27) la faisabilité de dé-identification d'une base de données issue des soins primaires. Pour ce faire, une extraction de la base de données a été réalisée en conservant les données personnelles. Un algorithme a été utilisé pour remplacer tous les identifiants dans les champs de données structurées. Les prénoms et noms de famille ont été remplacés par d'autres ayant une fréquence voisine dans la population danoise, les jours et mois de naissance ont été remplacés par des nombres aléatoires mais constants, le nombre du siècle a été conservé, un numéro a été attribué au genre. D'autres identifiants, comme le numéro de téléphone, ont été remplacés par des chiffres aléatoires. Les champs de texte libre sont inspectés par l'algorithme qui recherche des informations identifiantes (nom, lieu, numéro de téléphone, …) afin de les dé-identifier si nécessaire.

Au Royaume-Uni, une équipe (28) a réalisé une preuve de concept de mise en place d'une base de données pseudonymisée. Les données sont issues de cabinets de médecins généralistes. Elles sont téléchargées en lieu sûr (« safe haven ») après avoir été dé-identifiées. L'identifiant unique du patient (Community Health Index) est stocké à part et remplacé par un autre numéro. Les données sont ensuite transférées dans une base de recherche pseudonymisée qui elle seule est accessible par les chercheurs.
Ce « safe haven » est à rapprocher du principe du tiers de confiance. Il s'agit d'un stockage sécurisé et dé-identifié des données sensibles.

3.4.2 Techniques mises en œuvre par projet

Deux procédés ont été retrouvés dans les projets de recueil des Etats-Unis.

Le projet CAPriCORN (24) utilise un procédé de pseudonymisation particulier (« HASH-ID », Cf. 3.4.1.3).

Une équipe de la Vanderbilt University à Nashville (29) a mené un projet expérimental d'anonymisation de données cliniques comprenant des informations génomiques et a proposé un procédé basé sur la k-anonymisation. En fournissant un jeu de données cliniques comprenant les codes de la classification internationale des maladies (CIM) associés à l'âge et à une séquence d'ADN, et en précisant le niveau d'anonymisation souhaité (nombre k, tel que au moins k patients sont présents dans chaque bloc de données), le système génère un bloc de données où le code CIM et l'âge sont moins précis, de façon à ce que la séquence d'ADN puisse être attribuée à au moins k patients. Les auteurs rappellent cependant qu'il est toujours possible de ré-identifier un individu malgré ce procédé, mais que cela est d'autant plus difficile que le nombre k est élevé.

Au Canada, le projet EMRALD (30) a testé la dé-identification de jeux de données à l'aide du programme « De-ID », un logiciel open source développé initialement pour le traitement des dossiers infirmiers hospitaliers. Un travail important a été nécessaire pour adapter cet outil aux données médicales de soins primaires et au contexte local. La conclusion des auteurs est que cet outil une fois adapté est une bonne solution de dé-identification des données en texte libre, malgré quelques limites, notamment : peu généralisable surtout dans le cas de traitement de documents numérisés car la plupart des programmes n'appliquent pas la reconnaissance de caractères, et le traitement nécessite une puissance de calcul importante. Les auteurs rappellent qu'une tentative d'adaptation de De-ID à la langue française (31) en 2009 n'a pas été concluante, notamment du fait des fortes différences avec la langue anglaise.

Dans le cadre du réseau CPCSSN (18,22,32) de nombreuses données cliniques, biologiques, thérapeutiques et démographiques sont collectées par les réseaux régionaux, mais les documents numérisés et les notes de consultations en texte libre ne le sont pas, en raison de la difficulté de traiter ces informations et de les dé-identifier. Les données des patients sont dé-identifiées en 3 phases.

Tout d'abord, les données des patients sont dé-identifiées par le retrait des identifiants présents dans les champs de données structurées comme le nom, l'adresse, le numéro de sécurité sociale, soit par les services du réseau régional soit par l'éditeur du logiciel de dossier médical électronique. Le code postal complet est conservé à des fins de statistiques cartographiques. Les données extraites ne comportent donc pas ces champs. Une table de correspondance, conservée au cabinet médical, permet de ré-identifier les données, si nécessaire, elle contient le code postal, la date de naissance, le sexe, le numéro de sécurité sociale des patients. Les données sont transmises via une connexion sécurisée et chiffrée à l'Université Queen's où les réseaux régionaux et le réseau national centralisent leurs données.

Dans un deuxième temps, les données extraites sont traitées par un algorithme qui remplace les informations identifiantes qui pourraient se trouver dans des champs non structurés par des caractères aléatoires.

Dans un troisième temps, l'outil PARAT de la société Privacy Analytics Inc. est utilisé pour analyser les données restantes. Si l'outil détecte un risque de ré-identification supérieur à la norme fixée, il tronque les données de certains champs pour les patients concernés afin de réduire le risque.

La ré-identification de patient n'est permise qu'avec l'accord du cabinet médical concerné et du comité d'éthique. Le cabinet source des informations reste maître des informations d'identification.

En Australie, l'équipe du projet GRAPHC (33) a mis au point une technique associant dé-identification et conservation de l'information géographique à des fins d'analyse statistique et de cartographie.

Jusqu'ici les données géographiques étaient agrégées afin de satisfaire au principe de k-anonymisation, et l'information géographique était donc rendue trop imprécise pour être utilisée.

La nouvelle technique mise en place consiste à extraire deux jeux de données, l'un comprenant les données cliniques dé-identifiées et l'autre contenant les informations d'identification et l'adresse. Ces adresses sont envoyées à un serveur sécurisé qui attribue un identifiant unique (GTAG) à chacune, sans qu'il y ait possibilité de retrouver l'adresse à partir de l'identifiant, puis chaque GTAG est affecté aux données cliniques dé-identifiées correspondantes.

Lorsque les chercheurs ont accès aux données, ils n'ont pas directement accès à la localisation géographique, mais le GTAG leur permet de faire ensuite une analyse spatiale des données.

Les données issues des dossiers-patients informatisés sont importées dans le CPRD (13,20,34) en étant dé-identifiées, qu'il s'agisse de texte libre ou de données structurées. Les données identifiantes sont conservées séparément des données de santé par NHS Digital, le service national d'informatique sanitaire et social au Royaume-Uni.

Le CPRD dispose d'outils permettant le traitement du texte libre afin de retirer les informations identifiantes. Les techniques utilisées pour la dé-identification ne sont pas précisées dans les ressources que nous avons consultées.

Depuis 2012, dans le cadre du projet INTEGO (19), des données partielles dé-identifiées issues des cabinets de médecins généralistes sont transmises sous forme de fichier texte chiffré à un tiers de confiance indépendant qui est chargé de coder les données. Elles sont ensuite adressées au Département de Médecine Générale de l'Université de Louvain où se trouve la base de données centralisée.

Le projet ACHIL (15) (Ambulatory Care Health Information Laboratory), , fait appel à un tiers de confiance pour pseudonymiser sa base de données, conformément à la réglementation Belge. Les données cliniques sont transmises sous forme de courrier électronique chiffré.

La base de données est séparée en deux niveaux d'utilisation : un « bas niveau » dédié à la gestion des envois (identification du destinataire, dé-identification de l'expéditeur), les données cliniques étant

chiffrées et donc inaccessibles. Le « haut niveau » donne accès aux données cliniques, il est réservé aux chercheurs autorisés.

Les médecins généralistes conservent l'accès aux informations d'identification (mais pas aux pseudonymes) des patients alors que côté chercheurs, les médecins et les patients sont pseudonymisés avec un identifiant propre à chaque projet de recherche. Seul le tiers de confiance dispose de la table de correspondance entre pseudonymes et identifiants, il dispose aussi des données cliniques non chiffrées.

