

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

NOUVELLES PRATIQUES DE RECHERCHE À L'ÈRE NUMÉRIQUE : LES ENJEUX
ÉPISTÉMIQUES ET SOCIOPOLITIQUES DE L'ÉPIDÉMIOLOGIE NUMÉRIQUE DANS LE
CONTEXTE CANADIEN

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE DU

DOCTORAT EN SCIENCE, TECHNOLOGIE ET SOCIÉTÉ

PAR

RENÉE CARPENTIER

MAI 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Mes études doctorales ont élargi mes horizons et m'ont fait croître énormément autant sur le plan personnel que sur le plan intellectuel et professionnel.

Merci beaucoup à ma directrice de thèse, professeure Majlinda Zhegu, et ma co-directrice de thèse, professeure Florence Millerand, pour leur aide précieuse lors de mon cheminement dans le domaine STS. Merci pour vos conseils, votre professionnalisme, votre patience et la rigueur de vos commentaires tout au long de mon apprentissage et de la rédaction de ma thèse.

Je tiens à remercier tous les chercheurs et toutes les chercheuses en épidémiologie qui ont accepté de participer à ma recherche de terrain. Leur contribution fut très importante pour atteindre les objectifs de ma thèse.

Je n'aurais pas pu réussir ma thèse sans le soutien de mes proches.

Un gros merci à ma maman, Louise, d'avoir cru en la réussite de mon doctorat. Une maman extraordinaire qui m'a toujours énormément encouragé lors de mes études et dans toutes les sphères de ma vie. C'est bien grâce à elle que j'ai développé de la ténacité, de la discipline et l'intérêt pour les études.

Un gros merci aussi à Abdou, mon conjoint, pour ses judicieux conseils et sa compréhension tout au long de mon parcours depuis 2015.

Merci à Bernard qui a toujours agi comme un père pour moi et à grand-maman, Marie-Lise, pour tout le réconfort et les mots d'encouragement lors de ce parcours rempli de défis.

DÉDICACE

À Maman, Abdou, Bernard et Grand-maman

TABLE DES MATIÈRES

REMERCIEMENTS	ii
DÉDICACE.....	iii
LISTE DES FIGURES.....	viii
LISTE DES TABLEAUX.....	ix
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES	x
RÉSUMÉ.....	xi
ABSTRACT	xiii
INTRODUCTION.....	1
CHAPITRE I LA PROBLÉMATIQUE.....	5
1.1 Le contexte de recherche lié à la pratique de l'épidémiologie numérique.....	5
1.1.1. La discipline de l'épidémiologie	6
1.1.1.1 Les façons de concevoir l'épidémiologie	7
1.1.1.2 Les sous-disciplines de l'épidémiologie.....	9
1.1.1.3 Des moments clés dans l'histoire de l'épidémiologie	13
1.1.2 Les pratiques de recherche.....	17
1.1.3 Les changements dans les pratiques de recherche	20
1.1.4 L'interdisciplinarité	25
1.1.5 La mise en données de la société	29
1.1.5.1 Les façons de concevoir la mise en données de la société.....	30
1.1.5.2 Les enjeux sociopolitiques de la mise en données	31
1.1.5.3 Les conséquences pour la recherche	33
1.1.5.4 Les enjeux épistémiques liés à la place centrale des données.....	35
1.2 L'état des recherches en épidémiologie numérique	37
1.2.1 La méthodologie employée pour la revue de la littérature.....	37
1.2.2 Un portrait des recherches en épidémiologie numérique.....	40
1.2.3 Les façons de concevoir l'épidémiologie numérique	45
1.2.4 Les technologies utilisées en épidémiologie numérique.....	47
1.2.5 Les acteurs impliqués.....	48
1.2.6 Les difficultés et les controverses	50
1.2.7 La synthèse de la revue de la littérature.....	53
1.3 Le but, les objectifs et les questions de recherche dans le cadre de la thèse	54
1.4 La conclusion sur la problématique.....	56
CHAPITRE II LE CADRE THÉORIQUE	59
2.1 Les cultures épistémiques	62

2.1.1	La culture et la diversité dans les sciences.....	63
2.1.2	Les signes et les apparences secondaires	65
2.1.3	Les mesures et les données expérimentales	67
2.1.4	Les pratiques organisationnelles dans les sciences	69
2.1.5	Les critiques, apports et limites des cultures épistémiques.....	70
2.2	Les enjeux sociétaux des méthodes de recherche et des données	74
2.2.1	La vie sociale des méthodes.....	74
2.2.2	Les dispositifs méthodologiques	78
2.2.3	Les enjeux politiques des données	80
2.3	La conclusion sur le cadre théorique	83
CHAPITRE III LA MÉTHODOLOGIE		91
3.1	La recherche de terrain	91
3.2	La méthode d'analyse des données	100
3.3	La conclusion sur la méthodologie	105
CHAPITRE IV LA PRATIQUE DE L'ÉPIDÉMIOLOGIE NUMÉRIQUE		107
4.1	Les caractéristiques de la pratique de l'épidémiologie numérique.....	108
4.1.1	La provenance des données.....	108
4.1.2	La notion de « numérique »	111
4.1.3	La notion de « technologie » et « d'outils électroniques »	112
4.1.4	La diversité des sources de données et des technologies	113
4.1.5	La nouveauté de la pratique de l'épidémiologie numérique	115
4.2	La diffusion de la pratique de l'épidémiologie numérique auprès de chercheurs en épidémiologie	117
4.2.1	Le niveau de connaissance de la pratique de l'épidémiologie numérique	118
4.2.2	Le niveau d'expérience de la pratique de l'épidémiologie numérique	119
4.2.3	Les attitudes envers la pratique de l'épidémiologie numérique.....	122
4.2.3.1	L'adoption	123
4.2.3.2	Les résistances.....	125
4.2.3.3	Les réserves	126
4.2.4	Les perceptions en épidémiologie clinique	132
4.2.5	La crédibilité de la pratique de l'épidémiologie numérique	137
4.2.5.1	Les points de vue des chercheurs pratiquant l'épidémiologie numérique.....	138
4.2.5.2	Les points de vue des chercheurs ne pratiquant pas l'épidémiologie numérique	140
4.3	La conclusion sur la pratique de l'épidémiologie numérique.....	150
CHAPITRE V LES ENJEUX ÉPISTÉMIQUES DE LA PRATIQUE DE L'ÉPIDÉMIOLOGIE NUMÉRIQUE.....		153
5.1	La signification des données.....	154
5.1.1	Les données utilisées dans la pratique de l'épidémiologie traditionnelle	156
5.1.2	Les données utilisées dans la pratique de l'épidémiologie numérique	162

5.1.3	Les différences entre les données traditionnelles et les données non traditionnelles ...	164
5.1.3.1	Les points de vue des chercheurs ne pratiquant pas l'épidémiologie numérique	164
5.1.3.2	Les points de vue des chercheurs pratiquant l'épidémiologie numérique.....	169
5.1.4	L'échantillonnage populationnel.....	172
5.2	La confiance envers les données numériques comme dispositifs méthodologiques	177
5.2.1	Les données générées par des moteurs de recherche Internet.....	178
5.2.1.1	Le niveau de confiance et de connaissance du dispositif méthodologique	179
5.2.1.2	L'utilité du dispositif méthodologique	181
5.2.1.3	Les difficultés liées au dispositif méthodologique	183
5.2.1.4	La confiance au sein de l'épidémiologie environnementale	184
5.2.1.5	La confiance au sein de l'épidémiologie clinique	186
5.2.1.6	La confiance au sein des chercheurs pratiquant l'épidémiologie numérique.....	190
5.2.2	Les données générées par des médias sociaux	193
5.2.2.1	Les réserves	194
5.2.2.2	L'utilité des données générées par les médias sociaux pour des recherches sociales	198
5.2.3	Les données générées par des téléphones cellulaires	201
5.2.3.1	Les perceptions.....	202
5.2.3.2	Les biais liés au dispositif méthodologique	206
5.3	La conclusion sur les enjeux épistémiques	211
CHAPITRE VI LES ENJEUX SOCIOPOLITIQUES DE LA PRATIQUE DE L'ÉPIDÉMIOLOGIE NUMÉRIQUE		217
6.1	Les enjeux d'expertise liés à la pratique de l'épidémiologie numérique.....	217
6.1.1	Les nouvelles compétences et les spécialités envisagées.....	218
6.1.1.1	Mise à jour de la discipline de l'épidémiologie	218
6.1.1.2	Les compétences en informatique et en science des données	219
6.1.1.3	Les compétences non techniques	223
6.1.2	La formation des chercheurs en épidémiologie	226
6.1.2.1	Les points de vue des chercheurs ne pratiquant pas l'épidémiologie numérique	226
6.1.2.2	Les points de vue des chercheurs ayant l'expérience de l'épidémiologie numérique	229
6.1.3	L'interdisciplinarité.....	233
6.1.3.1	La discipline de l'informatique	234
6.1.3.2	Les autres disciplines	234
6.2	La légitimité de la pratique de l'épidémiologie numérique.....	238
6.2.1	La reconnaissance institutionnelle de la pratique de l'épidémiologie numérique	239
6.2.1.1	Les recommandations de formation des chercheurs en épidémiologie.....	239
6.2.1.2	Les programmes d'études en épidémiologie.....	241
6.2.1.3	Les liens avec l'informatique lors des enseignements en épidémiologie.....	245
6.2.2	Les barrières à la légitimité de la pratique de l'épidémiologie numérique	249
6.2.2.1	Les cadres juridiques liés aux données	249
6.2.2.2	Les ressources technologiques	251
6.2.2.3	Le manque de confiance envers les détenteurs de données et les gouvernements.....	253
6.2.3	Les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique	255

6.2.3.1	Les risques vis-à-vis de la sécurité et de la confidentialité des données numériques	256
6.2.3.2	Les pouvoirs de grandes entités privées détentrices des données numériques	259
6.3	La conclusion sur les enjeux sociopolitiques	263
CHAPITRE VII LA DISCUSSION		265
7.1	Les caractéristiques et les enjeux épistémiques et sociopolitiques de la pratique de l'épidémiologie numérique	265
7.2	L'écosystème de la pratique de l'épidémiologie numérique	270
7.2.1	Les institutions, les acteurs et les technologies	271
7.2.2	La confiance au sein des chercheurs	273
7.2.3	Les tensions	275
7.3	Les contextes de la coproduction des connaissances à l'ère numérique	279
7.3.1	Les modes de la coproduction des connaissances	279
7.3.2	La complémentarité entre l'épidémiologie numérique et l'épidémiologie traditionnelle	282
7.4	Vers la numérisation des disciplines?	285
CONCLUSION		289
La démarche de la thèse		289
Les contributions sur le plan empirique		292
Les contributions sur le plan théorique		296
Les principaux éléments de la discussion		298
Les limites de cette recherche et des pistes pour des recherches futures		299
ANNEXE A LE GUIDE D'ENTRETIEN		301
ANNEXE B LA CODIFICATION		314
BIBLIOGRAPHIE		324

LISTE DES FIGURES

Figure 1. 1 La démarche de la revue de la littérature datée de 2019.....	38
Figure 1. 2 La requête utilisée dans Scopus datée de 2022.....	39
Figure 1. 3 Le nombre de publications en épidémiologie numérique par année dans Scopus entre 2010 et 2021	41
Figure 1. 4 La distribution des publications en épidémiologie numérique par discipline ou spécialité à partir de Scopus entre 2010 et 2022	42
Figure 1. 5 Les affiliations des chercheurs ayant au moins 8 publications en épidémiologie numérique à partir de Scopus entre 2010 et 2022	43
Figure 1. 6 Les pays ayant publié au moins 15 articles en épidémiologie numérique à partir de Scopus entre 2010 et 2022	44
Figure 3. 1 La distribution des participants selon le sexe	98
Figure 3. 2 La distribution des participants selon leur sous-discipline de l'épidémiologie.....	99
Figure 3. 3 La distribution des participants selon leur expérience de la pratique de l'épidémiologie numérique.....	99
Figure 3. 4 Les étapes de la démarche d'analyse inspirée par l'analyse par théorisation ancrée développée par Paillé (1994).....	105

LISTE DES TABLEAUX

Tableau 2. 1 Le sommaire des travaux théoriques retenus pour la thèse	84
Tableau 2. 2 La cartographie des thèmes retenus en fonction des chapitres de la thèse	86
Tableau 3. 1 Le nombre de chercheurs contactés, le nombre de participants et le pourcentage de participation total et par province lors de la recherche de terrain	95
Tableau 3. 2 Le nombre de participants par institution	96
Tableau 3. 3 Les catégories des propos des participants à la recherche de terrain	101

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

Acronyme	Description
CDC	Centers for Disease Control and Prevention
CRSCC	Centre for Research on Socio-Cultural Change
DEL	Digital Epidemiology Lab
DEPHL	Digital Epidemiology and Population Health Laboratory
EC	L'Encyclopédie canadienne
INSPQ	Institut national de santé publique du Québec
IRD	Institut de recherche pour le développement
MGU	McGill University
NRC	National Research Council
OMS	Organisation mondiale de la santé
RKI	Robert Koch Institute
RSI	Règlement sanitaire international
SD	Science Direct
WU	Western University

RÉSUMÉ

Cette thèse en Science, technologie et société analyse l'émergence de la pratique de l'épidémiologie numérique, c'est-à-dire une nouvelle façon de faire des analyses en épidémiologie s'appuyant notamment sur des données de sources non traditionnelles comme des médias sociaux, des moteurs de recherche Internet et des téléphones cellulaires. Le but de notre recherche doctorale est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de l'épidémiologie numérique en tant que nouvelle pratique de recherche. De plus, nous cherchons à savoir ce que signifie la pratique de l'épidémiologie numérique pour les chercheurs, afin de saisir en quoi cette nouvelle pratique est différente des pratiques de recherche de l'épidémiologie traditionnelle et pour comprendre comment elle se diffuse au sein de la discipline de l'épidémiologie.

Pour ce faire, nous avons recueilli et analysé les points de vue de chercheurs en épidémiologie sur les différentes facettes des dimensions sociales, politiques, scientifiques et technologiques de cette nouvelle pratique. La méthodologie s'appuie sur une recherche qualitative basée sur des entretiens menés auprès de 36 chercheurs canadiens en épidémiologie. Les analyses ont été guidées par les questionnements suivants : comment est perçue la pratique de l'épidémiologie numérique et quels sont les enjeux épistémiques et sociopolitiques caractérisant le contexte d'émergence de cette nouvelle pratique? Notre recherche révèle que les enjeux épistémiques de la pratique de l'épidémiologie numérique sont liés à la question de la signification des données et à la confiance envers des dispositifs méthodologiques, tandis que les enjeux sociopolitiques concernent surtout l'expertise et l'interdisciplinarité en lien avec la pratique de l'épidémiologie numérique, ainsi que la légitimité de cette pratique émergente.

Nos analyses montrent que cette nouvelle pratique est peu connue et peu mise en pratique par les communautés de chercheurs canadiens en épidémiologie. Sur le plan épistémique, les chercheurs considèrent la pratique de l'épidémiologie numérique comme étant utile pour des recherches exploratoires et des recherches sociales analysant des opinions, mais ils la considèrent moins pertinente pour des recherches cliniques, pour déterminer l'incidence réelle des problèmes de santé et pour la surveillance des maladies. Sur le plan sociopolitique, les constats de notre recherche doctorale font ressortir des dimensions éthiques, de légalité et de pouvoir relatives à cette nouvelle pratique de recherche, ainsi que des écarts importants entre les ressources disponibles au sein de la discipline de l'épidémiologie et celles requises pour pratiquer l'épidémiologie numérique.

Cette étude montre que les nouvelles pratiques de recherche à l'ère numérique impliquent autant des acteurs humains (organismes et chercheurs) que des acteurs non-humains (dispositifs numériques et algorithmes) interagissant sous forme de réseaux fluides et hétérogènes lors des recherches. Les contextes de la coproduction des connaissances sont caractérisés à la fois par une combinaison de divers modes de production des connaissances, la complémentarité entre les connaissances algorithmiques et les pratiques traditionnelles et le début de la numérisation des disciplines. Des questions restent en suspens quant au rôle des organismes produisant des données officielles, de la transdisciplinarité et de la diversification des sites de production des connaissances, notamment en ce qui concerne la division du travail entre les chercheurs, experts professionnels et profanes.

Mots clés : épidémiologie numérique, mise en données, données numériques, dispositif méthodologique, coproduction des connaissances, cultures épistémiques, interdisciplinarité, discipline, pratique de recherche, diffusion d'une pratique de recherche, moteur de recherche Internet, médias sociaux, téléphones cellulaires, expertise, légitimité, institutionnalisation.

ABSTRACT

This Science, Technology and Society dissertation analyzes the emergence of the practice of digital epidemiology, a new way of doing epidemiological analysis that relies on data from non-traditional sources such as social media, Internet search engines and cell phones. The goal of our doctoral research is to document and understand the epistemic and sociopolitical issues that shape the emergence of digital epidemiology as a new research practice. In addition, we seek to learn what the practice of digital epidemiology means to researchers, to understand how this new practice is different from the research practices of traditional epidemiology, and to understand how it is diffusing within the discipline of epidemiology.

To do this, we collected and analyzed the views of epidemiology researchers on the various facets of the social, political, scientific and technological dimensions of this new practice. The methodology is based on qualitative research through interviews with 36 Canadian epidemiology researchers. The analyses were guided by the following questions: how is the practice of digital epidemiology perceived and what are the epistemic and sociopolitical issues characterizing the context of the emergence of this new practice. Our research reveals that the epistemic issues of the practice of digital epidemiology are related to the question of the meaning of data and confidence in methodological devices, while the sociopolitical issues relate mainly to expertise and interdisciplinarity in connection with the practice of digital epidemiology, as well as the legitimacy of this emerging practice.

Our analyses show that this new practice is little known and little used by the Canadian epidemiology research community. At the epistemic level, researchers consider the practice of digital epidemiology to be useful for exploratory research and social research analyzing opinions, but they consider it to be less relevant for clinical research, for determining the actual incidence of health problems and for disease surveillance. At the sociopolitical level, the findings of our doctoral research point to ethical, legal and power dimensions related to this new research practice, as well as significant gaps between the resources available within the discipline of epidemiology and those required to practice digital epidemiology.

This study shows that new research practices in the digital age involve both human actors (organizations and researchers) and non-human actors (digital devices and algorithms) interacting as fluid and heterogeneous networks during research. The contexts of knowledge coproduction are characterized both by a combination of various modes of knowledge production, the complementarity between algorithmic knowledge and traditional practices, and the beginning of the digitization of disciplines. Questions remain about the role of official data producing organizations, trans disciplinarity, and the diversification of knowledge production sites, particularly with respect to the division of labor between researchers, professional experts, and laypersons.

Keywords : digital epidemiology, datafication, digital data, methodological device, coproduction of knowledge, epistemic cultures, interdisciplinarity, discipline, research practice, diffusion of a research practice, Internet search engine, social media, cell phones, expertise, legitimacy, institutionalization

INTRODUCTION

Cette thèse porte sur l'émergence d'une nouvelle pratique de recherche en épidémiologie désignée comme l'épidémiologie numérique. Cette pratique serait une nouvelle façon de faire des analyses en épidémiologie s'appuyant notamment sur des données de sources non traditionnelles comme celles générées à partir des médias sociaux, des moteurs de recherche internet et de téléphones cellulaires (Salathé *et al.*, 2012; Makri, 2017). La pratique de l'épidémiologie numérique est devenue possible grâce au phénomène de la mise en données de la société à l'ère numérique. Certains affirment que ce phénomène réfère à la transformation d'informations dans un format de données permettant de les quantifier afin d'effectuer des analyses et ainsi, déverrouiller la valeur implicite et latente de l'information (Mayer-Schönberger et Cukier, 2013). Le phénomène de la mise en données peut amener des transformations dans presque tous les domaines où des données sont générées et analysées comme dans la sphère économique (Mayer-Schönberger et Cukier, 2013; Olleros et Zhegu, 2016; Beer, 2017), les sciences humaines (Vinck, 2016), la santé (Lupton, 2016; Thoër et Millerand, 2016) et les sciences sociales (Lupton, 2015; Marres, 2012; Boullier, 2015).

Le développement de la pratique de l'épidémiologie numérique est directement lié au phénomène de la mise en données de la société. Il s'agit d'une pratique émergente de l'ère numérique proposant d'utiliser des sources de données différentes par rapport à celles utilisées couramment en épidémiologie, afin d'étudier des phénomènes épidémiologiques. Or cette pratique soulève des enjeux épistémiques, parce que les connaissances, objectifs et pratiques liés à la vérité sont les éléments centraux des domaines scientifiques d'une part (Knorr-Cetina et Reichmann, 2015), et les méthodes d'analyse de données et les instruments utilisés peuvent influencer les façons de concevoir les phénomènes étudiés d'autre part (Knorr-Cetina, 1999).

Nous nous attendons aussi à ce que la pratique de l'épidémiologie numérique implique des enjeux sociopolitiques, car les méthodes de recherche et les données numériques peuvent amener de nouvelles façons de travailler en recherche (Marres, 2012; Mattern, 2013; Rogers, 2015; Boullier, 2015; Kitchin, 2014; Ruppert *et al.*, 2013), avoir un impact important sur la vie des individus (Mayer-Schönberger et Cukier, 2013; Kitchin, 2014; Lupton, 2015) et entraîner une concentration

de pouvoir auprès de grands détenteurs de données comme Google et Facebook (Mayer-Schönberger et Cukier, 2013; Zwitter, 2014; Ruppert *et al.*, 2015).

Comme toute nouveauté, certains chercheurs pourraient adopter la pratique de l'épidémiologie numérique alors que d'autres pourraient la considérer comme inappropriée pour des recherches en santé publique. La littérature sur le sujet porte sur les technologies utilisées et évoque certaines initiatives de recherches ainsi que des problèmes éthiques. Cependant, il n'y a pas d'analyse approfondie des enjeux épistémiques et sociopolitiques qui façonnent l'émergence de cette pratique, alors qu'elle fait l'objet d'un nombre grandissant de publications dans plusieurs pays, dont le Canada.

Le but de cette recherche doctorale est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. Nous voulons aussi comprendre ce que signifie la pratique de l'épidémiologie numérique afin de saisir en quoi cette nouvelle pratique est différente par rapport à l'épidémiologie traditionnelle et comment celle-ci est diffusée dans la discipline de l'épidémiologie. À la lumière de ces propos, cette recherche poursuit les objectifs suivants :

- Comprendre ce que signifie la pratique de l'épidémiologie numérique.
- Identifier les enjeux épistémiques et sociopolitiques qui accompagnent l'émergence de la pratique de l'épidémiologie numérique.

Pour atteindre ces objectifs, nous avons analysé les points de vue de chercheurs en épidémiologie sur les différentes facettes des dimensions sociales, politiques, scientifiques et technologiques de cette nouvelle pratique. La méthodologie employée s'appuie sur une recherche de terrain réalisée auprès de 36 chercheurs canadiens en épidémiologie, afin d'analyser leurs perspectives sur la pratique de l'épidémiologie numérique. Nous avons choisi une méthodologie qualitative fondée sur des entretiens individuels semi-dirigés, afin de mettre en œuvre la recherche de terrain. Les résultats de la recherche de terrain ont été analysés en s'inspirant de l'analyse par théorisation ancrée proposée par Paillé (1994). Nous avons aussi effectué une analyse thématique pour résumer et analyser les propos des participants aux entretiens.

Ce travail d'analyse a abouti à une théorisation de nos résultats fondée sur les cultures épistémiques de Knorr-Cetina (1999), la vie sociale des méthodes élaborée dans les travaux de Savage (2013), Law *et al.* (2011), Law (2010) et Lury *et al.* (2020) et les enjeux politiques des données élaborés par Ruppert *et al.* (2017). Cette démarche nous amène à théoriser la pratique de l'épidémiologie numérique comme une culture épistémique émergente soulevant des enjeux de nature épistémique et sociopolitique directement liés à l'omniprésence des données et au renouvellement des méthodes de recherche par l'utilisation de nouvelles sources de données. Cette théorisation s'appuie sur notre modélisation des résultats de la recherche qui prend la forme de deux types d'enjeux sur la pratique de l'épidémiologie numérique, à savoir les enjeux d'ordre épistémique et les enjeux d'ordre sociopolitique, selon les entretiens avec les chercheurs en épidémiologie ayant participé à cette recherche.

La thèse est structurée en sept chapitres, à savoir la problématique (Chapitre I), le cadre théorique (Chapitre II), la méthodologie (Chapitre III), la pratique de l'épidémiologie numérique (Chapitre IV), les enjeux épistémiques de la pratique de l'épidémiologie numérique (Chapitre V), les enjeux sociopolitiques (Chapitre VI) de la pratique de l'épidémiologie numérique et la discussion (Chapitre VII). Dans la problématique (Chapitre I), nous exposons le contexte de recherche lié à la pratique de l'épidémiologie numérique et présentons une revue de la littérature sur cette nouvelle pratique. Dans le cadre théorique (Chapitre II), nous abordons la coproduction des connaissances (Jasanoff, 2004), les cultures épistémiques (Knorr-Cetina, 1999), la vie sociale des méthodes (Law, 2010; Savage, 2013; Law et Ruppert, 2013; Ruppert *et al.*, 2013; Lury *et al.*, 2020) et les enjeux politiques des données (Ruppert *et al.*, 2017). Dans la méthodologie (Chapitre III), nous expliquons comment nous avons procédé pour conduire la recherche de terrain et l'analyse.

Dans le chapitre sur la pratique de l'épidémiologie numérique (Chapitre IV), nous questionnons ce que signifie cette nouvelle pratique et la diffusion de cette nouvelle pratique auprès de chercheurs en épidémiologie. Le Chapitre V explique les enjeux épistémiques de la pratique de l'épidémiologie numérique selon les points de vue des chercheurs canadiens en épidémiologie ayant participé à notre recherche, en faisant ressortir la question de la signification des données et du niveau de confiance envers les dispositifs méthodologiques que sont les moteurs de recherche Internet, les

médias sociaux et les téléphones cellulaires. Le Chapitre VI aborde les enjeux sociopolitiques en analysant les questions d'expertise et de légitimité de la pratique de l'épidémiologie numérique. Dans le Chapitre VII, nous discutons nos résultats de recherche en lien avec l'écosystème de cette nouvelle pratique, les contextes de la coproduction des connaissances et la numérisation des disciplines à l'ère numérique.

CHAPITRE I

LA PROBLÉMATIQUE

Cette recherche s'intéresse à une pratique de recherche émergente au sein de la discipline de l'épidémiologie, à l'ère numérique. Cette pratique, désignée comme l'épidémiologie numérique, constitue une nouvelle façon de faire des analyses en épidémiologie s'appuyant notamment sur des données de sources non traditionnelles comme des données générées à partir des médias sociaux, des moteurs de recherche Internet et de téléphones cellulaires (Salathé *et al.*, 2012; Makri, 2017). La pratique de l'épidémiologie numérique est devenue possible avec le phénomène de la mise en données de la société. Nous serions désormais dans une société numérique où les technologies numériques ont littéralement imprégné la vie quotidienne à travers l'utilisation massive d'Internet, des téléphones intelligents et des applications mobiles (Lupton, 2015). Or, ces technologies génèrent des données qui peuvent être utilisées pour faire de la recherche en épidémiologie (Salathé *et al.*, 2012; Makri, 2017).

La discipline de l'épidémiologie est axée sur les données depuis longtemps et celle-ci repose sur une approche systématique et valide de la collecte, l'analyse et l'interprétation des données (Dicker *et al.*, 2006). Il devient alors intéressant de comprendre les enjeux épistémiques et sociopolitiques que soulève l'épidémiologie numérique, afin de développer un regard critique vis-à-vis cette nouvelle pratique utilisant des sources de données différentes par rapport à celles utilisées couramment en épidémiologie. Pour ce faire, nous allons d'abord situer le contexte de recherche lié à la pratique de l'épidémiologie numérique pour ensuite expliquer l'état de la recherche sur le sujet. À la suite de cette démarche, nous préciserons le but, les objectifs de notre recherche et formulerons les questions de recherche qui guideront les analyses tout au long de cette thèse, pour terminer avec une synthèse de la problématique.

1.1 Le contexte de recherche lié à la pratique de l'épidémiologie numérique

La pratique de l'épidémiologie numérique est un phénomène complexe qui doit être analysé en regard de la discipline de l'épidémiologie, des pratiques de recherche en science, des

transformations dans les pratiques, de l'interdisciplinarité et du phénomène de mise en données de la société. Nous partons de la discipline de l'épidémiologie (Frérot *et al.*, 2018; Giroux, 2012; Dicker *et al.*, 2006) puisque la pratique de l'épidémiologie numérique propose des façons différentes de faire des recherches fondées sur cette discipline sous l'impulsion des transformations numériques telles que la socialisation à travers des médias sociaux, l'utilisation massive d'Internet et l'adoption des téléphones cellulaires comme moyen de communication. Tout au long de cette thèse, nous conceptualisons l'épidémiologie numérique comme une pratique de recherche, c'est-à-dire une façon de faire de la recherche épidémiologique. Cette façon de conceptualiser l'épidémiologie numérique nécessite de s'intéresser aux pratiques de recherche scientifique pour étudier cette nouvelle pratique. Nous nous intéressons également aux transformations dans les pratiques de recherche puisque la pratique de l'épidémiologie numérique peut impliquer des manières différentes de faire de l'épidémiologie, et ainsi illustrer certaines transformations contemporaines dans les sciences (Gibbons *et al.*, 1994; Knorr-Cetina, 1999; Louvel, 2015). En effet, la pratique de l'épidémiologie numérique peut faire émerger des collaborations interdisciplinaires (Louvel, 2015; Darbellay, 2011), notamment parce qu'elle requiert certaines compétences en informatique.

De plus, la pratique de l'épidémiologie numérique implique des méthodes d'analyse de données capables d'influencer les façons de travailler des chercheurs ainsi que leurs manières de percevoir les phénomènes étudiés (Marres, 2012; Ruppert *et al.*, 2013; Marres et Weltevrede, 2013; Marres et Gerlitz, 2016; Rogers, 2015), qui sont directement en lien avec l'accès à de grandes quantités de données numériques. Ces méthodes couplées à l'utilisation de certaines données peuvent avoir des impacts sur la vie des gens qui divulguent leurs informations personnelles sur Internet, et venir renforcer ainsi le pouvoir de certains acteurs de la science des données (Beer, 2009; Mayer-Schönberger et Cukier, 2013; Kitchin, 2014; Zwitter, 2014; Ruppert *et al.*, 2015; Ruckenstein et Granroth, 2020).

1.1.1. La discipline de l'épidémiologie

La discipline de l'épidémiologie est établie depuis longtemps avec ses approches, ses ressources et ses méthodologies bien connues et stabilisées au sein des communautés scientifiques du domaine.

L'épidémiologie est pratiquée dans les universités, mais aussi au sein de plusieurs ministères et organismes gouvernementaux. Les liens entre l'épidémiologie et les services gouvernementaux sont nombreux (Dussault, 1995). Il y a d'abord le domaine des politiques de santé publique où l'épidémiologie aide à définir les priorités, les objectifs et les stratégies. L'épidémiologie peut aussi contribuer à la reconfiguration des services, en examinant entre autres les conséquences de la décentralisation (Dussault, 1995). Dans la même perspective, Fontanet (2019) explique que les résultats des études épidémiologiques visent à contrôler les maladies par la surveillance et la modélisation des maladies pour l'épidémiologie descriptive ou à analyser certaines caractéristiques notamment les gènes, les comportements, l'environnement, pour l'épidémiologie analytique. Tous ces efforts sont orientés vers l'action pour gérer la santé publique amenant ainsi l'épidémiologie au cœur de nombreux débats de société faisant la une des journaux quotidiens (Fontanet, 2019). Afin de mieux comprendre cette discipline, nous aborderons les façons de concevoir l'épidémiologie, ses sous-disciplines et les moments clés de son histoire.

1.1.1.1 Les façons de concevoir l'épidémiologie

Selon Giroux (2012), « L'épidémiologie est généralement définie comme l'étude de la distribution et des déterminants des états de santé et des maladies dans les populations humaines ». D'autres sources en santé publique définissent l'épidémiologie comme l'étude de la distribution et des déterminants des états ou événements liés à la santé dans des populations spécifiques, et l'application de cette étude au contrôle des problèmes de santé (Dicker *et al.*, 2006). La version originale dans la langue anglaise est formulée comme ceci:

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems. (Dicker *et al.*, 2006, p.1-2)

Selon Dicker *et al.* (2006), le terme « distribution » tire son importance de l'idée que l'épidémiologie s'intéresse à la fréquence et aux schémas des événements sanitaires dans une population. Le terme « déterminants » est pertinent, car l'épidémiologie est utilisée pour rechercher les déterminants, c'est-à-dire les causes et autres facteurs influençant l'apparition de maladies et d'autres événements

liés à la santé. Quant aux termes « états ou événements liés à la santé », ils soulignent que l'épidémiologie étudie des maladies comme la grippe ou le cancer, mais aussi des comportements comme l'activité physique et l'utilisation de la ceinture de sécurité en voiture (Dicker *et al.*, 2006).

Le terme « population spécifique » renvoie à la façon dont l'épidémiologiste et le clinicien (c'est-à-dire le fournisseur de soins de santé) voient le patient (Dicker *et al.*, 2006). Le clinicien est préoccupé par la santé de l'individu alors que l'épidémiologiste est préoccupé par la santé collective des gens au sein d'une communauté ou d'une population. En d'autres mots, le clinicien se concentre généralement sur le traitement et les soins de l'individu alors que l'épidémiologiste se concentre sur l'analyse de l'exposition à une maladie, l'identification de la source ayant provoqué la maladie, le nombre de personnes ayant pu être exposées, le potentiel de propagation dans la communauté et les interventions pour prévenir d'autres cas ou les récurrences (Dicker *et al.*, 2006).

Enfin, le terme « application » met en évidence l'idée selon laquelle l'épidémiologie ne se limite pas à l'étude de la santé dans une population, mais qu'elle implique aussi l'application des connaissances à la pratique communautaire (Dicker *et al.*, 2006). À la lumière de ces termes, la définition de l'épidémiologie peut aussi se formuler comme ceci:

Epidemiology is the study (scientific, systematic, data-driven) of the distribution (frequency, pattern) and determinants (causes, risk factors) of health-related states and events (not just diseases) in specified populations (patient is community, individuals viewed collectively), and the application of (since epidemiology is a discipline within public health) this study to the control of health problems. (Dicker *et al.*, 2006, p.1-2)

Dicker *et al.* (2006) soulignent par ailleurs que les données sont importantes en épidémiologie, d'abord parce que cette discipline est axée sur les données et ensuite parce qu'elle repose sur une approche systématique et valide de la collecte, l'analyse et l'interprétation des données:

Epidemiology is a scientific discipline with sound methods of scientific inquiry at its foundation. Epidemiology is data-driven and relies on a systematic and unbiased approach to the collection, analysis, and interpretation of data. (Dicker *et al.*, 2006, p.1-2)

Ces sources affirment aussi que l'épidémiologie est une discipline quantitative s'appuyant sur une connaissance pratique des probabilités, des statistiques et des bonnes méthodes de recherche (Dicker *et al.*, 2006):

...epidemiology is a quantitative discipline that relies on a working knowledge of probability, statistics, and sound research methods. (Dicker *et al.*, 2006, p.1-2)

1.1.1.2 Les sous-disciplines de l'épidémiologie

L'analyse des définitions de l'épidémiologie conduite par Frérot *et al.* (2018) souligne l'existence de plusieurs sous-disciplines en épidémiologie. Celles-ci incluent notamment l'épidémiologie étudiant une maladie comme l'épidémiologie du cancer, l'épidémiologie sociale et l'épidémiologie environnementale (Frérot *et al.*, 2018). Frérot *et al.* (2018) proposent de définir l'épidémiologie sociale comme la sous-discipline de l'épidémiologie étudiant la distribution sociale et les déterminants sociaux des états de santé, en se concentrant sur des phénomènes sociaux spécifiques tels que la stratification socio-économique, les réseaux sociaux, la discrimination, l'organisation du lieu de travail, et les politiques publiques plutôt que sur les conséquences d'une maladie spécifique. Merrill (2008) cité dans Frérot *et al.* (2018) considère l'épidémiologie environnementale comme sous-discipline de l'épidémiologie qui étudie la fréquence et la structure des maladies et des événements liés à la santé et qui tente d'expliquer les facteurs environnementaux influençant ces conditions.

L'épidémiologie computationnelle fait référence à l'application de méthodes de simulation basées sur la science des systèmes fondées sur des modèles mathématiques d'épidémies (Li Vigni, 2021). L'épidémiologie computationnelle aurait été évoquée pour la première fois en 1988, mais sans connaître de succès (Li Vigni, 2021). Ce domaine a pris de l'ampleur depuis le début des années 2000 avec l'émergence de plusieurs épidémies zoonotiques ayant engendré la formation d'une nouvelle communauté de scientifiques européens et américains (Li Vigni, 2021). Se revendiquant être une nouvelle spécialité désignée comme l'« épidémiologie computationnelle », cette communauté serait composée de médecins, mathématiciens et autres scientifiques ayant des connaissances computationnelles des réseaux complexes utilisant tout un ensemble d'outils

d'analyse et de prédiction des épidémies. Il semble que cette nouvelle spécialité apporte des contributions significatives à l'évaluation des risques de contagion, des populations vulnérables et des effets des mesures d'intervention visant à arrêter la propagation (Marathé et Vullikanti, 2013 cités dans Li Vigni, 2021). Malgré des critiques importantes de la part de médecins, d'infectiologues, de biostatisticiens et de journalistes scientifiques, cette spécialité a été très médiatisée pendant la pandémie de la COVID-19 et elle fait désormais partie des mécanismes de préparation sanitaire de nombreux gouvernements. Les modèles proposés par l'épidémiologie computationnelle seraient pris au sérieux par les décideurs politiques et les citoyens (Li Vigni, 2021).

Frérot *et al.* (2018) mentionnent également l'épidémiologie numérique comme une des sous-disciplines émergentes en épidémiologie. Frérot *et al.* (2018) utilisent la définition de Salathé (2018) qui considère l'épidémiologie numérique comme une épidémiologie utilisant des données générées en dehors du système de santé publique, c'est-à-dire des données n'ayant pas été générées dans le but premier de faire de l'épidémiologie (Salathé, 2018 cité dans Frérot *et al.*, 2018). Cette perspective suscite des questions sur la nature de l'épidémiologie numérique elle-même. S'agit-il d'une nouvelle sous-discipline de l'épidémiologie ou d'une autre façon de faire des recherches épidémiologiques?

L'émergence de l'épidémiologie numérique au sein de l'épidémiologie est évoquée par l'ajout de l'expression « *Digital Epidemiology* » dans le Dictionnaire d'Oxford de l'épidémiologie (Porta, 2008, 2014). Or, l'épidémiologie numérique n'était pas incluse dans ce dictionnaire avant l'édition de 2014. Cette omission indiquerait que la discipline a commencé à tenir compte de cette nouvelle pratique très récemment, en l'occurrence dans le cadre de certaines recherches épidémiologiques et de la surveillance de certaines maladies (Porta, 2008, 2014). Pour définir l'épidémiologie numérique, Porta (2014) insiste, entre autres, sur l'exploitation des données d'Internet, des médias sociaux et des téléphones cellulaires ainsi que sur l'intelligence informatique et épidémiologique pour extraire des informations significatives des ensembles de données qualifiés comme extrêmement volumineux et bruyants. Cet auteur propose de définir l'épidémiologie numérique comme ceci:

The use of epidemiologic knowledge and digital technologies to enable disease surveillance and epidemiological research. For instance, in some outbreaks, digital disease surveillance has supplemented the critical laboratory studies and work in the trenches by public health officials and epidemiologists, by leveraging the use of the Internet, mobile phones, and social media. Some of the added insights come from the general population. Both computational and epidemiological intelligence are required to extract meaningful information from data sets that are both extremely large and noisy. (Porta, 2014)

L'épidémiologie clinique est une autre sous-discipline de l'épidémiologie mentionnée par Frérot *et al.* (2018). Certains voient son émergence comme un phénomène important dans l'histoire de l'épidémiologie et cette sous-discipline semble avoir des caractéristiques particulières (Giroux, 2012).

Selon Miettinen (2010), l'épidémiologie clinique serait la plus importante des sciences cliniques. Elle est définie comme la science consistant à faire des prédictions sur des patients individuels en comptant les événements cliniques dans des groupes de patients similaires et en utilisant des méthodes scientifiques solides pour s'assurer que les prédictions sont exactes :

Foremost among [the clinical sciences] is clinical epidemiology . . . the science of making predictions about individual patients by counting clinical events in groups of similar patients and using strong scientific methods to ensure that the predictions are accurate. (Miettinen, 2010, p.11)

Certaines sources comme celle de Conseils nationaux de la recherche (NRC, 1997) considèrent que l'épidémiologie clinique fait partie de l'épidémiologie moderne, c'est-à-dire l'étude des schémas pathologiques au sein des populations comprenant un large éventail d'objets de recherche étudiés notamment lors d'essais cliniques. Ces objets de recherche comprennent les maladies infectieuses, les maladies chroniques ainsi que les facteurs de risque comme la profession, la nutrition, et l'alcoolisme et des populations particulières. Giroux (2012) complète ce portrait en soulignant que le développement de l'épidémiologie clinique constitue un tournant dans l'histoire de l'épidémiologie puisque traditionnellement cette discipline se distinguait de la médecine clinique.

La discipline de l'épidémiologie est rattachée au domaine de la santé publique. En analysant des statistiques administratives et démographiques, cette discipline étudie les populations et non des individus (Giroux, 2012). Selon Giroux (2012), le milieu du XX^e siècle marque une étape importante dans l'histoire de l'épidémiologie puisque certains chercheurs commencent à développer des techniques d'enquête portant sur des populations bien délimitées pour l'analyse étiologique des maladies en se basant sur des cohortes d'individus bien définies et non sur la population générale en provenance des recensements comme c'est le cas traditionnellement en épidémiologie. Le but de ces recherches consiste à identifier des facteurs de risque individuels et à orienter les efforts de recherche vers une prévention individualisée des risques de maladie. Ce phénomène semble donner lieu à une nouvelle épidémiologie centrée sur l'individu et plus près de la médecine clinique (Giroux, 2012). En se référant à l'ouvrage d'Alvan Feinstein (1985), Giroux (2012) explique comment il est possible de faire la différence entre l'épidémiologie statistique (ou de santé publique) et l'épidémiologie clinique:

Si, dans l'épidémiologie de santé publique, le dénominateur est une population générale issue des recensements ou une communauté géographique particulière...dans l'épidémiologie clinique, on a affaire à un ensemble d'individus étudiés dans des conditions cliniques bien précises, soit une cohorte bien définie. (Giroux, 2012, p. 27)

À la lumière de ces constats, la prise en compte des différentes sous-disciplines en épidémiologie peut s'avérer pertinente pour comprendre le contexte de recherche lié à la pratique de l'épidémiologie numérique. Des chercheurs comme Frérot *et al.* (2018) affirment qu'il s'agit d'une sous-discipline de l'épidémiologie. Il serait intéressant de savoir dans quelle mesure ce point de vue est partagé par d'autres chercheurs. L'épidémiologie numérique est-elle une nouvelle pratique ou une nouvelle sous-discipline de l'épidémiologie? De plus, il est possible de s'interroger sur la pluralité des dénominations utilisées pour désigner l'épidémiologie numérique. Plusieurs aspects de l'épidémiologie numérique sont utilisés dans les différentes spécialités à différentes étapes des recherches. L'épidémiologie numérique constitue une pratique émergente ou plusieurs pratiques émergentes?

1.1.1.3 Des moments clés dans l'histoire de l'épidémiologie

Bien que l'épidémiologie en tant que discipline se soit épanouie à partir du XX^e siècle (Dicker *et al.*, 2006; Rothman, 2002), plusieurs sources indiquent que la pensée épidémiologique remonte à Hippocrate, 400a avant J.-C (Dicker *et al.*, 2006; Morabia, 2004; Rothman, 2002). Il semble que ce médecin grec (Rothman, 2002) a tenté d'expliquer l'apparition des maladies d'un point de vue rationnel plutôt que surnaturel (Dicker *et al.*, 2006). Dans son essai intitulé « Sur les airs, les eaux et les lieux », Hippocrate suggère que des facteurs liés à l'environnement et à l'hôte, tels que les comportements, peuvent influencer le développement des maladies (Dicker *et al.*, 2006).

En 1662, John Graunt (un mercier et conseiller municipal londonien) publia une analyse des données de mortalité qui semble avoir été la première étude ayant produit des modèles de naissance, de décès et d'apparition de maladies, en notant les disparités entre les hommes et les femmes, la mortalité infantile élevée, les différences entre les villes et les campagnes et les variations saisonnières. Au Canada, l'histoire de l'épidémiologie est marquée par l'accroissement de l'immigration européenne en Amérique dans les années 1700 due à l'arrivée des navires au port de Québec amenant des passagers souvent atteints de maladies (Rutty et Sullivan, 2010). Par la suite, l'histoire canadienne de l'épidémiologie suit son cours de manière similaire à celle des pays d'Europe; elle est marquée par les épidémies de choléra dans les années 1800 et par la création et le renforcement des infrastructures en santé publique, notamment la mise sur pied de bureaux de santé locaux permanents et le resserrement des lois sur la quarantaine (Rutty et Sullivan, 2010).

Considéré comme le père des statistiques vitales et de la surveillance moderne, William Farr aurait développé dans les années 1800 un grand nombre des pratiques de base utilisées aujourd'hui dans les statistiques vitales et la classification des maladies (Dicker *et al.*, 2006). Viennent ensuite les travaux de l'anesthésiste John Snow au milieu des années 1800 (Dicker *et al.*, 2006). John Snow amené à Londres une série d'enquêtes de grande qualité faisant de lui le père de l'épidémiologie de terrain. Vingt ans avant le développement du microscope, Snow aurait mené des études sur les épidémies de choléra, à la fois pour découvrir la cause de la maladie et pour prévenir sa récurrence (Dicker *et al.*, 2006).

Toujours selon Dicker *et al.* (2006), les méthodes épidémiologiques ont commencé à être appliquées à l'étude de l'apparition des maladies au milieu et à la fin du XIX^e siècle. Pendant cette période, la plupart des chercheurs étudiaient les maladies infectieuses aiguës. Dans les années 1930 et 1940, les épidémiologistes ont commencé à étudier les maladies non infectieuses (Dicker *et al.*, 2006). La période après la Seconde Guerre mondiale a vu une explosion du développement des méthodes de recherche et des fondements théoriques de l'épidémiologie. L'épidémiologie a été appliquée à toute la gamme des résultats, des comportements et même des connaissances et des attitudes liés à la santé (Dicker *et al.*, 2006). Dans cette vague de développements, le Canada a mis sur pied un programme en épidémiologie de terrain en 1975 afin d'offrir aux médecins canadiens l'opportunité d'acquérir une expérience pratique dans l'application des principes épidémiologiques au cours d'un apprentissage de deux ans dans le domaine de la santé publique (Hockin, 1991).