Les auteurs admettent toutefois des failles de sécurité potentielles : le risque que le tiers de confiance établisse un lien entre deux études au sujet d'un patient, le risque qu'il ait connaissance d'un lien thérapeutique entre un médecin et un patient.

Cependant, dans ce projet, il est prévu de ne confier aucune donnée sensible au tiers de confiance, les données cliniques ne sont pas accessible par l'opérateur humain et elles ne font que transiter. Un comité de surveillance veille au respect de la confidentialité.

Les données recueillies par le projet NIVEL sont extraites des bases de données de soins primaires et pseudonymisées sur place avant envoi à NIVEL. Une deuxième pseudonymisation est effectuée lorsque les données sont extraites à la demande de chercheurs. Le deuxième pseudonyme est spécifique à chaque étude, afin d'éviter les recoupements. La ré-identification reste possible si nécessaire à partir du pseudonyme en interrogeant la base de données.

3.5 Modalités d'accès aux données

3.5.1 Conditions d'accès

Les données sont accessibles aux chercheurs. Dans les réseaux de recueil analysés, aucune possibilité d'accès direct par des sociétés commerciales n'a été mise en évidence.

Les chercheurs souhaitant accéder aux données doivent rédiger un protocole de recherche et faire valider leur demande auprès d'un comité d'éthique.

Par exemple, l'utilisation des données de CAPriCORN dans le cadre de projets de recherche est soumise à l'approbation du Chicago Area Institutional Review Board (CHAIRb), un comité d'éthique soumis au règlement du Department of Health and Social Security et de la Food and Drugs Agency.

Le projet CPRD (20) dispose d'une autorisation éthique global du National Research Ethics Service Committee (NRES) en ce qui concerne les recherches observationnelles utilisant les données issues de soins primaires avec chainage pour les recherches internes au NHS. Le procédé standard d'accès aux données du CPRD nécessite une approbation par un comité d'éthique et l'ISAC. Hors de ce cadre, tout projet de recherche utilisant les données du CPRD nécessite une autorisation éthique spécifique de la part d'un comité d'éthique (Research Ethics Commitee).

Dans le cadre du projet GRAPHC (33), le niveau de précision géographique dépend d'une décision d'un comité d'éthique pour chaque étude.

3.5.2 Forme des données délivrées

Les informations fournies aux chercheurs selon leurs demandes sont variées :

- Nombre de sujets éligibles pour une étude
- Jeux de données (« datamart ») dé-identifiées répondant à leur requête avec ou sans traitement de l'information préalable
- Informations géographiques dé-identifiées à des fins de cartographie par exemple

3.5.3 Outils d'accès

Le projet CAPriCORN (24) est accessible notamment via PopMedNet (35), une plateforme open-source destinée à faciliter le fonctionnement des réseaux de données de santé multi-sites. Elle permet d'interroger les bases de données des différents sites participants au réseau tout en les laissant maitres de leurs propres données.

Lorsqu'un chercheur effectue une requête sur PopMedNet, les institutions concernées la consultent et l'évaluent avant son exécution. Si elles le souhaitent, elles donnent leur accord, exécutent la requête sur leur base de données et envoient leurs résultats. Le chercheur peut alors télécharger le jeu de données.

PopMedNet est développé depuis 2007 au sein du Department of Population Medicine du Harvard Pilgrim Health Care Institute.

Le réseau CPCSSN (32) permet à des chercheurs de soumettre une demande en ligne comportant un résumé du projet de recherche, puis, si la réponse est favorable, l'intégralité du protocole de recherche et la lettre d'approbation d'un comité d'éthique. Un sous-comité de surveillance et de recherche au sein du CPCSSN est chargé d'étudier ces demandes. Une fois le dossier complet et approuvé, le CPCSSN envoie le jeu de données aux chercheurs via une connexion sécurisée. La mise à disposition des données est tarifée à prix coûtant pour les chercheurs universitaires, une remise est consentie aux étudiants. Le CPCSSN propose également des services payants de traitement et d'analyse des données.

Le projet CPRD (20) donne accès aux données des patients dans le cadre de projets de recherche en santé après approbation du protocole d'étude par le Independent Scientific Advisory Committee (ISAC) de la Medicines and Healthcare products Regulatory Agency (MHRA) après validation du comité d'éthique. L'exploitation du CPRD nécessite des connaissances avancées en traitement de données. Les chercheurs disposent de dictionnaires et de guides leur permettant de déterminer les « Read codes » qui les intéressent. Le CPRD propose également les services d'une équipe de recherche interne qui traite l'information et la livre adaptée aux besoins du chercheur.

L'accès aux données de NIVEL se fait à la demande des chercheurs, une extraction de la base de données principale est effectuée en fonction des projets de recherche.

3.6 Réglementation et éthique

3.6.1 Comparaison des données identifiantes

Aux USA, la réglementation définit très précisément les informations considérées comme identifiantes (36), alors que dans l'Union Européenne elles sont définie à l'article 4-1 du RGPD. Ces deux définitions se recoupent globalement.

Elles regroupent les informations qui permette d'identifier directement ou indirectement une personne physique comme les nom et prénom, les dates de naissance, les adresses, les numéros identifiants (téléphone, immatriculation d'un véhicule, adresse IP…), les éléments de biométrie, et des éléments de la vie privée. (Cf. Tableau 1)

Tableau 1. Informations identifiantes au sens des réglementations des USA et de l'UE

Nom
Adresse, code postal, ville, état
Téléphone, Télécopie, Adresse électronique
Numéro de sécurité sociale
Numéro dossier médical, de police d'assurance, de compte bancaire
Numéro de permis de conduire, identifiants et numéro de série de véhicule
Adresse URL, adresse IP
Photographie du visage, identifiants biométriques (dont empreintes digitales et vocale)
Éléments d'identité physique, physiologique, génétique, psychique, économique, culturelle, sociale, syndicale, politique

3.6.2 Réglementation des Etats-Unis

La règlementation concernant la confidentialité des données de santé aux USA est issue d'une loi fédérale nommée Health Insurance Portability and Accountability Act (HIPAA) datant de 1996 (36–38). Elle est complétée par le document nommé The Standards for Privacy of Individually Identifiable Health Information ("Privacy Rule") qui définit des informations de santé protégées (PHI).

Les organismes qui détiennent ces PHI ne peuvent les communiquer sans l'accord écrit de la personne qu'à elle-même ou au Departement of Health and Human Services. Il existe 13 grandes exceptions à cette règle, permettant généralement avec une autorisation, de divulguer certaines informations.

Les PHI sont les informations y compris démographiques reliées à :

- des problèmes de santé physique ou mentale d'un individu, passé, présent ou futur
- des prestations de soins d'un individu
- des paiements passés, présents ou futurs des prestations de soin d'un individu

et qui identifient l'individu ou qui peuvent laisser penser raisonnablement qu'elles puissent être utilisées pour identifier un individu (coordonnées, date de naissance…).

Lorsque les informations sont dé-identifiées, ces règles ne s'appliquent pas. Pour cela, soit elles doivent être traitées et certifiées par un statisticien qualifié, soit les identifiants doivent être retirés et respectant la méthodologie dite du « Safe Harbor ». Cette méthodologie, détaillée dans la Privacy Rule, consiste à supprimer des données une liste de 18 types d'informations identifiantes (36) et à s'assurer que les informations résiduelles ne permettent pas la ré-identification des personnes.

L'une des exceptions à cette règle permet l'usage de ces PHI à des fins d'évaluation et d'amélioration de la qualité des soins. Cependant les activités de recherche (qui contribuent à un savoir généralisable) ne sont pas concernées par cette exception.