Ensuite, dans les années 1980, l'épidémiologie a été étendue à l'étude des blessures et de la violence (Dicker *et al.*, 2006). Les années 1990 auraient été marquées par l'expansion de l'épidémiologie moléculaire et génétique et par les défis de maladies infectieuses telles que le virus Ebola, le virus de l'immunodéficience humaine (VIH) et le syndrome d'immunodéficience acquise (SIDA) (Dicker *et al.*, 2006). Dicker *et al.* (2006) expliquent aussi que depuis les années 1990, et encore plus après les attaques terroristes du 11 septembre 2001, les épidémiologistes ont dû prendre en compte non seulement la transmission naturelle d'organismes infectieux, mais aussi la propagation délibérée par la guerre biologique et le bioterrorisme.

En 2003 arrive l'épidémie du syndrome respiratoire aigu sévère (SRAS) au Canada dans la ville de Toronto, un virus d'un nouveau genre ayant atteint environ 330 cas et causé 32 décès (Rutty et Sullivan, 2010). Cette épidémie aurait causé des pertes estimées à 35 millions de dollars par jour à Toronto lorsque l'OMS émet un avis préventif aux voyageurs les informant du risque de maladie encouru par le SRAS (Rutty et Sullivan, 2010). Suite aux recommandations du rapport du Comité consultatif national sur le SRAS en 2003, l'Agence de santé publique du Canada voit le jour en septembre 2004 afin de créer un organisme responsable de communiquer directement avec le public et le gouvernement sur les questions importantes en santé (Rutty et Sullivan, 2010).

En 2008, commence l'histoire de l'épidémiologie numérique avec le lancement de *Google Flu Trends* (GFT) afin d'aider à prédire les épidémies de grippe (Salathé, 2015). GFT était un outil de prévision à court terme de l'activité grippale hebdomadaire, proposant une méthode complémentaire de surveillance de la grippe (Guo *et al.*, 2021). Ce système utilisait une méthode de découverte automatique des requêtes de recherche sur Google liées à la grippe pour traiter des centaines de milliards de recherches individuelles sur le web, pendant une période de cinq ans aux États-Unis en vue d'une meilleure surveillance de ce virus (Ginsberg *et al.*, 2009). Les données historiques accessibles au public provenant du réseau de surveillance de la grippe aux États-Unis ont aussi été utilisées pour construire les modèles de GFT (Ginsberg *et al.*, 2009). L'idée était que lorsque les gens commençaient à se sentir malade, ils se tournaient souvent vers Google pour rechercher des informations sur les symptômes de la grippe, les traitements et d'autres sujets connexes. En regroupant et en analysant ces données, GFT a tenté de fournir des informations en temps réel sur la prévalence de la grippe (Salathé, 2015).

Par la suite en 2009, des chercheurs de Google et des Centers for Disease Control and Prevention (CDC) ont publié un article proposant une méthode d'utilisation des requêtes de recherche Internet pour reconnaître les épidémies de grippe (Ginsberg *et al.*, 2009). À cause de la surestimation de l'activité de la grippe par GFT en 2013 (Lazer *et al.*, 2014), les chercheurs se sont tournés vers d'autres sources de données numériques comme celle de Twitter afin d'étudier le virus de la grippe H1N1, de suivre les activités des maladies, d'analyser les sentiments vis-à-vis de la vaccination et de surveiller la sécurité des médicaments (Salathé, 2015). GFT est devenu inactif en 2015 (Duclos, 2019; Guo *et al.*, 2021). En plus de la surestimation de l'activité grippale, les enjeux rencontrés par GFT incluent : le manque de précision (les estimés étaient surtout fournis au niveau national et beaucoup moins au niveau local) ; l'hypothèse souvent implicite selon laquelle les big data sont un substitut, plutôt qu'un complément, à la collecte et à l'analyse traditionnelles des données; et les dynamiques propres aux algorithmes, c'est-à-dire les changements dans l'algorithme de recherche de Google et dans le comportement des utilisateurs qui sont grandement susceptibles d'affecter le suivi de GFT, notamment lors de la panique provoquée par les médias pendant la saison de la grippe en 2013 (Lazer *et al.*, 2014). S'ajoutent aussi le manque de transparence sur la façon dont les algorithmes fonctionnent et l'incapacité à répliquer les résultats obtenus pour valider les tendances observées lors de l'utilisation de GFT (Lazer *et al.*, 2014).

Dans les années 2010 et 2020, l'histoire de l'épidémiologie numérique prend un nouveau tournant avec l'utilisation des big data (Lee *et al.*, 2016; Guernier *et al.*, 2016; Gianfredi *et al.*, 2018; Hassan *et al.*, 2019). Viennent ensuite les techniques d'apprentissage profond (*deep learning*) appliquées entre autres lors d'une recherche portant sur la détection des maladies des plantes à partir d'images (Mohanty *et al.*, 2016; Zhang *et al.*, 2020) et lors de travaux de classification multilingue des tweets pour la surveillance des maladies à partir d'Internet (Magumba et Nabende, 2021).

À partir de 2019, l'épidémiologie est confrontée à la pandémie de la COVID-19 ayant débuté en décembre de cette année lorsque des cas de pneumonie inconnue ont été découverts chez des adultes dans la ville de Wuhan, de la région de Hubei, en Chine (Siddiqui *et al.*, 2022). Ce contexte pandémique est plus que jamais propice à la pratique de l'épidémiologie numérique qui devient possible grâce au potentiel des technologies récentes de télécommunication. Ces technologies permettent la surveillance et la modélisation des maladies infectieuses telles que la grippe et la COVID-19, en extrayant des informations des dossiers électroniques de santé, des paiements électroniques, de l'Internet et des médias sociaux (He *et al.*, 2020). Selon He *et al.* (2020), l'utilisation de technologies numériques pour la surveillance des mouvements des personnes pendant les épidémies offre le potentiel d'estimer en temps quasi réel des activités liées à la maladie et d'identifier rapidement des sujets potentiellement infectés tout en protégeant les informations et la vie privée des sujets.

D'autres recherches comme celle de Higgins *et al.* (2020) voient des corrélations entre les tendances détectées par les moteurs de recherche en ligne et l'incidence de la COVID-19. Cependant, une étude italienne plus récente apporte des nuances en recommandant aux agences gouvernementales d'utiliser *Google Trends* comme outil d'info-vigilance complémentaire pour surveiller et prévoir l'adhésion aux vaccins (Rovetta, 2022); l'utilisation appropriée et les limites de cette technologie pour les suivis de la vaccination requérant davantage de recherche. Il reste que ces études illustrent un autre tournant dans l'histoire de l'épidémiologie avec les succès et les limites de la pratique de l'épidémiologie numérique.

En somme, une des définitions les plus répandues de l'épidémiologie propose de la considérer comme l'étude de la distribution et des déterminants des états ou événements liés à la santé dans des populations spécifiques, et comme l'application de cette étude au contrôle des problèmes de santé (Dicker *et al.*, 2006). L'épidémiologie est une discipline axée sur les données et elle repose sur une approche systématique et valide de la collecte, de l'analyse et de l'interprétation des données. Cette discipline repose aussi sur une connaissance pratique des probabilités, des statistiques et des bonnes méthodes de recherche (Dicker *et al.*, 2006).

Dans le cadre de notre recherche, nous conceptualisons l'épidémiologie numérique comme une pratique de recherche plutôt que comme une sous-discipline, principalement parce que selon les caractéristiques de l'épidémiologie numérique identifiées dans la littérature, il s'agirait d'une nouvelle façon de faire de l'épidémiologie s'appuyant notamment sur des données de sources non traditionnelles (Salathé *et al.*, 2012; Makri, 2017). De plus, l'épidémiologie numérique ne se concentre pas sur un enjeu de santé particulier comme l'épidémiologie du cancer ni sur certains types d'enjeux comme l'épidémiologie sociale. Dans cette perspective, l'épidémiologie numérique constituerait une pratique de recherche, c'est-à-dire un ensemble d'activités renvoyant à une manière de faire de l'épidémiologie basée sur de nouvelles sources de données numériques.

1.1.2 Les pratiques de recherche

L'épidémiologie numérique étant conceptualisée comme une pratique de recherche, il nous paraît nécessaire de définir maintenant ce que sont des pratiques de recherche. La définition que nous proposons considère à la fois les nouveautés, les activités ainsi que les acteurs humains et non-humains impliqués dans les pratiques de recherche à l'ère numérique.

Plusieurs conceptions de la notion de pratique mettent l'accent sur le fait que la pratique est régie par des règles (Knorr-Cetina, 2001). D'après Knorr-Cetina (2001), la plupart des auteurs ayant conceptualisé les pratiques conviennent qu'il s'agit de processus récurrents régis par des schémas de préférences et de prescriptions dictant le cours des actions. De tels processus seraient présents dans de nombreux domaines de la vie sociale et refléteraient des manières coutumières ou routinières de se comporter (Knorr-Cetina, 2001). À l'époque actuelle, les organisations et les

professions seraient nombreuses à disposer d'une importante base de connaissances amenant les praticiens à continuer d'apprendre et les spécialistes à constamment réinventer leurs propres pratiques d'acquisition des connaissances (Knorr-Cetina, 2001). Dans ce contexte, les pratiques prendraient une tout autre signification et soulèveraient d'autres questions que celles soulevées par les activités habituelles, par exemple : comment les caractéristiques des pratiques peuvent-elles inclure le potentiel de changement (Knorr-Cetina, 2001) ?

Les pratiques de recherche semblent particulières sur ce plan dans la mesure où la définition des choses, la conscience des problèmes, etc., est considérablement façonnée par les objets et les réactions des chercheurs envers ces objets (Knorr-Cetina, 2001). Dans un tel contexte, il y aurait des moments de réflexion lors des activités de recherche, au cours desquels les efforts pour lire et interpréter les réactions des objets jouent un rôle décisif lors de la production des connaissances (Knorr-Cetina, 2001).

Pour définir les pratiques de recherche, nous adoptons la perspective proposée par Knorr-Cetina (2001) puisque celle-ci tient compte des particularités du travail de recherche et des nouveautés en recherche. Les nouvelles pratiques de recherche à l'ère numérique pourraient impliquer des apprentissages continus et des potentiels de changements. Il devient alors pertinent de conceptualiser les pratiques de recherche comme des pratiques constructives et créatives centrées sur les connaissances.

En ce qui concerne les activités composant les pratiques de recherche, nous soutenons qu'elles incluent le fait de poser des questions et de définir des problèmes, de développer et d'utiliser des modèles, de planifier et réaliser des investigations, d'analyser et interpréter des données, d'utiliser la pensée mathématique et informatique, de construire des explications et concevoir des solutions, de s'engager dans une argumentation à partir de preuves, d'obtenir, évaluer et communiquer des informations (Osborne, 2014). Le tout étant centralisé autour de la critique puisque toute nouvelle prétention à la connaissance doit toujours faire l'objet d'examen critiques pour aboutir à des connaissances fiables (Osborne, 2014).

Au niveau des acteurs, nous considérons que les pratiques de recherche sont conditionnées par des réseaux d'acteurs humains et non-humains incluant des personnes, des textes, des termes, des matériaux et des instruments qui s'associent et s'articulent (Vinck, 2007) lors des travaux scientifiques. Nous ajoutons à cette liste d'acteurs des technologies, des données, des institutions et des sociétés privées afin de prendre en compte le contexte actuel de la mise en données de la société.

Dans ce contexte, les nouvelles pratiques de recherche peuvent impliquer des objets (infrastructures de serveurs, dispositifs et câbles), des langages (code, programmation et algorithmes) et des personnes (scientifiques, entrepreneurs, ingénieurs, informaticiens, concepteurs) susceptibles de créer, ensemble, de nouveaux mondes (Ruppert *et al.*, 2017). Cette perspective fait ressortir l'importance des entités non-humaines (comme des instruments de mesure et des données) influençant une partie importante du travail des experts (Knorr-Cetina, 2001) dans le cadre des pratiques de recherches.

Comme le souligne Vinck (2007), les réseaux d'acteurs impliqués dans des pratiques de recherche entrent en relation en travaillant en équipes « au sein de laboratoires, d'organisations et de réseaux de coopération scientifique ». Les manières par lesquelles s'associent les acteurs ne sont pas nécessairement fondées sur des frontières disciplinaires ni sur des domaines de spécialité; les réseaux d'acteurs sont hétérogènes (Vinck, 2007). Les chercheurs appartiennent souvent à plusieurs réseaux (Vinck, 2007). De plus, la production des connaissances se caractérise par une multitude d'interactions entre des acteurs appartenant à la fois au monde scientifique et à d'autres secteurs de la société, impliquant aussi l'État, des partenaires privés, des médias et des citoyens (Vinck, 2007). Une dernière caractéristique importante des réseaux d'acteurs des pratiques de recherche est la fluidité de ces réseaux (Vinck, 2007) parce que ces réseaux naissent, croissent et meurent. La composition de ces réseaux change au fil du temps (Vinck, 2007).

En nous intéressant à ce qui change dans les pratiques scientifiques à l'ère numérique et en conceptualisant les pratiques de recherche comme des pratiques constructives et centrées sur les connaissances (Knorr-Cetina, 2001), nous définissons les pratiques de recherche comme des activités consistant, de manière critique, à poser des questions et définir des problèmes, développer

et utiliser des modèles, planifier et réaliser des enquêtes, analyser et interpréter des données, utiliser la pensée mathématique et informatique, construire des explications et concevoir des solutions, s'engager dans une argumentation à partir de preuves ainsi qu'obtenir, évaluer et communiquer des informations (Osborne, 2014). Ces activités sont menées au sein de réseaux fluides, temporaires et hétérogènes d'acteurs humains et non-humains situés à l'intérieur et à l'extérieur du monde académique qui interagissent et entrent en relations lors des travaux scientifiques (Vinck, 2007).

1.1.3 Les changements dans les pratiques de recherche

La pratique de l'épidémiologie numérique peut être vue comme une nouveauté dans la discipline de l'épidémiologie, apportant ainsi des changements dans les façons de produire les connaissances en épidémiologie. Ces changements peuvent susciter une tension centrale en épidémiologie entre, d'une part, un engagement conservateur à persister dans l'utilisation d'un cadre de recherche particulier, même face à des difficultés, et, d'autre part, une disposition novatrice à rechercher de nouvelles approches (D'Agostino, 2008). Selon cette perspective, les chercheurs en épidémiologie peuvent vouloir continuer à utiliser les méthodes et les données traditionnelles parce que celles-ci sont reconnues et bien établies dans leur discipline, mais ils peuvent aussi vouloir pratiquer l'épidémiologie numérique afin de tester une nouveauté et ainsi avoir la possibilité d'innover en apportant d'autres contributions à leur discipline.

Certains chercheurs en épidémiologie peuvent décider de pratiquer l'épidémiologie numérique alors que d'autres peuvent choisir de ne pas la pratiquer. Il nous paraît intéressant de savoir dans quelle mesure la pratique de l'épidémiologie numérique est adoptée en épidémiologie et quels sont les motifs expliquant les choix des chercheurs en épidémiologie vis-à-vis de cette nouvelle pratique. En reprenant des concepts de la théorie de la diffusion des innovations de Rogers (2003), nous envisageons que la pratique de l'épidémiologie numérique puisse se diffuser facilement en épidémiologie à cause de ses avantages relatifs comme l'apport de nouvelles connaissances par l'utilisation de sources de données non traditionnelles. Mais cette nouvelle pratique peut aussi ne pas se diffuser, à cause notamment de sa complexité au niveau technique et de son incompatibilité avec le système social dans lequel la discipline de l'épidémiologie évolue, entre autres, en raison des enjeux éthiques soulevés par l'utilisation des données numériques.

Gibbons *et al.* (1994) nous enseignent qu'il est possible de constater des changements notables dans les façons de produire les connaissances, en lien avec la diversification des pratiques de recherche. Ces auteurs soutiennent que des changements importants façonnent la manière de produire les connaissances et montrent comment ces changements marquent un tournant vers un nouveau mode de production des connaissances susceptible de remplacer ou réformer les institutions, les disciplines, les pratiques et les politiques établies (Gibbons *et al.*, 1994). Cette perspective contraste deux modes de production des connaissances en désignant le Mode 1 comme étant le mode traditionnel de production des connaissances en contexte disciplinaire et fondamentalement cognitif et le Mode 2 comme étant le nouveau mode de production des connaissances en contextes sociaux et économiques plus larges, transdisciplinaires et appliqués (Gibbons *et al.*, 1994).

La thèse développée par Gibbons *et al.* (1994) a suscité de nombreuses critiques (Nowotny *et al.*, 2003). Pour répondre à ces critiques et continuer à développer leur thèse générale, ces auteurs ont publié un autre ouvrage en 2001 intitulé « *Rethinking Science: Knowledge and the Public in an Age of Uncertainty* ». Selon Nowotny *et al.* (2003), les principales critiques adressées à Gibbons *et al.* (1994) sont les suivantes : l'ouvrage contribue à légitimer des tendances actuelles (particulièrement la subordination de la recherche à la politique et aux marchés), l'argumentation des auteurs ne s'appuie pas sur des preuves adéquates, aucune tentative réelle n'aurait été faite pour discuter des problèmes de la société, l'ouvrage se limite à une description des épiphénomènes sociaux, le cœur de la science reste le même et enfin, l'ouvrage offre une vision post-moderne de la science.

De plus, Bonneuil et Joly (2013) soulignent que l'ouvrage de Gibbons *et al.* (1994) privilégie une vision technocratique de la science en faveur de la recherche axée sur des finalités. Ces auteurs soulignent aussi que, pour certains chercheurs, la thèse développée par Gibbons *et al.* (1994) fournit une image erronée du Mode 2 du point de vue de l'histoire des sciences, car ce mode peut caractériser un bon nombre de disciplines à divers moments historiques, incluant autant la période de la révolution scientifique que la période après la Seconde Guerre mondiale (Bonneuil et Joly, 2013). Selon cette perspective, la recherche orientée vers l'action s'appliquerait autant aux mathématiques mixtes du XVI^e siècle qu'aux domaines plus récents, tels que le nucléaire et

l'aviation – et donc pas seulement aux nouveaux contextes de recherche comme le soutiennent Gibbons *et al.* (1994). Bonneuil et Joly (2013) expliquent aussi qu'au cours de l'histoire, la science a toujours été connectée à la réalité par l'usage de la technologie ou de la connaissance pour modifier le monde. Cette perspective diffère de celle de Gibbons *et al.* (1994) qui affirme que le lien entre la réalité et la science serait présent surtout dans le Mode 2 de la production des connaissances.

Si la théorisation proposée par Gibbons *et al.* (1994) prête le flanc à la critique (Bonneuil et Joly, 2013; Nowotny *et al.*, 2003), certains éléments restent intéressants pour saisir les changements importants dans les pratiques de recherche. Par exemple, certains attributs des modes de production des connaissances expliqués par Gibbons *et al.* (1994), comme les différentes manières de résoudre des problèmes de recherche ainsi que l'hétérogénéité et la diversité des organisations, des habiletés, des expériences et des sites potentiels de création des connaissances, peuvent être utiles pour comprendre la diversité dans les pratiques de recherche. Ces attributs peuvent illustrer la coexistence de différentes façons de produire les connaissances et ainsi, aider à mieux saisir le contexte de la pratique de l'épidémiologie numérique.

Selon Gibbons *et al.* (1994), la résolution des problèmes de recherche dans le Mode 1 s'effectue en suivant des codes de pratique pertinents pour une discipline alors dans le Mode 2, la résolution des problèmes est organisée autour d'une application particulière. Dans cette optique, le contexte de la production des connaissances du Mode 1 n'implique pas nécessairement des objectifs précis et se définit en relation avec des normes cognitives et sociales gouvernées par la recherche fondamentale ou la science académique. Par contraste, la production des connaissances dans le Mode 2 se veut utile pour l'industrie, les gouvernements ou pour la société en général et cet impératif est présent dès le début des initiatives de recherche (Gibbons *et al.*, 1994).

Le cas de la pratique de l'épidémiologie numérique soulève des questions sur le mode de production des connaissances proposé par Gibbons *et al.* (1994) qui caractérise la résolution de problème dans le cadre de cette nouvelle pratique. S'agit-il du Mode 1 ou du Mode 2 ou encore d'une combinaison de ces deux modes de production des connaissances ? Cette nouvelle pratique pourrait résoudre les problèmes en suivant les codes de pratique de la discipline de l'épidémiologie (Mode 1) et

s'organiser autour d'une application particulière comme celle de la surveillance des maladies pour informer les décisions en santé publique (Mode 2). La pratique de l'épidémiologie numérique illustrerait-elle la coexistence de ces deux manières de résoudre des problèmes de recherche et ainsi, la diversité dans les pratiques de recherche en épidémiologie ? Serait-il plus approprié de concevoir la production des connaissances dans le cadre de la pratique de l'épidémiologie numérique comme un mélange d'attributs du Mode 1 et du Mode 2, tel qu'ils sont conceptualisés par Gibbons *et al.* (1994) ?

Gibbons *et al.* (1994) affirment que le Mode 2 serait hétérogène et diversifié en termes organisationnels, d'habiletés et d'expériences. La composition des équipes travaillant à résoudre un problème change au fil du temps, selon l'évolution des exigences, et le tout n'est pas coordonné par un organisme central ni structuré de manière hiérarchique (Knorr-Cetina, 1999). Une telle situation s'illustre non seulement par l'augmentation du nombre moyen d'auteurs par article scientifique, mais aussi par la diversité des spécialités et disciplines impliquées dans l'écriture d'un même article ainsi que la pluralité des institutions et organisations d'appartenance des auteurs (Gibbons *et al.*, 1994). La pratique de l'épidémiologie numérique s'apparenterait-elle aux caractéristiques du Mode 2 tel que développé par Gibbons *et al.* (1994) ? Il ne semble pas y avoir d'organisme central coordonnant les activités. Les équipes sont très diversifiées en matière d'expertise et leur composition semble changer au fil du temps en fonction du problème à résoudre et de l'évolution des projets de recherche.

Le mode 2 serait aussi marqué par une augmentation du nombre de sites potentiels où les connaissances peuvent être créées (Gibbons *et al.*, 1994). L'expansion du nombre, de la nature et de la portée des communications entre les différents sites conduit non seulement à produire plus de connaissances, mais aussi à la production de différents types de connaissances. Les universités et les collèges ne sont plus les seuls sites de production des connaissances puisque des instituts non universitaires, agences gouvernementales et laboratoires industriels produisent aussi des connaissances. Gibbons *et al.* (1994) avancent aussi que les liens entre les sites s'établissent de façons différentes en fonctionnant à travers des réseaux de communication. Le Mode 2 aurait tendance à mettre plus d'emphasis sur les connaissances tacites par rapport aux connaissances codifiées (Gibbons *et al.*, 1994). Les groupes de recherche dans le Mode 2 seraient moins

fermement institutionnalisés que dans le Mode 1, car les gens forment des équipes de travail et des réseaux temporaires d'abord en fonction des problèmes à résoudre. Par ailleurs, les formes d'organisations seraient très diversifiées dans le Mode 2 puisque celles-ci incluent des entités de secteurs différents, dont des firmes multinationales, gouvernements et universités (Gibbons *et al.*, 1994). Dans quelle mesure toutes ces caractéristiques concernant la diversification des sites potentiels de création des connaissances, les réseaux de communication, l'emphase sur les connaissances tacites et le manque d'institutionnalisation s'appliqueraient-elles à la pratique de l'épidémiologie numérique ?

Gibbons *et al.* (1994) soulignent l'importance de l'informatique et de la modélisation des données dans le contexte du Mode 2. Selon Gibbons *et al.* (1994), l'ordinateur est devenu ce puissant nouvel outil en science générant un nouveau langage ainsi que de nouvelles visualisations permettant des analyses de données de plus en plus sophistiquées lors des recherches.

En somme, les changements dans les pratiques de recherche peuvent constituer un enjeu important dans le cadre de la pratique de l'épidémiologie numérique. Puisque la pratique de l'épidémiologie numérique peut être vue comme une nouveauté, dans quelle mesure cette nouveauté pourrait-elle amener une tension en épidémiologie entre les dispositions conservatrices et les dispositions novatrices par rapport à la recherche scientifique (D'Agostino, 2008) ?

En s'inspirant de certains concepts de la théorie de la diffusion des innovations de Rogers (2003), il serait pertinent de connaître le niveau de diffusion de la pratique de l'épidémiologie numérique en épidémiologie. Les avantages relatifs de cette nouvelle pratique, comme l'apport de nouvelles connaissances par l'utilisation de données de sources non traditionnelles, pourraient-ils faciliter sa diffusion au sein des communautés de chercheurs en épidémiologie ? La complexité au niveau technique de la pratique de l'épidémiologie numérique et son incompatibilité avec le système social dans lequel la discipline de l'épidémiologie évolue (à cause notamment des enjeux éthiques liés à l'utilisation des données numériques) pourraient-elles freiner l'adoption de cette nouvelle pratique ?

Les perspectives développées par Gibbons *et al.* (1994) sur le Mode 1 et le Mode 2 de la production des connaissances soulèvent des questions quant au mode de production des connaissances

caractérisant la pratique de l'épidémiologie numérique. Gibbons *et al.* (1994) expriment la diversité des modes de production des connaissances surtout à travers une dichotomie entre un mode traditionnel (théorique et disciplinaire) et un mode nouveau (appliqué et transdisciplinaire). Quel mode de production des connaissances correspond le mieux à la pratique de l'épidémiologie numérique ? Cette nouvelle pratique reflète-t-elle un mode de production des connaissances ou la coexistence de plusieurs modes ? Il semble que la pratique de l'épidémiologie numérique utilise les concepts et les théories de la discipline de l'épidémiologie. Cette caractéristique du Mode 1 est-elle suffisante pour associer un mode de production des connaissances de type disciplinaire à cette nouvelle pratique ? D'un autre côté, cette nouvelle pratique semble poursuivre des objectifs appliqués, car celle-ci s'organise autour d'une application particulière comme celle de la surveillance des maladies pour informer les décisions en santé publique. Cette caractéristique est-elle suffisante pour assigner le Mode 2 à la pratique de l'épidémiologie numérique ?

Certains attributs des modes de production des connaissances expliqués par Gibbons *et al.* (1994) comme les différentes manières de résoudre des problèmes de recherche ainsi que l'hétérogénéité et la diversité des organisations, des habiletés, des expériences et des sites potentiels de création des connaissances peuvent être utiles pour comprendre la pratique de l'épidémiologie numérique. Gibbons *et al.* (1994) soulignent aussi l'importance de l'informatique et de la modélisation des données dans le contexte du Mode 2. Cette caractéristique semble très présente à la fois dans la pratique de l'épidémiologie numérique et dans la discipline de l'épidémiologie elle-même, qui est axée sur les analyses de données et qui utilise l'informatique depuis longtemps. Ce lien étroit avec l'informatique nous amène à réfléchir sur l'interdisciplinarité qui semble constituer un enjeu important dans le contexte de la pratique de l'épidémiologie numérique.

1.1.4 L'interdisciplinarité

Selon Darbellay (2011), « l'interdisciplinarité apparaît comme une voie nouvelle dans le développement de l'enseignement et de la recherche universitaires ». Gibbons *et al.* (1994) poussent encore plus loin cette perspective en affirmant que la production des connaissances aurait évolué d'un mode disciplinaire (Mode 1) vers un mode transdisciplinaire (Mode 2). Dans le Mode

2, le travail de recherche serait guidé par le contexte d'application des connaissances, en évoluant en fonction des problèmes à résoudre.

Gibbons *et al.* (1994) soulignent que dans le Mode 2, les découvertes sont construites à l'extérieur des frontières d'une discipline alors que dans le Mode 1, une découverte peut se construire par rapport à une autre (Gibbons *et al.*, 1994). Certains chercheurs interprètent cette perspective comme une thèse soutenant « la dissolution des frontières disciplinaires » (Louvel, 2015) présente dans le Mode 2 de la production des connaissances. Heilbron et Gingras (2015) affirment que cette conception du Mode 1 et du Mode 2 de la production des connaissances proposée par Gibbons *et al.* (1994) manque de fondements tant historiques qu'empiriques. Heilbron et Gingras (2015) évoquent les travaux de Jerry Jacobs qui concluent que les disciplines ne constituent pas des silos isolés, mais qu'elles sont plutôt des points de rencontre dans un réseau remarquablement dynamique de bourses d'études en particulier. Jacobs (2013) souligne en effet que, dans les grandes universités de recherche, les disciplines sont reliées par un vaste réseau de programmes, de centres et d'instituts de recherche interdisciplinaire.

D'après Darbellay (2011), la notion d'interdisciplinarité peut se définir en rapport avec la disciplinarité. La disciplinarité impliquerait l'arrangement de plusieurs communautés de spécialistes divisés en différentes unités organisationnelles autonomes (facultés, départements, laboratoires) au sein d'une institution (Darbellay, 2011). En ce sens, une discipline serait vue comme une unité organisationnelle caractérisée par la division et la spécialisation du travail et orientée vers l'autonomie, en délimitant ses frontières avec un langage propre, des techniques - qu'elle développe ou qu'elle utilise - et ses propres théories (Darbellay, 2011). En s'inspirant de Bourdieu (2001) qui considère une discipline comme un « champ scientifique », Heilbron et Gingras (2015) définissent les disciplines ainsi:

Dans l'univers académique, les « disciplines » représentent une structure institutionnelle incontournable parce que ce champ repose sur une division de travail fortement institutionnalisée. Les « disciplines » constituent des univers relativement stables et délimités ; elles se sont mises en place au cours du XIXe siècle, ont acquis un nom connu et reconnu et structurent un ordre institutionnel avec des départements, des procédures de certification, des revues, ainsi que des instances nationales et internationales. Les disciplines se définissent ainsi par l'existence « d'un capital collectif de méthodes et de concepts spécialisés dont la maîtrise

constitue le droit d'entrée tacite ou implicite dans le champ ». Elles produisent des habitus disciplinaires spécifiques, c'est-à-dire des « systèmes de schèmes de perception et d'appréciation » incorporés, qui sont constitutifs de « styles » plus ou moins distinctifs. (Heilbron et Gingras, 2015, p.8)

La disciplinarité incite à réfléchir sur la place que pourrait occuper la pratique de l'épidémiologie numérique en milieu universitaire. Serait-elle une structure institutionnelle autonome délimitant ses frontières avec un langage propre, des techniques et des théories, comme proposé par Darbellay (2011) ? Dans quelle mesure cette pratique pourrait-elle former une discipline ? Gingras (1991) propose de distinguer trois aspects de la formation des disciplines, c'est-à-dire l'émergence d'une pratique nouvelle, l'institutionnalisation de cette pratique et la formation d'une identité sociale. La pratique de l'épidémiologie numérique existe, mais est-elle reproduite et diffusée de façon systématique (Gingras, 1991) ? Son identité sociale prend-elle la forme d'une profession ou d'une discipline (Gingras, 1991) ?

À propos de l'interdisciplinarité, Darbellay (2011) souligne :

[L'interdisciplinarité] vise la collaboration entre spécialistes d'horizons disciplinaires différents et complémentaires, voire l'intégration entre les disciplines. L'interdisciplinarité est fondée sur un partage consenti des objectifs, des méthodes et des langages spécifiques à chacune des disciplines impliquées dans le processus de la coproduction des connaissances. (Darbellay, 2011, p.74)

En ce qui concerne la transdisciplinarité, elle peut être vue comme un processus de connaissances fondé sur l'intégration et une vision globale venant restructurer les connaissances des disciplines pour analyser un objet d'études complexe (Darbellay, 2011). La transdisciplinarité irait au-delà des interactions et des réciprocitys entre des recherches spécialisées en situant ces liaisons au sein d'un système entier sans frontière et établi de manière stable entre les disciplines (Darbellay, 2011). Certains travaux font ressortir une autre conception de la transdisciplinarité, comme une approche de recherche critique et autoréflexive dont l'objectif est de contribuer au progrès sociétal et scientifique (Popa *et al.*, 2015). La transdisciplinarité met en relation des problèmes sociétaux et scientifiques en produisant de nouvelles connaissances fondées sur différentes perspectives scientifiques et extrascientifiques (Popa *et al.*, 2015). À titre d'exemple, il semble que les scientifiques du domaine du développement durable ont reconnu le besoin d'aller d'une approche

interdisciplinaire vers des collaborations transdisciplinaires réunissant des expertises scientifiques et extrascientifiques (Popa *et al.*, 2015).

La pratique de l'épidémiologie numérique va-t-elle plus loin que l'interdisciplinarité? La littérature révèle que la science citoyenne¹ (Benyei *et al.*, 2021; Dias Da Silva *et al.*, 2017) et la production participative (*crowdsourcing*) (Ranard *et al.*, 2014)² peuvent faire partie du paysage de la pratique de l'épidémiologie numérique (Hamer *et al.*, 2018; Katapally *et al.*, 2021; Müller et Salathé, 2019), démontrant ainsi de la transdisciplinarité puisque cette façon de faire de la science implique des formes de coopération entre chercheurs et communauté en vue de faire progresser la science et la société. D'ailleurs, il existe des laboratoires de recherche entièrement dédiés à la pratique de l'épidémiologie numérique, dont un laboratoire canadien et un laboratoire suisse.

Les chercheurs affiliés au laboratoire canadien *Digital Epidemiology and Population Health Laboratory* (DEPHL) font appel à des citoyens scientifiques et à des patients par l'entremise de leurs propres appareils numériques (p. ex., des téléphones intelligents) et de la plateforme intelligente *Smart Platform*, afin de mettre en œuvre des études sur la santé des populations, des initiatives communautaires permettant la participation des citoyens tout au long de leur vie, ainsi que des initiatives basées sur le couplage de données comportementales, administratives et de santé (Katapally, 2023). Le laboratoire suisse *Digital Epidemiology Lab* utilise l'apprentissage automatique ainsi que les technologies en ligne et mobiles pour créer des applications de santé numérique dans le but de relever les défis de santé du 21^e siècle (DEL, 2021). Une des façons de faire de la science au sein de ce laboratoire est la production participative (*crowdsourcing*) à travers la plateforme ouverte *AIcrowd* (AIcrowd, 2023). Les chercheurs utilisant cette plateforme

¹ Benyei *et al.* (2021) proposent de définir la science citoyenne comme une approche transdisciplinaire éliminant les hiérarchies de connaissances et encourageant la coproduction de connaissances par les profanes et les scientifiques dans le but de faire de la science par et pour les gens. Dans une perspective similaire, Dias Da Silva *et al.* (2017) mentionnent que la science citoyenne peut se définir entre autres comme une science faisant participer des membres du public lors de la production des connaissances scientifiques.

² En contexte de recherche, Ranard *et al.* (2014) affirment que la recherche par la production participative permet aux chercheurs d'engager des milliers de personnes pour fournir des données ou des analyses de données. Ranard *et al.* (2014) soulignent aussi que les projets mis de l'avant avec une approche fondée sur la production participative sont réalisés avec les contributions d'un large public externe avec peu de restrictions sur les personnes susceptibles de participer.

cherchent à résoudre des problèmes en réunissant les efforts et l'expertise de spécialistes et de passionnés de l'apprentissage automatique et de la science des données (AICrowd, 2023).

En revenant à l'interdisciplinarité, il est important de souligner que des besoins technologiques (Knorr-Cetina, 1999) et des besoins d'expertise spécialisée pour comprendre les phénomènes étudiés (Knorr-Cetina, 1999) peuvent favoriser l'interdisciplinarité. Dans le contexte de la pratique de l'épidémiologie numérique, des chercheurs peuvent être incités à collaborer avec des collègues en informatique dans le but d'accéder aux technologies et aux compétences requises pour traiter des données numériques. De plus, des collaborations interdisciplinaires peuvent aussi se former afin d'aller chercher de l'expertise spécialisée sur les phénomènes étudiés. Les chercheurs faisant déjà partie d'une discipline, ces collaborations interdisciplinaires seraient temporaires, en fonction des ressources requises pour le projet de recherche.

En somme, ces réflexions amènent des questionnements sur l'interdisciplinarité et la transdisciplinarité en lien avec la pratique de l'épidémiologie numérique. Cette nouvelle pratique est-elle interdisciplinaire ou transdisciplinaire? Jusqu'où pourrait aller l'institutionnalisation de cette nouvelle pratique ? De plus, ces réflexions incitent à questionner la pratique de l'épidémiologie numérique du point de vue des collaborations interdisciplinaires auxquelles elle donne lieu. Dans quelle mesure ces collaborations engendrent-elles des tensions avec les pratiques de recherche épidémiologiques existantes ? Si oui, jusqu'à quel niveau ? Ces tensions sont-elles suffisantes pour créer les conditions nécessaires à la constitution d'une nouvelle sous-discipline en épidémiologie ? Au-delà de la discipline de l'épidémiologie et de la diversité dans les pratiques de recherche, le contexte dans lequel se déploie la pratique de l'épidémiologie numérique est aussi influencé par le phénomène de la mise en données de la société.

1.1.5 La mise en données de la société

Nous considérons la mise en données de la société comme un phénomène important du contexte de la pratique de l'épidémiologie numérique puisque les données sont au cœur de cette nouvelle pratique et des pratiques courantes en épidémiologie. Nous aborderons ici les façons de concevoir

la mise en données de la société, les enjeux sociopolitiques associés, les conséquences pour la recherche ainsi que les enjeux épistémiques liés à la place centrale des données.

1.1.5.1 Les façons de concevoir la mise en données de la société

La mise en données de la société semble prendre de l'ampleur suite aux perturbations engendrées par la révolution des technologies de l'information (Castells, 2001) et l'émergence du paradigme techno-informationnel (Mattelart, 2009). Selon Mattern (2013), le phénomène de la mise en données peut être vu comme la présomption que tous les flux et activités significatifs peuvent être détectés et mesurés. D'autres affirment que ce phénomène renvoie à la prise d'informations sur « toutes les choses du monde » - y compris celles n'ayant jamais été considérées comme des informations telles que l'emplacement d'une personne, les vibrations d'un moteur ou la tension d'un pont - et de les transformer dans un format de données pour les quantifier (Mayer-Schönberger et Cukier, 2013, p.17). Selon ces chercheurs, cela permet d'utiliser les informations d'une nouvelle manière, de faire des analyses prédictives et ainsi, de déverrouiller la « valeur implicite et latente de l'information » (Mayer-Schönberger et Cukier, 2013, p.17).

Mayer-Schönberger et Cukier (2013) soulignent que le phénomène de la mise en données (« datafication ») a été rendu devenu possible grâce à des innovations technologiques qui ont permis de mieux traiter et analyser des données. Mais ces innovations se sont accompagnées d'un autre changement majeur renvoyant à la façon de conceptualiser les données. Selon ces chercheurs, les données ne sont plus considérées comme statiques et donc potentiellement périmées, lorsque leur utilité n'existe plus une fois que l'objectif pour lequel elles ont été collectées a été atteint. Au contraire, les données seraient devenues la matière première des entreprises, un apport économique vital, utilisé pour créer une nouvelle forme de valeur économique (Mayer-Schönberger et Cukier, 2013; Sadowski, 2019).

Ce phénomène de mise en données crée des perturbations dans de nombreux secteurs comme celui de la recherche en santé, qui s'y traduisent, entre autres, par les promesses de la médecine personnalisée (Younesi et Hofmann-Apitius, 2013; Hoeyer, 2019) ou par le potentiel d'analyses des données issues des appareils médicaux (Harezlak et Di, 2019; Karas *et al.*, 2019). De plus, les

grandes quantités de données générées à partir des moteurs de recherche Internet, des téléphones cellulaires et des médias sociaux peuvent maintenant être traitées, stockées et analysées en temps réel ou quasi réel grâce aux développements des technologies informatiques et de l'intelligence artificielle (Salathé, 2018; Armstrong *et al.*, 2021). L'analyse de ces données peut contribuer à l'avancement des connaissances dans le cadre de travaux en santé et apporter des transformations dans les pratiques de nombreux domaines de recherche, notamment en épidémiologie, la discipline fondatrice de la santé publique. Ainsi, certaines équipes de recherche interdisciplinaire dans le secteur de la santé adhèrent à cette perspective en pratiquant l'épidémiologie numérique et contribuent à des recherches sur la nutrition (Lazzari *et al.*, 2018), les maladies des plantes (Mohanty *et al.*, 2016) et la surveillance de la grippe (Clemente *et al.*, 2019).

Dans une perspective critique, Boyd et Crawford (2012) soulignent des questions importantes de la mise en données de la société en analysant les big data. Selon Boyd et Crawford (2012), les big data sont un phénomène culturel fondé sur l'interaction entre la technologie, l'analyse et les façons de concevoir le monde qui suscite des discours de grande ampleur. Boyd et Crawford (2012) mettent de l'avant des mises en garde en ce qui a trait entre autres à la qualité des données, aux difficultés d'accès des chercheurs à ce type de données et aux questions éthiques. Ces perspectives montrent que la mise en données de la société n'est pas seulement un phénomène technologique, mais aussi sociopolitique. Nous allons donc continuer notre analyse avec les enjeux sociopolitiques de la mise en données de la société pour ensuite expliquer les conséquences pour la recherche et les enjeux épistémiques liés à ce phénomène.

1.1.5.2 Les enjeux sociopolitiques de la mise en données

Nous soutenons que la pratique de l'épidémiologie numérique suscite aussi un ensemble des questionnements sur le renouvellement des méthodes de recherche et des types de données, qui soulèvent des enjeux de nature sociopolitiques, directement liés à l'omniprésence des données et à leur très grande taille. Pour certains chercheurs, les méthodes de recherche et les données renvoient à des phénomènes plus larges qui ne peuvent se réduire à des opérations purement techniques, et celles-ci peuvent avoir un impact important sur la vie des individus (Mayer-Schönberger et Cukier, 2013; Kitchin, 2014; Lupton, 2015), entraîner une concentration de pouvoir chez les grands détenteurs de données comme Google et Facebook (Mayer-Schönberger et Cukier, 2013; Zwitter,

2014; Ruppert *et al.*, 2015) et favoriser de nouvelles façons de travailler en recherche (Marres, 2012; Mattern, 2013; Rogers, 2015; Boullier, 2015; Kitchin, 2014; Ruppert et al., 2013).

Le phénomène de la mise en données atteint les individus dans leur quotidien, notamment lorsque ceux-ci font usage des réseaux sociaux afin d'échanger et de rester en contact avec leurs réseaux de connaissances. Les réseaux sociaux prennent des éléments intangibles de la vie quotidienne et les transforment en données, qui peuvent ainsi être utilisées dans différents buts (Mayer-Schönberger et Cukier, 2013) comme surveiller les réactions des usagers à des fins de marketing (Ruckenstein et Granroth, 2020) ou établir des corrélations entre différentes sources pour faire des prédictions (Lupton, 2015). Ces nouvelles utilisations des données pourront à leur tour créer de nouvelles connaissances sans que les usagers de ces technologies en soient conscients (Beer, 2009).

En lien avec des objectifs de justice sociale (Dencik et Sanchez-Monedero, 2022), cet aspect du phénomène de la mise en données soulève de nombreux défis quant aux enjeux éthiques de l'usage des données (Zwitter, 2014; Ruppert *et al.*, 2015), aux modes de surveillance, aux risques pour l'atteinte à la vie privée et au profilage (Mayer-Schönberger et Cukier, 2013; Kitchin, 2014). De plus, les effets perturbateurs liés à la mise en données ont de nombreuses implications pour les gouvernements, les entreprises et la société civile, car ils créent de nouvelles connaissances et pratiques susceptibles d'apporter des changements importants dans la gouvernance des personnes, la gestion des organisations et de l'économie ainsi que dans la manière de pratiquer et de penser la science (Kitchin, 2014).

Pour certains chercheurs, les méthodes de recherche impliquent des enjeux sociopolitiques dans le sens où elles ne sont pas une boîte à outils neutre que l'on peut choisir d'utiliser ou non (Savage, 2013). Elles sont des entités vivantes situées dans le social (Law *et al.*, 2011). Cette perspective amène des questionnements sur la façon dont les méthodes de recherche incarnent certaines préoccupations d'acteurs et sur la manière dont celles-ci subsistent dans des contextes ou des environnements particuliers (Law *et al.*, 2011). De tels questionnements peuvent engendrer des réflexions critiques vis-à-vis de ces méthodes et des analyses auxquelles elles conduisent, également sur le statut des données qu'elles produisent et sur les diverses préoccupations qu'elles

reflètent, comme la sécurité publique, le marketing ciblé, le contrôle des maladies et la génomique des populations (Law *et al.*, 2011).

1.1.5.3 Les conséquences pour la recherche

La place centrale des données dans la société peut apporter des changements importants dans des pratiques de recherche et susciter diverses attentes sur l'avenir des disciplines. Par exemple, dans les sciences humaines, les humanités numériques ont émergé avec la mise en données de la sphère culturelle. Selon Vinck (2016), « les humanités numériques reviennent, de fait, à saisir les phénomènes culturels et sociaux (comme des textes, des images et des films) en les transformant en des séries de données numériques » (Vinck, 2016, p.39) de plus en plus souvent à l'aide de l'informatique. Les humanités numériques peuvent être vues comme la quantification des sciences humaines (Vinck, 2016). Cette quantification des sciences humaines présente plusieurs avantages, comme les nouvelles possibilités en matière de traitement automatique des textes (comptage de mots pour identifier les plus utilisés ou associations entre les mots sur de très grands volumes de textes (Vinck, 2016)). Cependant, cette quantification est controversée, notamment en histoire à cause des difficultés à saisir la complexité et la singularité des événements historiques (Vinck, 2016). Les historiens utilisent tout de même des technologies numériques, non pas pour quantifier les phénomènes observés, mais pour faciliter l'accès à un grand nombre de documents rassemblés dans des bases de données et ainsi pouvoir les rapprocher et les comparer facilement (Vinck, 2016).

En sociologie, certains affirment que les sociologues devraient développer d'autres façons de faire de la sociologie, qui seraient plus en phase avec l'ère numérique, en particulier si les tenants de la discipline veulent conserver leur position prééminente d'experts en recherche sociale (Lupton, 2015). Ces débats mettent en confrontation des questions plus larges sur la nature de la discipline elle-même, notamment l'avenir de la recherche sociologique et de la théorisation à l'ère numérique (Lupton, 2015). Déjà en 2007, certains chercheurs en sociologie empirique soutiennent que les nouvelles générations de sociologues empiriques devront analyser des données transactionnelles « sociales » collectées, traitées et analysées de façon routinière par un large éventail d'institutions privées et publiques (Savage et Burrows, 2007). Ce type de recherche impliquant entre autres l'utilisation de données numériques serait différent par rapport à d'autres techniques traditionnelles

de recherche sociale comme celle du sondage basé sur la collecte de données auprès de populations dans le but d'étudier l'opinion publique et constituer des données officielles publiées par des agences statistiques nationales (Tremblay, 1991).

Dans une perspective historique, un autre chercheur en sociologie propose de regarder la mutation numérique comme une troisième génération pour les sciences sociales à travers une grille de lecture faisant ressortir les différentes « époques » de réflexivité sur les façons de concevoir les recherches sociales au fil du temps (Boullier, 2015). Selon Boullier (2015), la matière première des recherches sociales dans l'âge numérique ne serait ni les personnes, ni les identités, ni les communautés, mais les traces laissées sur les plateformes, qu'il devient possible d'exploiter pour étudier la société ou les opinions. Toujours selon Boullier (2015), la troisième génération des sciences sociales ne parlerait plus de personnes, mais de comptes, plus de communautés, mais de clusters, plus de sociabilité, mais de connectivité.

D'autres travaux soulignent les difficultés qu'il y a à différencier ce qui relève de la dynamique des médias eux-mêmes, de ce qui concerne les dynamiques sociales étudiées lorsque l'on réalise des analyses sur des données générées par Internet (Marres et Weltevrede, 2013; Marres et Gerlitz, 2016). Ces défis semblent liés au fait que des tendances observées à partir de données numériques comme celles issues des médias sociaux ne signifient pas nécessairement que de telles tendances s'observent dans la vie réelle. De plus, des enjeux techniques liés aux données du web peuvent venir renforcer les incertitudes face aux méthodes numériques en recherche. De tels enjeux incluent la durée limitée des services web et l'instabilité des flux de données (Rogers, 2015).

Marres (2012) propose le concept de redistribution des méthodes pour contribuer aux débats sur les implications du numérique dans les transformations de la recherche. Dans le contexte de la numérisation, Marres (2012) soutient que la recherche devient un accomplissement distribué, dans le sens où les plateformes en ligne, les utilisateurs, les dispositifs et les pratiques informationnelles contribuent tous activement à la réalisation de la recherche. Marres et Gerlitz (2016) proposent quant à eux le concept de méthodes d'interface (*interface methods*) pour souligner, entre autres, que de multiples traditions méthodologiques se croisent dans les dispositifs et la recherche numérique, comme celles de la recherche sociale numérique, le domaine STS et la sociologie. Pour

leur part, Bastin et Francony (2020) examinent les interactions et plus particulièrement les conflits d'interprétation entre différents chercheurs, dans le cadre d'un projet d'enquête exploitant des données du réseau social professionnel LinkedIn. Dans ce contexte, le processus de mise en données leur est apparu être beaucoup plus qu'une opération technique de la recherche. Ils définissent la mise en données comme un « processus de transformation des inscriptions individuelles sur le réseau en données » qui passe par des confrontations entre les différents acteurs du laboratoire de recherche (Bastin et Francony, 2020).

Il semble donc que les sites de réseaux sociaux, les algorithmes et les bases de données peuvent servir de dispositif méthodologique pour des recherches et ainsi être conceptualisés à la fois comme des entités pouvant influencer sur les relations sociales et comme des objets de recherche intéressants (Savage, 2013; Ruppert *et al.*, 2013). C'est pourquoi certains chercheurs s'intéressent à ce que les méthodes font à la recherche et au social, et comment elles sont façonnées par le monde social (Savage, 2013).

1.1.5.4 Les enjeux épistémiques liés à la place centrale des données

Nous considérons que le phénomène de la mise en données dans le monde de la recherche peut impliquer des enjeux épistémiques parce que les connaissances, les objectifs et les pratiques liés à la vérité constituent les éléments centraux des domaines scientifiques (Knorr-Cetina et Reichmann, 2015). De plus, il est possible que certaines méthodes d'analyse de données puissent influencer les façons de concevoir les phénomènes étudiés (Knorr-Cetina, 1999).

Les travaux de Knorr-Cetina (1999) mettent en évidence des enjeux épistémiques qu'il importe de prendre en compte pour analyser des pratiques de recherche liées à la mise en données, comme la pratique de l'épidémiologie numérique. Ces enjeux incluent les signes et les apparences secondaires ainsi que les façons de considérer les mesures dans les mondes scientifiques.

Les instruments utilisés lors des recherches produisent des signes et font apparaître des choses perçues par les chercheurs lors de l'étude des phénomènes. Ces signes et ces apparences peuvent être utiles pour comprendre les phénomènes étudiés, mais ils peuvent aussi être source de

complications, à cause des distorsions et des bruits générés par les instruments (Knorr-Cetina, 1999). Ainsi, les instruments utilisés dans le cadre de la pratique de l'épidémiologie numérique peuvent susciter une meilleure compréhension des phénomènes épidémiologiques, mais ils peuvent aussi générer des incertitudes sur la qualité des signes et des apparences.

De plus, les façons de concevoir les mesures, telles que développées par Knorr-Cetina (1999), font ressortir deux perspectives différentes. La première perspective met en évidence les attentes élevées envers les mesures, qui sont vues comme étant indispensables pour valider des connaissances, prouver ou réfuter des théories, observer de nouveaux phénomènes et représenter des résultats intéressants et potentiellement publiables (Knorr-Cetina, 1999). La deuxième perspective apporte des nuances en soulignant que les mesures sont parfois sujettes à interprétation et que celles-ci peuvent varier selon l'appareil de mesure et les critères utilisés pour en extraire de l'information (Knorr-Cetina, 1999). Quant aux données, Knorr-Cetina (1999) met l'accent sur l'idée qu'elles sont contingentes à l'appareil de mesure et que celles-ci restent des représentations de cet appareil et non pas du monde réel.

Les perspectives quant à la pratique de l'épidémiologie numérique peuvent être grandement influencées par les façons de concevoir les mesures et les données numériques. La discipline de l'épidémiologie étant axée sur les données et les mesures, il semble logique que la pratique de l'épidémiologie numérique soit confrontée à des exigences élevées sur les mesures et données générées par les méthodes utilisées par cette nouvelle pratique. Ces mesures peuvent être sujettes à l'interprétation des chercheurs pratiquant l'épidémiologie numérique et elles peuvent varier en fonction de l'instrument utilisé lors des analyses ou du traitement des données.

En somme, nous considérons la mise en données de la société comme un phénomène important, lié au contexte de la pratique de l'épidémiologie numérique puisque les données sont au cœur de cette nouvelle pratique. Nous affirmons aussi que les études sur la pratique de l'épidémiologie numérique nécessitent d'examiner les enjeux sociopolitiques et épistémiques des méthodes utilisées par les chercheurs. Les méthodes de recherche et les données ne peuvent pas être réduites à de simples opérations techniques, car celles-ci peuvent avoir un impact important sur la vie des individus, entraîner une concentration de pouvoir auprès de grands détenteurs de données et amener

de nouvelles façons de travailler en recherche. La pratique de l'épidémiologie numérique peut aussi impliquer des enjeux épistémiques liés d'une part, au fait que les connaissances, les objectifs et les pratiques liés à la vérité sont les éléments centraux des domaines scientifiques et d'autre part, au fait que les méthodes d'analyse de données peuvent influencer les façons de concevoir les phénomènes étudiés.

Pour terminer cette section, nous rappelons que la pratique de l'épidémiologie numérique est un phénomène complexe qui doit être analysé en tenant compte de la discipline de l'épidémiologie, des pratiques de recherche, des changements dans les pratiques de recherche, de l'interdisciplinarité et de la mise en données de la société. Le contexte de recherche lié à la pratique de l'épidémiologie numérique étant cerné, la section suivante présentera l'état des recherches sur cette nouvelle pratique scientifique.

1.2 L'état des recherches en épidémiologie numérique

Selon les résultats de nos recherches à partir des bases de données PubMed, Springer et Scopus, le nombre de publications liées à la pratique de l'épidémiologie numérique a augmenté de manière significative depuis les dernières années. Il s'avère donc pertinent d'en savoir davantage sur cette nouvelle pratique qui suscite un intérêt grandissant dans les recherches en épidémiologie.

Le but de cette section consiste à dresser un portrait des recherches sur l'épidémiologie numérique. Nous commencerons par expliquer la méthodologie employée pour réaliser la revue de la littérature. Puis nous présenterons une analyse quantitative des recherches. Ce travail fera ressortir les différentes perspectives sur l'épidémiologie numérique, les technologies et acteurs concernés, ainsi que les difficultés et les controverses liées à cette nouvelle pratique.

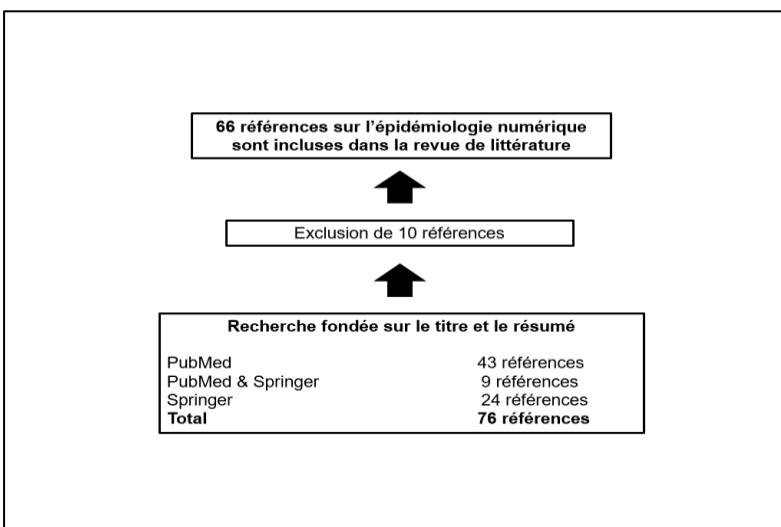
1.2.1 La méthodologie employée pour la revue de la littérature

L'analyse de la revue de la littérature a reposé sur la lecture complète de 66 articles en épidémiologie numérique. Cette démarche a été complétée par un inventaire de 535 articles afin de dresser un portrait plus large des recherches en épidémiologie numérique.

Nous avons réalisé une première recherche en date du 30 avril 2019, dans les bases de données PubMed et Springer, avec l'expression « *Digital Epidemiology* » recherchée dans le titre et le résumé. Cette recherche a fait ressortir 76 références dont 43 tirées de PubMed, 24 de Springer et 9 communes aux deux bases de données. Un total de 10 références ont été exclues par manque de pertinence ou manque d'accès aux articles (dont 4 références de PubMed et 6 de Springer). La revue de littérature est par conséquent fondée sur la lecture de 66 références (Figure 1.1). Ces références traitent de l'épidémiologie numérique selon les constats émanant de diverses disciplines et domaines comme l'informatique, la médecine et la santé publique.

L'analyse de la littérature visait à fournir un aperçu des caractéristiques de la pratique de l'épidémiologie numérique, comme les principales technologies et sources de données utilisées ainsi que les apports et limites de cette approche dans le cadre de recherches épidémiologiques.

Figure 1. 1 La démarche de la revue de la littérature datée de 2019



Or, nous avons réalisé lors de la recherche de terrain que plusieurs participants ne connaissaient pas l'expression « épidémiologie numérique », mais qu'ils étaient au courant que certains travaux de recherche en épidémiologie utilisaient des données en provenance de sources non traditionnelles comme les médias sociaux, les téléphones cellulaires et les moteurs de recherche Internet. Il est donc possible que d'autres recherches relèvent de l'épidémiologie numérique sans pour autant

mentionner l'expression « épidémiologie numérique ». Par exemple, certains articles pourraient traiter de recherche épidémiologique fondée sur des données en provenance des médias sociaux, sans mentionner qu'il s'agit de l'épidémiologie numérique.

Nous avons donc décidé d'utiliser la base de données Scopus pour faire une nouvelle recherche. Cette base de données contient une grande quantité de travaux provenant de plusieurs domaines et disciplines et elle est reconnue pour ses fonctions d'analyse automatisée des citations et des résultats obtenus à travers son moteur de recherche (Falagas *et al.*, 2008). En recherchant dans la base de données Scopus, nous avons utilisé une autre requête afin d'inclure les recherches liées à la pratique de l'épidémiologie numérique n'ayant pas mentionné spécifiquement l'expression « épidémiologie numérique ». La requête utilisée est celle incluse dans la Figure 1.2.

Figure 1. 2 La requête utilisée dans Scopus datée de 2022

```
( TITLE-ABS-KEY ( epidemiology ) AND TITLE ( facebook ) OR TITLE ( twitter ) OR TITLE ( tweet ) OR TITLE ( social AND media ) OR TITLE ( sentiment ) OR TITLE ( youtube ) OR TITLE ( cell AND phone ) OR TITLE ( google ) OR TITLE ( digital ) OR TITLE ( intelligent AND phone ) OR TITLE ( non AND traditional ) OR TITLE ( real AND time AND data ) OR TITLE ( search AND engine AND data ) OR TITLE ( Internet AND research AND data ) OR TITLE ( real AND world AND data ) ) AND ( LIMIT-TO ( PUBYEAR , 2022 ) OR LIMIT-TO ( PUBYEAR , 2021 ) OR LIMIT-TO ( PUBYEAR , 2020 ) OR LIMIT-TO ( PUBYEAR , 2019 ) OR LIMIT-TO ( PUBYEAR , 2018 ) OR LIMIT-TO ( PUBYEAR , 2017 ) OR LIMIT-TO ( PUBYEAR , 2016 ) OR LIMIT-TO ( PUBYEAR , 2015 ) OR LIMIT-TO ( PUBYEAR , 2014 ) OR LIMIT-TO ( PUBYEAR , 2013 ) OR LIMIT-TO ( PUBYEAR , 2012 ) OR LIMIT-TO ( PUBYEAR , 2011 ) OR LIMIT-TO ( PUBYEAR , 2010 ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) OR LIMIT-TO ( LANGUAGE , "French" ) )
```

Cette requête a été effectuée dans la recherche avancée de la base de données Scopus (Elsevier, 2023) le 7 janvier 2022 et celle-ci a généré un total de 1,277 références. Notre critère de sélection des références dans le cadre de cette revue de littérature est que les articles doivent porter sur la pratique de l'épidémiologie numérique. Nous avons donc exclu un total de 741 références parce que celles-ci ne portent pas sur l'épidémiologie numérique au sens entendu dans le cadre de cette thèse, c'est-à-dire l'utilisation de données de sources non traditionnelles comme des données

générées à partir des médias sociaux, des moteurs de recherche Internet et de téléphones cellulaires (Salathé *et al.*, 2012; Makri, 2017). Celles-ci traitaient d'infodémiologie, d'images numériques en dentisterie, de sondages diffusés sur les médias sociaux, de méthodes de recrutement à travers les médias sociaux ou encore de données administratives et registres. L'inventaire des recherches en l'épidémiologie numérique extrait à partir de Scopus a donc comporté 535 références.

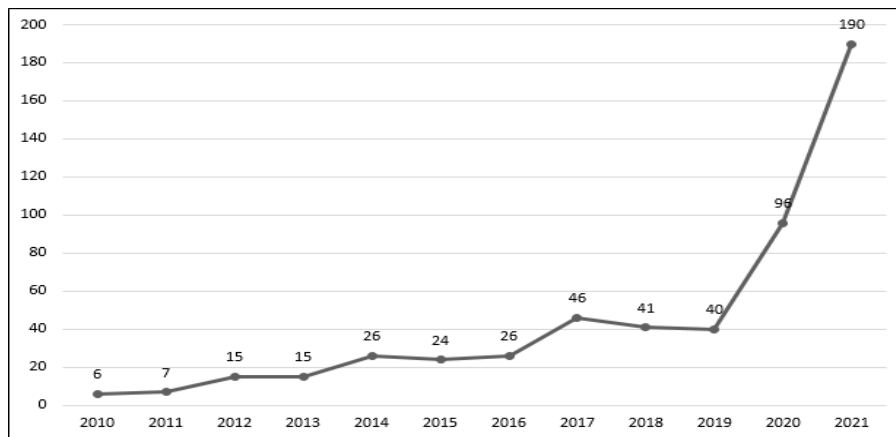
Il est important de souligner que la lecture complète des contenus des articles sur l'épidémiologie numérique s'est faite sur les 66 références résultant de la revue de la littérature à partir des bases de données de PubMed et Springer. Nous avons utilisé cet inventaire de 535 références issu de la recherche dans Scopus seulement pour dresser le portrait des recherches en épidémiologie numérique au niveau quantitatif. La base de données Scopus a permis d'avoir un meilleur estimé du nombre de recherches sur l'épidémiologie numérique, en comparaison avec les bases de données PubMed et Springer. De plus, Scopus a été particulièrement utile pour analyser les données sur les 535 articles publiés en lien avec l'épidémiologie numérique grâce à ses fonctions automatiques d'analyse permettant de traiter facilement un nombre important de références.

1.2.2 Un portrait des recherches en épidémiologie numérique

Comme nous l'avons expliqué précédemment, ce portrait des recherches en épidémiologie numérique consiste en une analyse quantitative des publications portant sur cette nouvelle pratique fondée sur l'inventaire des 535 références extraites à partir de la base de données Scopus. Cette analyse fournit le nombre de publications en épidémiologie numérique par année, par discipline ou spécialité, par affiliation et par pays ou territoire.

La Figure 1.3 montre le nombre de publications par année, liées à l'épidémiologie numérique en date du 7 janvier 2022. Le nombre de publications était inférieur à 30 entre 2010 et 2016 pour ensuite augmenter à 46 publications en 2017. La recherche indique 41 publications en 2018 et 40 publications en 2019. L'année 2020 a connu une augmentation importante avec 96 publications et cette augmentation a continué en 2021 avec 190 publications, correspondant au plus grand nombre de publications en épidémiologie numérique.

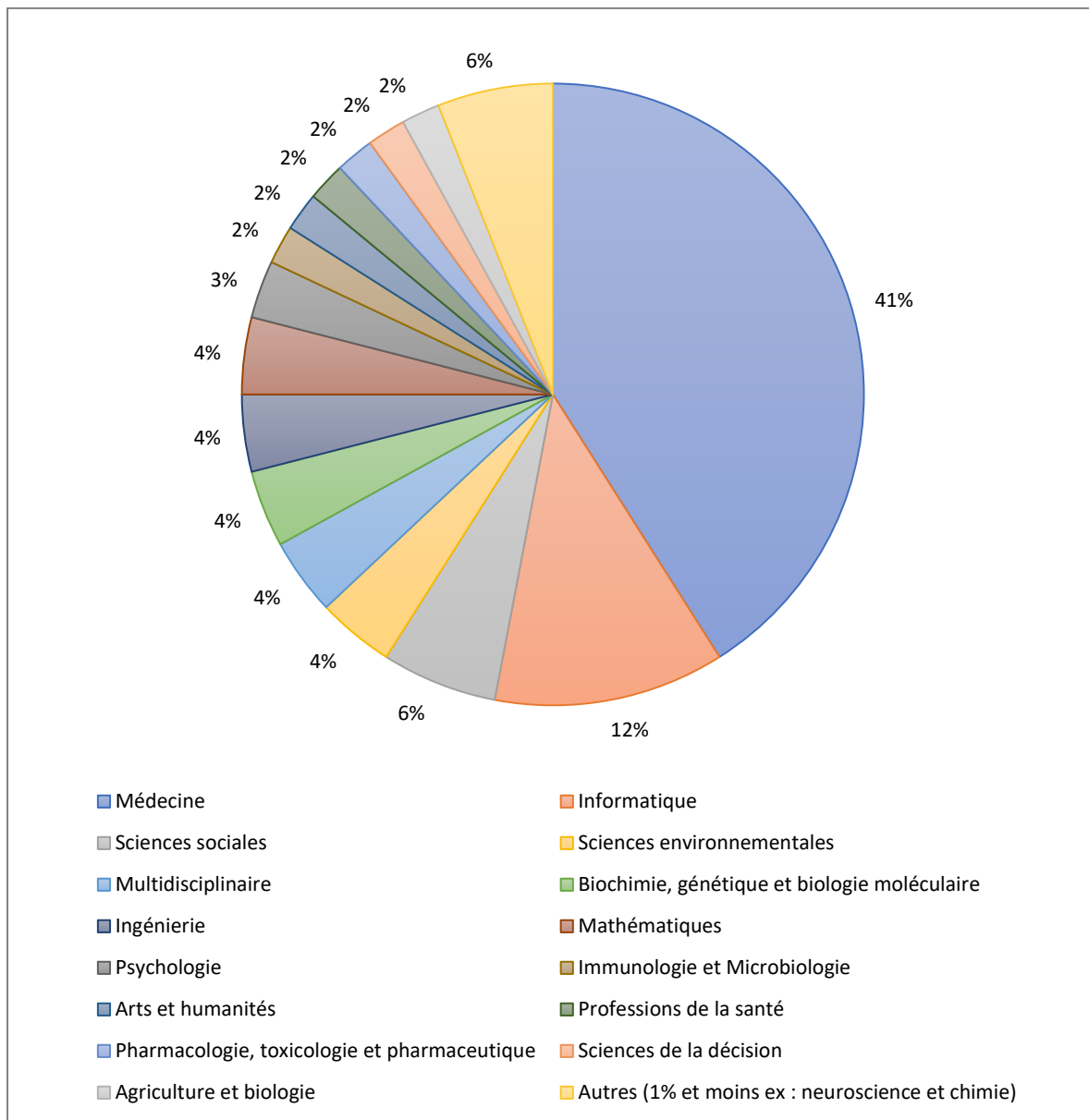
Figure 1. 3 Le nombre de publications en épidémiologie numérique par année dans Scopus entre 2010 et 2021



Source : Scopus en date du 7 janvier 2022

La Figure 1.4 indique le nombre de publications en épidémiologie numérique selon la discipline ou la spécialité des revues. La discipline la plus représentée est la médecine (41%), suivie de l'informatique (12%). Les sciences sociales représentent 6% des publications sur l'épidémiologie numérique. Les sciences environnementales, les spécialités multidisciplinaires, la catégorie « Biochimie, génétique et biologie moléculaire », l'ingénierie et les mathématiques représentent 4% des publications sur l'épidémiologie numérique.

Figure 1. 4 La distribution des publications en épidémiologie numérique par discipline ou spécialité à partir de Scopus entre 2010 et 2022

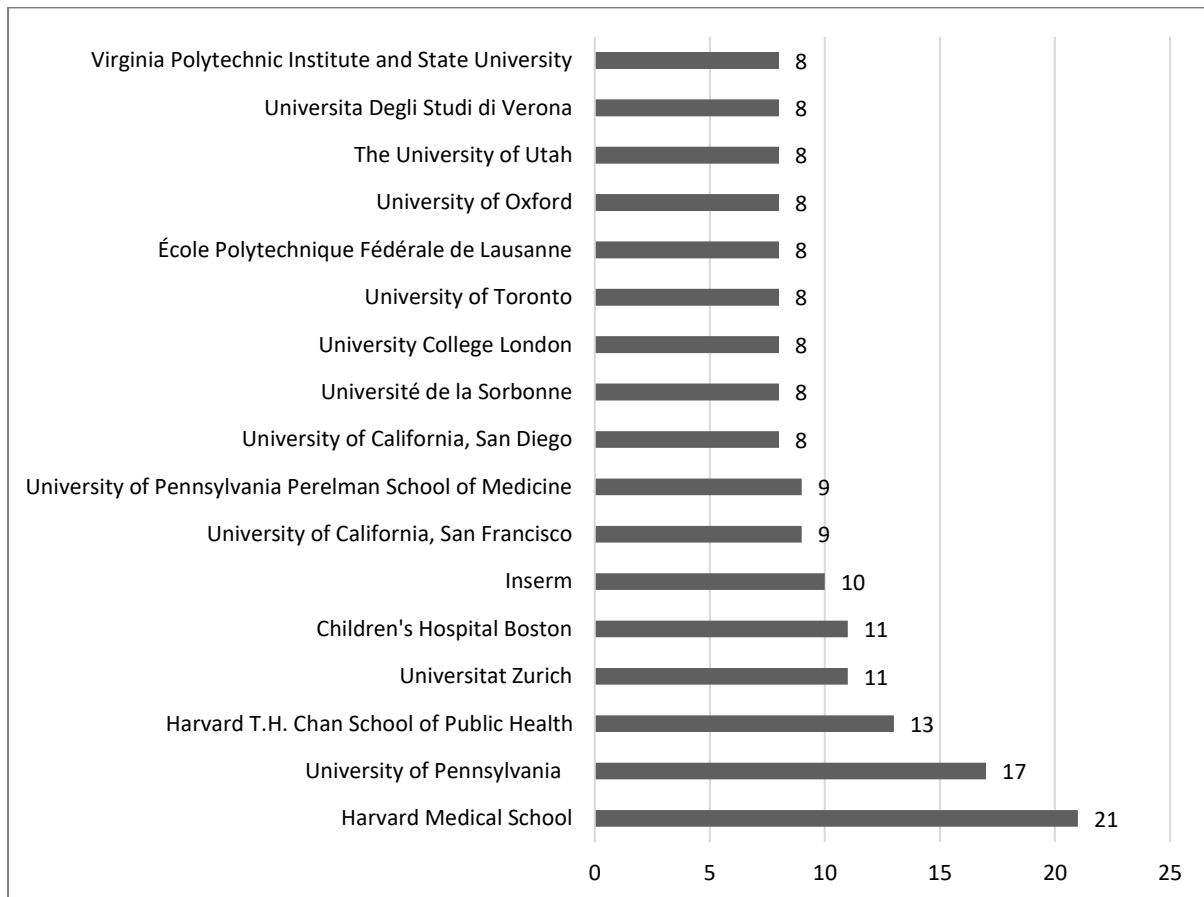


Source: Scopus en date du 7 janvier 2022

Ce portrait des recherches sur l'épidémiologie numérique nous indique aussi un total de 160 affiliations différentes des chercheurs ayant contribué à une publication sur l'épidémiologie numérique entre 2010 et 2022. La Figure 1.5 montre les affiliations des chercheurs ayant au moins 8 publications. Le nombre de publications le plus élevé correspond à la Faculté de médecine de Harvard avec 21 publications, suivie par l'Université de Pennsylvanie avec 17 publications et

l'École de santé publique de Harvard avec 13 publications. University of Toronto figure parmi celles ayant publié 8 publications en épidémiologie numérique.

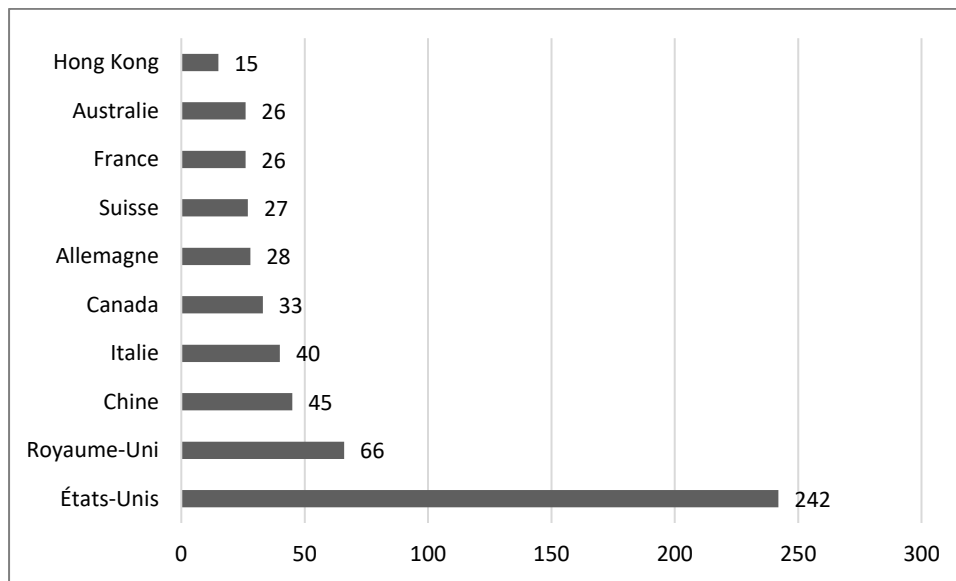
Figure 1. 5 Les affiliations des chercheurs ayant au moins 8 publications en épidémiologie numérique à partir de Scopus entre 2010 et 2022



Source : Scopus en date du 7 janvier 2022

De plus, les données indiquent un total de 67 pays et territoires de provenance pour les chercheurs ayant publié sur cette nouvelle pratique entre 2010 et 2022. La Figure 1.6 présente le nombre de publications en épidémiologie numérique par pays et territoires pour ceux ayant publié au moins 15 articles au cours de la période de référence. Selon ces données, le plus grand nombre de publications en épidémiologie numérique provient des États-Unis avec un total de 242 publications depuis 2010, suivis par le Royaume-Uni avec 66 publications, la Chine avec 45 publications, l'Italie avec 40 publications et le Canada avec 33 publications.

Figure 1. 6 Les pays ayant publié au moins 15 articles en épidémiologie numérique à partir de Scopus entre 2010 et 2022



Source : Scopus en date du 7 janvier 2022

Au Canada, les tendances semblent similaires à celles observables dans le monde avec une augmentation récente du nombre de publications en épidémiologie numérique passant de 7 articles en 2020 à 10 articles en 2021. Concernant la distribution du nombre de publications en fonction de la discipline et la spécialité des revues, c'est la médecine qui est la plus représentée (46%), puis les sciences sociales (8%) et l'informatique (6%). Les institutions canadiennes ayant publié le plus grand nombre d'articles liés à cette nouvelle pratique de recherche sont University of Toronto avec 8 articles et McGill University avec 4 articles, entre 2011 et 2021. Les chercheurs affiliés à University of Toronto ayant contribué à ces publications proviennent de centres de recherche de différents domaines comme celui de l'épidémiologie (Dalla Lana School of Public Health), la médecine, la psychiatrie, la psychologie, des sciences pharmaceutiques (Leslie Dan Faculty of Pharmacy), du travail social (Factor-Inwentash Faculty of Social Work), des sciences de l'information, des systèmes (Middleware System Research Group), de l'évaluation des technologies et de l'économie de la santé (Toronto Health Economics and Technology Assessment (THETA) Collaborative) ainsi que l'évaluation et la gestion des politiques de santé (Institute of Health Policy Management and Evaluation). En ce qui concerne McGill University, les auteurs

ayant contribué aux articles recensés travaillent dans le domaine des politiques sociales et en santé (Institute for Health and Social Policy), des sciences politiques et de l'informatique en santé (Clinical and Health Informatics Research Group). Ce contexte de recherche montre déjà l'interdisciplinarité de cette nouvelle pratique ainsi qu'une grande dispersion des praticants de l'épidémiologie numérique au sein de divers départements et centres de recherche.

En somme, cette analyse quantitative réalisée à partir de la base de données Scopus entre 2010 et 2022, indique que les recherches en épidémiologie numérique sont récentes, avec un sommet en 2021 atteignant 190 publications et que celles-ci sont publiées principalement dans des revues en médecine (42%) et en informatique (12%). L'affiliation la plus fréquente est celle de l'Université Harvard (34 publications) et les pays ayant publié le plus grand nombre d'articles sont les États-Unis avec 242 publications et le Royaume-Uni avec 66 publications. Le Canada figure en cinquième position avec 33 publications, une quantité non négligeable. Dans la section suivante, nous examinons les contenus des publications en épidémiologie numérique à partir de la lecture des 66 articles extraits à partir des bases de données PubMed et Springer.

1.2.3 Les façons de concevoir l'épidémiologie numérique

Une des définitions de l'épidémiologie numérique la plus présente dans la littérature est celle proposée par Salathé (2018) affirmant qu'il s'agit d'une épidémiologie utilisant des données numériques générées en dehors du système de santé publique, c'est-à-dire des données n'ayant pas été générées dans le but principal de faire de l'épidémiologie (Salathé, 2018 cité dans Eckmanns *et al.*, 2019). Plusieurs autres sources indiquent que les données externes au système de santé publique incluent des données numériques générées à partir des médias sociaux, tel que Twitter (Aramaki *et al.*, 2011; Gianfredi, 2018; Riordan *et al.*, 2022), des moteurs de recherche Internet (Clemente *et al.*, 2019; Liu *et al.*, 2021), des plateformes et outils à accès ouvert (Lazzari *et al.*, 2018; Hors-Fraile *et al.*, 2018), ou encore les données produites par les technologies mobiles (Smolinski, 2017; Salathé *et al.*, 2013; Cot *et al.*, 2021), les données et applications institutionnelles (Roche *et al.*,

2017; Lau *et al.*, 2014), les données spatiales³ (Lee *et al.*, 2016; Whitehead *et al.*, 2021) et les données d'imagerie⁴ (Galway *et al.*, 2012; Mohanty *et al.*, 2016).

Salathé *et al.* (2012) et Hasan *et al.* (2017) précisent que l'épidémiologie numérique vise à développer des modèles informatiques, des analyses et des outils d'aide à la décision afin de contribuer à la science épidémique. Makri (2017) ajoute la notion de temps réel en mentionnant que le but de l'épidémiologie numérique consiste à identifier les maladies en temps réel lorsqu'il s'agit d'épidémies de maladies infectieuses et de suivre les changements de comportement dans la communauté.

D'autres chercheurs voient le potentiel de l'épidémiologie numérique pour des recherches sur la surveillance, le suivi (Pollett *et al.*, 2017a; Ho *et al.*, 2022) et la détection des maladies et épidémies en temps réel (Young *et al.*, 2014) tant au niveau de la distribution spatiale qu'au niveau de l'activité. Les analyses des sentiments (Gruebner *et al.*, 2018) et des comportements (Hochreiter et Waldhauser, 2013; Roche *et al.*, 2017) constituent d'autres pratiques en épidémiologie qui peuvent bénéficier de l'épidémiologie numérique, selon certains chercheurs. Cette nouvelle façon de faire de l'épidémiologie peut aussi contribuer aux travaux sur les prédictions et modélisations des épidémies, incluant les facteurs environnementaux et la saisonnalité (Zhao *et al.*, 2018; Giordano *et al.*, 2019; Perra et Gonçalves, 2015) ainsi qu'aux analyses de données spatiales (Whitehead *et al.*, 2021) et en imagerie (Mohanty *et al.*, 2016).

La littérature examinée montre aussi que, dans plusieurs cas, les sources des connaissances utilisées incluent les travaux, données et systèmes traditionnels en santé. Bien que l'épidémiologie numérique constitue une approche émergente, cela ne signifie pas que les connaissances et approches utilisées sont entièrement nouvelles. En effet, tout un écosystème de connaissances et de données en santé est déjà en place (OMS, 2016 cités dans Vayena *et al.*, 2018). En fait, la littérature examinée illustre plutôt une combinaison entre les approches traditionnelles et les

³ Les données spatiales décrivent des informations géométriques, telles que l'emplacement d'une entité dans l'espace ou sa position par rapport à d'autres entités (SD, 2023a).

⁴ Les données d'imagerie sont utilisées pour l'analyse d'images, c'est-à-dire l'extraction d'informations significatives à partir d'images, principalement d'images numériques, au moyen de techniques de traitement d'images numériques (SD, 2023b).

approches numériques dans les pratiques en épidémiologie numérique (Lau *et al.*, 2014; Tourassi *et al.*, 2016). Cette revue de littérature montre aussi que plusieurs auteurs s'attendent à ce que l'épidémiologie numérique complète les approches et systèmes traditionnels en santé et non à ce que celle-ci remplace ce qui est déjà établi (Hartley *et al.*, 2013; Salathé *et al.*, 2016).

1.2.4 Les technologies utilisées en épidémiologie numérique

Au niveau des technologies, l'épidémiologie numérique semble impliquer l'utilisation de données et technologies numériques, de même que l'intelligence artificielle. Cette revue de la littérature met en lumière une utilisation importante d'entités informatiques, de technologies et d'objets numériques producteurs de données dans les pratiques en épidémiologie numérique, incluant des médias sociaux (Salathé *et al.*, 2016; Chary *et al.*, 2017; Bar-Lev *et al.*, 2021; Mavragani et Ochoa, 2019; Basteris *et al.*, 2020 ; Armstrong *et al.*, 2021), des technologies mobiles (Mohanty *et al.*, 2016; Lazzari *et al.*, 2018; Cot *et al.*, 2021; Vigfusson *et al.*, 2021), des algorithmes (Hassan *et al.*, 2019; Hochreiter et Waldhauser, 2013; Durazzi *et al.*, 2021) ainsi que l'intelligence artificielle et l'apprentissage automatique (*machine learning*) (Sau, 2017; Kaveh-Yazdy et Zareh-Bidoki, 2018; Magumba et Nabende, 2021).

Comme pour l'épidémiologie traditionnelle, les données spatiales continuent d'occuper une place importante en épidémiologie numérique. Entre autres, les travaux de Lee *et al.* (2016) expliquent comment des données spatiales massives (*big data*) ont été utilisées dans les analyses épidémiologiques. Zhao *et al.* (2018) examinent quant à eux les effets des conditions météorologiques sur l'incidence de l'infection grippale dans des villes aux États-Unis.

Certains travaux affirment par ailleurs qu'il est possible que les technologies utilisées en épidémiologie numérique amènent des changements importants dans les approches de recherche, en passant d'une production fondée sur les connaissances des experts à une production fondée sur les connaissances des algorithmes. Dans cette perspective, les algorithmes seraient des instruments de prédiction purement rationnels et objectifs indifférents à toute cause de phénomène, et ils chercheraient seulement à accroître au maximum les réservoirs de données pour élucider l'incertain (Eckmanns *et al.*, 2019).

Ces chercheurs soulignent qu'il est possible que les informations et connaissances sur le monde physique ainsi que sur les maladies infectieuses soient extraites de plus en plus sous forme de signaux et de signes provenant du numérique et de moins en moins à partir des processus statistiques fondés sur les analyses humaines. Toujours selon Eckmanns *et al.* (2019), une telle tendance signifierait que les connaissances et les informations générées par les techniques algorithmiques dépasseraient les méthodes traditionnelles humaines d'évaluation, d'analyse, d'hypothèse et d'essais cliniques qui sont essentielles aux calculs statistiques utilisés dans les recherches épidémiologiques traditionnelles. Deux questions se posent en lien avec les propos de Eckmanns *et al.* (2019) : premièrement, l'épidémiologie numérique est-elle complémentaire ou transformationnelle par rapport aux pratiques courantes en épidémiologie ? Deuxièmement, dans quelle mesure la pratique de l'épidémiologie numérique engendre-t-elle des ruptures avec les pratiques courantes en épidémiologie?

1.2.5 Les acteurs impliqués

Les principaux acteurs repérables dans la littérature sur l'épidémiologie numérique comprennent des chercheurs, des organismes internationaux, des non-experts ainsi que des grandes entités privées comme Google et Twitter. Les acteurs impliqués dans cette nouvelle pratique sont en bonne partie situés dans la sphère de la recherche en santé publique et plus particulièrement en milieu universitaire. Notre revue de la littérature à partir de PubMed et Springer montre que la plupart des travaux sont effectués par des chercheurs affiliés à une université. De plus, l'analyse de la littérature fait ressortir la présence d'organisations œuvrant au niveau international comme l'OMS et RKI qui communiquent chacun leur vision quant aux initiatives numériques en santé et leur appui de différentes initiatives, incluant l'épidémiologie numérique (Hill-Cawthorne et Sorrell, 2016; Makri, 2017).

À partir de 2007, l'OMS a probablement joué un rôle crucial dans la coopération internationale pour le contrôle de la propagation mondiale des maladies par la mise en œuvre du RSI de 2005 (OMS, 2016). Ce règlement reconnaît explicitement la nécessité d'une coopération intersectorielle et interdisciplinaire afin de gérer les risques potentiellement importants pour la santé publique au

niveau international. Cette vision a conduit à l'établissement de partenariats d'experts regroupant plus de 120 institutions et réseaux, y compris l'Organisation des Nations Unies et des organisations intergouvernementales (Merianos et Peiris, 2005). Hartley *et al.* (2013) mettent en évidence les bienfaits de ce cadre réglementaire pour la bio surveillance alors que d'autres soulignent ses limites en mentionnant que le RSI (2005) se concentre sur les infections humaines malgré le fait que la majorité des infections émergentes soient zoonotiques (Hill-Cawthorne et Sorrell, 2016).

Pour sa part, l'institut de santé publique Robert Koch en Allemagne (RKI, 2023a) contribue aussi à la protection de la santé dans le monde entier en collaborant étroitement avec l'OMS et en mettant sur pied des projets comme celui de la modélisation épidémiologique des maladies infectieuses basée entre autres sur des approches en épidémiologie numérique (RKI, 2023b). Certains auteurs s'inscrivent dans la vision du RKI sur l'épidémiologie numérique (Makri, 2017) et d'autres au sein du RKI ont discuté ses défis pratiques et éthiques (Velasco, 2018).

L'épidémiologie numérique peut impliquer des acteurs non-experts en plus des experts spécialistes affiliés aux institutions officielles comme les universités, les organismes de soins de santé et l'OMS. Cette situation est souvent désignée comme la médecine participative où les patients ou autres acteurs de la société civile apportent des contributions basées sur leurs expériences et connaissances personnelles afin de contribuer à la gestion des maladies et autres enjeux liés à la santé. La contribution et la participation des acteurs non-experts sont facilitées par les communautés de patients en ligne. Pour certains, dont Younesi et Hofmann-Apitius (2013), ces communautés représentent un vrai modèle de médecine participative, car l'Internet constitue un lieu de partage des expériences par l'intermédiaire de divers médias sociaux.

Certains travaux discutent de l'approche participative en l'associant à la surveillance participative (Salathé, 2016; Smolinski *et al.*, 2017; Hartley *et al.*, 2013; Hamer *et al.*, 2018; Leal-Neto *et al.*, 2021), à l'épidémiologie participative (Hartley *et al.*, 2013; Hamer *et al.*, 2018) ou encore à la santé publique participative (Salathé, 2016). Plusieurs auteurs tels que Hartley *et al.* (2013) et Salathé (2016) mentionnent que les acteurs non-experts incluent aussi les utilisateurs de plateformes participatives sur Internet telles que *FluNearYou*, Twitter et Facebook. Cette catégorie d'acteurs contribue à l'épidémiologie numérique à travers les données générées lors de leurs

recherches Internet et des informations qu'ils communiquent dans les forums de discussion sur les réseaux sociaux.

Hamer *et al.* (2018) complètent ce portrait en soulignant les contributions de citoyens pour la collecte des données lors d'études épidémiologiques sur les maladies à transmission vectorielle⁵. Ce type de science citoyenne offre l'occasion au grand public de prendre part à des travaux scientifiques en collaboration avec les experts scientifiques professionnels (Hamer *et al.*, 2018; Katapally *et al.*, 2021). Une telle pratique semble peu adoptée en santé publique, mais certains chercheurs en santé des populations soutiennent que la science citoyenne devrait être davantage utilisée en raison de son potentiel de fournir de nouvelles perspectives sur la nature des problèmes et sur les façons de mettre en œuvre leurs solutions (Rowbotham *et al.*, 2019).

L'utilisation des données numériques semble propice à l'influence des grands acteurs d'Internet comme Google et Twitter dans les pratiques de recherche en santé. Pour la recherche en épidémiologie numérique, la littérature laisse entrevoir cette influence par l'utilisation de *Google Trends* (Giordano *et al.*, 2019; Guernier *et al.*, 2016; Sato *et al.*, 2021) et Twitter (Roche *et al.*, 2017; Fung *et al.*, 2013; Cheng *et al.*, 2021). *Google Trends* est un outil en ligne disponible publiquement, destiné à analyser la popularité des mots clés de recherche Google au fil du temps et selon l'origine géographique, permettant ainsi d'évaluer le rôle de certains facteurs tels que le pays, la région et les sujets associés (Giordano *et al.*, 2019; Zhao *et al.*, 2018; Bates, 2017) pour des études épidémiologiques.

1.2.6 Les difficultés et les controverses

Les difficultés et les controverses liées à l'épidémiologie numérique peuvent se regrouper en deux catégories, à savoir : les difficultés et controverses inhérentes à cette nouvelle pratique de recherche et celles liées à leurs implications sociopolitiques. Les controverses et les difficultés inhérentes à la pratique de l'épidémiologie numérique incluent les disparités entre les usagers d'Internet (Salathé *et al.*, 2012; Bakker *et al.*, 2016), les ressources disponibles selon la situation

⁵ Selon l'OMS, « les maladies à transmission vectorielle sont des maladies humaines provoquées par des parasites, des virus ou des bactéries transmis par des vecteurs » (OMS, 2020).

des pays (Lippi et Cervellin, 2019), les différences entre les informations publiées sur le web et la réalité (Salathé *et al.*, 2013; Towers *et al.*, 2015; Génois et Barrat, 2018), la désinformation (Tourassi *et al.*, 2016), les faux positifs et faux négatifs au niveau statistique (Hill-Cawthorn et Sorrell, 2016; Hohle, 2017; Sato *et al.*, 2021), le manque de mécanismes de validation et de recherche (Mohanty *et al.*, 2016 ; Makri, 2017; Lazzari *et al.*, 2018), le manque de données sur l'emplacement des usagers d'Internet (Gruebner *et al.*, 2018) ainsi que la complexité de l'expertise requise par la pratique de l'épidémiologie numérique. (Sau, 2017).

De plus, il y a les difficultés à comprendre ce que signifient réellement certains concepts du domaine informatique souvent utilisés dans le contexte de la pratique de l'épidémiologie numérique. Par exemple, l'algorithme constitue un des concepts fondamentaux étudiés en informatique alors qu'il n'y a pas de consensus sur sa définition dans la littérature (Valdi, 2012; Hill, 2016; Angius *et al.*, 2024). Selon Angius *et al.* (2024), un algorithme est un ensemble d'instructions permettant l'accomplissement d'une tâche donnée. Les règles que l'on suit pour calculer les opérations arithmétiques de base telles que la multiplication ou la division seraient des exemples quotidiens d'algorithmes (Angius *et al.*, 2024). Hill (2016) propose de définir un algorithme comme une structure de contrôle finie, abstraite, efficace et composée, impérativement donnée, qui accomplit un objectif précis dans des conditions définies. Ces difficultés peuvent amener des défis pour bien saisir les phénomènes inhérents à la pratique de l'épidémiologie numérique.

La littérature indique que GFT a été l'objet de certaines controverses engendrées par des incertitudes sur la validité de ses prédictions de la grippe (Pollett *et al.*, 2017b; Salathé, 2018; Duclos, 2019). Comme mentionné précédemment, GFT a cessé d'être utilisé en 2015 (Duclos, 2019; Guo *et al.*, 2021). Par contre, d'autres recherches plus récentes ont démontré le potentiel de GFT par l'utilisation de méthodologies améliorées au niveau des modélisations (Clemente *et al.*, 2019) et la mise en place de *Weibo COVID-19 Trends*, un nouvel algorithme de prédiction des cas confirmés améliorant celui de GFT (Guo *et al.*, 2021). Certaines recherches ont aussi évalué la fiabilité de *Google Trends* pour l'épidémiologie numérique. Cervellin *et al.* (2017) sont arrivés à la conclusion que *Google Trends* est peu fiable pour les travaux en épidémiologie numérique, car celui-ci serait plus influencé par les bruits médiatiques que par le fardeau réel de la maladie.

D'autres chercheurs partagent les mêmes conclusions (Knipe *et al.*, 2021), mais soulignent le potentiel de *Google Trends* pour étudier les réactions du public vis-à-vis des épidémies (Adawi *et al.*, 2017).