Cette réglementation fédérale est complétée par d'autres textes :

- The Security Rule
- The Common Rule
- The Genetic Information Nondisclosure Act of 2008
- The Privacy Act of 1974 and the U.S. Freedom of Information Act
- Forty-two C.F.R. Part 2 (Part 2, 2013)

Les lois spécifiques aux Etats et la jurisprudence viennent également s'y ajouter.

Le projet CAPriCORN (24) dispose en complément d'un programme de protection des sujets humains (HSPP) destiné à s'assurer que les projets de recherches correspondent aux standards éthiques, obéissent aux réglementations, et soient contrôlés régulièrement et collégialement. Ce HSPP fonctionnent d'une manière classique aux USA avec plusieurs Institutional review board (IRB).

3.6.3 Réglementation Canadienne

Au Canada (32), les bureaux institutionnels d'éthique en recherche (Research Ethics Board (REB) en anglais) se basent sur ces principaux éléments pour rendre leurs avis : les règles définies par le Tri-Council Policy Statement on research ethics (TCPS2), les recommandations de bonne pratique clinique, le Health Information Protection Act (HIPA), la loi sur les services de santé et les services sociaux, la loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels, le Personal Information Protection and Electronic Documents Act, et le Personal Health Information Protection Act.

3.6.4 Réglementation Européenne

3.6.4.1 Directive Européenne sur la protection des données de 1995

Selon la Directive Européenne sur la protection des données de 1995(39) qui s'appliquait jusqu'en 2018 aux projets européens, le traitement des données personnelles sensibles était interdit, mais des dérogations étaient prévues notamment *« lorsque le traitement des données est nécessaire aux fins de la médecine préventive, des diagnostics médicaux, de l'administration de soins ou de traitements ou de la gestion de services de santé et que le traitement de ces données est effectué par un praticien de la santé soumis par le droit national ou par des réglementations arrêtées par les autorités nationales compétentes au secret professionnel, ou par une autre personne également soumise à une obligation de secret équivalente. »* Le consentement des patients participants n'était donc pas requis selon la réglementation européenne dans ce contexte.

3.6.4.2 Règlement Général sur la Protection des Données

Depuis le 28 mai 2018, est entré en vigueur le Règlement Européen 2016/679, couramment nommé « Règlement Général sur la Protection des Données » (RGPD) (6).

L'article 4-1 définit les informations identifiantes (Tableau 1). L'article 25 consacre le principe de protection des données dès la conception et la protection des données par défaut.

L'article 9 interdit le traitement de données à caractère personnel sensibles, notamment les données de santé, et prévoit bien sûr une série d'exceptions à cette règle.

Dans le cas de la collecte de données de soins primaires à des fins de recherche ou de statistiques le traitement est permis comme défini dans l'article 89-1 :

« Le traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique, ou à des fins statistiques est soumis, conformément au présent règlement, à des garanties appropriées pour les droits et libertés de la personne concernée. Ces garanties garantissent la mise en place de mesures techniques et organisationnelles, en particulier pour assurer le respect du principe de minimisation des données. Ces mesures peuvent comprendre la pseudonymisation, dans la mesure où ces finalités peuvent être atteintes de cette manière. Chaque fois que ces finalités peuvent être atteintes par un traitement ultérieur ne permettant pas ou plus l'identification des personnes concernées, il convient de procéder de cette manière. »

Ce règlement a été transposé dans le droit français par la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, qui vient modifier la Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

Le décret n° 2018-687 du 1er août 2018 vient en préciser les modalités d’application en modifiant le décret n°2005-1309 du 20 octobre 2005 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

4 Discussion

Cette revue de la littérature a permis de comparer les modalités de protection des données personnelles au sein des projets de recueil et d'analyse de dossiers médicaux de soins primaires. Les questions de sécurisation informatique des entrepôts de données n'ont pas été traitées. Elles représentent cependant une problématique à part entière.

Nous avons suivi les critères PRISMA et évalué la qualité méthodologique de notre étude par le biais de la grille AMSTAR2.

Notre étude a permis d'étudier 7 projets de recueil de données en Occident. Lors de notre première revue de littérature (1), nous avions identifié 36 projets à l'échelle mondiale. Cette dernière revue retrouve des projets répartis sur différents pays et qui sont représentatifs de différents contextes réglementaires. Plusieurs projets nationaux de grande ampleur sont représentés. On note cependant les informations concernant les bases privées (Qresearch, THIN, ResearchOne…) ne sont pas publiées dans le domaine de la confidentialité des données.

4.1 Des acteurs de gouvernance communs

Notre objectif était de définir précisément les composantes des projets de recueil existants afin de transposer ce modèle dans notre contexte national. Nous avons pu identifier des acteurs de gouvernance communs.

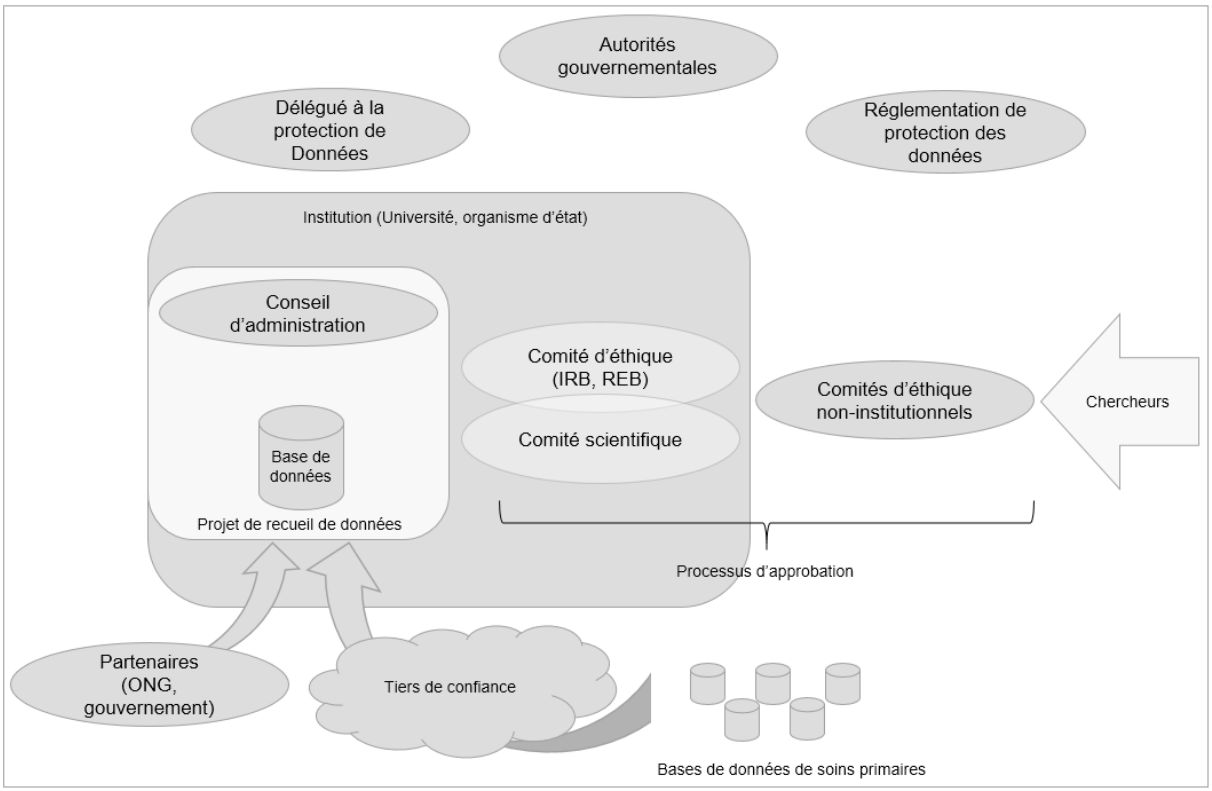

Figure 3. ***Schéma de synthèse de la gouvernance des projets de recueil de données :*** *les projets de recueil de données de soins primaires sont généralement pilotés par un conseil d'administration, sous le contrôle des autorités nationales. Les comités d'éthiques responsables de l'approbation des recherches effectuées à partir de la base de données peuvent être institutionnels, ou non-institutionnels.*

Le conseil d'administration des différents projets de recueil de données intègre le plus souvent des médecins généralistes participants et des universités, dont leur département de médecine générale.