En ce qui concerne les difficultés et les controverses liées aux implications sociopolitiques de cette nouvelle pratique de recherche, les chercheurs utilisant l'épidémiologie numérique seraient confrontés à des enjeux éthiques et de confidentialité des données (Gopichandran et Elango, 2018; Loi, 2021) liés à des risques d'atteinte à la vie privée des individus lors du partage de données personnelles notamment (Kostkova, 2018; Bour *et al.*, 2021; Chen et Wang, 2021). Plusieurs travaux de recherche ont soulevé ces enjeux, incluant Smolinski *et al.* (2017) dans leur étude sur la surveillance participative des maladies, ainsi que Denecke (2017) qui a proposé un modèle d'évaluation éthique des technologies de détection numérique des maladies. Mittelstadt *et al.* (2018) questionnent quant à eux l'obligation morale de la participation aux travaux de l'épidémiologie numérique, un questionnement lié à l'équilibre entre intérêts publics et individuels. Enfin, la notion de phénotype numérique a été développée par Loi (2019) pour conceptualiser les données numériques comme le prolongement du phénotype de l'humain, c'est-à-dire comme information numérique produite par l'humain affectant les comportements et la culture.

La pertinence de tels travaux de recherche en éthique dans le contexte de l'utilisation de données numériques est difficilement contestable, surtout lorsque les enjeux politiques autour des données peuvent faciliter une concentration de pouvoir par certains acteurs comme Google, Facebook et Twitter. Des chercheurs en sciences sociales tels que Ruppert *et al.* (2017) ont fourni un aperçu des enjeux politiques autour des données, qu'ils considèrent comme un champ de pouvoir et de connaissance traversant les mondes virtuels et réels. Ces mondes produisent des sujets dont les droits deviennent des objets de luttes (Ruppert *et al.*, 2017).

Pour Zwitter (2014), les aspects éthiques relatifs à l'utilisation de données massives en épidémiologie numérique renvoient à la difficulté de déterminer la propriété des données, la question de la réutilisation des données dans divers contextes ainsi que les problèmes liés aux réseaux de pouvoirs entre les collecteurs, utilisateurs et générateurs de données. Il semble que ce contexte spécifique aux données massives et à la mise en données exige de nouvelles façons de

penser l'éthique, comme le proposent Ruppert *et al.* (2015) avec la notion d'éthique de la sollicitude (*ethic of care*), afin de développer des modèles de propriété sociale des données axés sur le partage et la collaboration et considérant les données comme une ressource collective plutôt que privée.

1.2.7 La synthèse de la revue de la littérature

La méthodologie employée pour la revue de la littérature comprenait la lecture complète des 66 articles repérés dans les bases de données PubMed et Springer, portant spécifiquement sur l'épidémiologie numérique, complétée par l'analyse quantitative des 535 articles identifiés dans Scopus pour dresser un portrait des recherches relevant de l'épidémiologie numérique plus généralement.

L'analyse quantitative réalisée à partir de la base de données Scopus entre 2010 et 2022 indique que les recherches sur l'épidémiologie numérique sont récentes avec un sommet de 190 publications en 2021 et que celles-ci sont publiées principalement par des revues en médecine (42%) et en informatique (12%). Le Canada figure en cinquième position parmi les pays ayant publié le plus grand nombre d'articles avec 33 publications, une quantité non négligeable.

L'analyse des articles portant spécifiquement sur l'épidémiologie numérique a révélé qu'il existe plusieurs façons de définir la pratique de l'épidémiologie numérique. Certaines définitions mettent l'emphase sur l'utilisation de données en provenance de l'extérieur des systèmes de santé pour faire de l'épidémiologie, sur la possibilité de faire des analyses de données en temps réel ou quasi réel, sur l'importance des technologies informatiques, des données numériques et de l'intelligence artificielle dans le cadre de cette nouvelle pratique.

Au niveau des technologies, il ressort de la littérature l'utilisation de données et technologies numériques de même que les technologies d'intelligence artificielle. Au niveau des principaux acteurs repérables dans la littérature, ceux-ci comprennent des chercheurs de diverses disciplines comme l'épidémiologie et l'informatique, des organismes internationaux, des non-experts

(incluant des communautés de citoyens lors des recherches fondées sur la science citoyenne notamment) ainsi que de grandes entités privées comme celle de Google et Twitter.

Toujours selon la littérature analysée, les difficultés et les controverses relatives à la pratique de l'épidémiologie numérique incluent des enjeux de nature sociopolitique, qui sont liés à la présence de disparités importantes entre les usagers d'Internet, à des questions éthiques relatives à la confidentialité des données et à des préoccupations autour de la possibilité d'une concentration du pouvoir par des acteurs privés comme Google, Facebook et Twitter. La pratique de l'épidémiologie numérique soulève aussi des difficultés et controverses au niveau scientifique, qui sont liées notamment à l'existence de faux positifs et faux négatifs au niveau statistique, au manque de mécanismes de validation des connaissances générées ainsi qu'à un manque de connaissances et d'expertise sur cette nouvelle pratique.

1.3 Le but, les objectifs et les questions de recherche dans le cadre de la thèse

La littérature nous informe peu sur les enjeux épistémiques et sociopolitiques à l'égard de l'utilisation des données numériques dans le cadre d'études épidémiologiques. Il n'est donc pas évident de comprendre les enjeux épistémiques et sociopolitiques de la pratique de l'épidémiologie numérique en s'appuyant simplement sur les articles publiés dans la littérature – qui n'inclue par ailleurs pas de revues spécialisées sur cette nouvelle pratique. Or, le Canada a publié un nombre de travaux relativement élevé en lien avec la pratique de l'épidémiologie numérique. Soulignons aussi l'existence d'un laboratoire canadien dédié à l'épidémiologie numérique (Katapally, 2023) et l'offre au Canada, à McGill University, de programmes d'études supérieures conjoints avec l'Université de Bordeaux en France, menant à l'obtention d'un diplôme en science des données de santé publique (MGU, 2023), ce qui pourrait marquer un début d'institutionnalisation de l'épidémiologie numérique.

Il devient alors important de se pencher sur cette nouvelle pratique susceptible de prendre de l'ampleur au cours des prochaines années. Une recherche sur les enjeux épistémiques et sociopolitiques de la pratique de l'épidémiologie numérique, du point de vue de chercheurs en épidémiologie, peut nous aider à cerner cette nouvelle pratique en émergence et à en saisir les

principaux enjeux. En effet, ces chercheurs connaissent leur discipline et ils ont une expérience approfondie des analyses de données.

Par conséquent, **le but de cette recherche** est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. Les analyses seront guidées par la question principale ci-dessous :

La question principale : Quels sont les enjeux épistémiques et sociopolitiques caractérisant le contexte d'émergence de la pratique de l'épidémiologie numérique ?

À la lumière de ces propos, notre recherche poursuit les **objectifs** suivants :

- Comprendre ce que signifie la pratique de l'épidémiologie numérique.
- Identifier les enjeux épistémiques et sociopolitiques qui accompagnent l'émergence de la pratique de l'épidémiologie numérique.

Ces objectifs de recherche se traduisent en trois sous-questions de recherche. Chacune de ces sous-questions inclut d'autres sous-questions de recherche afin de préciser les analyses.

1. Comment la pratique de l'épidémiologie numérique est-elle perçue par les chercheurs en épidémiologie ?

1.1 Que signifie la pratique de l'épidémiologie numérique ?

1.2 Comment la pratique de l'épidémiologie numérique est-elle diffusée dans la communauté scientifique ?

2. Quels sont les enjeux épistémiques de la pratique de l'épidémiologie numérique ?

2.1 Quelle est la crédibilité de la pratique de l'épidémiologie numérique ?

2.2 Quelles sont les significations associées aux mesures et aux données en épidémiologie traditionnelle et dans le cadre de la pratique de l'épidémiologie numérique ?

2.3 Quel est le niveau de confiance envers les dispositifs méthodologiques utilisés en épidémiologie numérique ?

3. Quels sont les enjeux sociopolitiques de la pratique de l'épidémiologie numérique ?

3.1 Quels sont les enjeux d'expertise en lien avec la pratique de l'épidémiologie numérique ?

3.2 Comment la légitimité de la pratique de l'épidémiologie numérique est-elle envisagée ?

1.4 La conclusion sur la problématique

En somme, cette recherche porte sur l'émergence de la pratique de l'épidémiologie numérique. Elle s'intéresse à cette nouvelle pratique qui s'appuie sur l'analyse de données de sources non traditionnelles comme des données générées à partir des médias sociaux, des moteurs de recherche Internet et de téléphones cellulaires, pour mener des recherches épidémiologiques. Notons que les big data figurent parmi les enjeux de cette nouvelle pratique, mais ce ne sont pas les seuls puisqu'une telle pratique utilise aussi d'autres types de données comme des données d'imagerie (Mohanty *et al.*, 2016) et celle-ci peut combiner des données de source traditionnelle avec des données de source non-traditionnelle lors des recherches (Roche *et al.*, 2017). Pour cette raison, les big data constituent un aspect parmi d'autres dans le cadre de cette thèse qui s'interroge sur l'ensemble de la pratique de l'épidémiologie numérique et non sur un type de données ni sur la façon dont les big data en particulier peuvent aider lors des recherches épidémiologiques. La pratique de l'épidémiologie numérique est un phénomène complexe qui implique de s'intéresser à la discipline de l'épidémiologie, aux pratiques de recherche, aux changements dans les pratiques de recherche et au phénomène de la mise en données de la société.

Comme nous l'avons expliqué précédemment, l'analyse quantitative des articles publiés dans la base de données Scopus indique que le nombre de publications liées à la pratique de l'épidémiologie numérique a augmenté de manière significative depuis les dernières années. Il s'avère donc pertinent d'en savoir davantage sur cette nouvelle pratique qui suscite un intérêt grandissant au sein des recherches en épidémiologie, et cela, dans plusieurs pays dont le Canada figurant en cinquième position parmi les pays ayant publié le plus grand nombre d'articles avec 33 publications (selon notre requête à partir de Scopus datée de 2022). Nous observons par ailleurs certains signes précurseurs d'une institutionnalisation de la pratique de l'épidémiologie numérique au Canada, à travers la création d'un laboratoire dédié à l'épidémiologie numérique et à la mise sur pieds de programmes d'études supérieures en science des données de santé publique. Il devient alors

important de se pencher de manière critique sur une telle pratique émergente, qui pourrait se développer dans les prochaines années.

Une recherche sur les enjeux épistémiques et sociopolitiques de la pratique de l'épidémiologie numérique à travers les points de vue de chercheurs en épidémiologie devrait nous aider à cerner cette nouvelle pratique, puisque ces acteurs connaissent leur discipline et qu'ils ont une expérience approfondie des analyses de données. Nous allons donc investiguer les enjeux épistémiques et sociopolitiques liés à l'émergence de la pratique de l'épidémiologie numérique en analysant les points de vue de chercheurs canadiens en épidémiologie. Selon les constats tirés de notre revue de la littérature, les publications sur la pratique de l'épidémiologie numérique en expliquent peu les enjeux sur les plans épistémique et sociopolitique. De plus, la littérature n'explique pas comment cette nouvelle pratique se diffuse au sein de la discipline de l'épidémiologie. Par conséquent, le but de cette recherche est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. Nous voulons en connaître les dimensions sociales, politiques, scientifiques et technologiques. Cette recherche vise aussi à comprendre ce que signifie la pratique de l'épidémiologie numérique afin de saisir en quoi cette nouvelle pratique est différente de l'épidémiologie traditionnelle et comment elle se diffuse au sein de la discipline de l'épidémiologie.

Nous adoptons une telle approche, car nous supposons que la pratique de l'épidémiologie numérique implique des enjeux épistémiques. En effet, la véracité des connaissances occupe une place centrale dans les domaines scientifiques et les façons d'examiner les phénomènes étudiés peuvent être influencées par les méthodes et les données utilisées. Nous envisageons aussi que la pratique de l'épidémiologie numérique implique des enjeux sociopolitiques parce que nous avançons l'idée que les méthodes de recherche et les données ne peuvent pas être réduites à de simples opérations et entités techniques. Au contraire, nous pensons que celles-ci peuvent avoir un impact important sur les manières de travailler en recherche, sur la vie des individus et sur la concentration du pouvoir chez les grands détenteurs de données. Une telle approche amène à formuler la question principale de la thèse, à savoir : Quels sont les enjeux épistémiques et sociopolitiques caractérisant le contexte d'émergence de la pratique de l'épidémiologie numérique ? À partir de la question principale et de ces objectifs, trois questions de recherche sont formulées

pour guider les analyses, à savoir : 1) Comment la pratique de l'épidémiologie numérique est-elle perçue par les chercheurs en épidémiologie ? 2) Quels en sont les enjeux épistémiques? Et 3) Quels en sont les enjeux sociopolitiques ?

Le prochain chapitre de la thèse (Chapitre II) présentera le cadre théorique utilisé pour encadrer les analyses. Ce cadre théorique inclut la coproduction des connaissances (Jasanoff, 2004), les cultures épistémiques (Knorr-Cetina, 1999) ainsi que les enjeux sociétaux des méthodes de recherche et des données (Savage, 2013; Law *et al.*, 2011; Law, 2010; Lury *et al.*, 2020; Law et Ruppert, 2013; Ruppert *et al.*, 2017; Lupton, 2015). Le Chapitre III expliquera la méthodologie utilisée pour mettre en œuvre notre recherche, c'est-à-dire une méthodologie qualitative fondée sur des entretiens individuels semi-dirigés auprès de chercheurs canadiens en épidémiologie et une analyse thématique inspirée par Paillé et Mucchielli (2016). Le Chapitre IV explicitera la pratique de l'épidémiologie numérique en effectuant une analyse de ses caractéristiques et de sa diffusion au sein des chercheurs en épidémiologie. Le chapitre V analysera les enjeux épistémiques de la pratique de l'épidémiologie numérique en se concentrant sur la signification des données et la confiance envers les données numériques comme dispositif méthodologique (les données générées par les moteurs de recherche Internet, les médias sociaux et les téléphones cellulaires). Le Chapitre VI fournira une analyse des enjeux sociopolitiques de la pratique de l'épidémiologie numérique en mettant l'accent sur les enjeux d'expertise et la légitimité de cette nouvelle pratique. Le Chapitre VII est la discussion de la thèse qui porte sur les caractéristiques et les enjeux épistémiques, l'écosystème et les contextes de la coproduction des connaissances de la pratique de l'épidémiologie numérique.

CHAPITRE II

LE CADRE THÉORIQUE

Ce chapitre présentera les concepts théoriques qui guideront l'analyse des enjeux relatifs à la pratique de l'épidémiologie numérique, à travers les perspectives mises de l'avant par des chercheurs en épidémiologie. Rappelons que le but de notre recherche est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. Elle nous permettra de connaître les dimensions sociales, politiques, scientifiques et technologiques de cette pratique. Notre recherche vise aussi à comprendre ce que signifie la pratique de l'épidémiologie numérique afin de saisir en quoi elle est différente de l'épidémiologie traditionnelle et comment celle-ci est diffusée dans la discipline de l'épidémiologie.

Dans une perspective STS, le cadre théorique de cette thèse se situe entre le déterminisme technologique et naturel et le déterminisme social puisque nous soutenons que la pratique de l'épidémiologie numérique implique des aspects technologiques, scientifiques, conceptuels et sociaux. Il s'agit d'une perspective alignée de manière générale avec le concept de la coproduction des connaissances défini par Jasanoff (2004) :

Coproductivist accounts, conceived in this way, avoid the charges of both natural and social determinism that have featured in recent academic debates around the field of science and technology studies, including the infamous "science wars" of the 1990s (Sokal and Bricmont 1998; Koertge 1998; Gross and Levitt 1994). Science, in the coproductivist framework, is understood as neither a simple reflection of the truth about nature nor an epiphenomenon of social and political interests. Rather, coproduction is symmetrical in that it calls attention to the social dimensions of cognitive commitments and understandings, while at the same time underscoring the epistemic and material correlates of social formations. (Jasanoff, 2004, p. 3)

Dans le cadre de la coproduction, les connaissances scientifiques de même que les technologies ne sont pas un miroir transcendant de la réalité (Jasanoff, 2004). Elles incorporent et sont aussi incorporées par les pratiques sociales, les identités, les normes, les conventions, les discours, les instruments et les institutions (Jasanoff, 2004).

Toujours selon Jasanoff (2004), la science n'est ni un simple reflet de la vérité sur la nature ni un épiphénomène des intérêts sociaux et politiques. Au contraire, la coproduction implique une vision symétrique dans la mesure où elle attire l'attention sur les dimensions sociales des engagements et les compréhensions cognitives, tout en soulignant les corrélations d'ordre épistémique et matériel des formations sociales. L'idée clé derrière cette notion de symétrie est qu'il faut accorder de l'attention, dans nos recherches, autant aux aspects épistémiques -reliés à la connaissance - qu'aux aspects sociopolitiques, pour bien comprendre comment la connaissance se crée. La coproduction considère aussi les préoccupations normatives de la théorie politique et de la philosophie morale. Elle révèle aussi des dimensions en termes d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales qui constituent les sciences et technologies (Jasanoff, 2004). Nous soutenons que des investigations sur les multiples dimensions sociales, politiques, scientifiques et technologiques des phénomènes seront utiles pour comprendre les différentes facettes des perspectives des chercheurs en épidémiologie envers la pratique de l'épidémiologie numérique. En effet, nous nous attendons à ce que leurs préoccupations concernent les aspects épistémiques des pratiques en épidémiologie, mais aussi des aspects d'ordre social car le chercheur en épidémiologie se situe aussi dans un milieu social impliquant une culture, des valeurs et des normes.

Jasanoff (2004) souligne que la coproduction ne doit pas être présentée comme une théorie à part entière, prétendant à la cohérence et au pouvoir prédictif. Il s'agit surtout d'une manière d'interpréter des phénomènes complexes, afin de rendre compte de la complexité de ces phénomènes. Donc nous utiliserons la coproduction des connaissances comme principe général en STS pour cadrer théoriquement notre recherche.

Par ailleurs, notre cadrage théorique doit aussi rendre compte de manière plus spécifique des enjeux d'ordre épistémique et sociopolitique en ce qui a trait à la pratique de l'épidémiologie numérique. Les enjeux d'ordre épistémique sont compris comme les perspectives des chercheurs vis-à-vis des différentes façons dont une science se fait (donc le travail quotidien de recherche). Les enjeux sociopolitiques se réfèrent aux perspectives des chercheurs quant aux aspects non-scientifiques d'une pratique de recherche tels que les enjeux sociétaux, éthiques et politiques liés à une pratique de recherche.

En effet, ces chercheurs auront certainement des questionnements d'ordre épistémique sur la pratique de l'épidémiologie numérique du fait qu'ils pratiquent la science de l'épidémiologie, mais ils seront aussi préoccupés par des réalités d'ordre sociopolitique. Nous considérons les chercheurs et les pratiques scientifiques comme étant des entités non-isolées, c'est-à-dire faisant partie de la société sous forme de divers assemblages technologiques, scientifiques, épistémiques, cognitifs et sociopolitiques. Il sera alors important de comprendre comment les chercheurs en épidémiologie voient l'émergence de la pratique de l'épidémiologie numérique dans le cadre d'études épidémiologiques, en fonction de leurs perspectives sur les questions épistémiques et sociopolitiques.

Pour ce faire, les perspectives théoriques que nous avons jugées les plus pertinentes pour notre recherche sont les cultures épistémiques développées par Knorr-Cetina (1999) ainsi que les enjeux sociétaux des méthodes de recherche et des données (Savage, 2013; Law *et al.*, 2011; Law, 2010; Lury *et al.*, 2020; Law et Ruppert, 2013; Ruppert *et al.*, 2017; Lupton, 2015). Les cultures épistémiques (Knorr-Cetina, 1999) permettront d'expliquer les considérations d'ordre épistémique des chercheurs en épidémiologie envers la pratique de l'épidémiologie numérique. En effet, la connaissance, les objectifs et les pratiques liés à la vérité sont des éléments centraux des domaines scientifiques (Knorr-Cetina et Reichmann, 2015). De plus, comme toute pratique de recherche, l'épidémiologie numérique est grandement influencée par la nature des technologies, des instruments, des mesures, des données et des équipes de recherche. S'appuyant sur des études de laboratoires, les travaux de Knorr-Cetina (1999) conceptualisent les interactions entre les chercheurs et les instruments ainsi que les enjeux par rapport aux mesures et aux données. Ces travaux sont donc utiles pour comprendre les enjeux d'ordre épistémique soulevés par des chercheurs en épidémiologie envers la pratique de l'épidémiologie numérique.

Le cadre théorique de cette thèse comprend également les enjeux sociétaux des méthodes de recherche et des données (Savage, 2013; Law *et al.*, 2011; Law, 2010; Lury *et al.*, 2020; Law et Ruppert, 2013; Ruppert *et al.*, 2017; Lupton, 2015) afin de mettre en évidence les perspectives d'ordre sociopolitique de ces chercheurs. Les méthodes et les données utilisées dans le cadre de la pratique de l'épidémiologie numérique ne sont pas des moyens purement techniques (Savage, 2013; Law, 2010; Law *et al.*, 2011; Lupton, 2015), car elles soulèvent d'importantes questions

d'ordre sociopolitique. On peut citer les enjeux éthiques par rapport à l'utilisation de données numériques dans les recherches, ainsi que les rapports de pouvoir liés à la propriété de ces données (Ruppert *et al.*, 2017; Zwitter, 2014). Les chercheurs en épidémiologie étant en santé publique et régis par le bien public, il y a de fortes chances que les enjeux sociopolitiques des méthodes de recherche et des données constituent des préoccupations importantes pour eux.

Donc, les analyses dans le cadre de cette thèse seront inspirées par la perspective de la coproduction des connaissances (Jasanoff, 2004). Elles seront guidées par les cultures épistémiques développées par Knorr-Cetina (1999) et les enjeux sociétaux des méthodes de recherche et des données (Savage, 2013; Law *et al.*, 2011; Law, 2010; Lury *et al.*, 2020; Law et Ruppert, 2013; Ruppert *et al.*, 2017; Lupton, 2015). Les sections suivantes présenteront les cultures épistémiques, puis les enjeux sociétaux des méthodes de recherche et des données.

2.1 Les cultures épistémiques

Selon Knorr-Cetina (1999), les processus et les systèmes de connaissances suscitent des questions sur leurs fonctionnements et sur les principes qui sous-tendent leurs orientations et leurs procédures (Knorr-Cetina, 1999). Knorr-Cetina (1999) explore les caractéristiques épistémiques telles que la signification de l'empirique, la mise en œuvre des relations d'objet, la construction et le façonnement des arrangements sociaux au sein de la science (Knorr-Cetina, 1999). Les cultures épistémiques mettent l'emphase sur deux problèmes : l'un concerne le manque de compréhension des mécanismes contemporains de production de la connaissance, de leur profondeur et surtout de leur diversité ; l'autre concerne la transition des sociétés contemporaines vers des sociétés de la connaissance, dont les cultures épistémiques sont une caractéristique structurelle (Knorr-Cetina, 1999).

Knorr-Cetina (1999) conçoit les cultures épistémiques comme des cultures de création et de justification des connaissances. Elle les considère aussi comme des amalgames de dispositions de mécanismes et d'éléments liés par affinité, nécessité et coïncidence historique qui, dans un domaine donné, constituent la manière dont nous connaissons ce que nous connaissons. Knorr-Cetina (1999) définit les cultures épistémiques comme suit :

Those amalgams of arrangements and mechanisms-bonded through affinity, necessity, and historical coincidence - which, in a given field, make up how we know what we know. Epistemic cultures are cultures that create and warrant knowledge, and the premier knowledge institution throughout the world is, still, science. (Knorr-Cetina, 1999, p. 1)

Le terme « épistémique » est significatif. Il indique que la connaissance, les objectifs et les pratiques liés à la vérité sont des éléments centraux des domaines scientifiques (Knorr-Cetina et Reichmann, 2015). La notion de culture vise le niveau de la pratique, plutôt que le niveau des symboles et des significations, que les étudiants en culture ont généralement tendance à examiner (Knorr-Cetina et Reichmann, 2015). Selon cette notion de culture, l'épidémiologie numérique pourrait être conceptualisée comme une nouvelle façon de faire de l'épidémiologie susceptible d'amener d'autres manières de créer des connaissances sur les maladies et les conditions de santé des populations. Les concepts rattachés aux cultures épistémiques les plus pertinents pour cette recherche sont la culture et la diversité dans les sciences, les signes et les apparences secondaires, les mesures et les données expérimentales ainsi que les pratiques organisationnelles dans les sciences. Après avoir expliqué ces concepts, nous identifierons les apports et les limites des cultures épistémiques.

2.1.1 La culture et la diversité dans les sciences

Knorr-Cetina (1999) souligne que le concept de culture est généralement utilisé pour faire référence à des formes historiques spécifiques, telles que l'État (« culture nationale »), l'économie (« culture de marché ») et l'entreprise commerciale (« culture organisationnelle »). De plus, la culture serait très probablement une caractéristique permanente de l'évolution de toute vie humaine (Knorr-Cetina, 1999). Cependant, il semble tout aussi probable qu'il n'existe pas de culture humaine universelle, caractérisant tous les êtres humains de manière générique, en tout temps et en tout lieu (Knorr-Cetina, 1999). Toujours selon Knorr-Cetina (1999), les spécificités culturelles naissent et se développent lorsque des domaines de la vie sociale sont séparés les uns des autres.

Dans les situations reliées aux sciences et aux systèmes experts, Knorr-Cetina (1999) considère la culture comme des modèles agrégés et des dynamiques qui se manifestent dans la pratique de l'expertise et qui varient dans différents contextes d'expertise. Donc il semble que cette perspective renvoie la culture à la pratique. L'accent mis sur la pratique fait descendre le niveau de l'analyse culturelle jusqu'au domaine des régularités matérielles, mais sans perdre de vue les régularités symboliques et la façon dont elles sont associées au matériel (Knorr-Cetina et Reichmann, 2015). Par exemple, les physiciens des hautes énergies perçoivent leur détecteur comme un être vivant et cette façon de concevoir le détecteur influence leur travail lors des expériences, selon Knorr-Cetina (1999). Cette emphase sur la pratique et sur les régularités matérielles pour étudier les sciences amènerait le chercheur à étudier les scientifiques au travail plutôt que d'analyser l'histoire des idées, la structure des théories scientifiques ou les cadres institutionnels de la science (Knorr-Cetina, 1999).

La science et les systèmes experts constitueraient des candidats évidents pour la division culturelle (Knorr-Cetina, 1999). Ils sont composés de groupes de spécialistes séparés des autres experts par des frontières institutionnelles profondément ancrées à tous les niveaux de l'éducation et dans les domaines d'expertise (Knorr-Cetina, 1999). En effet, Knorr-Cetina (1999) souligne que la notion de discipline et ses dérivés sont importants pour définir les principes d'organisation qui répartissent la science et la technologie en sous-unités et sous-sous-unités. Cependant Knorr-Cetina (1999) fait remarquer que ces concepts se sont avérés moins utiles pour saisir les stratégies et les politiques d'apprentissage qui ne sont pas codifiées dans des manuels, mais qui avisent pourtant la pratique des experts. Pour comprendre cette dimension tacite dans les sciences, Knorr-Cetina (1999) affirme qu'il faut agrandir l'espace de la connaissance et de l'action, plutôt que de simplement observer les disciplines ou les spécialités comme des structures d'organisation.

C'est ce qui incite Knorr-Cetina (1999) à remplacer les notions de discipline par celle de culture épistémique. Cette approche vise à mettre en évidence les machineries des savoirs des sciences contemporaines afin d'expliquer les dimensions techniques, sociales et symboliques des systèmes experts complexes. D'après Knorr-Cetina (1999), l'étude des machineries de savoirs révèle la fragmentation des sciences contemporaines.

Ce genre d'études peut alors permettre de comprendre la diversité des cultures épistémiques (Knorr-Cetina, 1999). Contrairement à certains travaux en philosophie des sciences soutenant l'unité de la science, Knorr-Cetina (1999) refuse la thèse affirmant qu'il n'existe qu'un seul type de connaissance, qu'une seule science et qu'une seule méthode scientifique. Knorr-Cetina (1999) avance l'idée que la science peut se faire de différentes façons et au sein de nouvelles formes d'institutionnalisation interconnectées avec les systèmes et processus de production des connaissances. Par exemple, la science citoyenne peut être considérée comme une approche différente pour faire des recherches puisque des citoyens non scientifiques sont impliqués dans les projets de recherche (Benyei *et al.*, 2021; Dias Da Silva *et al.*, 2017). Ils apportent ainsi de la diversité dans les façons de faire de la science.

Ces perspectives sur la notion de culture et la diversité dans les sciences, telles que développées par Knorr-Cetina (1999) seront utiles pour comprendre la pratique de l'épidémiologie numérique. Conceptualiser la pratique de l'épidémiologie numérique en tant que culture épistémique pourra aider à saisir la diversité dans les façons de faire de l'épidémiologie. Cette approche pourra peut-être ainsi illustrer différents types de connaissance et diverses méthodes en épidémiologie. La pratique de l'épidémiologie numérique peut engendrer des ruptures avec les façons courantes de faire de l'épidémiologie. Il est possible qu'elle permette aussi de construire des connaissances en impliquant une multitude de cadres instrumentaux, linguistiques, théoriques et organisationnels.

2.1.2 Les signes et les apparences secondaires

Knorr-Cetina (1999) a analysé le monde des signes et des apparences secondaires au sein d'un laboratoire de physique des hautes énergies (Knorr-Cetina, 1999). Ce monde serait particulièrement marqué par l'incorporation des signes générés par des détecteurs lors des expérimentations. Les particules ne sont pas directement accessibles à travers les sens humains puisque celles-ci sont trop petites pour être observées autrement qu'indirectement par des détecteurs. Elles sont trop rapides pour être capturées et contenues dans un espace de laboratoire, trop dangereuses en tant que faisceaux de particules pour être manipulées directement (Knorr-Cetina, 1999). C'est pourquoi les physiciens manipulent des représentations du détecteur de

particules plutôt que les particules elles-mêmes. Ils travaillent sur des manipulations « hors ligne » des signaux extraits des détecteurs une fois les données prises (Knorr-Cetina, 1999). Les représentations reconstruisent les événements dans le détecteur et façonnent lentement ces signaux dans une forme permettant d'identifier les particules ayant un intérêt pour les physiciens. À partir de ces représentations, les physiciens créent des « variables » n'étant plus interprétées en termes de signes enregistrés dans les matériaux du détecteur, mais sont conçues et analysées en termes de distributions et de modèles de la physique (Knorr-Cetina, 1999).

Knorr-Cetina (1999) tente aussi d'illustrer les complications qui surviennent lors des expérimentations. En effet, les signes marquant des événements intéressants peuvent être étouffés et brouillés par les signes d'autres événements du détecteur. Dans ces expériences, l'univers des signes et des traces est recouvert par un univers de simulations et de distorsions des signes et des traces. Ces effets se réfèrent à l'arrière-plan, à l'événement sous-jacent, au bruit (du détecteur et de l'électronique) et à l'étalement des distributions. Elles proviennent de parties inintéressantes d'événements, d'autres classes d'événements ou de l'appareil lui-même (Knorr-Cetina, 1999). Toujours selon Knorr-Cetina (1999), tous ces phénomènes constituent une menace pour la capacité des scientifiques à reconnaître les événements intéressants et ils peuvent falsifier la signature des événements, déformer leur caractère, ou mettre en péril leur identification (Knorr-Cetina, 1999). De plus, les physiciens doivent également faire face à une grande variété de bruits : les signaux aléatoires, imprévisibles et indésirables dans un détecteur et dans l'électronique de l'appareil qui masquent l'information souhaitée (Knorr-Cetina, 1999).

Ces perspectives développées par Knorr-Cetina (1999) sur le monde des signes et des apparences secondaires pourront expliquer certaines perspectives mises de l'avant par des chercheurs en ce qui a trait aux phénomènes observables à travers les analyses de données numériques. En effet, il est possible que certains chercheurs soient incertains vis-à-vis des signes et des apparences visibles à travers les analyses de données numériques. Comme le mentionne Knorr-Cetina (1999) pour le cas de la physique des hautes énergies, il y a peut-être des bruits tels que les signaux aléatoires, imprévisibles et indésirables qui surviennent avec l'exploitation des données numériques dans le cadre d'études épidémiologiques. Ces bruits constitueraient une menace pour la capacité des chercheurs en épidémiologie à reconnaître les événements intéressants sous un angle

épidémiologique. De tels bruits peuvent falsifier la signature des événements, déformer leur caractère, ou mettre en péril leur identification (Knorr-Cetina, 1999). Cette recherche devra alors orienter ses efforts pour bien comprendre la nature et les sources des bruits liés à l'exploitation de données numériques.

2.1.3 Les mesures et les données expérimentales

Knorr-Cetina (1999) affirme que pour de nombreux domaines, les mesures comptent comme des preuves si celles-ci sont correctement effectuées et enregistrées par les expérimentateurs. Les mesures sont considérées comme capables de prouver ou de réfuter des théories, de discerner de nouveaux phénomènes, ainsi que de représenter des résultats relativement intéressants et potentiellement publiables (Knorr-Cetina, 1999). Knorr-Cetina (1999) souligne aussi que les mesures se font assigner un rôle puissant dans la validation des connaissances.

Cependant, les mesures sont parfois sujettes à interprétation, elles sont chargées de théories et en conséquence, elles sont aussi susceptibles de soulever des critiques (Knorr-Cetina, 1999). Il semble que certains chercheurs en physique des hautes énergies pensent que même que les données purement expérimentales ne signifient rien en elles-mêmes (Knorr-Cetina, 1999). La recherche de Knorr-Cetina (1999) indique aussi que des physiciens auraient souligné que non seulement il y a peu de quantités qui peuvent être mesurées directement, mais aussi que celles qui ont pu être prises directement ne doivent pas être enregistrées telles quelles. Ces participants croient que les résultats expérimentaux dépendent d'une configuration particulière du détecteur et des critères utilisés pour extraire l'information du détecteur. Selon ces participants, un autre détecteur qui utilisera un ensemble différent de critères donnera d'autres mesures (Knorr-Cetina, 1999).

Donc il est possible d'observer deux perspectives vis-à-vis des mesures, selon Knorr-Cetina (1999). La première perspective est caractérisée par des attentes élevées envers les mesures. Ces dernières sont vues comme indispensables pour valider des connaissances et capables de prouver ou de réfuter des théories. Elles permettent aussi d'entrevoir de nouveaux phénomènes, de représenter des résultats relativement intéressants et potentiellement publiables (Knorr-Cetina, 1999). La deuxième perspective apporte des nuances en soulignant que les mesures sont parfois sujettes à

interprétation et que celles-ci peuvent varier selon l'appareil et les critères utilisés pour en extraire l'information (Knorr-Cetina, 1999).

En ce qui concerne les données, Knorr-Cetina (1999) met l'accent sur l'idée que les données sont contingentes à l'appareil de mesure et que celles-ci sont des représentations de cet appareil plutôt que du monde réel. Knorr-Cetina (1999) affirme que le détecteur agit comme un dispositif médiateur interposé entre l'expérience et les phénomènes. L'ensemble de l'appareil de mesure incluant le signal obtenu du détecteur peut et doit lui-même être correctement compris par le chercheur (Knorr-Cetina, 1999). Lors des travaux scientifiques, les chercheurs se préoccupent de l'expérience elle-même, de l'observation, du contrôle, de l'amélioration et de la compréhension de ses composantes et processus (Knorr-Cetina, 1999). Selon Knorr-Cetina (1999), la compréhension fait donc référence à une approche globale de l'apprentissage de ce qui se passe dans chaque partie pertinente de la matière, de ce qui se passe dans le temps et de la raison pour laquelle ces choses se produisent. En fait, selon les recherches de Knorr-Cetina (1999), toutes les données sont considérées comme des mélanges de composantes incluant le signal, le fond, le bruit et l'événement sous-jacent. Ces composantes sont déformées et tronquées par les pertes et les étalements des détecteurs.

Les idées développées par Knorr-Cetina (1999) sur les façons de conceptualiser les mesures et les données expérimentales pourront être utiles pour expliquer les façons de concevoir la pratique de l'épidémiologie numérique et de comprendre certains questionnements d'ordre épistémique des chercheurs en épidémiologie envers cette nouvelle pratique. Il se peut que certains chercheurs voient les données traditionnelles, les données numériques ainsi que la mesure des données de manière différente en fonction de leurs pratiques courantes de recherche et de leur expérience en matière d'épidémiologie numérique. Il sera alors intéressant de considérer les données comme des mélanges de composantes qui incluent notamment le signal, le fond, le bruit et l'événement sous-jacent ainsi suggéré par Knorr-Cetina (1999) et d'observer comment ces mélanges de composantes sont perçus au sein de la communauté de chercheurs étudiée.

2.1.4 Les pratiques organisationnelles dans les sciences

En ce qui concerne les pratiques organisationnelles dans les sciences, Knorr-Cetina (1999) insiste sur l'influence des objets des expériences et des technologies dans la formation des groupes d'acteurs lors de la naissance d'une nouvelle collaboration. Les travaux de recherche de Knorr-Cetina (1999) ont fait ressortir d'une part, une nouvelle configuration des acteurs fondée sur la gestion orientée envers l'objet des expériences et d'autre part, l'influence des technologies dans la formation des groupes d'acteurs lors de la naissance d'une nouvelle collaboration.

Selon Knorr-Cetina (1999), la gestion orientée envers l'objet des expériences constitue une nouvelle configuration des acteurs qui s'ajoute aux structures communautaires et collectives déjà présentes dans le domaine de la physique des hautes énergies. Knorr-Cetina (1999) définit la gestion orientée envers l'objet des expériences comme une gestion fondée sur le contenu des recherches plutôt que sur des moyens bureaucratiques ou hiérarchiques. L'idée du management par le contenu peut être saisie par deux principes selon Knorr-Cetina (1999). Selon le premier principe, il maintient la proximité des participants avec les objets ou avec la substance du travail scientifique. Pour le deuxième principe, il substitue, dans la mesure du possible, des structures orientées objet aux structures d'autorité sociale.

Knorr-Cetina (1999) soutient aussi que la notion de gestion orientée envers l'objet des expériences vise à mettre en évidence la richesse des processus et des relations entre les experts, les formats organisationnels et le monde des objets lors des travaux de recherche. De plus, la gestion orientée envers l'objet des expériences impliquerait des systèmes mixtes de soins et d'attention se développant autour d'entités épistémiques dans lesquelles la gouvernance organisationnelle fonctionne principalement par la manipulation du contenu du problème plutôt que seulement par les personnes et les structures (Knorr-Cetina, 1999).

En termes d'organisation sociale, les lignes de commandement verticales entre les personnes sont remplacées par des liens horizontaux entre les scientifiques, ou les groupes de scientifiques, et les objets (Knorr-Cetina, 1999). Ces liens s'établissent lors d'une pratique organisationnelle fondée sur la gestion orientée envers l'objet des expériences. Toujours selon Knorr-Cetina (1999), les expériences sont construites à partir de tels circuits horizontaux, où circule l'essentiel de la

compétence. Dans le cadre de cette pratique organisationnelle, l'information n'a pas besoin d'être recueillie et traitée par une hiérarchie de contrôle centralisé. Au lieu de cela, Knorr-Cetina (1999) affirme que l'information réside, et reste, dans l'environnement immédiat des objets techniques, où elle est transportée par les scientifiques qui travaillent avec ces objets.

Selon Knorr-Cetina (1999), la naissance d'une nouvelle expérience fait partie de l'émergence d'une nouvelle collaboration avec une nouvelle génération d'instruments adaptés aux nouveaux régimes d'énergie (plus élevés). Cette étape de la naissance rassemble des groupes d'origine différente avec des investissements technologiques différents (Knorr-Cetina, 1999). Knorr-Cetina (1999) affirme aussi que la taille croissante des expériences oblige chaque génération d'expériences à rassembler de nombreux nouveaux membres, à forger une coalition avec des instituts et des physiciens extérieurs à l'expérience de départ. Les groupes qui participeront à une collaboration seront sélectionnés une fois la technologie choisie en fonction de la physique à explorer (Knorr-Cetina, 1999).

Les idées développées par Knorr-Cetina (1999) sur les pratiques organisationnelles pourront expliquer certains résultats de cette recherche quant aux enjeux sociopolitiques de la pratique de l'épidémiologie numérique. En effet, cette pratique peut impliquer des pratiques organisationnelles différentes par rapport à celles de l'épidémiologie traditionnelle, dont l'interdisciplinarité. Il sera alors intéressant de comprendre comment les chercheurs épidémiologistes anticipent les possibles reconfigurations du travail scientifique en épidémiologie, comme les nouveaux arrangements et relations entre les acteurs impliqués dans la pratique de l'épidémiologie numérique ? Ces perspectives sur les pratiques organisationnelles développées par Knorr-Cetina (1999) aideront à comprendre les liens qui peuvent se former entre les acteurs d'une part, et la place des technologies dans les stratégies des acteurs d'autre part.

2.1.5 Les critiques, apports et limites des cultures épistémiques

Les cultures épistémiques sont bien reconnues en études des sciences, en sociologie des connaissances scientifiques et en théorie sociale (Knorr-Cetina et Reichmann, 2015). Plusieurs critiques de l'ouvrage de Knorr-Cetina (1999) affirment que les cultures épistémiques offrent au

lecteur un regard comparatif et précieux sur les cultures épistémiques (Cutcliffe, 2001; Pinch, 2000). Selon Pinch (2000), il s'agit d'une étude sophistiquée qui, grâce à son optique comparative, fournit au lecteur des informations utiles sur deux domaines scientifiques (la physique des hautes énergies et la biologie moléculaire). Ces domaines seraient difficilement discernables s'ils étaient examinés séparément. Toujours selon Pinch (2000), l'étude de Knorr-Cetina (1999) contribue à la compréhension globale des multiples façons de pratiquer la science. Elle suggère en outre une étroite relation entre les techniques, les instruments technologiques et la création de connaissances scientifiques dans le travail de laboratoire contemporain (Pinch, 2000).

Selon l'article de Knorr-Cetina et Reichmann (2015), les cultures épistémiques ont contribué à plusieurs études qui peuvent se regrouper en cinq catégories. Premièrement, les cultures épistémiques ont été utilisées pour étudier en détail les caractéristiques épistémiques d'autres spécialités et de domaines de connaissance à l'intérieur et à l'extérieur des contextes universitaires. Deuxièmement, certaines recherches ont élaboré davantage la notion de cultures épistémiques et l'ont développée. Troisièmement, la notion a été utilisée pour expliquer les difficultés de la recherche interdisciplinaire. Quatrièmement, la recherche sur les organisations a utilisé le concept pour comprendre et optimiser les organisations complexes. Cinquièmement, il existe une catégorie hétérogène d'études portant sur les caractéristiques des domaines de connaissance qui peuvent être comprises en termes de notion de cultures épistémiques (Knorr-Cetina et Reichmann, 2015).

Par exemple, Cole (2013) cité dans Knorr-Cetina et Reichmann (2015) a étudié la culture épistémique des sciences médico-légales en la qualifiant de « culture médico-légale ». Dans le cadre de cette étude, le domaine des sciences médico-légales est compris en termes de cultures épistémiques. Un autre exemple de recherche utilisant la notion de cultures épistémiques serait Wagner et Newell (2004) cités dans Knorr-Cetina et Reichmann (2015). Cette recherche explique les diverses utilisations de progiciels standardisés dans les universités par les différences entre la culture épistémique des membres de l'administration de rang inférieur, de la direction supérieure de l'université (le niveau du président ou du recteur de l'université et du vice-chancelier) et des chercheurs.

En ce qui concerne l'interdisciplinarité, Knorr-Cetina et Reichmann (2015) affirment que les cultures épistémiques s'avèrent utiles pour souligner les spécificités des différentes cultures scientifiques et des pratiques de recherche. Certaines recherches sur l'interdisciplinarité comme celle de Kastenhofer (2007) cité dans Knorr-Cetina et Reichmann (2015) montrent comment la biotechnologie est née d'un processus de convergence entre plusieurs domaines aux cultures épistémiques très différentes. Selon Knorr-Cetina et Reichmann (2015), Kastenhofer affirme que de tels processus de convergence peuvent encourager la coopération entre les disciplines, l'intégration des groupes de recherche et des organisations, et l'assimilation des cultures épistémiques.

Cependant, la notion de cultures épistémiques comporte des limites. En citant Albert *et al.* (2008), Knorr-Cetina et Reichmann (2015) soulignent que certains pensent les cultures épistémiques comme étant en quelque sorte statiques, car elles ne permettent pas de comprendre les relations de pouvoir entre les groupes de scientifiques. Selon cette perspective, des cultures épistémiques différentes sont susceptibles de dévaloriser les autres cultures épistémiques, surtout lorsqu'elles ne communiquent pas directement entre elles (Albert *et al.*, 2008 cités dans Knorr-Cetina et Reichmann, 2015).

De plus, Markovsky (2000) pense que l'analyse empirique qui est à la base de cette étude est subjective et qu'elle aurait dû être considérée comme un exercice exploratoire de génération d'idées. Markovsky (2000) évoque le problème de la surgénéralisation. Plus l'analyse est nuancée et détaillée, moins elle est applicable au-delà des cas uniques. Un autre problème soulevé par Markovsky (2000) concerne la désunion de la science démontrée par Knorr-Cetina (1999) en expliquant que les disciplines impliquent parfois des façons de faire différentes d'un chercheur à l'autre. Markovsky (2000) n'est pas d'accord avec l'affirmation selon laquelle il s'agit d'une preuve de la désunion de la science. Il soutient plutôt que le tissu de la science est unifié à un niveau plus abstrait, ses pièces étant tissées ensemble non pas par les activités concrètes des chercheurs individuels, mais par la logique sous-jacente des méthodes théoriques et empiriques qu'ils emploient collectivement.

Pour notre part, nous soutenons que les cultures épistémiques développées par Knorr-Cetina (1999) sont très pertinentes pour étudier les enjeux épistémiques des pratiques de recherche. Il est possible que les cultures épistémiques développées par Knorr-Cetina (1999) comportent des limites, mais ce travail fait ressortir des enjeux épistémiques très précis et utiles pour saisir ce qui peut se passer lors des activités de recherche. Par exemple, Knorr-Cetina (1999) fournit une bonne illustration des complications survenant lors des expérimentations en soulignant entre autres que des signes marquant des événements intéressants peuvent être étouffés et brouillés par des bruits en provenance des instruments eux-mêmes. Ces bruits constitueraient une menace pour la capacité des scientifiques à reconnaître les événements intéressants. Ils peuvent aussi falsifier la signature des événements, déformer leur caractère, ou mettre en péril leur identification (Knorr-Cetina, 1999). Ces perspectives peuvent servir de guide pour analyser des problèmes dans le cadre de nouvelles pratiques de recherche.