Les partenaires sont pour la plupart des structures publiques. En France, le Ministère de la Santé et l'Union Nationale des Caisse d'Assurance Maladie (UNCAM) pourraient promouvoir ce type de projets. La création de l'Institut National des Données de Santé (INDS) issue de la Loi de Santé de 2016 semble être un premier pas dans cette direction.

Notre étude a mis en évidence différentes dénominations des comités d'éthique en recherche. Elle a révélé également des organisations différentes avec une forme soit non-institutionnelle, dirigée par les autorités administratives, soit institutionnelle, au sein des institutions de recherche, avec toujours un agrément de l'administration nationale.

En France, les comités d'éthique de la recherche sont les Comités de Protection des Personnes (CPP), instances dont le fonctionnement est régi par le Code de la Santé Publique (40). Ils sont chargés de donner un avis sur la validité des projets de recherche impliquant la personne humaine. Ils évaluent la protection des personnes, la pertinence du projet de recherche et sa qualité méthodologique. Leurs membres sont nommés par l'Agence Régionale de Santé pour 3 ans renouvelables. Leur composition doit répondre à une obligation de pluralisme avec la participation de chercheurs expérimentés, de médecins généralistes, de pharmaciens hospitaliers, d'infirmiers, de personnes qualifiées en éthique, de psychologues, de travailleurs sociaux, de juristes, de représentants des usagers du système de santé.

La Commission Nationale de l'Informatique et des Libertés (CNIL) (41,42) est garante de la protection des données personnelles, ses missions ont été renforcées par les récentes modifications législatives et réglementaires. Sur les sujets concernant les données de santé, elle travaille désormais en collaboration avec l'INDS.

Dans le cadre des projets de recherche n'impliquant pas la personne humaine mais comprenant un traitement de données personnelles de santé (article R1121-1-3 du Code de la Santé Publique), la demande d'autorisation doit être formulée auprès de l'INDS. C'est alors le comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé (CEREES) qui rend son avis sur la méthodologie, la nécessité du recours à des données à caractère personnel, la pertinence et sur la qualité scientifique du projet.

Les traitements de données qui sont conformes à une méthodologie de référence établie par la CNIL sont dispensés d'autorisation préalable, mais le responsable du projet doit adresser à la CNIL une attestation de conformité.

La CNIL a homologué dans une délibération du 13 juillet 2018 une méthodologie de référence (MR-004) (43) qui simplifie la procédure d'autorisation et qui parait correspondre aux caractéristiques des projets de recueil de données de soins primaires, cependant, en se limitant à une réutilisation de données.

Afin de garantir une utilisation des données pertinente et conforme à l'éthique, on peut imaginer que le projet de recueil de données soit piloté par un comité scientifique incluant des médecins participants et des chercheurs du Département de Médecine Générale (comme dans le cas du CPCSSN) ainsi que des représentants d'usagers du système de santé.

Le traitement des données et notamment leur pseudonymisation implique des compétences techniques qui nécessitent certainement une sous-traitance à un tiers de confiance. L'Agence Française de la Santé Numérique (ASIPSanté), une agence d'état déjà impliquée dans les services numériques en santé en France, semble être un partenaire de choix pour ce type de projet. C'est d'ailleurs le cas de ses homologues Belge (eHealth) et Anglais (NHS Digital) qui mettent leur expertise au service des projets de ces pays.

En France, la fonction de Délégué à la Protection des Données a été instaurée avec l'entrée en vigueur du RGPD (Cf. 3.2.5). Il sera donc probablement nécessaire de nommer un délégué spécifique pour le projet. D'autre part, les droits numériques sont garantis par la CNIL.

4.2 Une gestion uniforme du recueil des consentements

Il est admis dans l'ensemble des projets que le consentement n'a pas à être recueilli dans le cas du traitement de données dé-identifiées. Or, le RGPD interdit le traitement sans consentement de certaines catégories de données personnelles en particulier celles concernant la santé (Article 9-1). Il est cependant prévu plusieurs exceptions, notamment dans le cas de la recherche scientifique (Article 9-2-j). Dans ce cas, le responsable du traitement des données doit mettre en œuvre des mesures techniques et organisationnelles pour respecter le principe de minimisation des données, la pseudonymisation étant une des possibilités (Article 89-1).

L'utilisation des données comportant des éléments d'identification n'est permise qu'avec le consentement des personnes ou dans certains cas particuliers. Le droit d'opposition est prévu à l'article 21-6. En France, l'article 38 du Décret n°2005-1309 prévoit que cette opposition peut se faire par « tout moyen ».

Le principe de non-opposition présente cependant des limites. Il parait difficile de s'assurer que tous les patients dont les données sont recueillies soient parfaitement informés de leurs droits. Quel que soit le procédé d'information retenu (affiche, courrier, e-mail…) il est fort probable qu'un nombre non négligeable de personnes ne reçoive pas l'information, ne serait-ce que par un défaut de compréhension. Ces données pourront potentiellement être réutilisées par la suite, comment en informer les personnes concernées ? Enfin, cette souplesse réglementaire accordée à la recherche dans le traitement de données personnelles sans consentement permet de poser cette question : ne pas s'opposer signifie-t-il consentir à tout ?

4.3 Des données dé-identifiées mais ré-identifiables

L'ensemble des projets de recueil de données travaillent sur des données dé-identifiées avec pour certains la possibilité de ré-identification dans un second temps, au cas par cas, pour les chercheurs. Cette ré-identification peut se faire après autorisation grâce un tiers de confiance dans la plupart des réseaux.

4.3.1 Modalités de la dé-identification

Plusieurs projets utilisent un système de double pseudonymisation, une première étape au sein de la base de données puis une seconde étape de dé-identification pour le jeu de données fourni aux chercheurs. Ceci permet de limiter le risque de ré-identification en croisant deux jeux de données.

Les dé-identifications sont souvent réalisées en plusieurs étapes. Une première dé-identification est souvent réalisée au cabinet du médecin généraliste (extraction partielle de l'adresse par exemple) avec ensuite une ou deux étapes complémentaires au sein de l'entrepôt de données pour compléter le premier processus.

Certains projets comme CAPriCORN font appel à un tiers de confiance pour mener à bien la pseudonymisation (Cf. 3.2.6).

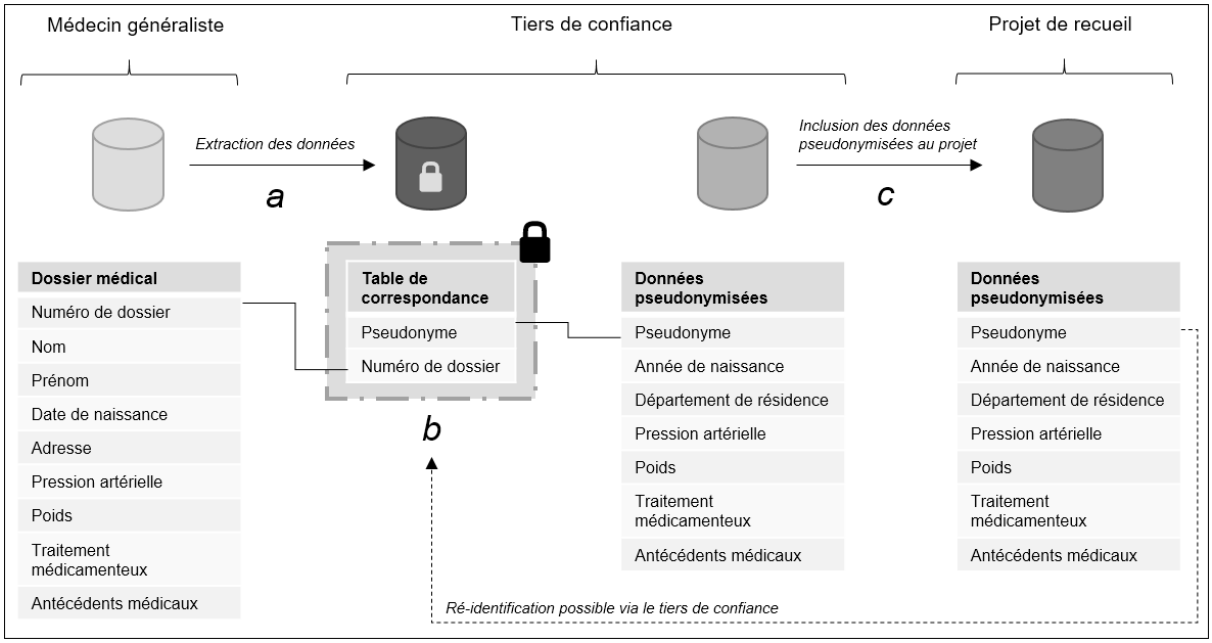

Figure 4. ***Schéma de synthèse des procédés de dé-identification*** *: a) les données sont extraites des bases de données des médecins généralistes, les données identifiantes sont retirées, cette phase peut être réalisée par un tiers de confiance. b) le tiers de confiance conserve en lieu sûr une table de correspondance à des fins de ré-identification ultérieure sous contrôle de la gouvernance du projet. c) seules les données dé-identifiées sont incluses dans l'entrepôt de données du projet de recueil.*

4.3.2 La dé-identification : un pré-requis indispensable

La dé-identification est une étape importante et retrouvée dans tous les projets de recueil de données. Plus elle est mise en place tôt dans le processus d'extraction des données et plus elle est performante, plus les données personnelles sont protégées mais à l'inverse il y a perte d'informations pour la recherche. Elle est rendue impérative par le concept de « Privacy by design » (Cf. 3.4.1.1).

La dé-identification est d'autant plus importante qu'il peut exister des failles de sécurité sur les entrepôts de données. En effet, celles-ci sont malheureusement fréquentes au sein des systèmes d'information, par exemple au Royaume-Uni : dans sa thèse le Docteur Angela Gibson-White (21) fait état de 7255 failles de sécurité découvertes au NHS entre 2011 et 2014, sur la base du rapport de l'organisation non gouvernementale Big Brother Watch. Plus récemment en 2017, il a été révélé (44) que les données de plus de 26 millions d'Anglais ont été rendues accessibles en ligne par erreur.

4.3.3 Limites de la dé-identification

Si la dé-identification semble un prérequis indispensable pour protéger les données personnelles, elle est loin d'être infaillible. Premièrement, les algorithmes de dé-identification comme la k-anonymisation sont d'autant plus fiable que le k est grand mais une ré-identification reste possible même si elle est techniquement compliquée. Par exemple, en 2017, des chercheurs de l'Université de Melbourne en Australie (45) ont montré qu'ils étaient capables de ré-identifier des personnes à partir d'un fichier de facturation dé-identifié concernant 2,9 millions de patients en le croisant avec des informations connues publiquement. Ils ont ainsi pu ré-identifier des informations concernant plusieurs personnalités publiques dont le Premier Ministre Australien. Deuxièmement, le chaînage des données accroit le risque

d'identification des données. Enfin, les clés de ré-identification sont gérées par un tiers de confiance, ce qui peut être source de faille de sécurité (15).

La dé-identification est un enjeu majeur car elle limite les données sources importées dans les projets, par exemple elle interdit généralement l'import des données de texte libre faute de dé-identification performante. Seul le CPRD extrait des données de texte libre, ce qui constitue une difficulté importante pour la dé-identification. Les procédés de dé-identification ne sont pas explicités sur le site ou les articles étudiés.

4.4 Un accès restreint aux données

Dans la plupart des réseaux de recueil de données, les accès semblent limités aux chercheurs. Les sociétés commerciales et industrielles ne sont pas citées. Ceci est peut-être dû à un biais de sélection des articles issus de notre équation de recherche. En effet, des réseaux d'investigations développés par des éditeurs logiciels ou industriels ont été identifiés mais n'exposaient pas dans leurs articles les modalités de protection des données.

Les accès aux données aux chercheurs se font en général après autorisation d'un comité d'éthique et d'un comité scientifique. Ces verrous sont essentiels à la protection des données mais ne devraient pas compliquer le travail des chercheurs. Pour simplifier ces étapes en France, les méthodologies de référence semblent être une solution adaptée. Il sera cependant nécessaire de confirmer si un projet de recueil de données en soins primaires est bien conforme à une méthodologie existante, et si les projets de recherche qui en découleront pourront en bénéficier à leur tour. Dans le cas contraire, il parait indispensable que la CNIL développe une méthodologie adaptée.

Deux modalités d'accès aux données sont souvent possibles : brutes ou agrégées. Il semble intéressant de conserver ces deux types d'accès. Une version agrégée, déjà travaillée sur le plan statistique, est plus facilement utilisable par certaines équipes. En revanche, l'accès aux données brutes permet de livrer une information plus riche et complète et de mener des études plus approfondies et plus solides méthodologiquement. Leur analyse nécessite cependant l'expertise des « data scientists ». Par exemple, la plateforme PopMedNet (35) développée à Harvard permet un large accès aux données tout en laissant les structures sources décisionnaires et propriétaires de leurs données.

Les modalités d'accès aux données sont actuellement très encadrées et restreintes. D'autres procédés de partage de l'information pourraient être développés. Par exemple, une offre « Open Data » permettrait à d'autres acteurs, non issus de la recherche, d'accéder à des échantillons informations recueillies en minimisant les risques de fuite de données personnelles(46).

La question de l'ouverture et du partage des données est importante et doit être très clairement définie et surveillée. Ce d'autant plus, à l'heure du machine learning, où les réseaux d'apprentissage peuvent assimiler des profils de patients à partir des bases de données. Un contrôle étroit doit donc être mené également concernant l'exploitation des sources de données.

4.5 Evolution de la réglementation

4.5.1 Impact du RGPD au niveau européen

Le Règlement Général de Protection des Données consacre notamment l'obligation de consentement au recueil de données personnelles, et ce pour toute l'Union Européenne. Il prévoit toutefois des exceptions en particulier dans le cadre de la recherche scientifique.

Les projets Européens seront amenés à se mettre en conformité avec les évolutions réglementaires récentes. Par exemple, le projet NIVEL (25) a été développé de façon à être conforme à la Directive Européenne de protection des données de 1995, et ce avant l'adoption de la nouvelle règlementation européenne sur la protection des données en 2016 (RGPD).

4.5.2 Evolution du cadre juridique Français

La France fut l'un des premiers pays à instaurer une réglementation sur le traitement des données personnelles en 1978 avec la mise en place de la Commission Nationale de l'Informatique et des Libertés (CNIL), dont est inspirée la Directive Européenne de 1995.