De plus, notre recherche s'appuie sur une méthodologie qualitative qui vise à comprendre le sens que les personnes donnent à un phénomène donné (Savoie-Zajc, 2000). Ce phénomène est celui de la pratique de l'épidémiologie numérique dans le cadre précis de recherches épidémiologiques. Nous ne cherchons pas à généraliser nos observations, comme expliqué précédemment par Markovsky (2000). En outre, la méthodologie de Knorr-Cetina (1999) étant aussi qualitative, il y a de fortes chances que cette recherche n'avait pas pour but de généraliser, mais de comprendre les perspectives individuelles, les tensions et contradictions en lien avec des pratiques de recherche (Savoie-Zajc, 2000).

Dans le cadre de notre recherche, les cultures épistémiques seront utiles pour comprendre les spécificités de la culture épistémique de l'épidémiologie numérique et des pratiques de recherche en épidémiologie. Les cultures épistémiques pourront aussi servir à mettre en évidence l'étroite relation entre les techniques, les instruments technologiques et la création de connaissances scientifiques dans la pratique de l'épidémiologie numérique.

En somme, Knorr-Cetina (1999) voit les cultures épistémiques comme des cultures de création et de justification des connaissances. Elle les conçoit aussi comme des amalgames de dispositions de mécanismes et d'éléments liés par affinité, nécessité et coïncidence historique qui, dans un domaine

donné, constituent la manière dont nous connaissons ce que nous connaissons. Les cultures épistémiques développées par Knorr-Cetina (1999) sont très pertinentes pour étudier de manière critique les enjeux épistémiques des pratiques de recherche comme celle de l'épidémiologie numérique. Cette notion peut faire ressortir des enjeux épistémiques très pertinents pour saisir ce qui peut se passer dans les activités de recherche. Dans le cadre de notre recherche, les concepts qui se rattachent le mieux aux cultures épistémiques sont la notion de culture et la diversité dans les sciences, le monde des signes et des apparences secondaires, les façons de conceptualiser les mesures et les données expérimentales ainsi que les pratiques organisationnelles dans les sciences. À l'exception des idées développées sur les pratiques organisationnelles dans les sciences, les cultures épistémiques s'avèrent plutôt limitées pour expliquer les enjeux sociétaux et politiques des méthodes et des données utilisées lors des recherches.

2.2 Les enjeux sociétaux des méthodes de recherche et des données

Le cadre théorique de cette thèse comprend aussi des travaux sur les enjeux sociétaux des méthodes de recherche et des données. Le but étant de rendre compte des dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales constituant les sciences et technologies tel que proposé par Jasanoff (2004). Pour ce faire, nous abordons la vie sociale des méthodes, les dispositifs méthodologiques et les enjeux politiques des données.

2.2.1 La vie sociale des méthodes

La vie sociale des méthodes est un agenda de recherche intéressé à ce que les méthodes font et comment elles sont façonnées par le monde social (CRSCC, 2022). Selon le *Centre for Research on Socio-Cultural Change* (CRSCC), ces deux questions ne sont pas seulement d'ordre technique. Elles impliquent aussi des enjeux politiques, car les méthodes ne sont pas une boîte à outils neutre que l'on peut simplement prendre et laisser tomber. L'efficacité et la valeur des méthodes varient au fil des années. C'est pourquoi il est important de reconnaître que les méthodes de recherche sont des acteurs essentiels du monde social. Elles ne se contentent pas de décrire la société, mais contribuent à la recréer (CRSCC, 2022). Selon cette perspective, il faut toujours situer une méthode par rapport à des questions plus larges (CRSCC, 2022). De telles questions englobent les enjeux

éthiques ainsi que l'interdisciplinarité, incluant les nouvelles opportunités qui émergent avec des formes différentes de collaboration et avec l'application et la combinaison de différentes méthodes. On peut citer aussi les questions soulevées par les expériences de participation illustrées notamment par l'essor de la science citoyenne, des projets de source ouverte et la remise en cause de certaines formes de connaissance experte (Lury *et al.*, 2020).

Au lieu de considérer les méthodes comme quelque chose d'ennuyeux ou encore comme une formation obligatoire que les nouvelles cohortes de chercheurs en sciences sociales doivent suivre, Savage (2013) soutient qu'elles constituent des objets d'études. Les sites de réseaux sociaux, les processus d'audit, les dispositifs de transparence, les algorithmes pour les transactions financières, les enquêtes, les cartes, les entretiens, les bases de données et les classifications peuvent être conceptualisés comme des modes d'instanciation des relations sociales et identifiés comme des objets de recherche intéressants (Savage, 2013).

Toujours selon Savage (2013), les méthodes de recherche sociale constituent des objets d'analyse critique et politiques parce qu'elles ont une vie sociale. Cette caractéristique des méthodes de recherche sociale incite à faire en sorte que le choix de la méthode n'est pas seulement une question purement technique ou instrumentale (Savage, 2013). Le choix d'investir dans des dispositifs méthodologiques (allant des essais cliniques aux politiques basées sur des preuves) est aussi fondé sur des enjeux économiques, sociaux et politiques (Savage : 2013). Ainsi, les méthodes deviennent à la fois l'objet de l'intérêt des sciences sociales et des véhicules indispensables aux chercheurs pour mener à bien leurs travaux de recherche (Savage, 2013).

La vie sociale des méthodes soutient aussi que les méthodes ont un but et des défenseurs et qu'elles s'appuient sur ce qui les a précédées (Law *et al.*, 2011; Law, 2010). Law *et al.* (2011) illustrent cet argument avec l'exemple des recensements produits et utilisés par des gouvernements afin de connaître leurs populations. Donc la méthode du recensement a un but et un contexte social. Le recensement a survécu comme méthode de recherche, car il a ses défenseurs, c'est-à-dire des gouvernements considérant le recensement comme étant efficace pour identifier et localiser leurs populations.

Une autre illustration développée par Law *et al.* (2011) est celle des méthodes d'analyses de données transactionnelles numériques préconisées par une variété d'individus et d'organisations appartenant à divers groupes dont l'État, le secteur privé ou le monde universitaire. Ces méthodes identifient des modèles et des associations constituant des intérêts et des objectifs pour ceux qui les développent et/ou les utilisent tels que la sécurité publique, le marketing ciblé, le contrôle des maladies et la génomique des populations (Law *et al.*, 2011).

Selon Law *et al.* (2011), ces préoccupations illustrent l'idée selon laquelle les méthodes sont du domaine social. Premièrement, la vie sociale des méthodes montre comment les méthodes de recherche incarnent les préoccupations de leurs défenseurs et subsistent dans des contextes ou des environnements particuliers. Deuxièmement, la vie sociale des méthodes illustre l'importance d'une réflexion critique sur les méthodes, sur ce que font les méthodes et sur le statut des données qu'elles produisent. Et puis, troisièmement, cela indique que les méthodes ne reflètent pas nécessairement les préoccupations de l'académie. Elles sont constituées, et souvent à grande échelle, par d'autres défenseurs ayant divers intérêts et objectifs (Law *et al.*, 2011).

Les perspectives apportées par la vie sociale des méthodes sont reconnues par plusieurs travaux de recherche comme ceux du *Centre for Interdisciplinary Methodologies* (Lury *et al.*, 2020). Selon Lury *et al.* (2020), l'intérêt pour la vie sociale des méthodes a été fondamental par rapport à l'élaboration des méthodes et cet agenda de recherche a fourni un contexte pour la création du Centre.

D'autres chercheurs ont aussi utilisé certaines idées et concepts proposés par la vie sociale des méthodes dans leurs travaux. Par exemple, Law *et al.* (2011) affirment que les méthodes constituent un triple mouvement composé par des chercheurs, des réalités (décrites et réalisées par les enquêtes) et des institutions. Cette perspective a été appliquée à la culture en expliquant comment les éléments clés de la politique culturelle ont des " vies sociales " très spécifiques et importantes (O'Brien et Lockley, 2015). Plus récemment en 2022, la vie sociale des méthodes continue à être utilisée dans des recherches comme celle de Cruz (2022), étudiant la vie sociale des données biomédicales. Cette recherche montre comment un même point de données peut, dans un contexte donné, démontrer la qualité des soins et, dans un autre contexte, l'occulter (Cruz, 2022). En

s'appuyant sur ce constat, cette recherche affirme que les données biomédicales n'ont pas une seule vie sociale, mais plusieurs vies sociales différentes selon le contexte du phénomène étudié (Cruz, 2022).

Les propos de Lury *et al.* (2020) ont cependant souligné que la notion de performativité proposée par John Law (Law *et al.*, 2011; Law, 2016) pourrait être remplacée par la notion de participation aux recherches. En effet, Lury *et al.* (2020) affirment préférer ne pas utiliser la terminologie de la performance (qui peut impliquer selon elle un public séparé de ce qui est mis en scène) et considérer la participation qui inclurait tous les acteurs impliqués lors des recherches sans qu'il y ait une séparation entre ces acteurs.

Toujours selon cette chercheuse, cette préférence serait liée à ce qu'elle désigne comme la mise en plateforme de l'infrastructure épistémique. L'un des traits distinctifs de l'infrastructure épistémique contemporaine est la manière dont l'auditoire participe à l'action des méthodes. La nature de cette participation peut être très diverse puisqu'il peut s'agir notamment de surveillance ou d'une science citoyenne (Lury *et al.*, 2020). La notion de participation serait alors plus appropriée que celle de performance selon cette perspective. Malgré cette critique, les chercheurs ayant produit ce genre de travaux ont reconnu l'idée que les méthodes ne sont pas seulement des boîtes à outils et qu'elles impliquent aussi des enjeux politiques importants à considérer lors des choix relatifs aux méthodes de recherche.

La vie sociale des méthodes sera utile pour notre recherche pour comprendre la légitimité de la pratique de l'épidémiologie numérique analysée sous ses aspects sociopolitiques. Nous considérons les enjeux liés aux méthodes utilisées dans la pratique de l'épidémiologie numérique comme dépassant les questions purement techniques, vu qu'elles ne sont pas une boîte à outils neutre que l'on peut simplement prendre et laisser tomber. Nous affirmons, à la suite de Lury *et al.* (2020), que les méthodes impliquées lors de la pratique de l'épidémiologie numérique peuvent être vues comme des objets épistémiques. Il est alors pertinent d'étudier ces méthodes dans les discours des chercheurs, de même que dans leurs manières de les utiliser et par rapport à des questions plus larges comme les enjeux éthiques, l'interdisciplinarité et autres questions liées à l'émergence de formes diversifiées de participations aux recherches. Nous envisageons aussi que les méthodes

liées à la pratique de l'épidémiologie numérique ont un but et des défenseurs et qu'elles s'appuient sur ce qui les a précédées (Law *et al.*, 2011; Law, 2010).

2.2.2 Les dispositifs méthodologiques

La vie sociale des méthodes est une réponse à l'importance croissante des dispositifs méthodologiques (Savage, 2013; Law et Ruppert, 2013; Beer, 2016). Selon la littérature, les dispositifs méthodologiques désignent les objets matériels et les idées immatérielles s'associant pour apporter des moyens de mener des recherches sociales (Lupton, 2015). Par exemple, les données générées par les médias sociaux comme Twitter deviennent des dispositifs méthodologiques lorsque les données qu'ils génèrent sont utilisées pour faire des recherches. Les méthodes de recherche ne sont pas seulement des artefacts socio-culturels, mais elles contribuent à comprendre et à influencer profondément les phénomènes étudiés (Lupton, 2015; Law et Ruppert, 2013). Dans ce sens, un dispositif méthodologique constitue un assemblage d'artefacts matériels, d'utilisateurs humains, de pratiques, d'idées et d'espaces constamment sujet à des changements (Lupton, 2015). De tels dispositifs ne sont pas seulement des méthodes de recherche, mais aussi des objets d'analyse (Lupton, 2015) pertinents pour des recherches sociales puisqu'ils influencent les façons de percevoir les phénomènes étudiés et les phénomènes eux-mêmes (Lupton, 2015).

D'après Lupton (2015), l'objectif des recherches sur les dispositifs méthodologiques (par exemple, les données de Twitter) n'est pas simplement de savoir s'ils sont appropriés, précis ou éthiques à des fins diverses, mais plutôt de connaître leurs potentialités, leurs capacités et leurs limites. Ces recherches visent à connaître la manière dont les dispositifs méthodologiques configurent les objets qu'ils tentent d'étudier et de mesurer ainsi la façon dont ils servent des objectifs politiques. C'est dans ce sens où les méthodes de recherche ont une double vie sociale. D'une part, elles sont des assemblages d'artefacts matériels, d'utilisateurs humains, de pratiques et d'idées apportant des moyens de mener des recherches. D'autre part, elles créent des assemblages en configurant les objets étudiés et les mesures et en servant des objectifs politiques (Savage, 2013; Lupton, 2015). Par conséquent, les méthodes de recherche et les dispositifs méthodologiques n'impliquent pas seulement des enjeux purement techniques, mais aussi des enjeux épistémiques et sociopolitiques.

Law et Ruppert (2013) soulignent que les dispositifs méthodologiques amènent des choses qui ne sont pas toujours prévues et évidentes. En l'occurrence, les conséquences de ce que font ou peuvent faire ces choses constituent souvent l'objet de contestations sociales et politiques (Law et Ruppert, 2013). Ainsi, toute analyse sociale des dispositifs méthodologiques sera marquée par un certain niveau de susceptibilité critique (Law et Ruppert, 2013). Par exemple, les caméras de sécurité peuvent porter atteinte à la vie privée, et en conséquence de ce risque, ces dispositifs sont susceptibles d'entraîner des contestations sociales ou politiques sur leur utilisation. Ce genre de phénomène peut être désigné comme des réalités collatérales (Law, 2011 cité dans Law et Ruppert, 2013), c'est-à-dire des réalités dont les effets arrivent de manière accidentelle lors des recherches (Rubio et Baert, 2012) que nous considérons comme des conséquences individuelles et collectives de l'utilisation généralisée de dispositifs méthodologiques numériques pour des recherches.

Law et Ruppert (2013) affirment que les arrangements constituant les dispositifs méthodologiques sont matériellement hétérogènes. Les manières dont se jouent ces arrangements sont dans une large mesure indéterminées et contingentes puisque les dispositifs méthodologiques sont porteurs d'incertitude et façonnés par la poursuite d'objectifs variés (Law et Ruppert, 2013). Les dispositifs méthodologiques sont matériels vu qu'ils comprennent entre autres des technologies et des instruments (Law et Ruppert, 2013). Ils sont hétérogènes selon Law et Ruppert (2013), car ils incluent (généralement) aussi des personnes (techniciens, utilisateurs, membres d'une communauté), des textes, des inscriptions, des représentations ou des symboles (statistiques sur les échantillons, des graphiques).

Il semble que le concept de dispositif méthodologique soit fréquemment utilisé dans le cadre de travaux de recherche. Beer (2016) a utilisé ce concept afin d'analyser comment la mesure est liée au pouvoir, à la gouvernance et au contrôle. Dans le cadre de ce travail, la mesure est conceptualisée comme un dispositif méthodologique, c'est-à-dire un mélange de composantes matérielles (des instruments permettant de lire les mesures) et d'enjeux sociaux (les liens entre les mesures, le pouvoir, la gouvernance et le contrôle) impliquant des personnes (notamment les analystes des mesures, des membres d'une communauté), des textes, des inscriptions et des représentations. Dans son ouvrage sur la sociologie numérique, Lupton (2015) s'inspire de la vie sociale des méthodes pour produire un chapitre sur la théorisation de la société numérique. Dans ce chapitre, elle souligne

entre autres que le terme « dispositif méthodologique » est grandement utilisé dans la littérature et qu'il existe un mouvement en sociologie et dans d'autres théories sociales étudiant comment les méthodes de recherche sociale sont elles-mêmes des objets socialement configurés (Lupton, 2015).

Dans le cadre de cette thèse, les perspectives développées par Law et Ruppert (2013) sur les dispositifs méthodologiques serviront à comprendre certains enjeux d'ordre épistémique soulevés par les chercheurs en épidémiologie envers la pratique de l'épidémiologie numérique. Ces enjeux seront analysés en déterminant le niveau de confiance de ces chercheurs envers les dispositifs méthodologiques utilisés dans cette nouvelle pratique de recherche. La notion de dispositif méthodologique sera aussi utile pour saisir les conséquences individuelles et collectives anticipées par les chercheurs en épidémiologie sur le plan sociopolitique, dans le contexte de la pratique de l'épidémiologie numérique. Dans cette perspective, les dispositifs méthodologiques comme les sites de réseaux sociaux, les algorithmes et les bases de données peuvent être conceptualisés comme des modes d'instanciation des relations sociales et identifiés comme des objets de recherche intéressants (Savage, 2013; Ruppert *et al.*, 2013).

2.2.3 Les enjeux politiques des données

Le cadre théorique de cette thèse inclut aussi les enjeux politiques des données, car tel que soutenu par Ruppert *et al.* (2017), les données sont des entités capables de reconfigurer les relations entre les gouvernements, les citoyens et les industries. Au lieu de conceptualiser les données comme des représentations inertes, Ruppert *et al.* (2017) soutiennent que les données ont des pouvoirs concrets sur la vie politique et que ce sont des objets investis de certains pouvoirs, influences et rationalités. En d'autres termes, les enjeux politiques des données rendent compte non seulement des luttes politiques autour de la collecte et du déploiement des données, mais aussi des façons dont les données génèrent de nouvelles formes de relations de pouvoir et de politique à des échelles différentes et interconnectées (Ruppert *et al.*, 2017).

Les enjeux politiques des données s'intéressent à la manière dont les choses (infrastructures de serveurs, dispositifs et câbles), les langages (code, programmation et algorithmes) et les personnes (scientifiques, entrepreneurs, ingénieurs, informaticiens, concepteurs) peuvent ensemble créer de

nouveaux mondes (Ruppert *et al.*, 2017). Ruppert *et al.* (2017) soulignent aussi que les données et plus précisément les données massives (big data) sont devenues un objet majeur d'investissements économiques, politiques et sociaux pour gouverner les gens. Ces développements seraient préoccupants par rapport aux enjeux liés à la surveillance, à la vie privée et à l'anonymat. Toujours selon Ruppert *et al.* (2017), les enjeux politiques des données font ressortir des menaces incluant le partage de données personnelles, la désinformation et les fausses annonces politiques. Un travail collectif est requis pour protéger les gens des dangers d'Internet et des manipulations, mais ultimement, il revient aux individus de changer leurs comportements pour se protéger des forces obscures d'Internet (Ruppert *et al.*, 2017). Les droits d'accès et d'usages des données font aussi partie des enjeux politiques des données parce que ceux qui possèdent, distribuent, vendent, accèdent, utilisent, s'approprient, modifient, partagent et réutilisent les données luttent pour revendiquer des droits sur ces données (Ruppert *et al.*, 2017).

Les perspectives développées par Ruppert *et al.* (2017) auraient inspiré plusieurs travaux de recherche sur les aspects politiques des données. Henman (2022), dans un article expliquant le passé, le présent et le futur des politiques sociales numériques, suggère des compétences à la fois conceptuelles et méthodologiques permettant aux chercheurs en politique sociale de mieux étudier les technologies numériques. En citant notamment Ruppert *et al.* (2017), Henman (2022) souligne qu'il est important de prendre conscience que les données numériques et les algorithmes ne sont pas ontologiquement, éthiquement, politiquement ou socialement neutres.

Beraldo et Milan (2019) proposent de passer des enjeux politiques des données (Ruppert *et al.*, 2017) à une « politique contestataire des données ». En particulier, Beraldo et Milan (2019) observent comment les pratiques émergentes d'activisme des données articulent des politiques en adoptant une position critique à l'égard de la collecte massive de données. Ces chercheurs soutiennent qu'un « champ d'action partagée » complexe est en train d'émerger dans le domaine de la société civile. Ce champ serait composé d'initiatives vaguement connectées intervenant dans la politique des données d'un point de vue critique (Beraldo et Milan, 2019).

Dans le contexte de notre recherche, les enjeux politiques des données développés par Ruppert *et al.* (2017) seront utiles pour expliquer les enjeux sociopolitiques liés à la pratique de

l'épidémiologie numérique. Nous orienterons nos analyses vers la question de la légitimité de la pratique de l'épidémiologie numérique telle que l'envisagent les chercheurs en épidémiologie, en nous inspirant des enjeux politiques des données proposés par Ruppert *et al.* (2017). Nous allons voir que les enjeux politiques des données peuvent engendrer des barrières à la légitimité de la pratique de l'épidémiologie numérique ainsi que des conséquences individuelles et collectives de l'usage généralisé d'une telle pratique de recherche.

En somme, le cadre théorique de cette thèse comprend des travaux sur les enjeux sociétaux des méthodes de recherche et des données afin de rendre compte des dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales qui constituent les sciences et technologies, tel que proposé par Jasanoff (2004). Pour ce faire, nous avons choisi d'utiliser la vie sociale des méthodes de recherche, le concept de dispositif méthodologique et les enjeux politiques des données.

La vie sociale des méthodes sera utile pour notre recherche en ce qui concerne la légitimité de la pratique de l'épidémiologie numérique analysée dans le chapitre sur les aspects sociopolitiques de cette nouvelle pratique. Nous soutenons que les méthodes appliquées lors de la pratique de l'épidémiologie numérique constituent des objets épistémiques dont les conséquences résultantes de leur utilisation vont au-delà du monde de la recherche impliquant entre autres des enjeux éthiques, l'interdisciplinarité et l'émergence de diverses formes de participations aux recherches.

Les dispositifs méthodologiques serviront à comprendre certains enjeux d'ordre épistémique liés à la pratique de l'épidémiologie numérique. Cette compréhension pourra se faire en analysant le niveau de confiance de ces chercheurs envers des dispositifs méthodologiques utilisés dans cette nouvelle pratique de recherche. Ces dispositifs seront aussi utiles pour saisir les conséquences individuelles et collectives anticipées par les chercheurs en épidémiologie, selon leurs préoccupations sociopolitiques à l'égard de la pratique de l'épidémiologie numérique. En effet, en tant que dispositifs méthodologiques, les sites de réseaux sociaux, les algorithmes et les bases de données impliquent des enjeux de société et constituent des objets de recherche en eux-mêmes.

Enfin, les analyses de la légitimité de la pratique de l'épidémiologie numérique seront également guidées par les enjeux politiques des données développés par Ruppert *et al.* (2017). Les enjeux politiques des données sont susceptibles de constituer des barrières à la légitimité et d'engendrer des conséquences individuelles et collectives dans le cadre de cette nouvelle pratique de recherche.

2.3 La conclusion sur le cadre théorique

À partir de la perspective STS sur la coproduction des connaissances développée par Sheila Jasanoff (2004), nous conceptualisons la pratique de l'épidémiologie numérique comme une culture épistémique (Knorr-Cetina, 1999) impliquant des méthodes et des données de recherche pouvant être analysées sous des angles épistémiques et sociopolitiques. Les analyses des angles épistémiques de cette nouvelle pratique seront donc guidées par les cultures épistémiques (Knorr-Cetina, 1999). Pour leur part, les angles sociopolitiques de cette nouvelle pratique seront analysés en se basant sur la vie sociale des méthodes (Savage, 2013; Law, 2010; Law *et al.*, 2011) et les enjeux politiques des données (Ruppert *et al.*, 2017).

Rappelons que notre question principale de recherche est la suivante : quels sont les enjeux épistémiques et sociopolitiques caractérisant le contexte d'émergence de la pratique de l'épidémiologie numérique ? Ce cadre théorique permettra d'apporter des réponses à cette question parce que le concept de cultures épistémiques (Knorr-Cetina, 1999) est pertinent pour expliquer les enjeux épistémiques de la pratique de l'épidémiologie numérique, de même que les enjeux sociétaux des méthodes de recherche et des données sont utiles pour comprendre les enjeux sociopolitiques de cette nouvelle pratique. Plus précisément, les cultures épistémiques (Knorr-Cetina, 1999) seront utiles pour analyser la crédibilité de la pratique de l'épidémiologie numérique, les significations associées aux mesures et aux données en épidémiologie traditionnelle et dans le cadre de la pratique de l'épidémiologie numérique ainsi que le niveau de confiance envers les dispositifs méthodologiques utilisés en épidémiologie numérique. La vie sociale des méthodes (Savage, 2013; Law, 2010; Law *et al.*, 2011) et les enjeux politiques des données (Ruppert *et al.*, 2017) seront pertinents pour analyser les enjeux d'expertise en lien avec la pratique de l'épidémiologie numérique et la légitimité d'une telle pratique.

Cette approche permettra de comprendre ce que signifie la pratique de l'épidémiologie numérique et de saisir les enjeux épistémiques et sociopolitiques de cette pratique émergente à travers les points de vue de chercheurs en épidémiologie. Pour chacune de ces approches théoriques, le Tableau 2.1 indique les auteurs principaux, les idées clés, les sources et les thèmes retenus pour la thèse.

Tableau 2. 1 Le sommaire des travaux théoriques retenus pour la thèse

Les travaux théoriques	Les auteurs principaux	Les idées clés	Sources(s)	Les thèmes clés retenus pour la thèse
La coproduction des connaissances	Sheila Jasanoff	<p>Les connaissances scientifiques et les technologies incorporent et sont aussi incorporées par les pratiques sociales, les identités, les normes, les conventions, les discours, les instruments et les institutions</p> <p>La coproduction est symétrique, car elle observe les dimensions sociales des engagements et les compréhensions cognitives, tout en soulignant les corrélations épistémiques et matérielles des</p>	Jasanoff (2004)	<p>La symétrie de la coproduction des connaissances</p> <p>Les dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales qui constituent les sciences et technologies</p>

Les travaux théoriques	Les auteurs principaux	Les idées clés	Sources(s)	Les thèmes clés retenus pour la thèse
		formations sociales		
Les cultures épistémiques	Karin Knorr-Cetina	Les cultures épistémiques sont des cultures et différents amalgames de création et de justification des connaissances qui constituent la manière dont nous connaissons ce que nous connaissons	Knorr-Cetina (1999)	<p>La notion de culture et la diversité dans les sciences</p> <p>Les signes et apparences secondaires</p> <p>Les conceptions des mesures et des données expérimentales</p> <p>Les pratiques organisationnelles dans les sciences</p>
Les enjeux sociétaux des méthodes de recherche et des données	Mike Savage John Law Evelyn Ruppert Celia Lury Evelyn Ruppert Engin Isin	<p>Les méthodes et les données ne se résument pas à des questions purement techniques</p> <p>Les méthodes et les données impliquent des enjeux sociopolitiques incluant des</p>	Savage (2013) Law (2010) Law <i>et al.</i> (2011) Law et Ruppert (2013) Lury <i>et al.</i> (2020) Ruppert <i>et al.</i> (2017)	<p>La vie sociale des méthodes de recherche</p> <p>Les dispositifs méthodologiques</p>

Les travaux théoriques	Les auteurs principaux	Les idées clés	Sources(s)	Les thèmes clés retenus pour la thèse
	Didier Bigo	conséquences individuelles et collectives Les enjeux sociopolitiques des données en tant que champ de pouvoir et de connaissance qui traverse les mondes virtuels et réels, et dont les droits deviennent des objets de lutte à travers ce champ		Les enjeux sociopolitiques des données

Le Tableau 2.2 présente la cartographie des approches théoriques et des thèmes retenus en fonction de chaque chapitre de la thèse. Le but de ce tableau est d'expliquer où ces approches théoriques et ces thèmes se situent dans le cadre de la thèse.

Tableau 2. 2 La cartographie des thèmes retenus en fonction des chapitres de la thèse

Les travaux théoriques	Les thèmes clés retenus pour la thèse	Chapitre sur la pratique de l'épidémiologie numérique	Chapitre sur les enjeux d'ordre épistémique	Chapitre sur les enjeux sociopolitiques
La coproduction	La symétrie de la coproduction des connaissances	X	X	X

Les travaux théoriques	Les thèmes clés retenus pour la thèse	Chapitre sur la pratique de l'épidémiologie numérique	Chapitre sur les enjeux d'ordre épistémique	Chapitre sur les enjeux sociopolitiques
des connaissances	Les dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales qui constituent les sciences et technologies			X
Les cultures épistémiques	La notion de culture et la diversité dans les sciences	X	X	
	Les signes et apparences secondaires	X	X	
	Les conceptions des mesures et des données expérimentales	X	X	
	Les pratiques organisationnelles dans les sciences			X
Les enjeux sociétaux des méthodes de recherche et des données	La vie sociale des méthodes de recherche	X	X	X
	Les dispositifs méthodologiques		X	X

Les travaux théoriques	Les thèmes clés retenus pour la thèse	Chapitre sur la pratique de l'épidémiologie numérique	Chapitre sur les enjeux d'ordre épistémique	Chapitre sur les enjeux sociopolitiques
	Les enjeux sociopolitiques des données			X

La thématique de la symétrie de la coproduction des connaissances traversera tous les chapitres. Elle guidera nos analyses des multiples dimensions sociales, politiques, scientifiques et technologiques de la pratique de l'épidémiologie numérique à travers les points de vue des chercheurs en épidémiologie. Nous nous attendons à ce que les enjeux soulevés portent sur les aspects épistémiques des pratiques en épidémiologie, mais aussi sur des questions d'ordre sociopolitique puisque le chercheur en épidémiologie se situe aussi dans un contexte social et politique impliquant une culture, des valeurs et des normes. Les dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales constituant les sciences et technologies seront utilisées dans le chapitre sur les enjeux d'ordre sociopolitique puisque le but de ce chapitre est de faire ressortir les enjeux sociopolitiques en lien avec la pratique de l'épidémiologie numérique.

Pour les cultures épistémiques, la notion de culture et la diversité dans les sciences seront utiles pour le chapitre sur la pratique de l'épidémiologie numérique et celui sur les enjeux d'ordre épistémique. Conceptualiser la pratique de l'épidémiologie numérique comme une culture épistémique aidera à saisir la diversité dans les façons de faire de l'épidémiologie et ainsi peut-être illustrer différents types de connaissance et de méthodes scientifiques en épidémiologie. Cette nouvelle pratique peut engendrer des ruptures avec les façons courantes de faire de l'épidémiologie et construire des connaissances en impliquant une multitude de cadres instrumentaux, linguistiques, théoriques et organisationnels.

Les signes et les apparences secondaires de même que les conceptions des mesures et des données expérimentales guideront certaines analyses du chapitre sur la pratique de l'épidémiologie numérique et celui portant sur les enjeux d'ordre épistémique. Les travaux de Knorr-Cetina (1999) font ressortir des enjeux épistémiques très précis permettant de saisir ce qui peut se passer lors des activités de recherche. Par exemple, Knorr-Cetina (1999) fournit une bonne illustration des complications survenant lors des expérimentations en soulignant entre autres que des signes marquant des événements intéressants peuvent être étouffés et brouillés par des bruits en provenance des instruments eux-mêmes. Ces bruits constituent une menace pour la capacité des scientifiques à reconnaître les événements intéressants et ils peuvent falsifier la signature des événements, déformer leur caractère, ou mettre en péril leur identification (Knorr-Cetina, 1999). Donc, les travaux de Knorr-Cetina peuvent servir de guide permettant de se rendre compte des problèmes qui peuvent apparaître dans le cadre de nouvelles pratiques de recherche comme celle de la pratique de l'épidémiologie numérique.

Les pratiques organisationnelles dans les sciences seront pertinentes pour analyser certains aspects du chapitre sur les enjeux sociopolitiques. La pratique de l'épidémiologie numérique peut impliquer des pratiques organisationnelles différentes par rapport à celles de l'épidémiologie traditionnelle, dont l'interdisciplinarité en recherche. Il sera alors intéressant de comprendre comment les chercheurs épidémiologistes prévoient les reconfigurations du travail scientifique en épidémiologie avec la pratique de l'épidémiologie numérique. Il sera aussi intéressant de connaître les arrangements et les relations entre les acteurs envisagés par les chercheurs épidémiologistes dans le cadre de la pratique de l'épidémiologie numérique.

En ce qui a trait aux enjeux sociétaux des méthodes de recherche et des données, rappelons que les thèmes clés retenus sont la vie sociale des méthodes de recherche, les dispositifs méthodologiques et les enjeux sociopolitiques des données. La vie sociale des méthodes de recherche guidera les analyses dans le chapitre sur la pratique de l'épidémiologie numérique à propos de l'influence de l'épidémiologie traditionnelle sur la pratique de l'épidémiologie numérique. Elle sera aussi utile dans le chapitre sur les enjeux épistémiques en ce qui concerne des similarités entre l'épidémiologie traditionnelle et la pratique de l'épidémiologie numérique. Finalement, nous l'évoquerons dans le chapitre sur les enjeux sociopolitiques, plus précisément sur la légitimité de la pratique de

l'épidémiologie numérique. Nous soutenons que les méthodes impliquées dans cette nouvelle pratique constituent des objets épistémiques impliquant des questions d'ordre sociétal comme les enjeux éthiques, l'interdisciplinarité et l'émergence de diverses formes de participations aux recherches.

Les travaux sur les dispositifs méthodologiques seront utiles à la fois dans le chapitre sur les enjeux d'ordre épistémique et dans celui sur les enjeux sociopolitiques. Pour l'analyse des aspects épistémiques, les dispositifs méthodologiques cadreront l'analyse des niveaux de confiance des chercheurs envers cette nouvelle pratique de recherche. Quant aux aspects sociopolitiques, la notion de dispositif méthodologique sera utile pour saisir les conséquences individuelles et collectives de la pratique généralisée de l'épidémiologie numérique anticipées par les chercheurs en épidémiologie.

Les enjeux sociopolitiques des données serviront de guide pour les analyses du chapitre sur les enjeux sociopolitiques. Ces enjeux concernent plus particulièrement la légitimité de la pratique de l'épidémiologie numérique, telle qu'elle est envisagée par les chercheurs en épidémiologie; lesquels incluent les enjeux politiques des données susceptibles de constituer des barrières à sa légitimité et d'amener des conséquences individuelles et collectives dans le contexte de cette nouvelle pratique.

Le chapitre suivant présentera la méthodologie mise en place afin de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique, selon des chercheurs canadiens en épidémiologie.

CHAPITRE III

LA MÉTHODOLOGIE

Comme expliqué précédemment dans la problématique, le but de cette recherche est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. Pour atteindre cet objectif, nous avons effectué une recherche de terrain auprès de 36 chercheurs canadiens en épidémiologie afin de collecter et d'analyser leurs points de vue sur cette nouvelle pratique. Ce chapitre présentera l'approche choisie pour aborder cette recherche de terrain ainsi que la méthode d'analyse des données obtenues.

3.1 La recherche de terrain

Pour mettre en œuvre notre recherche de terrain, nous avons choisi une méthodologie qualitative fondée sur des entretiens individuels semi-dirigés auprès de chercheurs canadiens en épidémiologie. Ainsi, les chercheurs canadiens en épidémiologie constituent les acteurs observés dans cette recherche. La discipline de l'épidémiologie est une composante importante dans le contexte de notre recherche, car la pratique de l'épidémiologie numérique est appliquée à cette discipline et celle-ci est une grande utilisatrice de données lors des recherches épidémiologiques. Il devient alors important de connaître les préoccupations de chercheurs en épidémiologie vis-à-vis des enjeux épistémiques et sociopolitiques de cette nouvelle manière de faire de la recherche en épidémiologie, axée sur les données. La compréhension de ces enjeux aidera à développer un regard critique vis-à-vis de la pratique de l'épidémiologie numérique qui pourra s'avérer utile pour la prise de décisions sur les méthodes utilisées dans les recherches en santé.

La recherche de terrain a été approuvée par le Comité d'éthique de la recherche pour les projets étudiants impliquant des êtres humains de l'UQAM le 24 janvier 2020. Le choix porté sur les entretiens semi-dirigés comme mode de collecte de données est fondé sur le principe selon lequel les entretiens visent à comprendre le sens que les personnes donnent à un phénomène donné (Savoie-Zajc, 2000), ici la pratique de l'épidémiologie numérique dans le cadre de recherches épidémiologiques. Notre recherche vise à comprendre à la fois les perspectives individuelles, les

tensions et contradictions ainsi que le monde de la recherche épidémiologique en lien avec l'émergence de la pratique de l'épidémiologie numérique. Notre objectif est d'accéder à la compréhension de comportements complexes et à la trame culturelle sous-jacente aux actions des acteurs (Savoie-Zajc, 2000).

Les entretiens ont visé l'étude des points de vue des chercheurs épidémiologistes sur les enjeux épistémiques et sociopolitiques de la pratique de l'épidémiologie numérique. Le guide d'entretien rédigé en français et en anglais inclut cinq sections, à savoir 1) les informations professionnelles (formation, projets, travail en cours) sur le participant, 2) les pratiques de recherche du participant, 3) les perceptions du participant sur l'épidémiologie numérique, 4) une discussion sur le cas pratique de *Google Flu Trends* (GFT) et de *FoodRepo*, 5) le futur de l'épidémiologie numérique. Nous avons choisi de discuter de GFT lors des entretiens même si cet outil n'est plus actif depuis 2015 (Duclos, 2019; Guo *et al.*, 2021) parce qu'il s'agit d'un cas déterminant ayant marqué les débuts de la pratique de l'épidémiologie numérique (Salathé, 2015) et qui continue à susciter de l'intérêt pour certaines recherches épidémiologiques sur la grippe (Clemente *et al.*, 2019; Guo *et al.*, 2021). Les perspectives apportées par des chercheurs en épidémiologie sur ce genre de dispositif peuvent certainement être utiles pour comprendre la pratique de l'épidémiologie numérique. L'annexe A contient le guide d'entretien en français et en anglais tel que présenté au Comité d'éthique de la recherche pour les projets étudiants impliquant des êtres humains de l'UQAM.

Le guide d'entretien a été légèrement modifié au cours de la période des entretiens afin d'adapter certaines questions aux participants ayant pratiqué l'épidémiologie numérique et à la suite d'un blocage dans les discussions sur GFT *et FoodRepo*. Les propos de la première moitié des participants étaient tous les mêmes vis-à-vis de GFT *et de FoodRepo*. Ces propos portaient principalement sur le potentiel de GFT pour la surveillance syndromique et les recherches exploratoires ainsi que sur les problèmes liés à la qualité des données et à la représentativité populationnelle. Quant à *FoodRepo*, les participants ne connaissaient pas ce dispositif. Leurs propos étaient axés sur l'importance de bien comprendre le fonctionnement de *FoodRepo* lors de son utilisation pour des études épidémiologiques. Nous avons alors décidé de discuter, avec la deuxième moitié des participants, de leur façon de voir la valeur ajoutée de la pratique de

l'épidémiologie numérique pour effectuer leurs études épidémiologiques les plus récentes. Aucune donnée ni aucune information sur les patients n'est requise pour ce projet, puisque la recherche s'intéresse aux perspectives professionnelles des chercheurs épidémiologistes sur cette nouvelle méthode scientifique qu'est l'épidémiologie numérique.

La communauté ciblée pour cette recherche de terrain sont des chercheurs canadiens en épidémiologie ayant intégré ou non la pratique de l'épidémiologie numérique dans leurs activités de recherches. En effet, cette recherche vise à connaître les perspectives de l'ensemble des chercheurs en épidémiologie à l'égard de cette nouvelle pratique. Le premier critère de sélection d'un participant pour cette recherche est donc d'être un chercheur en épidémiologie. Le choix s'est porté sur le Canada pour des raisons pratiques de ressources, au cas où les entretiens nécessiteraient des déplacements et en raison des langues parlées par la doctorante : le français et l'anglais. En outre, comme nous l'avons déjà expliqué, le Canada figure en cinquième position parmi les pays ayant publié le plus grand nombre d'articles sur cette nouvelle pratique avec un total de 33 publications selon notre requête à partir de Scopus datée de 2022.

Nous avons recruté des chercheurs canadiens en épidémiologie en utilisant les listes publiées sur les sites web de départements d'épidémiologie et de santé publique des universités canadiennes abritant une faculté de médecine ou un programme lié au domaine de la santé publique. La plupart des provinces du Canada ont des programmes d'études en épidémiologie au sein d'Écoles de santé publiques souvent attachées à une Faculté de médecine. Les cinq principales universités canadiennes (selon le nombre d'effectifs à temps plein aux études supérieures), c'est-à-dire University of Toronto, Université de Montréal, University of British Columbia, Université Laval et McGill University⁶ offrent toutes des programmes d'études aux cycles supérieurs en épidémiologie au niveau de la maîtrise et du doctorat (Universities Canada, 2023). L'Université d'Ottawa offre aussi des programmes d'études aux cycles supérieurs en épidémiologie au niveau de la maîtrise et du doctorat (Universities Canada, 2023). Lors de cette phase de recrutement, les chercheurs ciblés ont été contactés directement par la doctorante Renée Carpentier en utilisant le

⁶ Association of Atlantic Universities, Council of Ontario Universities, Individual institutions, and Bureau de coopération interuniversitaire (2022). *Enrolment by university*. Universities Canada. <https://www.univcan.ca/universities/facts-and-stats/enrolment-by-university/>. Consulté le 4 novembre 2023

courrier électronique de l'UQAM, en français pour les universités francophones et en anglais pour les universités anglophones, pendant les mois de février et mai 2020.

Les entretiens ont été conduits à distance entre février et juin 2020, certains par téléphone et d'autres par les médias *Zoom*, *Skype* ou *Hangouts* en raison du contexte sanitaire de la pandémie de la COVID-19 de 2020 et de la situation géographique des participants puisque plusieurs résident dans différentes provinces canadiennes allant de la Colombie-Britannique jusqu'en Nouvelle-Écosse. Les propos des participants ont été enregistrés au moyen d'une enregistreuse numérique pour les entretiens par téléphone et au moyen de l'enregistrement en ligne pour les entretiens conduits par *Zoom*, *Skype* et *Hangouts*. Les entretiens se déroulaient en anglais ou en français selon la préférence du chercheur participant. Le contenu de chaque entretien a été transcrit manuellement par la doctorante dans un document en format *Word* accessible uniquement par celle-ci.

Le Tableau 3.1 fournit le nombre de chercheurs contactés, le nombre de participants et le pourcentage de participation pour l'ensemble du Canada et pour chaque province canadienne. Au cours de cette recherche de terrain, 279 chercheurs canadiens en épidémiologie au sein de 20 institutions ont été contactés pour participer à un entretien individuel. Parmi ces chercheurs, un total de 36 a donné suite à notre demande d'entretien ce qui représente un pourcentage de participation global de 13%. Le plus grand nombre de chercheurs participants provient de la province de l'Ontario avec 18 participants et du Québec avec 8 participants. Cependant, le plus grand pourcentage de participation provient de la province de Terre-Neuve avec un pourcentage de participation de 67% suivie par l'Alberta (50%), la Saskatchewan (33%) et le Manitoba (33%).

Notre recherche s'appuie donc sur des entretiens avec 36 chercheurs canadiens en épidémiologie, incluant 3 chercheurs ayant l'expérience de l'épidémiologie numérique (Figure 3.1). Ce nombre de participants n'est pas représentatif de l'ensemble de la population des chercheurs en épidémiologie au Canada. Les perspectives constatées dans le cadre de cette thèse ne sont donc pas généralisables pour tous les chercheurs en épidémiologie. Ces perspectives sont considérées comme des suppositions à confirmer en effectuant d'autres recherches fondées sur des échantillons populationnels représentatifs. Cette thèse vise à supposer certains phénomènes, à soulever des questions et à ouvrir d'autres pistes de recherche pour comprendre les aspects épistémiques et

sociopolitiques de la pratique de l'épidémiologie numérique dans le contexte canadien. Nous considérons que les entretiens sont le meilleur outil pour collecter les données requises afin de répondre à notre question de recherche, puisque le but de notre recherche est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. D'autres formes d'observation, comme des études de laboratoires, auraient été intéressantes, mais ce travail constitue un autre cadre de recherche et le recrutement aurait impliqué une autre stratégie. Une fois avoir évalué les informations obtenues suite aux entretiens, nous avons jugé qu'une telle perspective n'était pas nécessaire parce que la richesse des perspectives apportées par les participants s'est avérée amplement suffisante pour atteindre le but de notre recherche.

Par ailleurs, le contexte sanitaire de 2020 n'était pas du tout favorable à des rencontres en personne sur les lieux de travail, à l'exception des situations où les travailleurs occupent des fonctions essentielles qu'il est impossible de remplir à distance ce qui n'est pas forcément le cas dans les laboratoires. Cela étant dit, lorsque nous discutons des pistes pour des recherches futures dans la conclusion de la thèse, nous proposons de poursuivre les réflexions sur la pratique de l'épidémiologie numérique en effectuant des études de laboratoires et en élargissant l'étendue des participants à d'autres pays dans le monde, à d'autres secteurs que le secteur universitaire (dont les organismes gouvernementaux étant nombreux à avoir des chercheurs en épidémiologie) et à d'autres spécialistes, comme ceux de l'informatique, afin d'inclure un plus grand nombre de chercheurs ayant l'expérience de la pratique de l'épidémiologie numérique, en plus d'ajouter des participants ayant des expertises différentes pour mieux saisir cette nouvelle pratique en elle-même.

Tableau 3. 1 Le nombre de chercheurs contactés, le nombre de participants et le pourcentage de participation total et par province lors de la recherche de terrain

Province	Nombre de chercheurs contactés	Nombre de chercheurs participants	Pourcentage de participation
Ontario	198	18	9%
Québec	40	8	20%
Alberta	4	2	50%
Saskatchewan	6	2	33%
Terre-Neuve	3	2	67%

Province	Nombre de chercheurs contactés	Nombre de chercheurs participants	Pourcentage de participation
Colombie-Britannique	14	2	14%
Nouvelle-Écosse	11	1	9%
Manitoba	3	1	33%
Total	279	36	13%

Le Tableau 3.2 indique que les participants à la recherche de terrain sont répartis dans 20 institutions. L'Université d'Ottawa et McMaster University sont les institutions avec le plus grand nombre de participants (5 participants) suivies par University of Toronto et l'Université de Montréal (4 participants).

Tableau 3. 2 Le nombre de participants par institution

Institution	Nombre de participants
Université d'Ottawa	5
McMaster University	5
University of Toronto	4
Université de Montréal	4
McGill University	2
Memorial University	2
University of Lethbridge	1
Queen's University	1
Dalhousie University	1
Simon Fraser University	1
University of British Columbia	1
Northern Ontario School of Medicine (NOSM)	1
University of Manitoba	1
Western University	1
University of Regina	1
University of Saskatchewan	1
University of Alberta	1
Université du Québec à Montréal	1
Agence de santé publique du Canada	1
Université de Sherbrooke	1
Total	36

Nous avons catégorisé les participants selon la sous-discipline de l'épidémiologie au sein de laquelle ils pratiquent et selon leur expérience en matière d'épidémiologie numérique. Comme nous l'avons vu, l'épidémiologie comprend de nombreuses sous-disciplines, comme l'épidémiologie du cancer ou celle de l'environnement (Frérot *et al.*, 2018). Nous avons regroupé les participants en trois sous-disciplines, à savoir l'épidémiologie sociale, l'épidémiologie environnementale et l'épidémiologie clinique. Nous avons fait ce choix parce que nous nous attendions à ce que les points de vue des chercheurs en épidémiologie varient selon le type d'étude épidémiologique qu'ils entreprennent et selon le type de données qu'ils utilisent et non selon les maladies étudiées.