Cependant, jusqu'en 2016, il n'existait pas de texte précis régulant les données de santé, et notamment leur utilisation à des fins de recherche ou d'amélioration du système de soins.

Dans un article (46) issu d'une table ronde de 2015 sur les enjeux de l'accès aux données de santé pour les chercheurs, les participants formulaient plusieurs recommandations :

- Réduire les délais d'instruction des demandes à la CNIL en allégeant ou en automatisant les procédures
- Définir clairement les règles de gouvernance pour le futur Institut National des Données de Santé et un travail de collaboration avec toutes les parties prenantes du domaine
- Médicaliser les données en interfaçant le PMSI avec les entrepôts de données
- Mettre à jour rapidement les nomenclatures, notamment la CCAM
- Former des « data scientists » (ou « expert en mégadonnées » selon la Commission d'enrichissement de la langue française)
- Passer du « Big Data » à « Open Data » en facilitant l'accès aux données avec par exemple des échantillons de données sans accord CNIL préalable
- Permettre le croisement des bases de données en simplifiant les procédures
- Confier à l'INSERM la gestion d'un « Open Data » de la santé afin d'en pérenniser l'accès

La Loi de Santé de 2016 en instaurant le Système National des Données de Santé (SNDS) est venue garantir un accès à ces données sous contrôle de la CNIL avec une traçabilité des accès et en détaillant les règles. La loi prévoit que les entreprises productrices de produits de santé et des assureurs en santé ne puissent pas faire usage des données du SNDS pour la promotion de produits de santé ou pour l'exclusion de garanties des contrats d'assurance et la modification de cotisations ou de primes d'assurance d'un individu ou d'un groupe d'individus présentant un même risque (Article L1461-1-V du Code de la Santé Publique). Le SNDS se limite en 2018 aux données de l'Assurance Maladie (SNIIRAM), des hôpitaux (PMSI), des causes de décès (CépiDC), du handicap (MDPH, CNSA).

Une nouvelle instance éthique, spécifique aux données de santé, a vu le jour à cette occasion : le Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé (CEREES). Il est composé de 21 membres nommés par arrêté conjoint du Ministre chargé de la Santé et du Ministre chargé de la Recherche. Il a pour mission d'étudier les projets de recherche sur les données de santé qui ne rentrent pas dans le cadre de la recherche sur la personne humaine au sens du code de la santé publique.

La loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, transpose en France les règles édictées par le RGPD. Elle consiste en une modification de la loi du 6 janvier 1978 relative à l'informatique et aux libertés.

Concernant la recherche en santé on peut retenir ces grands principes :

- L'obligation de prévoir la protection des données par défaut dès la conception d'un projet
- La généralisation des analyses d'impact sur la protection des données (Cf. ci-dessous)
- L'obligation d'assurer la minimisation des données, en particulier par la pseudonymisation
- L'utilisation des données est possible sans consentement si elles sont anonymisées

Le principe de l'analyse d'impact relative à la protection des données (Privacy Impact Assessment, PIA) a été développé initialement aux USA. Il s'agit de repérer les risques d'atteinte à la vie privée qui pourraient être associés à l'utilisation de données personnelles. Le E-Government Act en 2002 a établi l'obligation aux USA de réaliser des PIA pour les collectes de données électroniques. Cette approche a été reprise au Canada, où le chainage des données avec d'autres bases n'est autorisé qu'après accord exprès d'un comité d'éthique. Chaque réseau régional du CPCSSN a réalisé au moins un PIA. Les critères de dé-identification (22) sont déterminés par le Tri-Council Policy Statement 2 (TCPS2).

En France, la CNIL propose une méthode de PIA qui consiste à :

- Délimiter et décrire le contexte du traitement de données considéré
- Analyser les mesures garantissant le respect des principes fondamentaux (proportionnalité et nécessité du traitement de données, protection des droits des personnes)
- Apprécier les risques sur la vie privée liés à la sécurité des données et vérifier qu'ils sont convenablement traités
- Formaliser la validation du PIA au regard des éléments précédents ou bien décider de réviser les étapes précédentes

4.6 Proposition d'une organisation d'un projet de recueil de données à l'échelle française

La figure 5 propose un modèle hypothétique pour un projet de recueil de données de soins primaires en France.

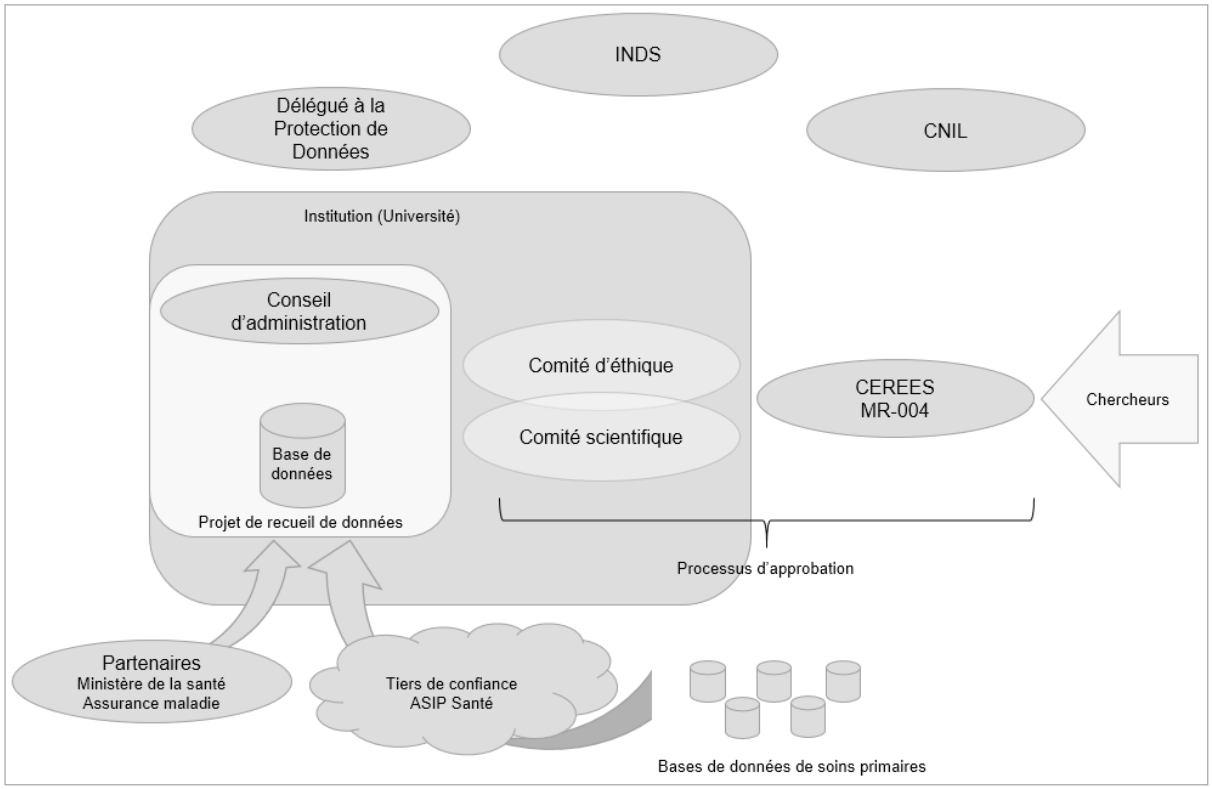

Figure 5. ***Hypothèse d'organisation d'un projet de recueil de données en France***

5 Conclusion

Cette revue systématique de la littérature internationale nous a permis de mettre en évidence des structures de gouvernance relativement similaires des différents projets de recueil de données en soins primaires, avec en particulier un pilotage éthique et scientifique. Le consentement des patients participants est pris en compte par l'intermédiaire de leur droit d'opposition à l'utilisation de leurs données. Nous avons constaté que chaque projet met en œuvre plusieurs mécanismes de sécurité afin de protéger les données des patients, à commencer par une dé-identification systématique des données, le plus souvent par un procédé de pseudonymisation. L'accès aux données est restreint aux chercheurs.