L'épidémiologie sociale s'intéresse aux déterminants sociaux ayant un impact sur les risques et la prévalence d'une mauvaise santé auprès de populations (Frérot *et al.*, 2018). Cette sous-discipline de l'épidémiologie nous apparaît pertinente pour notre étude, car pour étudier la santé des populations sous l'angle des déterminants sociaux, certains chercheurs peuvent avoir des points de vue particuliers sur l'utilisation de données générées par les médias sociaux ou par les moteurs de recherche Internet.

Quant à l'épidémiologie environnementale, elle s'intéresse aux enjeux écologiques et environnementaux en lien avec la santé des populations (Frérot *et al.*, 2018). Cette sous-discipline vise à déterminer les facteurs de risque pour la santé en établissant des liens entre les données de diverses sources. De telles données peuvent provenir des systèmes de santé ou d'organismes environnementaux. On pourrait alors aboutir à des perspectives intéressantes de la part de ces chercheurs puisque la pratique de l'épidémiologie numérique propose d'utiliser les données provenant d'autres sources.

Finalement, l'épidémiologie clinique s'applique au contexte clinique (WU, 2023). Cette sous-discipline de l'épidémiologie est généralement axée sur la prise de décision dans le but d'améliorer les soins de santé aux patients (WU, 2023). Les chercheurs en épidémiologie clinique peuvent avoir des avis différents sur l'utilisation de données numériques (comme celles générées par Internet) pour conduire des recherches épidémiologiques. En effet, leurs recherches, qui se situent en milieu clinique, peuvent impliquer l'utilisation de tests de laboratoires et de prélèvements effectués sur des patients ainsi que des interactions en personne et régulières avec les participants aux recherches

cliniques. Cette façon de faire de l'épidémiologie pourrait influencer non seulement les perceptions des chercheurs en épidémiologie clinique vis-à-vis des enjeux épistémiques et sociopolitiques de la pratique de l'épidémiologie numérique, mais aussi leur expérience par rapport à cette nouvelle pratique.

Les participants à la recherche de terrain incluent 14 femmes (39%) et 22 hommes (61%) d'après la Figure 3.1. La Figure 3.2 fournit la distribution des 36 participants à la recherche selon leur sous-discipline d'appartenance et la Figure 3.3 selon leur expérience de la pratique de l'épidémiologie numérique. Plus de la moitié des participants sont des chercheurs en épidémiologie sociale (53%), 25% sont des chercheurs en épidémiologie clinique et 22% sont des chercheurs en épidémiologie environnementale. La Figure 3.3 indique que parmi les 36 participants, 3 participants pratiquent l'épidémiologie numérique et 33 participants n'ont pas fait l'expérience de cette nouvelle pratique.

Figure 3. 1 La distribution des participants selon le sexe

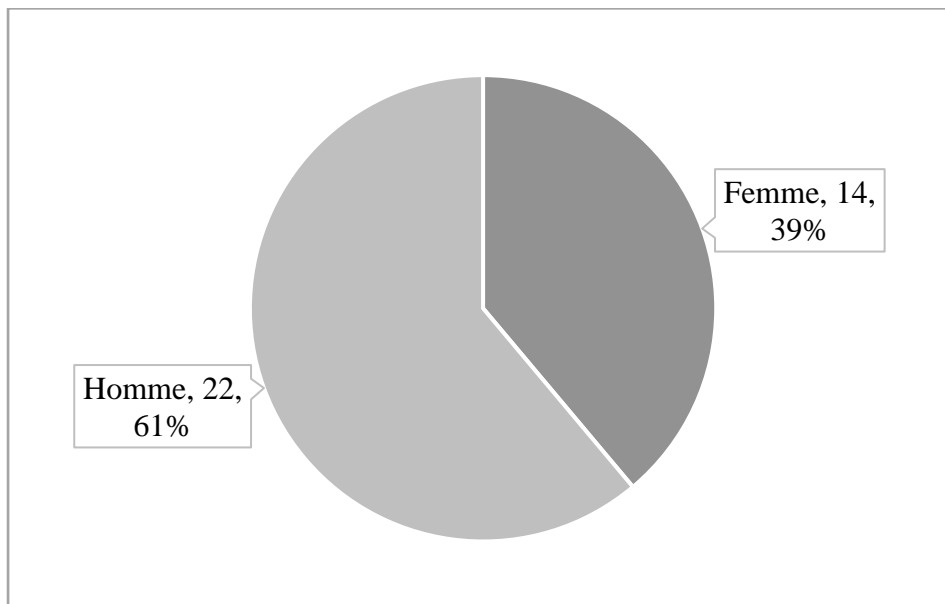


Figure 3. 2 La distribution des participants selon leur sous-discipline de l'épidémiologie

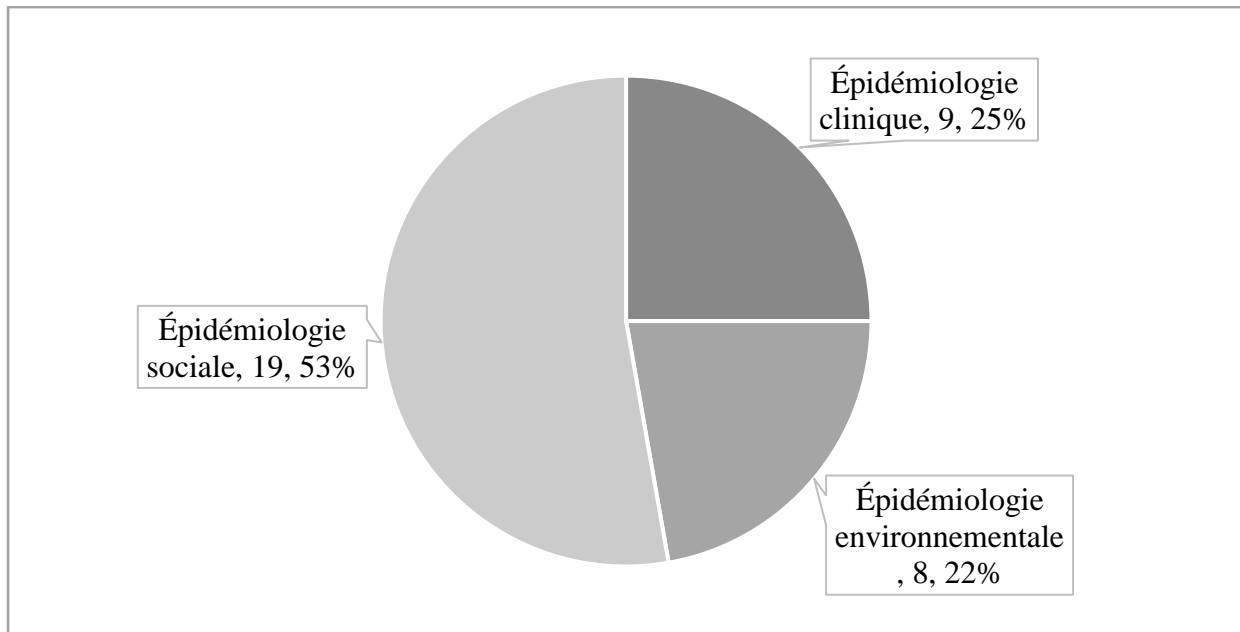
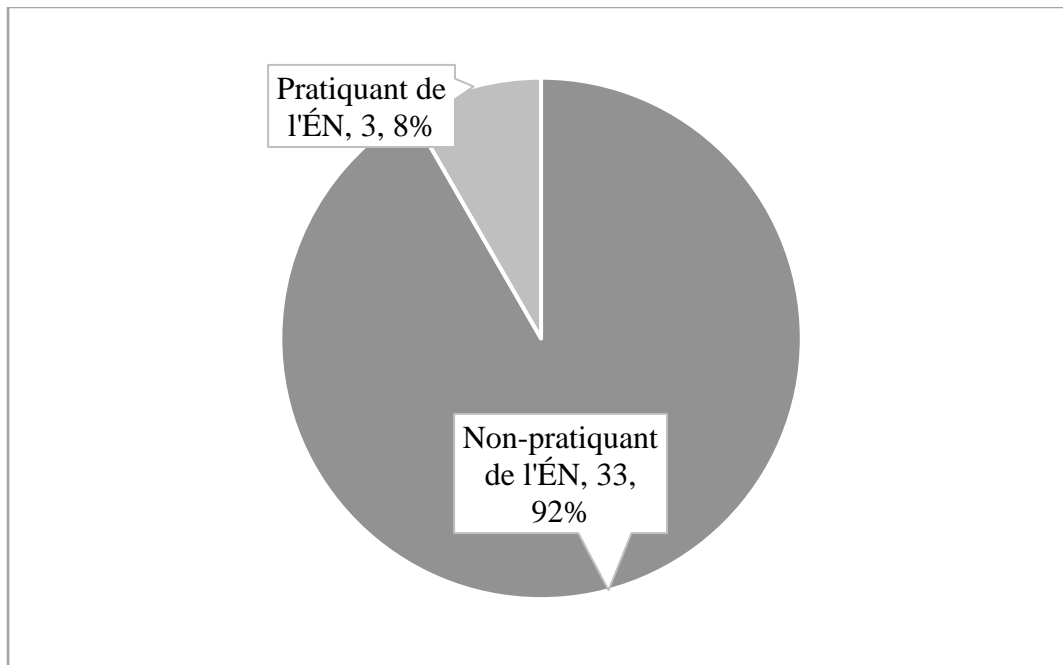


Figure 3. 3 La distribution des participants selon leur expérience de la pratique de l'épidémiologie numérique



Nous avons assigné un code individuel sous la forme d'une combinaison de chiffres et de lettres pour chaque participant en fonction du nom du participant, de son université d'appartenance et de la province de son université. Ces codes sont consignés dans un document protégé par un mot de passe accessible uniquement par la doctorante Renée Carpentier.

Lorsqu'un participant a fait l'expérience de la pratique de l'épidémiologie numérique, nous avons ajouté le terme « pratiquant de l'ÉN » dans la catégorie du participant pour lequel les propos ont été cités. Par exemple, si un participant en épidémiologie sociale a pratiqué l'épidémiologie numérique, sa catégorie est désignée comme « Chercheur- épidémiologie sociale - pratiquant de l'ÉN- 10Ixxviii ». Si le participant n'a pas pratiqué l'épidémiologie numérique, il est désigné par le type d'épidémiologie qu'il pratique et son code individuel. Par exemple, nous avons catégorisé un épidémiologiste clinique n'ayant pas pratiqué l'épidémiologie numérique comme « Chercheur-épidémiologie clinique - 10Ixxviii ».

3.2 La méthode d'analyse des données

Nous avons analysé les données recueillies suite à la recherche de terrain en nous inspirant de l'analyse par théorisation ancrée proposée par Paillé (1994). Le but de l'analyse par théorisation ancrée est de théoriser un phénomène empirique en suivant un processus d'analyse en six étapes : la codification, la catégorisation, la mise en relation, l'intégration, la modélisation et finalement, la théorisation. Paillé (1994) voit l'analyse par théorisation ancrée comme une démarche de théorisation. Selon Paillé (1994), la théorisation consiste à faire ressortir le sens d'un événement, construire un schéma explicatif reliant les divers éléments d'une situation et à mettre à jour la compréhension d'un phénomène en l'exposant différemment.

Comme proposé par Paillé (1994), nous avons commencé l'analyse des verbatim d'entretien en faisant leur codification. Les différentes étiquettes identifiées dans le corpus des entretiens sont incluses dans le tableau en annexe B. Pour faire cette codification, nous avons procédé en faisant une analyse thématique inspirée par Paillé et Mucchielli (2016). Paillé et Mucchielli (2016) expliquent l'analyse thématique comme ceci :

Avec l'analyse thématique, la thématization constitue l'opération centrale de la méthode, à savoir la transposition d'un corpus donné en un certain nombre de thèmes représentatifs du contenu analysé, et ce, en rapport avec l'orientation de recherche (la problématique). L'analyse thématique consiste, dans ce sens, à procéder systématiquement au repérage, au regroupement et, subsidiairement, à l'examen discursif des thèmes abordés dans un corpus, qu'il s'agisse d'une transcription d'entretiens, d'un document organisationnel ou de notes d'observation. (Paillé et Mucchielli, 2016, p. 3)

Le but de cette démarche est d'identifier les thèmes représentatifs à partir des discours des participants et en fonction de la problématique de recherche afin d'aboutir avec des catégories. Nous avons utilisé le logiciel d'analyse qualitative NVivo lors de ce travail. Notons que ce travail d'analyse est influencé par la subjectivité de l'auteure de cette thèse pouvant entraîner des biais dans la manière dont les enjeux sont présentés et compris par les chercheurs ayant participé aux entretiens.

Une fois l'analyse thématique complétée, nous avons décelé cinq grandes catégories en lien avec la pratique de l'épidémiologie numérique, à savoir l'adoption et les résistances, les acteurs, les pratiques, les perceptions, ainsi que les difficultés et les controverses. Le Tableau 3.3 rassemble les grandes catégories, c'est-à-dire les thèmes identifiés lors de l'analyse des propos des participants à la recherche de terrain.

Tableau 3. 3 Les catégories des propos des participants à la recherche de terrain

Catégorie	Signification
L'adoption et les résistances envers l'épidémiologie numérique (ÉN)	Il s'agit du niveau et des critères d'adoption et de résistance envers la pratique de l'ÉN selon les participants. Certains participants seraient prêts à adopter la pratique de l'ÉN si elle est capable de répondre à la question de recherche et si elle respecte de la méthodologie du domaine. D'autres participants considèrent le type de recherche dans leur appréciation de l'ÉN (clinique versus sociale). Quelques participants adoptent entièrement la pratique de l'ÉN tandis qu'une petite proportion la refuse pour des recherches en santé.

Catégorie	Signification
Les acteurs	Cette catégorie inclut les propos des participants sur le rôle de l'épidémiologiste et celui des nouveaux acteurs liés à la pratique de l'ÉN. En l'occurrence, les propos des participants ont indiqué que les chercheurs en épidémiologie ne voient pas de disparition d'acteurs avec la pratique de l'ÉN, mais un ajout d'acteurs nouveaux.
Les pratiques	Cette catégorie regroupe les propos des participants sur les ruptures et continuités dans les pratiques de recherche en épidémiologie et en ÉN, sur les aspects techniques des pratiques en épidémiologie, sur la surveillance des maladies et le contenu des recherches en épidémiologie avec l'émergence de l'ÉN, sur les interactions entre les acteurs dans les pratiques en épidémiologie avec l'émergence de l'ÉN et sur l'acquisition de l'expertise en épidémiologie avec l'émergence de l'ÉN.
Les perceptions	Cette catégorie porte sur les façons de concevoir les relations entre l'ÉN et l'épidémiologie traditionnelle, les différences entre les données numériques et les données traditionnelles ainsi que la place de technologies numériques dans le paysage de la recherche en épidémiologie.
Les difficultés et les controverses	Les difficultés et les controverses concernent les problèmes liés à l'échantillonnage, les controverses sur GFT, les difficultés de repérage de la désinformation, les limites des données écologiques pour informer sur les états des individus, la complexité des enjeux éthiques ainsi que la concentration des pouvoirs informationnels au sein de grandes entités privées comme Google.

Ensuite, nous avons entamé une autre phase consistant à effectuer les mises en relation et l'intégration des résultats obtenus à la suite de la recherche de terrain. Les mises en relation et l'intégration de nos résultats ont fait ressortir trois grands thèmes, à savoir les enjeux épistémologiques, l'organisation du travail scientifique et les enjeux sociétaux.

Nous avons identifié le thème des enjeux épistémologiques, car plusieurs participants ont laissé entrevoir des incertitudes sur la signification réelle des phénomènes observables à travers les

analyses de données numériques. Nous soutenons que cette situation soulève des questionnements sur les façons de percevoir les phénomènes épidémiologiques dans la pratique de l'épidémiologie numérique. Nous avons donc intégré les propos des participants liés aux différentes façons de concevoir les phénomènes épidémiologiques sous les enjeux épistémologiques, puisque nous considérons que ces phénomènes relèvent de l'épistémologie en épidémiologie et en épidémiologie numérique. Le thème sur l'organisation du travail scientifique concerne les fondements des collaborations entre chercheurs, c'est-à-dire comment et pourquoi les chercheurs collaborent dans leur pratique de l'épidémiologie numérique. Quant au thème sur les enjeux sociétaux, il renvoie aux conséquences de l'application des méthodes de l'épidémiologie numérique dans la société. Les propos des participants ont souvent souligné certaines préoccupations sociopolitiques, telles que la complexité des enjeux éthiques et la concentration des pouvoirs informationnels par les grandes entités privées comme Google. Nous avons donc intégré ces préoccupations dans les enjeux sociétaux liés aux méthodes utilisées en épidémiologie numérique.

Ces éléments d'intégration ont été modélisés sous la forme de deux types d'enjeux, à savoir les enjeux épistémiques et les enjeux sociopolitiques à l'égard de la pratique de l'épidémiologie numérique. Nous associons les enjeux épistémiques aux points de vue d'ordre épistémologique des participants. Ces aspects de la pratique de l'épidémiologie numérique portent sur comment cette pratique influence les façons de faire et de connaître dans la discipline de l'épidémiologie. Les enjeux sociopolitiques sont associés quant à eux aux points de vue d'ordre sociétal des participants. Ils concernent l'organisation des équipes de recherche et renvoient à des questions qui ne relèvent pas de la science en soi comme les problèmes d'éthique liés à l'utilisation des données numériques.

Nous avons abouti à cette modélisation en nous appuyant sur l'idée que les chercheurs en épidémiologie auront certainement des préoccupations d'ordre épistémique sur la pratique de l'épidémiologie numérique. En effet, comme nous l'avons expliqué, ces derniers doivent conduire des recherches valides selon les exigences de leur discipline, mais aussi pour le bien public étant donné qu'ils font partie de la société et que celle-ci influence leurs choix et perspectives en tant que chercheurs. Par ailleurs, comme nous l'avons mentionné dans le chapitre sur le cadre théorique, nous soutenons que les chercheurs et les pratiques de recherche font partie de la société sous forme de divers assemblages technologiques, scientifiques, cognitifs et sociopolitiques. Il est alors

important de comprendre comment les chercheurs en épidémiologie entrevoient l'émergence de la pratique de l'épidémiologie numérique dans le cadre d'études épidémiologiques, en fonction de leurs perspectives d'ordre épistémique et sociopolitique.

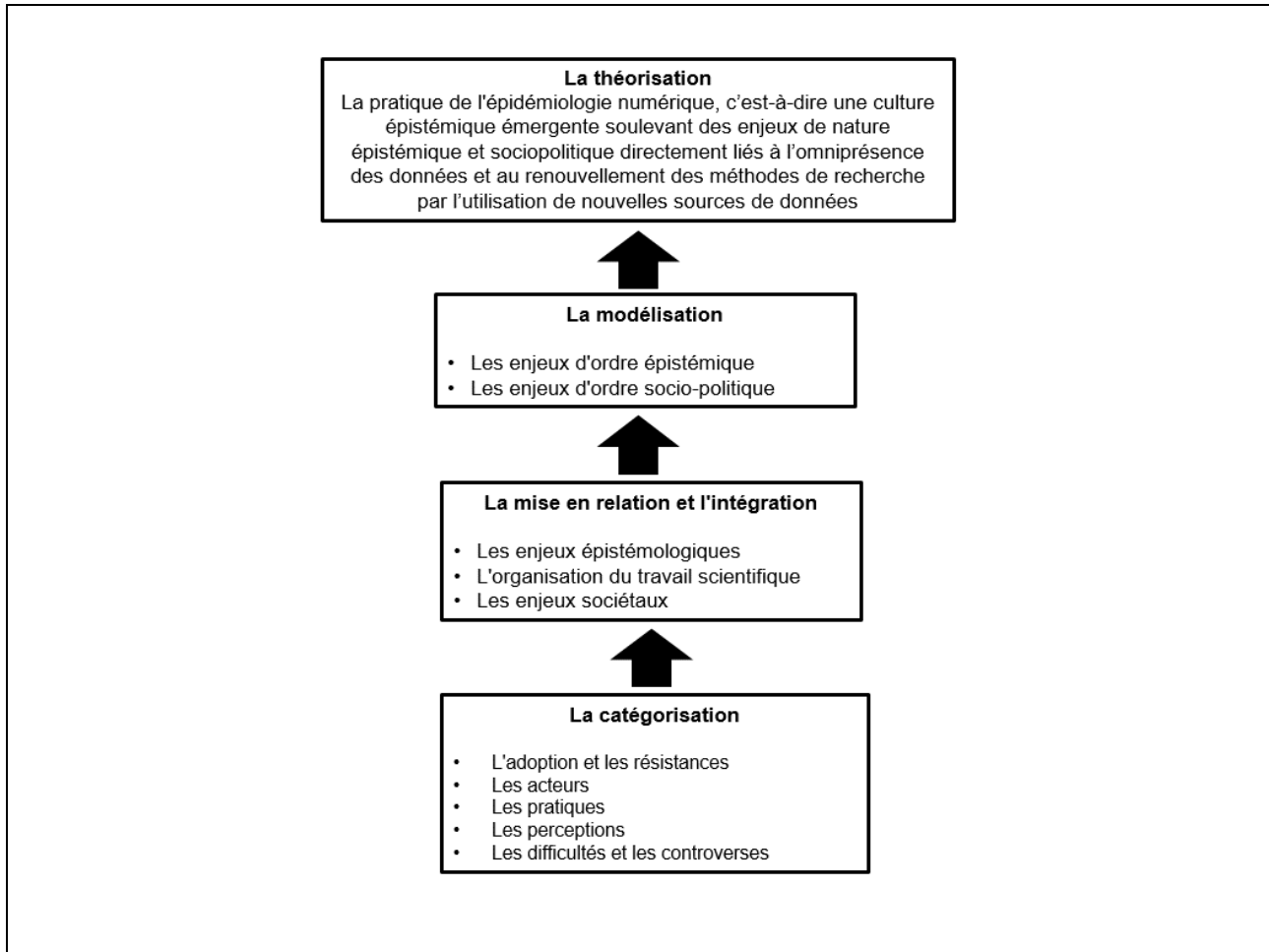
Comme expliqué précédemment dans le cadre théorique, les cultures épistémiques sont utiles pour expliquer les enjeux d'ordre épistémique des chercheurs en épidémiologie. La raison est que celles-ci s'intéressent entre autres à la culture, à la diversité et aux pratiques organisationnelles dans les sciences. Elles s'intéressent aussi aux signes et aux apparences secondaires liées aux dispositifs scientifiques ainsi qu'aux manières de considérer les mesures et les données expérimentales. Les propos des chercheurs ayant participé à la recherche de terrain peuvent ainsi être conceptualisés en utilisant ces différents aspects des cultures épistémiques.

Pour sa part, la vie sociale des méthodes et les enjeux politiques des données s'avèrent pertinents pour analyser les enjeux d'ordre sociopolitique soulevés par des chercheurs en épidémiologie quant à la pratique de l'épidémiologie numérique. En effet, les méthodes et les données sont importantes dans le cadre de notre recherche sur la pratique de l'épidémiologie numérique puisque celles-ci ne représentent pas seulement des boîtes à outils neutres fournissant simplement un support technique, mais font intervenir des acteurs essentiels qui façonnent le monde social et qui sont à leur tour façonnés par ce monde (Savage, 2013; Law *et al.*, 2011; Law, 2010). La vie sociale des méthodes et les enjeux politiques de données sont des agendas de recherche s'intéressant à ce que les méthodes et les données font et comment elles sont façonnées par le monde social. Notre recherche s'intéresse aussi à comprendre l'impact des méthodes et des données utilisées lors de la pratique de l'épidémiologie numérique sur la société et à saisir comment l'adoption de ces méthodes et l'utilisation de ces données sont influencées par les acteurs lors des recherches. C'est pourquoi les méthodes comme objet d'études, la double vie sociale des méthodes, les dispositifs méthodologiques et les enjeux politiques des données seront utiles pour investiguer les points de vue d'ordre sociopolitique relatifs à la pratique de l'épidémiologie numérique.

Cette démarche nous amène à théoriser la pratique de l'épidémiologie numérique comme une culture épistémique émergente soulevant des enjeux de nature épistémique et sociopolitique directement liés à l'omniprésence des données et au renouvellement des méthodes de recherche

par l'utilisation de nouvelles sources de données. Les étapes de cette démarche d'analyse sont résumées à l'aide de la Figure 3.4.

Figure 3. 4 Les étapes de la démarche d'analyse inspirée par l'analyse par théorisation ancrée développée par Paillé (1994)



3.3 La conclusion sur la méthodologie

La méthodologie employée dans le cadre de cette thèse s'appuie sur une recherche de terrain de type qualitative menée auprès de 36 chercheurs canadiens en épidémiologie. Les données ont été collectées en conduisant des entretiens individuels semi-dirigés et l'analyse de ces données a été effectuée en suivant l'approche par théorisation ancrée développée par Paillé (1994).

Cette démarche a abouti à une théorisation des résultats fondée sur les cultures épistémiques de Knorr-Cetina (1999), la vie sociale des méthodes élaborée dans les travaux de Savage (2013), Law *et al.* (2011), Law (2010) et Lury *et al.* (2020) et les enjeux politiques des données élaborés par Ruppert *et al.* (2017). Cette théorisation s'appuie sur la modélisation des résultats de la recherche de terrain ayant pris la forme de deux types d'enjeux sur la pratique de l'épidémiologie numérique, à savoir ceux étant d'ordre épistémique et ceux relevant de préoccupations d'ordre sociopolitique.

La modélisation proposée s'appuie quant à elle sur l'idée que les chercheurs et les pratiques de recherche font partie de la société sous forme de divers assemblages technologiques, scientifiques, cognitifs et sociopolitiques. Il s'agit alors de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique à travers les points de vue de chercheurs en épidémiologie.

Les chapitres suivants interprètent le matériel empirique de cette thèse afin de décrire la pratique de l'épidémiologie numérique (Chapitre IV) et de déceler les enjeux épistémiques (Chapitre V) et sociopolitiques (Chapitre VI) de cette pratique émergente.

CHAPITRE IV

LA PRATIQUE DE L'ÉPIDÉMIOLOGIE NUMÉRIQUE

Dans le cadre de ce chapitre, l'objectif est de comprendre la manière dont la pratique de l'épidémiologie numérique est perçue par les chercheurs épidémiologistes canadiens. Cette nouvelle pratique semble représenter une nouveauté considérable puisque celle-ci propose d'utiliser des sources de données et des technologies différentes comparativement à celles utilisées couramment lors des recherches. Il devient alors pertinent de définir en quoi consiste la pratique de l'épidémiologie numérique afin de développer un regard critique vis-à-vis de cette nouvelle pratique.

La pratique de l'épidémiologie numérique peut être conceptualisée comme une nouvelle culture épistémique, c'est-à-dire une nouvelle façon de faire de l'épidémiologie en ayant recours à différents types de connaissance et à des méthodes de recherche innovantes pouvant être utilisées pour des recherches épidémiologiques. De plus, la culture épistémique de la pratique de l'épidémiologie numérique peut amener des façons différentes de connaître les phénomènes épidémiologiques par l'utilisation de données de sources non traditionnelles à l'extérieur des systèmes de santé. D'après Knorr-Cetina (1999), les cultures épistémiques sont des cultures de création et de justification des connaissances liées par affinité, nécessité et coïncidence historique qui, dans un domaine donné, constituent la manière dont nous connaissons ce que nous connaissons.

La pratique de l'épidémiologie numérique peut aussi occasionner des ruptures avec les façons courantes de faire de l'épidémiologie et de construire des connaissances en impliquant une multitude de cadres instrumentaux, linguistiques, théoriques et organisationnels étant donné qu'il s'agit d'une nouvelle pratique. Toute nouveauté en recherche comme celle de la pratique de l'épidémiologie numérique peut sembler bénéfique pour certains chercheurs, tandis qu'elle peut apparaître comme incertaine ou inappropriée pour d'autres. Cette nouvelle pratique émergente qui ne s'avère pas encore stabilisée peut créer des divergences d'opinions entre les chercheurs pratiquant l'épidémiologie numérique et ceux ne l'ayant pas expérimentée, élargir l'espace

épistémique de la discipline de l'épidémiologie et susciter des confrontations lors de la rencontre de cultures épistémiques différentes.

Pour analyser ces enjeux, nous nous interrogerons sur les caractéristiques de la pratique de l'épidémiologie numérique et la diffusion de la pratique de l'épidémiologie numérique auprès des chercheurs en épidémiologie ayant participé à cette recherche de terrain. À l'exception de la section sur la crédibilité de la pratique de l'épidémiologie numérique où les résultats sont présentés en fonction de l'expérience de cette pratique, les résultats de cette recherche inclus dans ce chapitre sont présentés en fonction des points de vue des chercheurs étant donné qu'aucune tendance particulière n'a été décelée dans les propos des participants selon la sous-discipline de l'épidémiologie du participant ni selon l'expérience de la pratique de l'épidémiologie numérique du participant.

4.1 Les caractéristiques de la pratique de l'épidémiologie numérique

Pour certains chercheurs canadiens en épidémiologie, la notion d'épidémiologie numérique leur semblait vague parce que l'épidémiologie utilise des données et des technologies informatiques depuis très longtemps. Pour ces chercheurs, cette caractéristique peut très bien représenter une pratique fondée sur le « numérique », suscitant ainsi des questionnements sur le caractère nouveau de l'épidémiologie numérique et sur ses différences vis-à-vis des pratiques courantes en épidémiologie. D'après nos analyses, les façons de concevoir la pratique de l'épidémiologie numérique auprès des participants à la recherche de terrain concernent la provenance des données, la notion de « numérique », la notion de « technologie » et « d'outils électroniques », la diversité des sources de données et des technologies ainsi que la nouveauté de cette pratique.

4.1.1 La provenance des données

Hormis trois participants, les chercheurs canadiens en épidémiologie ayant participé à la recherche de terrain n'ont pas fait allusion à la provenance des données (à l'extérieur versus à l'intérieur du système de santé publique) lorsqu'ils ont expliqué leur définition de la pratique de l'épidémiologie numérique au début des entretiens contrairement à la définition proposée par Salathé (2018). Deux

des trois participants ayant communiqué une définition semblable à celle de Salathé (2018) possèdent en fait une expérience de la pratique de l'épidémiologie numérique dans le cadre de leurs travaux. Les propos de ces participants pratiquant l'épidémiologie numérique sont plus précis et plus élaborés que ceux des autres participants n'ayant pas expérimenté cette nouvelle pratique.

Par exemple, l'un de ces participants en épidémiologie environnementale a fait allusion à des dispositifs précis comme Twitter et aux recherches sur Google lorsqu'il a expliqué comment il voyait la pratique de l'épidémiologie numérique (entretien 10Ixxviii). Ce participant a aussi parlé de l'utilisation de données à l'extérieur des systèmes de santé en affirmant qu'il est d'avis que la pratique de l'épidémiologie numérique implique l'exploitation de données générées lorsqu'il n'y a aucune intention liée à la santé, et que ces données sont utilisées à des fins épidémiologiques. Il a aussi souligné qu'il ne considère pas que ses travaux de recherche fondés sur des données du ministère de la Santé reflètent la pratique de l'épidémiologie numérique. Cette perspective peut illustrer une nouvelle façon de faire de l'épidémiologie et des manières différentes de connaître les phénomènes épidémiologiques, ce qui reflète ainsi une nouvelle culture épistémique au sens de Knorr-Cetina (1999).

« I consider ... you know... mining Twitter to be digital epidemiology or mining Google searches. Those types of things when there is no or any health-related intention to the data, but the data are being used for epidemiological purposes... I guess for me it means conducting epidemiology with data that are purely digitally derived. So, there is no... you know I do a lot of research with the Ministry of Health data, and I don't consider that to be digital epidemiology. I consider you know mining Twitter to be digital epidemiology or mining Google searches. » (Chercheur - épidémiologie environnementale pratiquant de l'ÉN - entretien 10Ixxviii)

L'autre participant en épidémiologie sociale pratiquant l'épidémiologie numérique affirme que cette pratique de recherche va au-delà des systèmes de santé et se prolonge à l'extérieur du domaine de la santé (entretien 8Sxx). C'est là qu'il voit un alignement sur la santé des populations et qu'il établit des liens avec les politiques sur la santé. Pour ce participant, la pratique de l'épidémiologie doit être un succès parce qu'il faut déterminer comment rassembler des big data de différents secteurs et de différentes disciplines afin de comprendre les réalités. Il donne l'exemple de la pandémie de la COVID-19 au cours de laquelle des données sont générées en temps réel et pour laquelle une proportion significative des solutions se trouve à l'extérieur des systèmes de santé.

D'après ce participant, la pratique de l'épidémiologie numérique s'avère bien plus large que l'épidémiologie traditionnelle, cette dernière étant beaucoup plus axée sur les soins de santé.

« When I think about digital epidemiology, I actually think about beyond health care and health systems. So, in my mind, digital epidemiology is doing epidemiology that extends outside of health care or the health system itself. So that's where I see the alignment with population health and linking it with health policies. To me, digital epidemiology has to succeed, we have to figure out a way to bring big data from different sectors and different disciplines and understanding what is going on. For example, what is happening with Covid-19 right now, it gives some real time data, I believe a significant proportion of solutions lay outside of the health care systems themselves. So, to me digital epidemiology is much broader than traditional epidemiology which is more health care focus. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

Les propos de ce participant peuvent illustrer l'émergence d'une nouvelle culture épistémique avec la pratique de l'épidémiologie numérique puisque cette pratique peut amener des façons différentes de connaître les phénomènes épidémiologiques par l'utilisation de données de sources non traditionnelles à l'extérieur des systèmes de santé. Il est également possible d'envisager que la pratique de l'épidémiologie numérique puisse amener un mode de production des connaissances s'éloignant des approches disciplinaires afin de se diriger vers un autre mode de production des connaissances incorporant des approches interdisciplinaires ou transdisciplinaires et des collaborations entre le monde académique et industriel comme proposé par Gibbons *et al.* (1994).

Un seul participant n'ayant pas expérimenté la pratique de l'épidémiologie numérique a fait allusion à l'utilisation d'information de source atypique lorsqu'il a formulé sa définition de cette nouvelle pratique (entretien 5Kxv). Lors de la discussion, ce participant a affirmé considérer la pratique de l'épidémiologie numérique comme les big data et des données générées par des sources atypiques d'information.

« Je pensais que vous parliez de big data, de données qui proviennent de sources atypiques d'information. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

4.1.2 La notion de « numérique »

Plusieurs chercheurs n'ayant pas l'expérience de la pratique de l'épidémiologie numérique ont mentionné des termes liés à la notion de « numérique » dans leur tentative de définir cette nouvelle pratique en insistant notamment sur l'utilisation d'outils numériques (entretien 5Cxxvii), de méthodes numériques (entretien 5Gxxv), de technologies numériques (entretien 6Biv) et de sources numériques (entretien 10Txxii) pour faire de l'épidémiologie, mais sans évoquer de dispositifs particuliers. Un chercheur en épidémiologie environnementale soutient que la pratique de l'épidémiologie numérique poursuit les mêmes buts que l'épidémiologie traditionnelle (entretien 5Gxxv). La seule différence pour ce chercheur est que la pratique de l'épidémiologie numérique utilise des méthodes numériques pour la collecte des données lors du processus de recherche. Un autre participant en épidémiologie sociale estime que cette pratique concerne l'utilisation de technologies numériques pour effectuer des recherches épidémiologiques (entretien 6Biv). Nous avons aussi entendu un autre chercheur en épidémiologie sociale affirmer que les données proviennent d'Internet dans le cadre de cette nouvelle pratique (entretien 10Txxii). Pour un autre chercheur en épidémiologie environnementale, il s'agit d'une épidémiologie fondée sur une lunette utilisant des outils numériques (entretien 5Cxxvii).

« D'après moi c'est l'interface entre l'épidémiologie d'abord, donc c'est une base sur la définition de l'épidémiologie qu'on peut prendre comme étant le domaine d'études qui se rapporte à la progression des maladies dans les populations, si on peut utiliser celle-là, mais avec aussi une lunette d'utilisation des outils numériques. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvii)

« I think that the way I would define it is sort of it's the same goals as we consider traditional epidemiology, but it uses digital methods for data collection through the research process. I think it is just using digital methods used from data collection all the way to your analysis. » (Chercheur - épidémiologie environnementale - entretien 5Gxxv)

« I would think that it has to do with using digital technology to do epidemiological research. I can also see the use of digital technology in epidemiology ». (Chercheur - épidémiologie sociale - entretien 6Biv)

«Well, I would guess that it has to do with using data that is stored or drawn from digital sources. So, I would assume you know data that's coming from the Internet or that is produced by the Internet » (Chercheur - épidémiologie sociale - entretien 10Txxii)

4.1.3 La notion de « technologie » et « d'outils électroniques »

D'autres chercheurs épidémiologistes canadiens ont apporté des perspectives plus générales lorsqu'ils ont expliqué leur définition de cette nouvelle pratique. Cette catégorie de chercheurs a défini la pratique de l'épidémiologie numérique comme l'utilisation de technologies (entretien 5Cxxxii, entretien 6Jv), de nouvelles technologies (entretien 6Dxii) ou encore d'outils électroniques (entretien 6Fix) pour réaliser des études épidémiologiques. Quelques chercheurs voient cette nouvelle pratique comme une autre méthode pour collecter les données lors des recherches épidémiologiques (entretien 8Hxvi, entretien 6Fxi). L'un de ces participants a spécifiquement mentionné le transfert effectif de mesures et d'informations d'un participant ou d'un sujet directement vers une base de données, au lieu de les enregistrer sur un questionnaire papier ou sur un formulaire de collecte de données dans une sorte d'interface numérique (entretien 6Fxi).

D'après un participant en épidémiologie sociale, la pratique de l'épidémiologie numérique consiste à utiliser de nouvelles technologies pour identifier des épidémies de maladies plus rapidement qu'avec les méthodes traditionnelles (entretien 6Dxii). Un autre participant en épidémiologie sociale pense que cette nouvelle pratique consiste en l'utilisation de la technologie pour collecter et analyser les données en utilisant des sources numériques ou électroniques (entretien 8Hxvi). Ce participant suppose aussi que les bases de données sur la santé, les dossiers médicaux de facturation électronique et les dossiers médicaux électroniques puissent relever du numérique. Un chercheur en épidémiologie clinique, pour sa part, a fait allusion à des infrastructures en « cloud » en affirmant que pour lui, la pratique de l'épidémiologie numérique est quelque chose sur le « cloud » (entretien 6Biii).

Toutes ces perceptions ont été observées auprès de chercheurs en épidémiologie n'ayant pas pratiqué l'épidémiologie numérique. Aucun participant pratiquant l'épidémiologie numérique n'a fait allusion à des technologies de manière générale puisque ces participants ont mentionné des technologies spécifiques dans leur définition de cette nouvelle pratique.

« Ma première impression, ça fait un peu référence à toute l'utilisation justement des technologies dans nos recherches, mais tu sais l'épidémiologie est tellement vague aussi là, mais oui effectivement, c'est l'image que j'ai eue tout de suite en tête. » (Chercheur - épidémiologie clinique - entretien 5Cxxxii)

« Application of technologies to perform the science of epidemiology ». (Chercheur - épidémiologie sociale - entretien 6Jv)

« I can only think about it in terms of ways of using new technologies to identify disease outbreaks faster than using traditional methods. » (Chercheur - épidémiologie sociale - entretien 6Dxii)

« Epidemiology that uses electronic tools. » (Chercheur - épidémiologie sociale - entretien 6Fix)

« I would look at that in using technology in some way to collect the data maybe possibly uniquely, collect the data, analyze it, basically doing this by using digital sources or electronic sources as the data and statistical analysis. I would probably consider things like health databases, electronic billing medical records, electronic medical records, these types of things would be considered digital as well... I suppose... » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« I would say that the first thing that comes in my mind when you say that work, digital epidemiology would be the effective transfer of measure and information from a participant or a subject straight to a database instead of recording on a hard copy questionnaire or a data collection form that would be going directly in some sort of digital interface. That's sort of what it would mean to me I would say. » (Chercheur - épidémiologie clinique - entretien 6Fxi)

« What does it mean to me is something on the cloud. » (Chercheur - épidémiologie clinique - entretien 6Biii)

4.1.4 La diversité des sources de données et des technologies

Selon la littérature, l'épidémiologie numérique implique l'utilisation de sources de données et de technologies très diversifiées lors de travaux en épidémiologie comprenant les données numériques générées à partir des médias sociaux tels que Twitter (Gianfredi, 2018), des moteurs de recherche Internet (Clemente *et al.*, 2019), des plateformes et outils à accès ouvert (Lazzari *et al.*, 2018), les données produites par les technologies mobiles (Smolinski, 2017; Bates, 2017), les données et applications institutionnelles (Roche *et al.*, 2017) et les données spatiales (Gruebner *et al.*, 2018). De manière similaire à ces travaux de la littérature, certains des participants à la recherche de terrain ont associé l'épidémiologie numérique à l'utilisation des données générées à partir de Twitter (10Ixxviii), aux données produites par des technologies mobiles (9Rxiv), au traçage numérique (entretien 5Cxxvi), à l'intelligence artificielle (entretien 6Axxiii), au big data (entretien 6Axxiii,

entretien 5Kxv, entretien 6Fviii), à l'apprentissage automatique (*machine learning*) (entretien 6Axxiii), aux données générées par Google (entretien 10Ixxviii, entretien 3Exxxiii) et Google Maps (entretien 5Cxxx) ainsi qu'à l'épidémiologie computationnelle (entretien 5Cxxxii) dans leur façon de définir cette pratique. Ces chercheurs impliquent les participants ayant l'expérience de la pratique de l'épidémiologie numérique et ceux n'ayant pas expérimenté cette pratique.

L'un des participants en épidémiologie sociale pratiquant l'épidémiologie numérique soutient que cette pratique apporte des améliorations en épidémiologie en utilisant des outils numériques, les big data, l'intelligence artificielle ou l'apprentissage automatique (entretien 6Axxiii). Nous avons aussi constaté qu'un des chercheurs en épidémiologie environnementale n'ayant pas l'expérience de cette nouvelle pratique a exprimé beaucoup d'incertitude lorsqu'il a expliqué sa définition de la pratique de l'épidémiologie numérique parce qu'il considère l'épidémiologie comme une science numérique (entretien 5Cxxx) tout comme la pratique de l'épidémiologie numérique. Une telle perspective peut indiquer qu'il s'avère difficile parfois d'établir une distinction nette entre l'épidémiologie traditionnelle et la pratique de l'épidémiologie numérique puisque l'utilisation d'outils et de technologie numériques est présente dans les deux manières de faire de l'épidémiologie.

« I consider you know mining Twitter to be digital epidemiology or mining Google searches. »
(Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

« I tough it means using digital tools in epidemiology. So, whatever it is a collection of big data or using artificial intelligence to help to analyse the data from machine learning, it is kind of using digital as a tool to enhance what we do in epidemiology. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 6Axxiii)

« Ça me fait penser à l'utilisation de nouvelles technologies par exemple, je vois un lien avec l'étude de Katapally qui utilisait beaucoup les applications mobiles sur les téléphones ». (Chercheur - épidémiologie sociale - entretien 9Rxiv)

« Pour moi ça me faisait plus penser à toutes les données qui sont générées allant par Internet, et donc toutes ces connaissances du comportement des gens, le traçage des gens qui sont accessibles via par exemple le suivi de leurs mouvements, avec leur téléphone cellulaire etc. C'est vrai que c'est plus ces données-là générées par Internet en fait par l'interconnexion des gens par Internet. Voilà. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvi)

« Je pensais que vous parliez de big data, de données qui proviennent de sources atypiques d'information. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

« Je pensais à d'autres termes comme big data, les mégadonnées, peut-être les collectes de données par les appareils comme les apps sur les cellulaires ou par des réseaux comme *I don't know* Google Maps ou Google, C'est ça... j'ai pensé que c'est par rapport à ça, mais j'étais un peu confus parce que j'ai pensé que l'épidémiologie est une science numérique non ? » (Chercheur - épidémiologie environnementale - entretien 5Cxxx)

« My impression of it is that this is essentially using big data methods in epidemiology. This is my conception of it. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

« What I would guess is that digital epidemiology is ... you know how we can observe trends in Googling different terms, and some people say if you look at trends that people google the flu, then you can predict a flu outbreak in a region because a lot of people are searching for flu on Google. That would be my guess. » (Chercheur - épidémiologie sociale - entretien 3Exxxiii)

« It could be interpreted in many ways. But in terms of my use and my exposure, it has really been about it's like computational epidemiology. » (Chercheur - épidémiologie sociale - entretien 5Cxxxii)

4.1.5 La nouveauté de la pratique de l'épidémiologie numérique

Enfin, quelques entretiens ont fait ressortir des nuances sur la nouveauté de la pratique de l'épidémiologie numérique, car certains participants ont affirmé que l'utilisation de bases de données numériques, de données géographiques et de données satellitaires est chose courante en épidémiologie depuis longtemps. Plusieurs de ces participants sont des chercheurs en épidémiologie environnementale. L'un des chercheurs en épidémiologie environnementale pense notamment qu'il existe deux types d'outils numériques, c'est-à-dire les outils numériques actuels et des outils numériques plus anciens comme les systèmes d'information géographique utilisés depuis longtemps en épidémiologie (entretien 5Cxxvii). Un autre chercheur en épidémiologie environnementale a affirmé qu'il n'avait aucune idée de ce que signifie la pratique de l'épidémiologie numérique parce que, selon lui, le terme numérique désigne des bases de données électroniques et des logiciels, c'est-à-dire des technologies utilisées depuis des décennies en épidémiologie (entretien 1Oxxxv).

Un chercheur en épidémiologie sociale a exprimé le même avis en déclarant qu'il a l'impression que les outils numériques comme les sources de données et les méthodes d'analyse de certains logiciels de géolocalisation font déjà partie de la pratique de l'épidémiologie depuis longtemps (entretien 5Uxxxvi). À la lumière de ces constats, certains chercheurs en épidémiologie semblent penser que la pratique de l'épidémiologie numérique n'est pas réellement une nouvelle pratique puisque plusieurs de ses méthodes feraient déjà partie des activités courantes en épidémiologie. Cette perception pourrait aussi illustrer l'idée que les méthodes de recherche s'appuient souvent sur ce qui les a précédées (Law *et al.*, 2011; Law, 2010) puisqu'il semble que la pratique de l'épidémiologie numérique s'appuie entre autres sur l'épidémiologie, celle-ci étant la discipline l'ayant précédée.

« Donc on va parler des outils numériques qui nous rend en relation rapide, mais on peut aussi parler d'outils numériques plus anciens comme j'ai mentionné tantôt par exemple l'utilisation des systèmes d'information géographiques, ça a commencé à faire un bout que ça existe et que c'est encore en développement rapide et ça fait partie de cette classe d'utilisation d'outils numériques pour faire de l'épidémiologie. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvii)

« I really have no sort of notion of what it means because to me epidemiology ... digital means electronic databases, the software, the computing side of things which is something epidemiology has been doing for quite a long time...decades and decades and decades. It is almost a qualifier that doesn't mean so much to me... it all depends on what kind of epidemiology we are doing, different aspects of what we sort of involve. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxv)

« Ben...je ne le sais pas...en fait...parce que j'ai l'impression moi que les outils numériques sont déjà intégrés dans la pratique de l'épidémiologie, les sources de données, les méthodes d'analyse de certains logiciels de géolocalisation. Donc je peux comprendre qu'on peut caractériser certaines pratiques étant numériques, mais dans l'ensemble, si on fait référence par exemple uniquement à l'utilisation de données, ben ça c'est une pratique courante là. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

En somme, ces discussions avec les participants à la recherche de terrain laissent supposer que la pratique de l'épidémiologie numérique n'est pas très présente dans la culture épistémique de la discipline de l'épidémiologie puisqu'une bonne partie des participants étaient incertains de leur définition de cette nouvelle pratique. Cette recherche a aussi mis en lumière des perspectives différentes d'un chercheur à l'autre sur la manière de concevoir l'épidémiologie numérique.

Certains ont affirmé que la pratique de l'épidémiologie numérique consiste à utiliser des technologies et des nouvelles technologies (pouvant améliorer la collecte des données notamment) pour faire de l'épidémiologie. D'autres chercheurs ont mis l'accent sur le qualificatif « numérique » en affirmant que la pratique de l'épidémiologie numérique consiste à utiliser des outils et technologies numériques pour les recherches en épidémiologie, y compris les données d'Internet, notamment. Ces deux types de points de vue sont partagés par les chercheurs en épidémiologie n'ayant pas pratiqué l'épidémiologie numérique.

Quelques chercheurs ont apporté des notions plus précises sur la pratique de l'épidémiologie numérique qui rejoignent des travaux dans la littérature en faisant allusion à l'utilisation des données de Twitter et celles produites par des technologies mobiles, au traçage numérique, à l'intelligence artificielle, à l'apprentissage automatique, aux big data, aux données générées par Google et à l'épidémiologie computationnelle, dans leur façon de définir cette approche. Ces caractéristiques sont relatées autant par des chercheurs en épidémiologie pratiquant l'épidémiologie numérique que par des chercheurs n'ayant pas expérimenté cette nouvelle pratique.

Enfin, des chercheurs en épidémiologie environnementale et en épidémiologie sociale ont exprimé des incertitudes concernant la nouveauté de la pratique de l'épidémiologie numérique puisque plusieurs méthodes utilisées par celle-ci feraient déjà partie des activités courantes en épidémiologie. Ces incertitudes quant à la nouveauté peuvent s'expliquer par l'idée que les méthodes de recherches s'appuient souvent sur ce qui les a précédées étant donné que la pratique de l'épidémiologie numérique implique des méthodes déjà utilisées en épidémiologie. La prochaine section de ce chapitre se concentrera sur la diffusion de cette nouvelle pratique afin d'aller plus loin dans l'analyse de la pratique de l'épidémiologie numérique.

4.2 La diffusion de la pratique de l'épidémiologie numérique auprès de chercheurs en épidémiologie

Cette section porte sur la diffusion de la pratique de l'épidémiologie numérique auprès des communautés de chercheurs canadiens en épidémiologie. Nous allons analyser le niveau de

connaissance, le niveau d'expérience, les attitudes et la crédibilité de la pratique de l'épidémiologie numérique au sein des participants à la recherche de terrain. Nous voulons analyser le niveau de connaissance, le niveau d'expérience, les attitudes et la crédibilité parce que nous soutenons que la diffusion d'une nouvelle pratique de recherche passe par la connaissance, l'expérience, les attitudes et la crédibilité de cette pratique auprès des chercheurs. Nous allons aussi analyser la diffusion de la pratique de l'épidémiologie numérique en épidémiologie clinique parce que cette sous-discipline de l'épidémiologie montre des particularités par rapport aux autres sous-disciplines comme l'épidémiologie sociale et l'épidémiologie environnementale. Dans cette section, nous commencerons par analyser le niveau de connaissance, le niveau d'expérience, les attitudes et les perceptions en épidémiologie clinique de la pratique de l'épidémiologie numérique pour ensuite analyser la crédibilité de cette nouvelle pratique en nous appuyant sur les entretiens avec les participants à la recherche de terrain.

4.2.1 Le niveau de connaissance de la pratique de l'épidémiologie numérique

La connaissance d'une pratique constitue un premier indice de la diffusion de cette pratique, car, pour être moindrement diffusée, une pratique doit d'abord être connue par la communauté étudiée. Le niveau de connaissance de la pratique de l'épidémiologie numérique au sein des communautés de chercheurs étudiées lors de la recherche de terrain apparaît comme étant plutôt faible.

En effet, plusieurs chercheurs ont affirmé qu'ils ne connaissent pas la pratique de l'épidémiologie numérique et plusieurs d'entre eux n'avaient jamais entendu l'expression « épidémiologie numérique ». Par exemple, un participant en épidémiologie sociale a affirmé qu'il ne sait pas ce que signifie la pratique de l'épidémiologie numérique et qu'il n'a jamais entendu cette expression avant qu'il soit contacté dans le cadre de notre recherche (entretien 6Pvi). S'appuyant sur sa manière de considérer l'épidémiologie comme l'étude des déterminants des maladies dans les populations, ce participant propose de considérer la pratique de l'épidémiologie numérique comme une épidémiologie étudiant des facteurs transférables sur la santé de l'individu selon une approche numérique. Des chercheurs ayant participé à la recherche de terrain ont aussi affirmé qu'ils ne connaissent pas *Google Flu Trends* (GFT) (entretien 6Axvii) même si cet outil a suscité beaucoup de réactions en épidémiologie lorsque celui-ci était actif entre 2008 et 2015. Selon un chercheur en

épidémiologie sociale, GFT pourrait avoir un certain potentiel pour des recherches épidémiologiques (entretien 6Axvii). Ce chercheur ne sait pas exactement comment ce dispositif pourrait fonctionner et il ne connaît pas ses faiblesses, mais il entrevoit GFT comme quelque chose d'intéressant en épidémiologie.

« I am not sure. Epidemiology is the study of determinants of disease in populations. So, I think Digital... I don't know what that means... we study factors that could be transferred digitally on someone's health. I am not an expert on health issues. That would be my best guess. I never heard of that term before I received your e-mail. » (Chercheur - épidémiologie sociale - entretien 6Pvi)

« No, I didn't hear of that at all (Google Flu Trends)... It sounds that it has some potential indeed. I feel rather ignorant. So, I don't know exactly how that might work out and I don't know its weaknesses, but it sounds interesting for sure. » (Chercheur - épidémiologie sociale - entretien 6Axvii)

4.2.2 Le niveau d'expérience de la pratique de l'épidémiologie numérique

Pour déterminer le niveau de diffusion d'une pratique de recherche, il faut aussi considérer le niveau d'expérience de cette pratique au sein des communautés étudiées. Pour le cas de la pratique de l'épidémiologie numérique, celle-ci apparaît comme étant peu expérimentée au sein des communautés de chercheurs étudiées puisque la plupart des participants à la recherche de terrain ont affirmé qu'ils n'ont jamais fait l'expérience de cette nouvelle pratique dans le cadre de leurs travaux. Seulement trois chercheurs parmi les 36 participants pratiquent l'épidémiologie numérique, dont un chercheur en épidémiologie environnementale et deux chercheurs en épidémiologie sociale dont un dirige un laboratoire en épidémiologie numérique.

Selon la courbe de la diffusion de l'innovation de Rogers (2003), nos analyses indiquent que la pratique de l'épidémiologie numérique serait peu diffusée au sein des communautés de chercheurs en épidémiologie dans le contexte canadien. La temporalité est importante pour cette analyse, car, comme mentionné précédemment, la pratique de l'épidémiologie numérique est en phase d'émergence et celle-ci n'est donc pas stabilisée. Seulement trois chercheurs en épidémiologie sont des adeptes précoces de la pratique de l'épidémiologie numérique au sein du groupe de participants

à la recherche de terrain. La plupart des chercheurs en épidémiologie semblent tarder à adopter la pratique de l'épidémiologie numérique.

Par exemple, un chercheur en épidémiologie sociale a affirmé qu'il n'a jamais utilisé des données des médias sociaux lors de ses recherches (entretien 8Hxvi). Un chercheur en épidémiologie clinique a mentionné qu'il utilise des données volumineuses, mais ce sont des données cliniques (entretien 5Cxxxi).

De plus, un chercheur en épidémiologie sociale a confirmé qu'il n'a jamais utilisé des données de Twitter, de téléphones cellulaires ou d'autres données de ce type pour ses recherches (entretien 6Bxviii). Ce chercheur n'est pas opposé à la pratique de l'épidémiologie numérique. Il n'a simplement pas expérimenté cette nouvelle pratique, car l'opportunité ne s'est pas présentée dans son domaine. Les travaux sur l'activité physique et l'obésité de ce chercheur sont fondés sur l'épidémiologie traditionnelle, c'est-à-dire des analyses avec des données collectées sur le terrain.

« I never actually used social media data in my research » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« Je n'ai jamais joué avec ça, moi je joue avec beaucoup de données qui sont gigantesques souvent, mais qui sont très très cliniques. » (Chercheur - épidémiologie clinique - entretien 5Cxxxi)

« Je n'ai jamais utilisé moi-même comme on a parlé tantôt de Twitter, de téléphones cellulaires ou des choses comme ça pour ma recherche. Je ne suis vraiment pas contre ça du tout. C'est juste que ça ne s'est pas présenté encore dans mon domaine. C'est la vieille méthode qu'on utilise encore donc on n'a pas encore vraiment fait aucun changement dans le domaine de l'activité physique ou d'obésité. Donc c'est vraiment des analyses avec des données qu'on collecte sur le terrain et qu'on analyse. Donc on n'a pas utilisé d'autres techniques comme Twitter ou des téléphones cellulaires ou réseaux sociaux pour la recherche. » (Chercheur - épidémiologie sociale - entretien 6Bxviii)

Nous avons aussi constaté que les trois participants ayant l'expérience de la pratique de l'épidémiologie numérique entretiennent une opinion favorable de celle-ci. Deux d'entre eux l'ont adoptée occasionnellement (entretien 6Axxiii, entretien 10Ixxviii) lors de certains de leurs projets de recherche et un participant l'utilise pour tous ses projets de recherche (entretien 8Sxx).

Le participant pratiquant constamment l'épidémiologie numérique a affirmé que toutes ses recherches impliquent une approche d'épidémiologie numérique qu'il croise avec la science citoyenne (entretien 8Sxx). Ce chercheur en épidémiologie sociale fait participer les citoyens à ses recherches par le biais de leurs appareils mobiles pour obtenir des données en temps réel. La science citoyenne semble peu pratiquée par les chercheurs canadiens en épidémiologie selon les constats établis lors de la recherche de terrain sur la pratique de l'épidémiologie numérique parce que seul ce chercheur y a fait allusion (entretien 8Sxx). Cependant, certains chercheurs en santé des populations soutiennent que la science citoyenne devrait être davantage utilisée en raison de son potentiel de fournir de nouvelles perspectives sur la nature des problèmes et sur les façons de mettre en œuvre leurs solutions (Rowbotham *et al.*, 2019).

L'autre participant en épidémiologie sociale ayant l'expérience de la pratique de l'épidémiologie numérique a mentionné qu'il a travaillé sur trois projets fondés sur cette nouvelle pratique (entretien 6Axxiii). L'un de ces projets en pharmacovigilance implique l'utilisation de données générées par des médias sociaux. Un autre projet constitue une synthèse de connaissances sur la COVID-19 à partir des études publiées sur la base de données *MEDLINE*. Un autre projet avait pour but de réfléchir à la manière dont il est possible d'utiliser des outils tels que l'apprentissage automatique ou l'intelligence artificielle lors de processus de synthèse de connaissances, et de déterminer comment des outils numériques peuvent faciliter la saisie des données.

Le chercheur en épidémiologie environnementale pratiquant l'épidémiologie numérique a affirmé que cette approche lui est accessible parce qu'il détient les compétences nécessaires pour travailler avec des données numériques (entretien 10Ixxviii). Ce chercheur serait autant un scientifique des données qu'un épidémiologiste. D'après ce participant, les chercheurs en épidémiologie environnementale utilisent des données de satellites pour évaluer l'exposition environnementale et pour comprendre l'exposition, mais ils n'essaient pas de générer des données sur la santé à partir des données satellitaires proprement dites. Ils essaient de relier les données satellitaires à d'autres ensembles de données liées à la santé, selon ce chercheur.

Les propos de ces participants font remarquer que la pratique de l'épidémiologie numérique peut s'avérer utile dans des contextes de recherche variés. En effet, chacun de ces chercheurs travaille

sur des projets très différents. L'un d'entre eux utilise l'approche de la science citoyenne (entretien 8Sxx), un autre chercheur l'utilise pour des synthèses de connaissances (entretien 6Axxiii) et un autre l'utilise dans le cadre d'analyses de données de satellites afin de les lier avec des données en santé (10Ixxviii).

« All my research projects involve digital epidemiology now. So, everything that I do involves digital epidemiology approach to it that I intersect with citizen science what I mean by that is from a population health perspective, by engaging citizens through their app mobile devices to get a real time data. So, all the projects, and all the study in my lab have a digital epidemiology approach. Each project might have a different design, different objective, but the approach is the same. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN – entretien 8Sxx)

« The one with the social media that we did on pharmacovigilance was done for Health Canada. So that was one project. And then we did a couple of projects recently on Covid where we worked with computer scientists who grab the all the studies that was published in MEDLINE, and they did like a computer algorithm and tell us which one would be related to our research question. So, we used that in the project. So, I would say that we've done about three projects that way and we also worked with computer scientists to work on two methods projects. One of them was to think about the flow of the systematic review process and where we can use tools such as machine learning or AI to help with that and then another one was looking specifically at data caption and how we can facilitate the process using digital tools. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 6Axxiii)

« I would say that the approach is very accessible to me as I have the skill set necessary to work with those types of data...I am as much as a data scientist as I am an epidemiologist... in environmental epidemiology we use these data for environmental exposure assessment and to understand exposure, but we are not trying to generate health data from the satellite data per say. We are trying to link those with other data sets that are related to health » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

4.2.3 Les attitudes envers la pratique de l'épidémiologie numérique

Les attitudes envers une nouvelle pratique peuvent aussi constituer un indice sur la diffusion de celle-ci. Dans le cadre de la recherche de terrain, les propos des participants ont démontré trois attitudes envers la pratique de l'épidémiologie numérique, à savoir l'adoption, la résistance et les réserves.

4.2.3.1 L'adoption

Parmi les trois chercheurs ayant adopté la pratique de l'épidémiologie numérique, seulement un participant utilise entièrement la pratique de l'épidémiologie numérique. Tous les travaux de recherche de ce participant sont fondés sur la pratique de l'épidémiologie numérique avec une approche basée sur la science citoyenne (entretien 8Sxx). Dans une perspective axée sur la santé des populations, ce participant fait participer des citoyens lors de ses recherches par le biais de leurs appareils mobiles en vue d'obtenir des données en temps réel.

Comme expliqué précédemment, ce chercheur affirme que la pratique de l'épidémiologie numérique va au-delà des systèmes de santé et se prolonge à l'extérieur du domaine de la santé (entretien 8Sxx). C'est là qu'il entrevoit un alignement sur la santé des populations et qu'il établit des liens avec les politiques sur la santé. Pour ce participant, cette nouvelle pratique doit être un succès parce qu'il est important de déterminer comment rassembler des big data de différents secteurs et de différentes disciplines pour comprendre le monde. Il donne l'exemple de la pandémie de COVID-19 au cours de laquelle des données sont générées en temps réel et pour laquelle une proportion significative des solutions se trouve à l'extérieur des systèmes de santé. D'après ce participant, la pratique de l'épidémiologie numérique s'avère bien plus large que l'épidémiologie traditionnelle, cette dernière étant beaucoup plus axée sur les soins de santé. Selon ce participant, la pratique de l'épidémiologie numérique s'applique à tout projet de recherche épidémiologique, y compris l'épidémiologie clinique, car tout le monde est connecté à des dispositifs numériques. Ce participant amène aussi l'idée que l'épidémiologie doit s'adapter au monde numérique si elle veut répondre aux besoins de recherche actuels.

De plus, ce chercheur pense que la pratique de l'épidémiologie numérique peut aider à résoudre les problèmes récurrents sur le plan des biais de sélection des échantillons dans les populations, car une large proportion des personnes a au moins un dispositif numérique (comme un téléphone intelligent), élargissant ainsi les échantillons à au moins 90 % des populations au lieu de rejoindre 50 à 60 % des populations avec les méthodes traditionnelles. Nos analyses sur les résistances envers la pratique de l'épidémiologie numérique (expliquées à la suite de ce paragraphe) montrent que certains chercheurs en épidémiologie ne sont pas nécessairement en accord avec ce point de vue. En effet, certains d'entre eux ont des doutes quant à la qualité de l'échantillonnage populationnel

avec les données numériques selon les constats établis lors de la recherche de terrain. L'échantillonnage populationnel sera analysé de manière plus approfondie dans le chapitre sur les enjeux épistémiques de la pratique de l'épidémiologie numérique puisqu'il s'agit d'un enjeu épistémique important en épidémiologie et dans le cadre de cette nouvelle pratique.

« All my research projects involve digital epidemiology now. So, everything that I do involves digital epidemiology approach to it that I intersect with citizen science. What I mean by that is from a population health perspective, by engaging citizens through their app mobile devices to get real time data. So, all the projects, and all the study in my lab have a digital epidemiology approach. Each project might have a different design, different objective, but the approach is the same. » (Chercheur – épidémiologie sociale – pratiquant de l'ÉN – entretien 8Sxx)

« So, in my mind, digital epidemiology is doing epidemiology that extends outside of health care or the health system itself. So that's where I see the alignment with population health and linking it with health policies. To me, digital epidemiology has to succeed, we have to figure out a way to bring big data from different sectors and different disciplines and understanding what is going on. For example, what is happening with Covid-19 right now, it gives some real time data, I believe a significant proportion of solutions lay outside of the health care systems themselves. So, to me digital epidemiology is much broader than traditional epidemiology which is more health care focus. » (Chercheur – épidémiologie sociale – pratiquant de l'ÉN – entretien 8Sxx)

« At the end of my PhD in epidemiology, I became more and more involved in digital epidemiology. I cannot really tell you really how I did it but I can tell you why I did it because it didn't make any sense that we are still conducting epidemiology and epidemiological research like it is the 1950s! The whole world has moved away, and all the data, all the relevant data is digital. I didn't make any sense to me to continue doing the same old fashion epidemiological research. » (Chercheur – épidémiologie sociale – pratiquant de l'ÉN – entretien 8Sxx)

« I would say that it's applicable to everything... is there anything that we do in our lives and days that are not emphasis into a digital tool? We are connected to our devices, right ? When they are so interconnected, how can we for any research whether is it clinical research whether is it social research, how can we ignore these tools? » (Chercheur – épidémiologie sociale – pratiquant de l'ÉN – entretien 8Sxx)

« So, the big thing that they are asking is what about selection bias? Well, I am like if you could recruit the whole population, where is the selection bias of that? And they compensate that by asking how do you do that? Because 90% of Canadians have a smartphone. That doesn't mean that everyone is going to appear in the data but theoretically, you need to use these tools to understand what's going on. So why do I want to do a study based on a research question by recruiting just 50 or 60% of a population when I can recruit thousands of people and take data in real time? I see that as a big value for epidemiology and we have to understand the power of digital tools. » (Chercheur – épidémiologie sociale – pratiquant de l'ÉN – entretien 8Sxx)

4.2.3.2 Les résistances

Deux chercheurs parmi les trente-six participants à la recherche de terrain résistent par rapport à la pratique de l'épidémiologie numérique. En donnant l'exemple des données générées par le moteur de recherche Google, l'un de ces chercheurs soutient notamment que la pratique de l'épidémiologie numérique n'est pas utile pour des recherches en santé (entretien 5Gxxix). Ce genre de pratique s'avérerait plutôt utile pour des motifs de marketing, selon ce chercheur en épidémiologie environnementale. Il semble donc que ce chercheur associe l'utilisation de données numériques à des domaines en dehors du champ scientifique ou du moins du domaine de la santé.

L'autre chercheur en épidémiologie sociale affirme qu'il doute que la qualité des données numériques soit suffisante pour répondre à des questions de recherche et qu'il est plutôt sceptique vis-à-vis des succès de cette pratique en raison des biais liés à ces données (entretien 6Dxii). Ce point de vue peut mettre en évidence des incertitudes sur la validité des signes générés par les données numériques. En effet, certains chercheurs sont incertains vis-à-vis des signes et des apparences visibles à travers les analyses de données numériques. Il existe peut-être des bruits tels que des signaux aléatoires, imprévisibles et indésirables survenant lors de l'exploitation des données numériques dans le cadre d'études épidémiologiques. Ces bruits constitueraient une menace pour la capacité des chercheurs en épidémiologie à reconnaître les événements intéressants sous l'épidémiologie. De tels bruits peuvent notamment falsifier la signature des événements, déformer leur caractère, ou mettre en péril leur identification.

« Honestly, not credible. You could use Google for marketing purposes. That's what basically it is useful but for health purposes? No. » (Chercheur – épidémiologie environnementale – entretien 5Gxxix)

« I have trouble believing that the data would be of sufficient quality to answer questions. I'm afraid that I am rather skeptical of the success that we are going to see because of the biases. » (Chercheur, épidémiologie sociale – entretien 6Dxii)

4.2.3.3 Les réserves

Les réserves envers la pratique de l'épidémiologie numérique se définissent comme l'attitude la plus souvent observée auprès des chercheurs ayant participé à la recherche de terrain. En effet, la plupart d'entre eux sont ouverts à pratiquer l'épidémiologie numérique, mais ils expriment de la prudence face à cette nouvelle pratique, car ils la considèrent comme une façon de faire de l'épidémiologie présentant du potentiel et aussi des lacunes. Selon ces participants, le chercheur doit évaluer les bienfaits et les inconvénients de la pratique de l'épidémiologie numérique comme toute autre pratique de recherche.

Un chercheur en épidémiologie clinique montre une attitude de réserve envers la pratique de l'épidémiologie numérique en donnant l'exemple du moteur de recherche Internet de Google (entretien 6Fx). Ce chercheur soutient qu'il faut comprendre comment le moteur de recherche fonctionne parce que, selon lui, il y a une logique derrière le moteur de recherche procédant à la classification et à la priorisation des informations. D'après ce participant, si le chercheur n'est pas conscient de la manière dont l'algorithme classe l'information, il se livrera à une mauvaise interprétation des données. Ce participant affirme aussi que si le chercheur comprend la manière dont le moteur de recherche fonctionne, la pratique de l'épidémiologie numérique devient comme toute autre approche en science exigeant toujours d'expliquer les forces et les faiblesses de la recherche. Pour ce participant, la force de cette nouvelle pratique réside dans les grandes quantités de données disponibles et la faiblesse concerne le risque que la logique de l'algorithme⁷ derrière le moteur de recherche ne soit pas alignée sur tous les types de recherche.

Cette attention sur l'algorithme derrière le moteur de recherche Google illustre l'idée que les données et les mesures sont contingentes à l'appareil de mesure et que celles-ci constituent des représentations de cet appareil plutôt que du monde réel. Par conséquent, le moteur de recherche Google (ou le détecteur) doit être correctement compris par le chercheur parce que c'est à travers cette technologie que le chercheur pourra analyser des tendances dans les recherches effectuées par

⁷ Comme défini précédemment selon Angius et al. (2024), un algorithme est un ensemble d'instructions permettant l'accomplissement d'une tâche donnée. Les règles que l'on suit pour calculer les opérations arithmétiques de base telles que la multiplication ou la division seraient des exemples quotidiens d'algorithmes (Angius et al., 2024). Hill (2016) propose de définir un algorithme comme une structure de contrôle finie, abstraite, efficace et composée, impérativement donnée, qui accomplit un objectif précis dans des conditions définies.

les gens sur Internet. Ainsi, le moteur de recherche Google peut être vu comme un détecteur agissant comme un dispositif médiateur interposé entre les chercheurs et les phénomènes. Les données et les mesures dépendent de l'appareil de mesure et celles-ci sont des représentations de cet appareil (Knorr-Cetina, 1999).

Selon ce raisonnement, la manière dont est construite une technologie numérique comme celle du moteur de recherche Google peut donc influencer les analyses du chercheur en fonction de la conception de l'algorithme derrière ce moteur de recherche. Le moteur de recherche Google et son algorithme deviennent alors un dispositif médiateur entre le chercheur et les phénomènes lors de la pratique de l'épidémiologie numérique. Ce faisant, il est possible que cette nouvelle pratique représente une culture épistémique différente par rapport à celle de l'épidémiologie puisque l'algorithme semble jouer un rôle dans les prises de décision et dans le déroulement des recherches lors de la pratique de l'épidémiologie numérique.

« Avec ça il faut comprendre comment le moteur de recherche fonctionne par exemple, si vous mettez votre nom sur Google, il va y avoir des sites qui vont apparaître en premier, mais il y a une logique derrière, donc il faut comprendre comment le moteur de recherche priorise et classe l'information pour... parce que si la personne n'en est pas consciente, elle va faire de la mauvaise interprétation des data parce que les résultats qu'elle va voir seront basés sur l'algorithme utilisé pour la classification des données. Donc je crois qu'à partir du moment où un chercheur comprend comment le moteur de recherche fonctionne ...bien comme toute approche en science, il faut toujours parler des forces et des faiblesses de la recherche donc je crois qu'il y a des forces. La force c'est les tonnes de data disponibles qui sont invraisemblables. La faiblesse c'est qu'il y a l'algorithme qui utilise une certaine logique et que du coup peut-être que la logique de l'algorithme ne va pas matcher avec tous les types de recherche forcément. » (Chercheur - épidémiologie clinique - entretien 6Fx)

Un chercheur en épidémiologie sociale montre aussi une attitude de réserve en affirmant que la pratique de l'épidémiologie numérique peut fournir des informations intéressantes parlant de tendances (5Kxv). Cependant, ce chercheur soutient qu'il faut être conscient des limites associées aux données numériques.

Un chercheur en épidémiologie clinique semble ouvert à la pratique de l'épidémiologie numérique si la recherche porte sur ce que les gens pensent suivant une approche quantitative parce qu'il

estime que ce genre de recherche vise à comprendre des perceptions auprès de populations (entretien 6Axix). Si une recherche fondée sur la pratique de l'épidémiologie numérique tente d'analyser des incidents comme la fréquence d'une maladie, ce chercheur aurait moins confiance envers cette recherche. Selon ce chercheur, il y a peu de chances qu'une recherche fondée sur la pratique de l'épidémiologie vise à répondre à ce genre de questions.

« C'est de l'information qui est intéressante, car ça nous parle de tendances, l'information est populaire, mais il faut prendre ça avec les limites qui sont associées à ces données-là. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

« It depends on what the research question is. If the research is just mainly on a quantitative side by looking generally at what people think, share then I would have a lot of confidence because that kind of research just try to understand a point of view understanding that it is not all points of view. If it was one trying to determine numbers that we are looking at incidents such as disease frequency and things like that. That's different. And those types of questions. I would have less confidence in that. But I don't think that those would be the kind of questions that would be asked when using these methods. » (Chercheur - épidémiologie clinique - entretien 6Axix)

Les résultats de la recherche de terrain indiquent aussi que si un chercheur pratique l'épidémiologie numérique, cela ne signifie pas automatiquement que ce chercheur adoptera entièrement cette nouvelle pratique, ni qu'il l'utilisera constamment, ni que celui-ci envisagera que cette nouvelle pratique devra remplacer l'épidémiologie traditionnelle. Les résultats de cette recherche indiquent qu'un seul participant a démontré une adoption complète de la pratique de l'épidémiologie numérique. Les deux autres chercheurs pratiquant l'épidémiologie numérique ont adopté une attitude très ouverte, mais prudente envers cette pratique.

Les propos du chercheur en épidémiologie environnementale pratiquant l'épidémiologie numérique sont très explicites sur la place des données numériques dans ses pratiques de recherche (entretien 10Ixxviii). S'il est entendu que la pratique de l'épidémiologie numérique implique l'utilisation des données de télédétection et la science de l'exposition, cette pratique se révèle fondamentale pour ce chercheur puisque ces données sont essentielles pour tous ses travaux. S'il est entendu que la pratique de l'épidémiologie implique uniquement l'utilisation de données numériques comme celles produites par les moteurs de recherche Internet et les médias sociaux,

cette pratique devient alors moins importante puisqu'elle découlerait surtout d'un intérêt personnel plutôt que de représenter quelque chose d'essentiel aux pratiques de ce chercheur. Ce participant se penche sur la pratique de l'épidémiologie numérique, mais elle n'est pas essentielle dans le cadre de ses travaux. Certains de ses projets utilisent cette approche, mais ce ne sont pas nécessairement ses projets les plus critiques. Les projets de ce chercheur se fondent généralement sur des approches utilisant des données environnementales, l'intelligence artificielle et des données de satellites.

« I guess it depends on the definition that we are going to take on digital epidemiology. If we are taking the definition that includes the use of remote sensing data and exposure science, it is fundamental, I couldn't do anything of what I do without those data. So, then it becomes less important, it becomes of a personal interest I guess than something that is critical to my practice. So, we do look at digital epidemiology, and we do consider it, but it hasn't been critical for me. I work with people at the center for whom it is not true, so it would be critical to the work that they do especially the mobility data and especially in the context of Covid. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

D'après ce participant, il n'y a pas de situation où la pratique de l'épidémiologie numérique devrait être évitée (entretien 10Ixxviii). Ce chercheur n'a pas identifié de limites du côté des recherches, mais il envisage des limites politiques avec la pratique de l'épidémiologie numérique. Ce chercheur ne serait jamais à l'aise de faire pleinement confiance envers des signaux numériques. Cependant, si les signaux provenant de l'épidémiologie numérique étaient cohérents avec les signaux conventionnels, il estime qu'ils seraient en mesure de renforcer l'argument politique. Ce chercheur ne pense donc pas que les signaux numériques à eux seuls soient suffisants pour prendre des décisions politiques.

« I don't think it should ever be avoided in different research honestly. I don't see research limitations. I see policy limitations. So, in the kind of work I do we often move research policy quite quickly, I would never be comfortable doing that based on digital signals alone, but if digital signals were consistent with conventional signals, I think they can strengthen the policy argument. Digital signals alone I don't think they are adequate to make policy decisions on. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN – entretien 10Ixxviii)

Ce chercheur ayant l'expérience de la pratique de l'épidémiologie numérique a mentionné que cette nouvelle pratique axée sur les données numériques (comme celles générées par Google et les

médias sociaux) constitue une science très intéressante (entretien 10Ixxviii). Si le chercheur a la possibilité d'aller très loin avec certaines de ces données, il peut poser beaucoup de questions différentes d'après ce participant. Les applications sont presque infinies si le chercheur réfléchit bien à la façon dont les données numériques peuvent être utilisées et ce n'est certainement pas l'épidémiologie traditionnelle qui recèle un tel potentiel selon ce participant. Selon ce chercheur, la pratique de l'épidémiologie numérique peut offrir une énorme possibilité de générer des hypothèses et de les tester, et le simple fait de jouer avec les données numériques constitue quelque chose d'innovant en soi et de précieux.

« On the good side, I foresee just some really interesting science. Again, you must have the opportunity to go so far with some of these data and you can ask so many different questions when you sit down and think about how these data can be used, the applications are almost endless and that's certainly not true with conventional epidemiology where you need to have a data set and collaborators and all those things. So, I think there is a huge opportunity for hypothesis generation and hypothesis testing and just play around which is innovative itself and a valuable thing. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

Ce participant a aussi exprimé une vision incertaine sur la pratique de l'épidémiologie numérique. D'une part, il souligne son potentiel pour soulever de nouvelles questions et des hypothèses de recherche épidémiologique, un potentiel plus grand selon lui avec la pratique de l'épidémiologie numérique qu'avec celle de l'épidémiologie traditionnelle. D'autre part, il exprime ses craintes concernant des incertitudes quant aux interprétations des données et au fait que certains chercheurs publient des études utilisant l'épidémiologie numérique sans être épidémiologistes et sans inclure un épidémiologiste dans l'équipe du projet de recherche (entretien 10Ixxviii). Il semble donc que ce chercheur apprécie les bons côtés de l'épidémiologie numérique tout en voulant conserver les fondements de la discipline de l'épidémiologie épidémiologique en vue de communiquer des informations valides dans les recherches sur les enjeux de santé publique et d'éviter les confusions liées au manque d'expertise en épidémiologie.

Where I see the pitfalls is incorrect interpretation of the data, often or at least sometimes I see an example of people doing digital epidemiology who are not epidemiologists, you will look at the team of people who published something and there is no epidemiologist on the team and draw conclusions in epidemiology about having the right interpretation or the right caveat and

check and balances are in place I think we will see more of that. And then it will cause confusion, you know, it is a new shiny thing and there are a lot of people who gravitate towards it because it is really cool, it is really fun. Without the adequate grounding in epidemiological principles and epidemiological training, there is danger there and where I think the real danger is when... I mean it is fine when digital methods and conventional methods agree. When you start to see disagreement between digital methods and conventional methods and you start seeing arguments when digital methods must be better because there is more data, that's where I think there is potential of concern. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

L'attitude de réserve envers la pratique de l'épidémiologie numérique peut aussi s'illustrer par la complémentarité entre les deux approches épidémiologiques. En effet, plusieurs participants considèrent cette nouvelle pratique comme un complément à l'épidémiologie traditionnelle (entretien 6Bxviii, entretien 6Biii, entretien 5Cxxxii). En effet, selon ces chercheurs, la pratique de l'épidémiologie numérique pourrait aider notamment à obtenir des échantillons populationnels plus larges (entretien 6Bxviii), à tester des relations entre les phénomènes observés dans les données des médias sociaux par exemple et ceux observés dans les données des systèmes de surveillance (entretien 5Cxxxii), à étudier des choses impossibles à mesurer autrement comme des normes et des valeurs au sein de certains groupes à l'aide de données extraites des discours sur Twitter notamment (5Uxxxvi) et à accélérer les processus de surveillance des maladies (entretien 6Fix). Cette complémentarité est aussi soulignée dans la littérature sur l'épidémiologie numérique, notamment dans le cadre de la pharmacovigilance combinant des données de sources traditionnelles et des données générées par les patients, contribuant ainsi à renforcer les systèmes de santé publique (Salathé, 2016).

« Donc je pense que ça ne va pas remplacer, ça va compléter qu'est-ce qu'on a présentement. Je vois ça comme un plus, mais ce n'est pas pour effacer ce qu'on a présentement, mais c'est vraiment pour ajouter et nuancer. » (Chercheur - épidémiologie sociale - entretien 6Bxviii)

« I think it is an important thing to epidemiology, but I don't see it taking over epidemiology and I don't see epidemiology taking it over. I think that perhaps it could be a very useful skill set tool to have as part of the epidemiology tool kit. » (Chercheur - épidémiologie clinique - entretien 6Biii)

« And then about social media, again it depends on your research question, but I think it can be a nice complement, like you know, you can even test a relationship that you're seeing on surveillance data and you can use social media data to see if you're finding like the same

direction, the magnitude would probably be different but again, dependently on your research question, it could be a nice way to test what you're seeing in surveillance data in terms of robustness. » (Chercheur - épidémiologie sociale - entretien 5Cxxxii)

« L'épidémiologie numérique pourrait rajouter une valeur à nos recherches qu'on n'a peut-être pas, de regarder plus ce qu'on appelle les big data, donc on fait beaucoup de recherche avec des échantillons qui sont très très petits qui ne sont pas représentatifs de l'ensemble de la population donc aller plus large. » (Chercheur - épidémiologie sociale - entretien 6Bxviii)

« Donc caractériser des choses qu'on ne peut pas mesurer autrement. Là je n'ai pas d'exemple précis. Par exemple, on pourrait parler de valeurs ou de normes ou... je ne sais pas, là vous me prenez un peu de court, je réfléchis à voix haute. Je me dis par exemple, des gens qui sont spécialisés dans l'analyse de discours pourraient aller chercher dans des Tweets ou dans des post Facebook des choses intéressantes. Pour justement voire comment certaines caractéristiques sociales ici par exemple les normes ou les attitudes qui prévalent dans certains groupes. Donc c'est ça, je n'ai pas d'exemple précis parce que je ne me suis jamais lancé là-dedans, mais j'ai l'impression qu'il y a énormément d'utilité de façon générale pour compléter des données qui ne sont pas accessibles autrement pour décrire des phénomènes qui se manifestent directement aussi à travers les réseaux sociaux. Donc on n'a pas de lecture possible à l'extérieur de ces médias-là. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

« The main problems of current surveillance systems: they are slow, it doesn't come in real time, and it is not tied to a location per say. We can have data from people in Hamilton but with cell phones, we can locate people exactly on what street they are. You can use that to identify people with the same condition in a specific area or if there is a spreading somewhere or follow movement. You can do that with digital technology. You can do this with actual surveillance, but it is much slower. » (Chercheur - épidémiologie sociale - entretien 6Fix)

4.2.4 Les perceptions en épidémiologie clinique

Quelques-uns des participants à notre recherche ont aussi fait allusion à une distinction entre l'épidémiologie clinique et l'épidémiologie sociale (désignée aussi comme l'épidémiologie de santé publique) lorsqu'ils ont partagé leurs perspectives envers la pratique de l'épidémiologie numérique. Par exemple, un participant en épidémiologie sociale se considère différent vis-à-vis des chercheurs en épidémiologie clinique (entretien 6Bii). Un participant en épidémiologie clinique a clairement fait entendre que cette nouvelle pratique a une place en épidémiologie sociale et non en épidémiologie clinique (entretien 6Fviii). En soulignant qu'il n'a jamais pratiqué l'épidémiologie numérique, ce participant a mentionné que cette nouvelle pratique pourrait produire des recherches valides si celles-ci portaient sur des questions de santé publique. Il a de la

difficulté à imaginer que la pratique de l'épidémiologie numérique puisse être appropriée pour des questions d'ordre biomédical et clinique.

« I am public health epidemiologist to distinguish myself from the clinical people. » (Chercheur - épidémiologie sociale - entretien 6Bii)

« Having not used them, I would expect that they would be reliable, and the validity would depend on probably more public health questions and their inference based on kind of questions like biomedical, clinical questions certainly. I can't imagine a use case for those. I think it would be for public health concerns. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

Cependant, l'un des chercheurs en épidémiologie sociale pratiquant l'épidémiologie numérique perçoit les choses différemment puisqu'il estime que cette nouvelle pratique peut toujours être utile, et ce, peu importe s'il s'agit d'une recherche sociale ou d'une recherche clinique (entretien 8Sxx). Ce participant appuie ce point de vue sur l'idée que tout le monde est constamment connecté à des outils numériques et que, selon lui, il s'avère inconcevable d'ignorer ces outils dans le cadre de recherches épidémiologiques.

De plus, ce participant affirme que les données des systèmes de santé et les données d'Internet ne sont pas mutuellement exclusives. Il illustre son idée en expliquant une situation où il constate qu'un problème du système de soin de santé peut être résolu en utilisant des données d'Internet ou des téléphones intelligents. Ce participant donne l'exemple des capacités de prédiction du stress, de l'anxiété et de la dépression des personnes à l'aide des données générées par des outils numériques et Internet. Selon ce participant, si les bons algorithmes d'intelligence artificielle sont développés, il y a des chances de prédire et de prévenir le suicide. Les données utilisées n'ont rien à voir avec le système de soins de santé puisque celles-ci proviennent de l'extérieur de ce système, mais ces données peuvent se révéler bénéfiques pour les soins de santé, selon ce participant. Une telle perspective apporte une vision différente sur cette séparation entre l'épidémiologie sociale et l'épidémiologie clinique par rapport à la pratique de l'épidémiologie numérique puisque certains chercheurs ayant l'expérience de cette nouvelle pratique voient des rapprochements importants entre des recherches cliniques et des recherches sociales.

« I think you've got to ask the question: is there anything that we do in our lives and days that are not linked to a digital tool? We are connected to our devices, right? When they are so interconnected, how can we ... for any research whether is it clinical research whether is it social research, how can we ignore these tools? » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

«... You can prevent people from getting into the health care system if you understand data that's coming from the Internet. So, they seem at this moment mutually exclusive, but I will give you an example where you can see an intersection between a health care system problem that could be solved using data from the Internet or data from smartphones. For example, there is a way to predict people's mood stress, anxiety and depression using data from digital tools and from the Internet. If we develop the right artificial intelligence algorithms, there is a chance to predict and prevent suicide. Now that's a benefit to the health care system, right? The data that are used has nothing to do with the health care system it's coming outside of it. So that's how I would put it. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

Dans une autre perspective, plusieurs chercheurs ayant participé aux entretiens sont incertains de la manière dont la pratique de l'épidémiologie numérique peut s'appliquer aux recherches biomédicales (entretien 6Fxi), aux recherches cliniques (entretien 5Cxxxi, entretien 10Txxii) et aux essais cliniques aléatoires (entretien 6Fvii). Par exemple, un chercheur en épidémiologie clinique a affirmé que les données qu'il utilise sont presque toujours liées à des mesures biologiques telles que le fonctionnement d'un organe et des tests de sang (entretien 6Fxi). Un autre chercheur en épidémiologie clinique étudiant le virus du papillome humain a exprimé le même point de vue de manière très élaborée en soulignant que la pratique de l'épidémiologie numérique lui semble moins importante pour l'épidémiologie clinique parce qu'il utilise des données cellulaires et des données microbiologiques à partir de prélèvements auprès de femmes enceintes (entretien 5Cxxxi).

Il a également été constaté qu'un chercheur en épidémiologie sociale est très incertain de la place de la pratique de l'épidémiologie numérique dans le cadre d'essais cliniques aléatoires (entretien 6Fvii). Un autre chercheur en épidémiologie sociale pense de manière similaire en expliquant la différence entre la recherche clinique et la pratique de l'épidémiologie numérique (entretien 10Txxii). La première se ferait en quelque sorte dans une bulle et la seconde analyserait plutôt le monde réel selon les propos de ce chercheur. Pour ce dernier, la recherche clinique étudierait les

gens comme ils se comportent alors que la recherche fondée sur la pratique de l'épidémiologie numérique étudierait ce que les gens disent sur leurs comportements.

« The survey data that I use is almost always tied to biological measures - different types of measures of people including organ function or blood test. How it would apply to biomedical research, that I am very hesitant to comment, I am not sure. » (Chercheur - épidémiologie clinique - entretien 6Fxi)

« C'est moins important pour l'épidémiologie clinique comme je fais je pense. En fait c'est le besoin qui est peut-être pas là pour moi, tu sais pas exemple, je peux penser aux données cellulaires, je ne sais pas ... parce que les paramètres sur lesquels que j'étudie ou les choses j'étudie par exemple, sur le virus du papillome humain, les charges virales au niveau du col ou au niveau des tissus, tu sais on regarde si ça va se transférer au placenta, on regarde pourquoi ça se transfère, pourquoi ça se transmet, tu sais on est très moléculaire, très tu sais je ne sais pas jusqu'à quel point ces données-là sont pertinentes pour moi... c'est tellement clinique puis microbiologique que je vois assez difficilement parce que moi je prends des patientes et enfants entre autres dans ce programme de recherche j'ai recruté 1500 femmes en début de grossesse dont au premier trimestre on prend les prélèvements, on mesure un virus quelques fois durant la grossesse puis on va voir si ça se transmet à l'enfant puis pourquoi ça se transmet à l'enfant et qu'est-ce que ça fait chez l'enfant. C'est des mesures aussi de charge virale, tu sais c'est très moléculaire, on va mesurer des variantes dans le code génétique du virus pour voir si les variantes ont un impact différent pour des personnes différentes, on va voir le développement d'anticorps, on regarde s'il y a un transfert d'anticorps, tu sais c'est très clinique, très moléculaire, c'est de l'épidémiologie clinique. » (Chercheur - épidémiologie clinique - entretien 5Cxxxix)

« A lot of my colleagues are doing very traditional clinical epidemiology where they are doing randomized control trials and I am not sure how digital epidemiology could benefit for randomized control trials but there might be a role. » (Chercheur - épidémiologie sociale - entretien 6Fvii)

« I think there is this question about research that's kind of done in a bubble, that kind of clinical kind of contained research versus this kind of real world and it could move us into that second category so studying people as they behave rather than as they tell us they behave. » (Chercheur - épidémiologie sociale - entretien 10Txxii)

Un des participants en épidémiologie clinique verrait le potentiel de la pratique de l'épidémiologie numérique pour des études en toxicologie et moins pour des recherches cliniques (entretien 5Cxxxix). Les travaux de ce chercheur portent sur le virus du papillome humain en examinant entre autres des charges virales. Ce chercheur pense que la pratique de l'épidémiologie numérique serait

moins pertinente pour ses recherches parce qu'il perçoit le genre de données qu'il utilise comme très différentes par rapport à celles utilisées dans le contexte de cette nouvelle pratique.

Un autre participant en épidémiologie clinique a déclaré qu'il est sceptique quant aux diagnostics de santé établis avec la pratique de l'épidémiologie numérique (entretien 7Mxxi). Pour avoir vraiment confiance envers cette nouvelle pratique, il suppose qu'il lui faudrait prendre connaissance de recherches comparatives. Ce chercheur soutient que pour avoir un diagnostic valide, les allergies doivent être testées, car beaucoup de gens se considèrent comme étant allergiques à quelque chose et qu'il existe de nombreux types d'allergies. D'après ce chercheur, la pratique de l'épidémiologie numérique ne fournira pas nécessairement les résultats de ces tests. Ce chercheur souligne aussi qu'il existe un risque d'un mauvais diagnostic, dans le sens où le fait de penser à être allergique pourrait influencer l'état de santé. La pratique de l'épidémiologie numérique pourrait alors donner lieu à une interprétation erronée selon ce chercheur si les tests sont pris au-delà de la portée et du contexte dans lesquels ils devraient être utilisés.

Ce regard envers l'épidémiologie clinique est aussi partagé par plusieurs chercheurs en épidémiologie sociale. Par exemple, l'un des participants en épidémiologie sociale ne pratiquant pas l'épidémiologie numérique a affirmé qu'il pense que cette nouvelle pratique peut être utile pour beaucoup de gens, mais tout dépend de l'orientation des travaux de recherche (entretien 8Hxvi). Selon ce participant, il y a encore de nombreuses études portant sur des phénomènes cliniques en épidémiologie dont les travaux de recherche ne sont pas orientés vers le numérique.