En France et en Europe, un nouveau cadre juridique vient à la fois renforcer les droits des citoyens en matière de données personnelles, mais aussi simplifier la réalisation de recherches à partir des données de santé.

L'équilibre entre le respect de la confidentialité et la puissance de la recherche doit cependant rester en constante réévaluation, d'autant que les nouvelles techniques de traitement de l'information facilitent la ré-identification des personnes a posteriori.

Imprimé n° 4

UNIVERSITE DE RENNES 1

Scolarité médecine 3ème cycle
2 avenue Professeur Léon Bernard
35043 RENNES Cedex

FACULTE DE MEDECINE

NOM et Prénom : LE BERRE Thomas

TITRE DE LA THESE d'EXERCICE

(Ce document sera à insérer dans les thèses définitives)

Titre :

Confidentialité des données au sein des projets de recueil et d'analyse de données en soins primaires : une revue systématique de la littérature.

Rennes, le

Docteur Marie-Line Gentil
10 rue de la Belle Epine
35132 Vezin Le Coquet
RPPS : 10100516276
AM : 351004023

Le Directeur de thèse

Rennes, le 27/8/2018

Le Président de jury

Vu et permis d'imprimer

Rennes, le 11 SEP. 2018

UNIVERSITE DE RENNES 1

Le Président de l'Université de Rennes1

Pour le Président et par délégation
le Vice-Président
D. ALIS

Maj : 23-janv -14

6 Bibliographie

1. Gentil M-L, Cuggia M, Fiquet L, Hagenbourger C, Le Berre T, Banâtre A, et al. Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. BMC Med Inform Decis Mak. 25 sept 2017;17(1):139.

2. Lacroix-Hugues V, Darmon D, Pradier C, Staccini P. Creation of the First French Database in Primary Care Using the ICPC2: Feasibility Study. Stud Health Technol Inform. 2017;245:462-6.

3. Riou C, Cuggia M, Garcelon N. Comment assurer la confidentialité dans les entrepôts de données biomédicaux ? /data/revues/03987620/v60sS1/S039876201100561X/ [Internet]. 22 févr 2012 [cité 2 sept 2018]; Disponible sur: http://www.em-consulte.com/en/article/694152

4. Papoutsi C, Reed JE, Marston C, Lewis R, Majeed A, Bell D. Patient and public views about the security and privacy of Electronic Health Records (EHRs) in the UK: results from a mixed methods study. BMC Med Inform Decis Mak [Internet]. 14 oct 2015 [cité 27 sept 2017];15. Disponible sur: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4607170/

5. Riordan F, Papoutsi C, Reed JE, Marston C, Bell D, Majeed A. Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. Int J Med Inf [Internet]. avr 2015 [cité 6 déc 2017];84(4):237-47. Disponible sur: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4344220/

6. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [Internet]. Disponible sur: https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679

7. Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ [Internet]. 21 juill 2009 [cité 20 juill 2018];339(jul21 1):b2535-b2535. Disponible sur: http://www.bmj.com/cgi/doi/10.1136/bmj.b2535

8. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ [Internet]. 21 sept 2017 [cité 8 août 2018];358:j4008. Disponible sur: https://www.bmj.com/content/358/bmj.j4008

9. CAPriCORN [Internet]. [cité 12 août 2018]. Disponible sur: http://capricorncdrn.org/?page_id=88

10. Electronic Medical Record Administrative data Linked Database (EMRALD) [Internet]. [cité 26 juill 2018]. Disponible sur: https://www.ices.on.ca/Research/Research-programs/Primary-Care-and-Population-Health/EMRALD

11. Réseau canadien de surveillance sentinelle en soins primaires [Internet]. [cité 26 juill 2018]. Disponible sur: http://rcsssp.ca/

12. GRAPHC [Internet]. [cité 26 juill 2018]. Disponible sur: http://graphc.anu.edu.au/

13. Clinical Practice Research Datalink - CPRD [Internet]. [cité 26 juill 2018]. Disponible sur: https://www.cprd.com/home/

14. Intego [Internet]. [cité 26 juill 2018]. Disponible sur: https://intego.be/en/Welcome

15. De Clercq E, Van Casteren V, Bossuyt N, Moreels S, Goderis G, Bartholomeeusen S, et al. Nation-wide primary healthcare research network: a privacy protection assessment. Stud Health Technol Inform. 2012;174:23-8.

16. NIVEL - Netherlands institute for health services research [Internet]. [cité 26 juill 2018]. Disponible sur: https://www.nivel.nl/en

17. Institute for Clinical Evaluative Sciences - ICES [Internet]. [cité 18 août 2018]. Disponible sur: https://www.ices.on.ca/About-ICES/Mission-vision-and-values

18. Kotecha JA, Manca D, Lambert-Lanning A, Keshavjee K, Drummond N, Godwin M, et al. Ethics and privacy issues of a practice-based surveillance system: Need for a national-level institutional research ethics board and consent standards. Can Fam Physician [Internet]. 1 oct 2011 [cité 24 juin 2018];57(10):1165-73. Disponible sur: http://www.cfp.ca/content/57/10/1165

19. Truyers C, Goderis G, Dewitte H, Akker M vanden, Buntinx F. The Intego database: background, methods and basic results of a Flemish general practice-based continuous morbidity registration project. BMC Med Inform Decis Mak [Internet]. 6 juin 2014 [cité 24 juin 2018];14:48. Disponible sur: https://doi.org/10.1186/1472-6947-14-48

20. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol [Internet]. 1 juin 2015 [cité 27 nov 2017];44(3):827-36. Disponible sur: https://academic.oup.com/ije/article/44/3/827/632531

21. Gibson-White A. Using information from electronic patient records for clinical, epidemiological and health services research [Thèse]. 2015 [cité 27 nov 2017]. Disponible sur: http://spiral.imperial.ac.uk/handle/10044/1/41839

22. Manca D. Data to inform primary care: The Canadian Primary Care Sentinel Surveillance Network (CPCSSN). 2014.

23. Solomonides A, Goel S, Hynes D, Silverstein JC, Hota B, Trick W, et al. Patient-Centered Outcomes Research in Practice: The CAPriCORN Infrastructure. Stud Health Technol Inform. 2015;216:584-8.

24. Kho AN, Hynes DM, Goel S, Solomonides AE, Price R, Hota B, et al. CAPriCORN: Chicago Area Patient-Centered Outcomes Research Network. J Am Med Inform Assoc JAMIA. août 2014;21(4):607-11.

25. Kuchinke W, Ohmann C, Verheij RA, van Veen E-B, Arvanitis TN, Taweel A, et al. A standardised graphic method for describing data privacy frameworks in primary care research using a

flexible zone model. Int J Med Inf [Internet]. 1 déc 2014 [cité 27 nov 2017];83(12):941-57. Disponible sur: http://www.sciencedirect.com/science/article/pii/S1386505614001634

26. Stevenson F. The use of electronic patient records for medical research: conflicts and contradictions. BMC Health Serv Res [Internet]. 29 mars 2015 [cité 24 juin 2018];15:124. Disponible sur: https://doi.org/10.1186/s12913-015-0783-6

27. Pantazos K, Lauesen S, Lippert S. De-identifying an EHR database - anonymity, correctness and readability of the medical record. Stud Health Technol Inform. 2011;169:862-6.

28. MacRury S, Finlayson J, Hussey-Wilson S, Holden S. Development of a pseudo/anonymised primary care research database: Proof-of-concept study. Health Informatics J [Internet]. 1 juin 2016 [cité 16 nov 2017];22(2):113-9. Disponible sur: https://doi.org/10.1177/1460458214535118

29. Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B. Anonymization of longitudinal electronic medical records. IEEE Trans Inf Technol Biomed Publ IEEE Eng Med Biol Soc. mai 2012;16(3):413-23.

30. Tu K, Klein-Geltink J, Mitiku TF, Mihai C, Martin J. De-identification of primary care electronic medical records free-text data in Ontario, Canada. BMC Med Inform Decis Mak. 18 juin 2010;10:35.

31. Grouin C, Rosier A, Dameron O, Zweigenbaum P. Testing tactics to localize de-identification. Stud Health Technol Inform [Internet]. 2009 [cité 7 juill 2018];150:735-9. Disponible sur: http://europepmc.org/abstract/med/19745408

32. Garies S, Birtwhistle R, Drummond N, Queenan J, Williamson T. Data Resource Profile: National electronic medical record data from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). Int J Epidemiol [Internet]. 1 août 2017 [cité 24 juin 2018];46(4):1091-1092f. Disponible sur: https://academic.oup.com/ije/article/46/4/1091/3058732

33. Mazumdar S, Konings P, Hewett M, Bagheri N, McRae I, Del Fante P. Protecting the privacy of individual general practice patient electronic records for geospatial epidemiology research. Aust N Z J Public Health [Internet]. 1 déc 2014 [cité 16 nov 2017];38(6):548-52. Disponible sur: http://onlinelibrary.wiley.com.passerelle.univ-rennes1.fr/doi/10.1111/1753-6405.12262/abstract

34. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Ther Adv Drug Saf [Internet]. avr 2012 [cité 15 nov 2017];3(2):89-99. Disponible sur: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4110844/

35. PopMedNet [Internet]. [cité 26 juill 2018]. Disponible sur: https://www.popmednet.org/

36. HHS.gov, U.S. Department of Health & Human Services, HHS.gov, Office for Civil Rights. Summary of the HIPAA Privacy Rule [Internet]. mai 7, 2008. Disponible sur: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html?language=es

37. Thorpe JH, Gray EA. Big data and ambulatory care: breaking down legal barriers to support effective use. J Ambulatory Care Manage. mars 2015;38(1):29-38.

38. Ohno-Machado L, Agha Z, Bell DS, Dahm L, Day ME, Doctor JN, et al. pSCANNER: patient-centered Scalable National Network for Effectiveness Research. J Am Med Inform Assoc JAMIA [Internet]. juill 2014 [cité 16 nov 2017];21(4):621-6. Disponible sur: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4078293/

39. Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données [Internet]. 281, 31995L0046 nov 23, 1995. Disponible sur: http://data.europa.eu/eli/dir/1995/46/oj/fra

40. Code de la santé publique - Comités de protection des personnes et autorité compétente [Internet]. Disponible sur: https://www.legifrance.gouv.fr/affichCode.do;jsessionid=B86849809D14704D7DC9DBD3274E5933.tplgfr31s_2?idSectionTA=LEGISCTA000006171003&cidTexte=LEGITEXT000006072665&dateTexte=20180815

41. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

42. Décret n°2005-1309 du 20 octobre 2005 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. 2005-1309 oct 20, 2005.

43. Délibération n° 2018-155 du 3 mai 2018 portant homologation de la méthodologie de référence relative aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches n'impliquant pas la personne humaine, des études et évaluations dans le domaine de la santé (MR-004).

44. Donnelly L. Security breach fears over 26 million NHS patients. The Telegraph [Internet]. 17 mars 2017 [cité 15 août 2018]; Disponible sur: https://www.telegraph.co.uk/news/2017/03/17/security-breach-fears-26-million-nhs-patients/

45. Melbourne DVT Dr Chris Culnane and Dr Ben Rubinstein, University of. The simple process of re-identifying patients in public health records [Internet]. Pursuit. 2017 [cité 15 août 2018]. Disponible sur: https://pursuit.unimelb.edu.au/articles/the-simple-process-of-re-identifying-patients-in-public-health-records

46. Chatellier G, Varlet V, Blachier-Poisson C, participants of Giens XXXI, Round Table No. 6. « Big data » and « open data »: What kind of access should researchers enjoy? Thérapie. févr 2016;71(1):97-105, 107-14.

U.F.R. DE **MEDECINE** DE RENNES

N°

LE BERRE, Thomas - Confidentialité des données au sein des projets de recueil et d'analyse de données en soins primaires : une revue systématique de la littérature.

56 feuilles, 5 graphiques, 1 tableau, 30 cm.- Thèse : Médecine ; Rennes 1 ; 2018 ; N°

INTRODUCTION : Les données issues des dossiers médicaux de soins primaires (DMSP) contiennent une synthèse de l'histoire médicale des patients et une vue globale de la santé de la population. Si elles sont particulièrement intéressantes et utilisées dans plusieurs domaines de recherche comme l'épidémiologie, elles restent des données sensibles et nécessitent une protection adaptée. **MÉTHODE** : Une revue systématique de la littérature à partir de PubMed et Google Scholar en suivant les critères PRISMA et AMSTAR2. Les sites web des projets ont également été inclus dans l'analyse ainsi que les textes réglementaires des USA et de l'Union Européenne les concernant. Un formulaire standardisé préétabli a été rempli à partir de toutes ces informations en suivant cinq axes : acteurs des projets, modalités de consentement des patients, modalités d'accès aux données, méthodes de dé-identification des données, cadre réglementaire et éthique autour de ces projets. **RÉSULTATS** : Sept projets de recueil de données de soins primaires ont été extraits et leurs sites web ont été analysés, 23 articles ont été étudiés en texte complet. Le consentement des patients n'est généralement pas recherché, ceux-ci pouvant faire valoir leur droit d'opposition. Les accès sont autorisés aux chercheurs après validation d'un comité d'éthique. La méthode de dé-identification la plus répandue est la pseudonymisation. Des comités d'éthique sont associés aux structures de recueil. Le Règlement Général sur la Protection des Données (RGPD) met en place un nouveau cadre réglementaire en Europe. **DISCUSSION** : Les projets de recueil et d'analyse de DMSP disposent de structures de gouvernance qui donnent une orientation au projet. Le consentement des patients participant n'est pas recueilli dans la mesure où les données sont dé-identifiées, cependant, les méthodes de dé-identification ne sont pas infaillibles et font courrir un risque de ré-identification. L'accès aux données est soumis à un processus d'approbation qui limite le risque de mésusage des données. Avec l'entrée en vigueur du RGPD, la France s'est dotée d'un cadre juridique protecteur qui facilite toutefois l'utilisation des DMSP pour la recherche en santé. **CONCLUSION** : L'équilibre entre le respect de la confidentialité et la puissance de la recherche doit rester en constante réévaluation, d'autant que les nouvelles techniques de traitement de l'information facilitent des réidentifications a posteriori.

Rubrique de classement : MEDECINE GENERALE

Mots-clés : Médecine générale, soins primaires, recueil de données, confidentialité des données, consentement, dé-identification, RGPD

Mots-clés anglais MeSH : General Practice, Primary Care, Electronic Health Records, Data Anonymization, Privacy, Informed Consent, Ethics Committees

JURY :

Président : Monsieur le Professeur Marc CUGGIA

Assesseurs : Madame le Docteur Marie Line GENTIL [directeur de thèse]

Monsieur le Professeur Bruno LAVIOLLE

Monsieur le Docteur Ronan GARLANTEZEC