« Tu sais je joue avec de grosses bases de données comme celles-là, mais ce n'est pas des bases de données comme la RAMQ, je n'ai jamais joué avec ça, moi je joue avec beaucoup de données qui sont gigantesques souvent, mais qui sont très très cliniques... En fait c'est le besoin qui est peut-être pas là pour moi, tu sais pas exemple, je peux penser aux données cellulaires, je ne sais pas ... parce que les paramètres sur lesquels que j'étudie ou les choses j'étudie par exemple, sur le virus du papillome humain, les charges virales au niveau du col ou au niveau des tissus, tu sais on regarde si ça va se transférer au placenta, on regarde pourquoi ça se transfère, pourquoi ça se transmet, tu sais on est très moléculaire, très tu sais je ne sais pas jusqu'à quel point ces données-là sont pertinentes pour moi... tu sais moi c'est peut-être pas le bon exemple pour ma recherche à moi, mais tu sais comme en toxicologie par exemple, qui utilise les données, tu sais c'est extrêmement important, C'est moins important pour l'épidémiologie clinique comme je fais je pense » (Chercheur - épidémiologie clinique - entretien 5Cxxxi)

« I have a skepticism for health diagnosis that are made in that way...In order to really have trust, I would have to see comparative research I guess... I think that allergies diagnosis, the type of allergies that I work with are very possible for a lot of people consider themselves having allergies, allergies need to be tested and there are a lot of different allergies. Digital epidemiology is not necessary to provide results of those tests. There is also the concern of over-diagnosis in the sense that thinking about being allergic might influence the health condition. There could be over-interpretation with digital epidemiology. The testing that we do could be over-interpreted if taken beyond the scope and context of where it should be. » (Chercheur - épidémiologie clinique - entretien 7Mxxi)

« I think it would be useful for a lot of people. It will depend on the focus of their specific research work... There is still a lot of people focusing on for example clinical epidemiology not based on the digital side doing work with ... and the hospital ones. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

À la lumière de ces analyses, il est possible qu'il y ait plus d'incertitudes sur la crédibilité de la pratique de l'épidémiologie numérique en épidémiologie clinique que dans d'autres sous-disciplines de l'épidémiologie en raison de la nature des données et des informations utilisées lors des recherches cliniques. Cette vision peut se retrouver autant auprès des chercheurs en épidémiologie clinique qu'auprès des chercheurs en épidémiologie sociale, et ce, surtout si les chercheurs n'ont pas l'expérience de cette nouvelle pratique. Il semble donc que la culture épistémique de la pratique de l'épidémiologie numérique pourrait être vue comme moins bien alignée sur la culture épistémique de l'épidémiologie clinique que sur celle d'autres sous-disciplines comme l'épidémiologie sociale ou environnementale. Il faut cependant noter que pour les chercheurs ayant participé à la recherche de terrain, la pratique de l'épidémiologie numérique semble être associée uniquement aux données. À l'exception d'un chercheur pratiquant l'épidémiologie numérique (entretien 8Sxx), l'usage d'autres dispositifs numériques comme l'intelligence artificielle n'a pas été abordé par ces chercheurs.

4.2.5 La crédibilité de la pratique de l'épidémiologie numérique

La diffusion de la pratique de l'épidémiologie numérique est aussi influencée par la crédibilité de cette nouvelle pratique auprès des chercheurs en épidémiologie. Dans cette section, nous visons à comprendre comment sont jugées les connaissances produites lors de la pratique de l'épidémiologie

numérique à travers les points de vue des chercheurs pratiquant l'épidémiologie numérique et ceux des chercheurs n'ayant pas l'expérience de cette nouvelle pratique.

4.2.5.1 Les points de vue des chercheurs pratiquant l'épidémiologie numérique

D'abord, les analyses effectuées dans le cadre de cette recherche indiquent que tous les participants pratiquant l'épidémiologie numérique considèrent cette nouvelle pratique comme crédible pour des recherches épidémiologiques. En effet, tous ces participants ont été fermes dans leurs propos sur la crédibilité de cette nouvelle pratique lors de la recherche de terrain.

L'un de ces participants en épidémiologie sociale affirme que la pratique de l'épidémiologie numérique est crédible et que le futur de la crédibilité de cette nouvelle pratique dépendra des problèmes résolus (entretien 8Sxx). Par exemple, si le chercheur travaille sur le virus de la COVID-19, la pratique de l'épidémiologie numérique est crédible selon ce participant puisque cet objet d'étude porte sur un enjeu d'actualité. Ce participant soutient que les défenseurs de cette nouvelle pratique devront se concentrer sur la résolution des problèmes de données et sur les enjeux de santé actuels comme la santé mentale, le VIH, la COVID-19 et la situation des communautés autochtones. Selon ce participant, les chercheurs pratiquant l'épidémiologie numérique devraient se concentrer sur ces problèmes s'ils veulent faire en sorte que cette pratique soit crédible.

Ces propos expliquent l'importance que ce chercheur accorde à la pertinence sociale et politique des recherches fondées sur la pratique de l'épidémiologie numérique. Une telle perspective ne démontre pas exactement comment ce chercheur évalue la crédibilité de cette nouvelle pratique. Les propos de l'autre chercheur en épidémiologie sociale pratiquant l'épidémiologie numérique mènent au même constat. Ce chercheur affirme qu'il existe davantage de subventions pour des recherches fondées sur cette nouvelle pratique et que de plus en plus de gens voient le pouvoir de celle-ci (entretien 6Axxiii). Les propos de ce chercheur en épidémiologie sociale indiquent que le pouvoir de la pratique de l'épidémiologie numérique est lié à son efficacité lors de synthèses de connaissances, car cette nouvelle pratique lui a paru très bénéfique pour économiser du temps lors de ses synthèses de connaissances. Ces propos illustrent des intérêts politiques et économiques

envers la pratique de l'épidémiologie numérique et non un point de vue sur la crédibilité de cette pratique.

Le chercheur en épidémiologie environnementale pratiquant l'épidémiologie numérique soutient que celle-ci est crédible si la recherche fondée sur cette nouvelle pratique tient compte de ses apports et de ses limites (entretien 10Ixxviii). Selon ce chercheur, il n'existe pas réellement de limites vis-à-vis de cette nouvelle pratique pour des recherches, mais il appréhende des problèmes en contexte politique. Ce chercheur ne serait jamais à l'aise de se fier à la base des seuls signaux numériques pour les prises de décision, mais si les signaux numériques étaient cohérents avec les signaux conventionnels, il pense qu'ils pourraient renforcer l'argument politique. Selon ce chercheur, les signaux numériques à eux seuls ne sont pas adéquats pour prendre des décisions politiques.

Ce regard envers les signaux numériques est intéressant pour comprendre comment ce chercheur évalue la crédibilité de la pratique de l'épidémiologie numérique parce qu'il implique un critère d'ordre épistémique et un critère d'ordre sociopolitique. Le critère d'ordre épistémique est visible à travers la nécessité à ce que la recherche fondée sur la pratique de l'épidémiologie numérique tienne compte des apports et des limites de cette pratique. Le critère d'ordre sociopolitique concerne la crédibilité de la pratique de l'épidémiologie numérique en contexte politique et plus particulièrement sur la confiance envers les signaux numériques lors des prises de décision en santé publique. Dans cette optique, il est possible de constater que les chercheurs peuvent évaluer la crédibilité d'une pratique de recherche s'appuyant sur des critères sociopolitiques et non pas seulement sur des critères épistémiques. Un chercheur peut donc accorder de l'importance autant à des enjeux sociopolitiques qu'à des enjeux épistémiques lorsqu'il évalue une pratique de recherche.

Ce phénomène peut illustrer la coproduction des connaissances scientifiques, c'est-à-dire que les connaissances scientifiques incorporent et sont aussi incorporées par les pratiques sociales, les identités, les normes, les conventions, les discours, les instruments et les institutions (Jasanoff, 2004). Cette conception de la production des connaissances scientifiques se situe entre le déterminisme technologique et naturel et le déterminisme social puisque celle-ci consiste à tenir

compte des aspects technologiques, scientifiques, conceptuels et sociaux lors des analyses des pratiques de recherche. Ainsi, en se basant sur la coproduction des connaissances scientifiques (Jasanoff, 2004), il serait logique de penser que les propos de ce chercheur pratiquant l'épidémiologie numérique montrent que la crédibilité de cette nouvelle pratique peut être évaluée en fonction de critères d'ordre épistémique, mais aussi selon des critères d'ordre sociopolitique.

« I think the credibility is pretty high as it is right now. The credibility will depend on the kinds of problems we solve. So, if I work on research on Covid-19, it will be credible right away. I think digital epidemiology should focus on the data problems that we have now and on the problems that we have now: mental health, HIV, Covid-19, indigenous communities, I think that this is where digital epidemiology should focus to be credible. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

« I think that the scientific credibility of digital epidemiology is good. I think that more grants are focused on it, and I think that people are seeing the power of it. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 6Axxiii)

« I see the scientific credibility of digital epidemiology quite high with all the checks and balances we've talked about... I don't think it should ever be avoided in different research honestly. I don't see research limitation. I see policy limitations. So, in the kind of work, I do we often move research policy quite quickly, I would never be comfortable doing that based on digital signals alone, but if digital signals were consistent with conventional signals, I think they can strengthen the policy argument. Digital signals alone I don't think they are adequate to make policy decisions on » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

4.2.5.2 Les points de vue des chercheurs ne pratiquant pas l'épidémiologie numérique

D'autres participants ont aussi un discours assez ferme sur la crédibilité de la pratique de l'épidémiologie numérique, même s'ils ne l'ont pas pratiquée dans le cadre de leurs travaux de recherche. Ces chercheurs considèrent cette nouvelle pratique comme étant aussi crédible que d'autres pratiques de l'épidémiologie traditionnelle si elle s'appuie sur une méthodologie adéquate en prenant compte de ses limites et de la nécessité de bien interpréter les données (entretien 6Fviii, entretien 6Biv, entretien 3Exxxiii, entretien 5Cxxx). Ce point de vue a été exprimé par des chercheurs autant en épidémiologie sociale et environnementale qu'en épidémiologie clinique. Par exemple, un des chercheurs en épidémiologie clinique souligne que les méthodes traditionnelles en épidémiologie sont elles aussi soumises à des réserves selon le domaine de recherche et le

contexte de la question posée (entretien 6Fxi). Un autre de ces chercheurs en épidémiologie clinique affirme que la pratique de l'épidémiologie numérique est une chose importante pour l'épidémiologie, mais il ne pense pas qu'elle surpasse l'épidémiologie et que l'épidémiologie prenne le dessus sur cette nouvelle pratique (entretien 6Biii). Ce chercheur pense que cette nouvelle pratique être un outil de compétences très utile à avoir dans la trousse à outils de l'épidémiologie.

« Same as any other data collection. it comes to how the data are interpreted, recognizing the limitation of the data as all epidemiological study. » (Chercheur - épidémiologie sociale - entretien 6Fviii)

« If people establish an approach to use digital epidemiology that is shown to be reliable, then I could see it as being very scientific reliable with the same caveat that any epidemiological approach would have. So, questionnaire data for example, whatever you are providing it via the Internet or over the phone, paper, or tablets are all having caveats and those caveats are dependent on the area of research and in the context of what is being asked. So, if people are able to show that an approach of digital epidemiology can be reliable and an effective tool then I could see as being reliable as any other epidemiological tool. » (Chercheur - épidémiologie clinique - entretien 6Fxi)

« I don't think that it will be any worse than any other science nor any better. I mean I think there will be methods for establishing the validity of any studies. These studies will have to be carried out with validation or calibration of their methods and their approach. In clinical epidemiology we have tools to appraise how good is a randomized trial. I suspect that in digital epidemiology there probably already is development of tools to appraise how these studies are. But those might become more well-known and widespread as more people begin to use these tools or need these tools. » (Chercheur - épidémiologie sociale - entretien 6Biv)

« On par with traditional methods. » (Chercheur - épidémiologie sociale - entretien 3Exxxiii)

« I would say something similar to the scientific credibility of epidemiology itself. I think it is an important thing to epidemiology, but I don't see it taking over epidemiology and I don't see epidemiology taking it over. I think that perhaps it could be a very useful skill set tool to have as part of the epidemiology tool kit. » (Chercheur - épidémiologie clinique - entretien 6Biii)

« I think the scientific credibility is robust if the studies are designed appropriately and account of all the issues related to digital epidemiology. The credibility will be as good as the methodology. » (Chercheur - épidémiologie sociale - entretien 6Fix)

« Ça dépend du devis d'étude, si c'est une étude épidémiologique qui est bien effectuée, les biais sont bien contrôlés, l'analyse est appropriée, donc la crédibilité reste toujours avec les mêmes concepts de base. » (Chercheur - épidémiologie environnementale - entretien 5Cxxx)

L'un des participants en épidémiologie sociale soutient que la pratique de l'épidémiologie numérique est totalement crédible et que celle-ci fournit un signal dans les données des recherches des gens sur Internet, notamment (entretien 6Bii). Selon lui, ce signal ne signifie pas que les gens effectuant des recherches Internet sur la COVID-19 ont ce virus, mais que ceux-ci ont des inquiétudes. Pour ce dernier, il s'agit d'un indicateur épidémiologique valide qu'il faut interpréter correctement en fonction du contexte. D'ailleurs, la nécessité de bien contextualiser les indicateurs n'est pas apparue avec l'émergence de la pratique de l'épidémiologie numérique puisque celle-ci a toujours fait partie de la science.

« 100% credible. It is a signal, it may not be the signal that you think it is, but it's a signal. So, for example, if you get an history in a population and you see everybody is querying Covid-19, it is not a signal that you have the virus, it is a signal that people are worried about that. But it is a valid epidemiological indicator that people are worried so for me this is the question to me that we must calibrate exactly what it is we are looking at but that is always the part of science, it is not new. We always must do that. So, to me it is very credible and still very scientific. It is just a question that we must see what it means. » (Chercheur - épidémiologie sociale - entretien 6Bii)

Plusieurs participants n'ayant pas l'expérience de la pratique de l'épidémiologie numérique sont incertains vis-à-vis de la crédibilité de cette nouvelle pratique. Ces participants affirment que la crédibilité de la pratique de l'épidémiologie numérique dépend de plusieurs facteurs comme la qualité de l'interprétation des données (entretien 5Cxxxii), le travail du chercheur sur les données (entretien 5Gxxv), les habiletés des gens à pratiquer adéquatement cette façon de faire de l'épidémiologie (entretien 9Qxxiv), les connaissances du chercheur sur les données utilisées, sur les biais et sur les limites de ces données (entretien 6Pvi) ainsi que la crédibilité des sources des bases de données et la manière dont les données peuvent répondre aux questions de recherche (entretien 8Hxvi).

Ces facteurs font ressortir des critères d'ordre épistémique dans l'évaluation de la crédibilité de la pratique de l'épidémiologie numérique auprès des chercheurs n'ayant pas l'expérience de cette nouvelle pratique. Il s'agit de critères d'ordre épistémique parce que ceux-ci concernent la discipline de l'épidémiologie et les façons de connaître les phénomènes étudiés. Le terme « épistémique » est significatif. Il indique que la connaissance, les objectifs et les pratiques liés à la vérité représentent les éléments centraux des domaines scientifiques (Knorr-Cetina et Reichmann,

2015), y compris les disciplines comme celle de l'épidémiologie. L'analyse de ces propos indique qu'une recherche fondée sur la pratique de l'épidémiologie numérique est crédible si les auteurs démontrent une bonne capacité à interpréter et à comprendre les données numériques utilisées, l'habileté à pratiquer adéquatement cette façon de faire de l'épidémiologie, une bonne connaissance des biais et des limites de ces données, la fiabilité des sources des bases de données et l'aptitude à répondre aux questions de recherche épidémiologique.

« I think it depends. The approach like it seems more widely accepted, I do find that we must do this with diligence which is normal like we should be doing with any kind of epidemiological approach (digital epidemiology or classical approach). The data, the methods, model checking, but I do find that it's becoming more and more a mainstream in terms of acceptability, but again we don't want to go too much on the other side, where as soon as we can find some fancy method, we publish that and loose the scrutiny of the approach. And interpretation is quite important too for us to foster and make sure that we have the right interpretation. » (Chercheur – épidémiologie sociale - entretien 5Cxxxii)

« It just depends... the quality is so dependent on the data work that the researcher does and what these data are being used for. It is hard to say exactly what I think about that because it is all dependent on all these factors. » (Chercheur – épidémiologie environnementale - entretien 5Gxxv)

« It depends, I think, on the skills and ability and honestly on which people approach it. There is nothing about it that says that it would have to be poorly researched. We still need to be exposed to the same kind of rigor as the other epidemiology that we've done. Did it work? Did other people get it to work? Does it work the same way every time? If it doesn't, do we understand why? These are kind of broad questions, but they are going to be critically important when people go in and treat the material they get from Facebook, Twitter or Instagram, and start making broad generalizations about something that is not a science, it will just quickly destroy the whole field. It doesn't have to be that way for everything we talked about. It could be as rigorous in its own way and potentially way more illuminating for some of the things that we need a better understanding of, or we need better interventions for. » (Chercheur – épidémiologie sociale - entretien 9Qxxiv)

« Well again if there are any existing databases like that FoodRepo that you just mentioned, I think that if it is a credible database you need to make sure who has access to enter information into that, where the information come from. Once we have a better understanding of that, then I think ... I think that there will also biases to think about ... it will allow us a lot of power because if the data is good, there will be a lot of valuable information that we could use. I think that is going to be one of the advantages for sure. Again, we need to be very careful in terms of understanding the limitations of the data. How people are entering the data and for what purposes I think that this is important. It shouldn't be just something that we look at and just use it. I think we really need to understand where that data is coming from. » (Chercheur – épidémiologie sociale - entretien 6Pvi)

« It will depend on where the data sources are coming from and on what kind of questions you are trying to answer. your credibility is going to be focused on that. If you are using some strong set of data from some other sources to answer your research question, I think you will have a strong credibility. If you are taking poorly collected data and trying to make some conclusions around that then it's going to be less credible. I think it will depend on data sources and how it matches your question. » (Chercheur – épidémiologie sociale - entretien 8Hxvi)

La recherche de terrain indique aussi que plusieurs participants considèrent la pratique de l'épidémiologie numérique comme crédible dans le cadre de recherches exploratoires (entretien 6Axix, entretien 9Rxiv) afin de repérer des régions où il peut y avoir des problèmes sur certaines maladies (entretien 9Rxiv), de recherches sur les opinions et préférences des gens (entretien 6Axxiii, entretien 5Cxxxii) et pour générer des hypothèses (entretien 6Axxiii). Certains chercheurs ne voient pas la pratique de l'épidémiologie numérique comme étant crédible pour déterminer l'incidence de maladies (entretien 7Mxxi, entretien 6Axix). L'un de ces chercheurs a affirmé qu'une recherche fondée sur la pratique de l'épidémiologie numérique nécessiterait d'être validée (entretien 7Mxxi). Ce chercheur serait sceptique à l'égard d'une recherche affirmant qu'il est possible de déterminer avec précision l'asthme chez les enfants ou les allergies alimentaires chez les enfants en fonction du type de recherche effectuée sur le Web. Ce chercheur serait ouvert à la possibilité qu'une telle recherche soit valide, mais pour lui donner le même type de crédibilité scientifique, il voudrait une validation avec des méthodes plus traditionnelles.

« D'après ce que vous décrivez, je pense que vois là un outil qui aurait un bon potentiel pour la recherche exploratoire pour voir si par exemple dans une population X il y a beaucoup de gens qui font des recherches sur certaines maladies donc ça permettrait de sensibiliser au fait que c'est peut-être un problème qui est sérieux dans cette région-là. » (Chercheur – épidémiologie sociale - entretien 9Rxiv)

« If we are relying on things like Twitter matter things, we will get some insights as more as exploratory type, insights and things like that, but we won't get really... you know, we will not know incidents of certain diseases. Without a denominate, you can't do those type of things. » (Chercheur – épidémiologie clinique - entretien 6Axix)

« Well, I think it depends on your question. If your question is looking at people's preferences, or their opinion, if you are thinking about how they experienced a health event or for example, the pharmacovigilance one... it is more like a hypothesis generating kind of thing then a high hypothesis trusting kind of thing. » (Chercheur – épidémiologie sociale - pratiquant de l'ÉN - entretien 6Axxiii)

« Tout dépend en fait de la recherche qui a été faite. Tu sais sur des trucs sociaux, on s'interroge sur des opinions, on veut plus de recherche qualitative, je pense que j'aurais tout à fait confiance en ça compte tenu de la généralisation qu'on peut en faire. » (Chercheur – épidémiologie clinique - entretien 5Cxxxii)

« I think that it would require validation as I said earlier and particularly to studies that apply to allergies or asthma if that's the question. I would be skeptical of a study saying that I can accurately determine asthma in children or food allergies in children based upon which type of research is done on the web. I would be open to the possibility that it could occur but in order to give it the same type of scientific credibility, I would want to see a validation with more traditional methods. » (Chercheur – épidémiologie clinique - entretien 7Mxxi)

Ces résultats peuvent indiquer une différence dans les points de vue des chercheurs en épidémiologie, selon le type de recherche épidémiologique. En effet, il se peut que les chercheurs en épidémiologie soient plus enclins à envisager la crédibilité la pratique de l'épidémiologie numérique dans le cadre d'études portant sur les aspects sociétaux et environnementaux des enjeux sur la santé plutôt que sur des études ayant pour but de déterminer la prévalence réelle d'une maladie. Dans une telle logique, la pratique de l'épidémiologie numérique serait associée à une culture épistémique particulière, c'est-à-dire à une façon différente de connaître les phénomènes étudiés ne permettant pas de connaître la prévalence réelle des maladies. Il est possible que la pratique de l'épidémiologie numérique soit vue comme crédible dans un contexte de recherche régi par une culture épistémique axée sur les aspects sociétaux et environnementaux des enjeux sur la santé.

Quelques participants n'ayant pas pratiqué l'épidémiologie numérique pensent que celle-ci pourrait devenir crédible dans le futur. Des chercheurs en épidémiologie sociale soutiennent que cette pratique a du potentiel, mais que celle-ci a besoin de démontrer quelques grands succès pour traiter certains types de questions (entretien 6Dxii, entretien 6Bxviii). Un autre participant en épidémiologie environnementale affirme que cette pratique pourrait devenir crédible si les limites des données sont considérées et que ces données sont utilisées de la manière la plus appropriée (entretien 1Oxxxv). Ce chercheur souligne aussi qu'il voudrait qu'il y ait davantage de moyens de relier les données de différents types de données aux données sur la santé, mais là encore, les liens

entre différentes sources de données sont difficiles à établir sans oublier les problèmes quant à la confidentialité et à l'éthique.

Un autre chercheur en épidémiologie clinique estime que la pratique de l'épidémiologie numérique sera critiquée lors de ses premières utilisations et qu'ensuite, celle-ci deviendra une pratique commune n'étant pas moins efficace que les autres pratiques courantes de recherche (entretien 6Fviii). Un autre participant en épidémiologie clinique a affirmé que cette nouvelle pratique produira d'autres ensembles de données s'ajoutant à ce qu'il y a maintenant et que celle-ci sera une des façons de traiter des problèmes complexes (entretien 6Bi). Une fois que les données et les connaissances seront mieux comprises dans les méthodes utilisées, elles ne seront plus qu'une source de données comme les autres selon ce chercheur qui semble associer la pratique de l'épidémiologie numérique uniquement avec les données.

« From the perspective that it is a tool, like any other tool over time epidemiologist were doing research by going from door to door and asking people questions and tracking trends of disease through the data they are collecting from door to door. As a new tool, it just needs to go through the same steps as with any other new tool and so it will grow in credibility over time as people recognize the unique biases of attributed to the technology and figure out ways around it or restricted use the situation where these biases are less likely to be present. So, I could see that there is potential for work in this area, but it needs demonstrating it needs some big successes to show that this is something we all need to at least think about using to answer certain types of questions. » (Chercheur – épidémiologie sociale - entretien 6Dxii)

« Ça pourrait être vraiment très crédible. Il y a juste beaucoup de chemin à faire encore pour démontrer la validité de cette approche-là, mais quand ça va être démontré, ça peut prendre plusieurs années et puis ça prend toujours une question de temps aussi pour changer les façons de faire. Pas seulement changer, mais aussi rajouter à ce que nous avons présentement. Je pense que ça va devenir une approche utile, crédible, mais ce ne l'est pas encore présentement. Donc je pense que c'est juste une question de temps et une question de faire les bonnes choses pour démontrer ça. Et quand ça va l'être, si on peut le démontrer avec des données, je pense que les scientifiques croient aux données et aux résultats. Donc si on peut démontrer que cette approche est bonne et que ça ajoute beaucoup de choses, je pense que personne ne va être contre ça. Après ça c'est une question de préférence aussi. Ce n'est pas tout le monde qui aime l'aspect numérique aussi. On est tous différents. Il faut être d'accord avec ça. Et ça c'est une autre question. » (Chercheur – épidémiologie sociale - entretien 6Bxviii)

« Moving forward, I think that a way to reconcile the limitations of the data and use the data in the most appropriate way, I think it quite useful. Personally, I would like to see more ways of being able to link data across different types of data to the health data but again it comes with

challenges on the linkage side but also on the privacy side and ethic side. » (Chercheur – épidémiologie environnementale - entretien 1Oxxxv)

« On his first use, it will be criticized and then it will become probably a common place. I think objectively depending on the use case and depending on the source it could vary but overall, it is probably not going to be any worse than existing measurement done in a research environment. » (Chercheur – épidémiologie clinique - entretien 6Fviii)

« It is going to be one set of data. It is going to be one way of addressing complex problems. So, my guess is that once the data become better understood in the methods used to extract knowledge from those data and that knowledge becomes more understood, it will be just another source of data. In terms of challenges to the validity of digital epidemiology, it is a matter of working through some of these initial validity issues and linking issues. » (Chercheur – épidémiologie clinique - entretien 6Bi)

Cette analyse révèle que les chercheurs ne pratiquant pas l'épidémiologie numérique semblent s'appuyer sur des critères d'ordre épistémique lorsqu'ils évaluent la crédibilité de cette nouvelle pratique. Par exemple, certains ont affirmé que la crédibilité de la pratique de l'épidémiologie numérique dépend entre autres de la qualité de l'interprétation des données de la recherche et du contrôle des biais. Ces critères paraissent différents par rapport à ceux mentionnés par un chercheur ayant l'expérience de la pratique de l'épidémiologie numérique. Les analyses des propos de ce chercheur font ressortir un critère d'ordre sociopolitique en plus d'un critère d'ordre épistémique lorsqu'il évalue la crédibilité de cette nouvelle pratique. Ces constats rappellent la symétrie dans la coproduction des connaissances (Jasanoff, 2004), c'est-à-dire que les connaissances produites lors des recherches impliquent des enjeux épistémiques, mais aussi des enjeux sociaux et politiques.

En somme, nos analyses sur la diffusion de la pratique de l'épidémiologie numérique montrent qu'il y a peu de connaissance et peu d'expérience de cette nouvelle pratique auprès des chercheurs en épidémiologie ayant participé à la recherche de terrain. En effet, moins de 10 % de l'échantillon des participants à la recherche de terrain pratiquent l'épidémiologie numérique, notamment un chercheur en épidémiologie environnementale et deux chercheurs en épidémiologie sociale dont un dirige un laboratoire en épidémiologie numérique. Trois attitudes envers cette nouvelle pratique ont été identifiées lors de ces analyses, à savoir l'adoption, les résistances et les réserves. Seulement un participant adopte entièrement la pratique de l'épidémiologie numérique parmi le groupe de chercheurs ayant participé à la recherche de terrain. Pour ce chercheur, cette nouvelle pratique doit

être un succès parce qu'il faut déterminer comment rassembler des big data de différents secteurs et de différentes disciplines pour comprendre la santé des populations. Ce chercheur pense aussi que la pratique de l'épidémiologie numérique peut aider à résoudre les problèmes récurrents en ce qui concerne les biais de sélection des échantillons populationnels, car une large proportion des personnes dispose d'au moins un dispositif numérique (comme un téléphone intelligent), permettant ainsi d'élargir les échantillons à au moins 90 % des populations au lieu de rejoindre seulement 50 à 60 % avec les méthodes traditionnelles.

Deux participants ont exprimé des résistances vis-à-vis de la pratique de l'épidémiologie numérique. Ces participants n'ont jamais expérimenté la pratique de l'épidémiologie numérique et ils la voient comme étant inappropriée dans le cadre d'études épidémiologiques. L'un de ces participants estime que les données numériques ne sont pas utiles pour des recherches en santé. L'autre participant éprouve des doutes quant à la qualité des données numériques et il est plutôt sceptique en raison des biais liés à ces données, démontrant ainsi de l'incertitude vis-à-vis des signes et des apparences visibles à travers les analyses de données numériques.

L'attitude observée le plus souvent au sein des participants à cette recherche est l'attitude de réserve puisque ces chercheurs voient la pratique de l'épidémiologie numérique comme une façon de faire de l'épidémiologie présentant du potentiel et aussi des lacunes. Selon ces participants, le chercheur doit évaluer les bienfaits et les inconvénients de la pratique de l'épidémiologie numérique comme toute autre pratique de recherche. Cette attitude envers la pratique de l'épidémiologie numérique se retrouve chez les participants n'ayant pas expérimenté cette nouvelle pratique et aussi chez deux participants ayant fait l'expérience de cette nouvelle pratique. Un seul participant ayant fait l'expérience de la pratique de l'épidémiologie numérique montre une adoption complète envers la pratique de l'épidémiologie numérique et les deux autres participants expérimentés ont véhiculé une attitude prudente envers cette nouvelle pratique. En donnant l'exemple du moteur de recherche Google, un chercheur ne pratiquant pas l'épidémiologie numérique a souligné qu'il faut comprendre comment les technologies numériques fonctionnent parce que, selon lui, il existe une logique derrière la façon dont les algorithmes de ces technologies classent et priorisent les informations. Ainsi, le moteur de recherche Google peut être vu comme un détecteur agissant comme un dispositif médiateur interposé entre l'expérience et les phénomènes. Il est donc possible

que la pratique de l'épidémiologie numérique reflète une culture épistémique différente par rapport à l'épidémiologie traditionnelle parce que l'algorithme semble jouer un rôle dans les prises de décision et dans le déroulement des recherches lors de la pratique de l'épidémiologie numérique.

Ces analyses révèlent aussi des attitudes différentes envers la pratique de l'épidémiologie numérique auprès des chercheurs en épidémiologie en fonction de la sous-discipline de l'épidémiologie. Il est possible que des chercheurs en épidémiologie considèrent la pratique de l'épidémiologie numérique comme étant moins appropriée pour l'épidémiologie clinique puisque celle-ci utilise des données cliniques comme des tests de sang des patients qui s'avèrent très différentes par rapport aux données numériques comme celles générées par les médias sociaux ou les moteurs de recherche Internet. La culture épistémique de l'épidémiologie clinique pourrait ainsi être perçue comme différente par rapport à celle de la pratique de l'épidémiologie numérique. Cette distinction entre l'épidémiologie clinique et la pratique de l'épidémiologie numérique serait fondée sur les différences d'une part dans les types de données et les sources de données utilisées et d'autre part dans les objets d'études (des opinions versus l'incidence d'une maladie).

Quant à la crédibilité, nos analyses révèlent que les chercheurs ne pratiquant pas l'épidémiologie numérique semblent s'appuyer sur des critères d'ordre épistémique lorsqu'ils évaluent la crédibilité de cette nouvelle pratique. Ces analyses indiquent que la crédibilité sur le plan épistémique d'une recherche fondée sur la pratique de l'épidémiologie numérique se caractérise par la capacité à interpréter et à comprendre les données numériques utilisées, l'habileté du chercheur à pratiquer adéquatement cette façon de faire de l'épidémiologie, une bonne connaissance des biais et des limites de ces données, la fiabilité des sources des bases de données et la capacité de la recherche à répondre aux questions de recherche épidémiologique. Pour la situation où le chercheur pratique l'épidémiologie numérique, nos analyses font ressortir un critère d'ordre sociopolitique en plus d'un critère d'ordre épistémique lors de l'appréciation de la crédibilité de cette nouvelle pratique. Ce constat renvoie de nouveau au principe de la symétrie dans la coproduction des connaissances (Jasanoff, 2004), c'est-à-dire l'importance de considérer à la fois les enjeux épistémiques et les enjeux sociaux et politiques pour comprendre comment les connaissances sont produites.

4.3 La conclusion sur la pratique de l'épidémiologie numérique

L'objectif de ce chapitre est de comprendre comment est perçue la pratique de l'épidémiologie numérique par les chercheurs épidémiologistes canadiens. Les discussions avec les participants à la recherche de terrain laissent supposer que la pratique de l'épidémiologie numérique n'est pas très présente dans la culture épistémique de la discipline de l'épidémiologie puisqu'une bonne partie d'entre eux étaient peu ou pas familiers avec cette pratique émergente et incertains de la manière de définir cette nouvelle pratique.

Sur la base des résultats de notre recherche, les points de vue sur la manière de concevoir la pratique de l'épidémiologie numérique diffèrent d'un chercheur à l'autre. Par exemple, certains chercheurs ont affirmé que l'épidémiologie numérique consiste à utiliser des technologies et des nouvelles technologies pour faire de l'épidémiologie. D'autres participants ont plutôt apporté des notions plus précises sur l'épidémiologie numérique de façon similaire à des travaux dans la littérature en faisant allusion à l'utilisation des données de Twitter et celles produites par des technologies mobiles, au traçage numérique, à l'intelligence artificielle, à l'apprentissage automatique, aux données massives, aux données générées par Google et à l'épidémiologie computationnelle dans leur façon de définir cette approche. Quelques participants ont exprimé des incertitudes sur la nouveauté de la pratique de l'épidémiologie numérique parce que certaines de ses méthodes sont déjà utilisées en épidémiologie. Ce constat s'explique par l'idée qu'une méthode de recherche est souvent fondée sur ce qui l'a précédée.

En ce qui concerne la diffusion, il y a peu de connaissance, peu d'expérience et beaucoup d'incertitudes vis-à-vis de la pratique de l'épidémiologie numérique auprès des chercheurs en épidémiologie ayant participé à la recherche de terrain. Seulement trois chercheurs parmi les 36 participants pratiquent l'épidémiologie numérique : un chercheur en épidémiologie environnementale et deux chercheurs en épidémiologie sociale dont un dirige un laboratoire en épidémiologie numérique. L'attitude envers la pratique de l'épidémiologie numérique observée le plus souvent est l'attitude de réserve puisque la plupart des participants voient cette nouvelle pratique comme une façon de faire de l'épidémiologie comportant des bienfaits et des inconvénients. Cette attitude se retrouve chez les participants n'ayant pas expérimenté cette nouvelle pratique et aussi chez deux participants ayant fait l'expérience de cette nouvelle pratique.

Un seul chercheur adopte entièrement la pratique de l'épidémiologie numérique et deux chercheurs ont une attitude de refus vis-à-vis de cette nouvelle pratique.

En outre, notre recherche montre que les chercheurs en épidémiologie se montrent plutôt sceptiques envers la pratique de l'épidémiologie numérique en raison des biais liés aux données numériques démontrant ainsi de l'incertitude vis-à-vis des signes et des apparences visibles à travers les analyses de données numériques. Certains chercheurs en épidémiologie expriment une attitude prudente envers la pratique de l'épidémiologie numérique parce qu'ils estiment qu'il faut saisir la façon dont les technologies numériques classent et priorisent les informations, démontrant ainsi l'idée que ces technologies puissent être vues comme des détecteurs agissant comme des dispositifs médiateurs interposés entre l'expérience et les phénomènes.

Les résultats de la recherche de terrain laissent entrevoir une nuance importante sur l'attitude envers l'épidémiologie numérique et la pratique de celle-ci. En effet, si un chercheur pratique l'épidémiologie numérique, cela ne signifie pas automatiquement que ce chercheur l'utilisera constamment ni que celui-ci envisagera que cette nouvelle pratique va remplacer l'épidémiologie traditionnelle. Un seul chercheur parmi nos participants a exprimé une attitude d'adoption complète. Les deux autres chercheurs ayant pratiqué l'épidémiologie numérique ont démontré une attitude ouverte, mais prudente envers cette nouvelle pratique.

Ces analyses montrent aussi des attitudes différentes envers la pratique de l'épidémiologie numérique auprès des chercheurs en épidémiologie en fonction de la sous-discipline de l'épidémiologie. En effet, la pratique de l'épidémiologie numérique serait moins appropriée pour l'épidémiologie clinique puisque cette sous-discipline de l'épidémiologie utilise des données cliniques (comme des tests de sang des patients) qui s'avèrent très différentes par rapport aux données numériques (comme celles générées par les médias sociaux ou les moteurs de recherche Internet). La culture épistémique de l'épidémiologie clinique pourrait ainsi être perçue comme différente par rapport à celle de la pratique de l'épidémiologie numérique d'après les constats de notre recherche. Cette distinction entre l'épidémiologie clinique et la pratique de l'épidémiologie numérique serait fondée sur les différences d'une part dans les types de données et les sources de

données utilisées et d'autre part dans les objets d'études (des opinions versus l'incidence d'une maladie).

Quant à la crédibilité, les chercheurs ne pratiquant pas l'épidémiologie numérique s'appuieraient sur des critères d'ordre épistémique (tels que la qualité de l'interprétation des données de la recherche et le contrôle des biais) lorsqu'ils évaluent la crédibilité de cette nouvelle pratique. Les chercheurs pratiquant l'épidémiologie numérique se baseraient sur des critères sociopolitiques (dont la confiance envers les signaux numériques lors des prises de décision en santé publique) et des critères épistémiques (y compris les apports et les limites des recherches fondées sur la pratique de l'épidémiologie numérique) pour évaluer la crédibilité de cette nouvelle pratique. Ce constat rappelle la symétrie dans la coproduction des connaissances (Jasanoff, 2004), c'est-à-dire que les connaissances produites lors des recherches impliquent des enjeux épistémiques, mais aussi des enjeux sociaux et politiques.

Dans ce chapitre, les perceptions de la pratique de l'épidémiologie numérique par des chercheurs canadiens en épidémiologie ont été analysées. Cependant, des analyses plus approfondies sur les enjeux épistémiques et sociopolitiques s'avèrent nécessaires pour mieux cerner le contexte de l'émergence de cette nouvelle pratique. Les méthodes de recherche choisies par les chercheurs ne sont pas seulement des outils ou des technicités, mais aussi des dispositifs méthodologiques soulevant des questions portant sur une multitude d'aspects scientifiques et sociétaux lors de la production des connaissances. Les chapitres suivants seront alors consacrés aux enjeux épistémiques et sociopolitiques caractérisant le contexte d'émergence de la pratique de l'épidémiologie numérique selon les points de vue de chercheurs en épidémiologie.

CHAPITRE V

LES ENJEUX ÉPISTÉMIQUES DE LA PRATIQUE DE L'ÉPIDÉMIOLOGIE NUMÉRIQUE

Ce chapitre a pour objectif de saisir les enjeux épistémiques qu'implique l'émergence de la pratique de l'épidémiologie numérique. Les analyses seront guidées par la coproduction des connaissances (Jasanoff, 2004), les cultures épistémiques (Knorr-Cetina, 1999) et la notion de dispositif méthodologique (Law *et al.*, 2013) expliquées dans le cadre théorique de la thèse. Il a été mentionné précédemment qu'un chercheur doit produire des connaissances valides selon les exigences de sa discipline. Il paraît donc logique que les points de vue de chercheurs en épidémiologie envers la pratique de l'épidémiologie numérique comportent des enjeux épistémiques importants à comprendre afin de réfléchir de manière critique sur cette nouvelle pratique. Les enjeux épistémiques concernent les façons d'effectuer les recherches épidémiologiques, les manières de connaître les phénomènes étudiés en épidémiologie ainsi que les systèmes de création et de justification des connaissances.

Le premier enjeu épistémique de la pratique de l'épidémiologie numérique mentionné par les chercheurs ayant participé à la recherche est sa capacité à répondre aux questions de recherche formulées dans les études épidémiologiques (entretien 6Pvi). Cet enjeu est revenu dans les propos de la plupart des chercheurs en épidémiologie ayant participé aux entretiens incluant un des participants refusant catégoriquement de pratiquer l'épidémiologie numérique parce que cette nouvelle pratique ne lui permet pas de répondre à la question de recherche (entretien 6Dxii).

« It depends on the question and a lot of it depends on the research question to try to determine whether the data is appropriate or not. That's difficult for me to answer without knowing what a specific research question might be. With every epidemiological study we need to determine the question and then we evaluate the data source that could respond to that question. I think that if you are asking questions about medical diagnostic. I would say that Twitter is probably not good because you are looking at self-reported stuff. But if you are asking things about self perceptions on some sort of treatment that people have gone and maybe that would be ok. I think it really depends on the question. » (Chercheur – épidémiologie sociale – entretien 6Pvi)

« I have trouble believing that the data would be of sufficient quality to answer questions. I'm afraid that I am rather skeptical of the success that we are going to see because of the biases. »
(Chercheur, épidémiologie sociale – entretien 6Dxii)

Ces réponses signifient que le chercheur en épidémiologie veut que les données utilisées lors de son étude puissent répondre à sa question de recherche en fournissant des résultats valides pour être en mesure de tirer des conclusions sur son objet d'étude. Le chercheur en épidémiologie aura confiance envers la pratique de l'épidémiologie numérique si les données peuvent répondre à sa question de recherche. Il n'est pas surprenant d'entendre cette réponse de la part de chercheurs puisque, comme mentionné précédemment, poser des questions est une des activités les plus importantes lors des pratiques de recherche (Osborne, 2014).

Les enjeux épistémiques constatés lors de nos analyses incluent aussi la signification des données et la confiance envers des dispositifs méthodologiques utilisés dans la pratique de l'épidémiologie numérique. Les enjeux épistémiques sont centrés sur les données parce que celles-ci jouent un rôle central en épistémologie et qu'il s'agit d'une phase émergente d'une pratique de recherche dont l'enjeu majeur semble être la source et la nature des données. Afin de brosser un portrait des enjeux épistémiques de la pratique de l'épidémiologie numérique, nous commencerons par analyser la signification des données dans le cadre des pratiques courantes de recherche en épidémiologie et dans le contexte de la pratique de l'épidémiologie numérique. Ensuite, nous analyserons la confiance à l'égard des données numériques comme dispositifs méthodologiques. Les constats expliqués dans ce chapitre sont présentés en fonction des points de vue des chercheurs en général, étant donné qu'aucune tendance particulière n'a été décelée dans leurs propos selon leur sous-discipline d'appartenance ni selon leur expérience de la pratique de l'épidémiologie numérique. Dans le cas contraire, nous préciserons l'approche choisie au début des analyses sur le thème concerné.

5.1 La signification des données

Comme mentionné précédemment dans la problématique de la thèse, les données sont importantes en épidémiologie puisque cette discipline est axée sur les données et qu'elle repose sur une approche systématique et non biaisée de la collecte, de l'analyse et de l'interprétation des données

(Dicker *et al.*, 2006). Ces caractéristiques sont clairement mentionnées dans la description de l'épidémiologie développée par l'Agence nationale de protection de la santé des États-Unis désignée comme Centers for Disease Control and Prevention (CDC) (Dicker *et al.*, 2006) :

Epidemiology is a scientific discipline with sound methods of scientific inquiry at its foundation. Epidemiology is data-driven and relies on a systematic and unbiased approach to the collection, analysis, and interpretation of data. (Dicker *et al.*, 2006, p. 1-2)

De plus, cette agence affirme que l'épidémiologie est une discipline quantitative reposant sur une connaissance pratique des probabilités, des statistiques et des bonnes méthodes (Dicker *et al.*, 2006) :

...epidemiology is a quantitative discipline that relies on a working knowledge of probability, statistics, and sound research methods. (Dicker *et al.*, 2006, p. 1-2)

Les participants à la recherche de terrain étant des chercheurs en épidémiologie, ils ont souvent fait allusion aux données lors des discussions sur la pratique de l'épidémiologie numérique et celle de l'épidémiologie traditionnelle. Leurs propos nous informent sur la manière dont ils voient les différences entre les données utilisées dans la pratique de l'épidémiologie traditionnelle et celles utilisées dans la pratique de l'épidémiologie numérique.

Notre analyse vise aussi à cerner les réserves vis-à-vis de l'utilisation des données non traditionnelles en recherche afin de développer un regard critique envers ce type de données. Pour ce faire, nous nous interrogerons sur la façon dont des chercheurs en épidémiologie perçoivent les données utilisées lors de la pratique de l'épidémiologie traditionnelle et celles utilisées lors de la pratique de l'épidémiologie numérique. S'ensuivra une analyse des différences entre les données traditionnelles et les données non traditionnelles vues par les participants à la recherche de terrain. Nous terminerons cette section avec l'échantillonnage populationnel, qui constitue un enjeu épistémique important en épidémiologie.

5.1.1 Les données utilisées dans la pratique de l'épidémiologie traditionnelle

La discipline de l'épidémiologie se situe dans le domaine de la santé publique. En fait, l'épidémiologie est souvent décrite comme la science fondamentale de la santé publique (Dicker *et al.*, 2006; Brunekreef, 2008). Il existe une grande variété de sources de données disponibles pour des recherches épidémiologiques selon l'écosystème de données en santé proposé par l'OMS en 2016 (OMS, 2016 cités dans Vayena *et al.*, 2018). D'après la figure retrouvée dans Vayena *et al.* (2018), cet écosystème indique que les données courantes en santé peuvent inclure les données en provenance des services de santé (dont les dossiers médicaux, les laboratoires et les assurances), des données sur la santé publique (dont la surveillance des maladies, les dossiers de vaccination et les registres) et les données de recherche (dont les données sur la génétique, les essais cliniques et les bio banques). Cet écosystème de données indique aussi plusieurs parties prenantes comme des agences de santé des gouvernements, des organismes de recherche et des universités (Vayena *et al.*, 2018), lesquelles utilisent ou produisent des données sur la santé.

La recherche de terrain effectuée dans le cadre de cette thèse apporte des précisions sur les données et les sources de données pouvant être utilisées en épidémiologie. Selon les propos des chercheurs ayant participé à la recherche de terrain, les types de données utilisés en épidémiologie comprennent des données collectées sur le terrain (entretien 6Bxviii, entretien 6Fx), des données d'enquêtes (entretien 9Rxiv, entretien 5Uxxxvi), des données en provenance d'organismes municipaux (entretien 5Uxxxvi), des bio banques (entretien 6Fx), des données administratives et des données de cohortes (entretien 7Mxxi). Des chercheurs en épidémiologie sociale ont souligné qu'ils utilisent dans le cadre de leurs recherches des données d'enquêtes de Statistique Canada (entretien 5Uxxxvi), dont l'Enquête canadienne sur la mesure de la santé (entretien 9Rxiv). Un chercheur en épidémiologie sociale a quant à lui mentionné qu'il utilise des données de gouvernements municipaux sur l'utilisation du sol, des données de géolocalisation et des fichiers administratifs médicaux informant entre autres sur les hospitalisations, les naissances et les décès (entretien 5Uxxxiv). Certains chercheurs en épidémiologie clinique, pour leur part, ont affirmé qu'ils se servent de données qu'ils collectent eux-mêmes, de données cliniques et génétiques de bio banques (entretien 6Fx), de données administratives et de données de cohortes (entretien 7Mxxi).

Cette analyse laisse entrevoir des données différentes selon la sous-discipline de l'épidémiologie puisque le type de données utilisé en épidémiologie clinique ne serait pas le même que celui utilisé en épidémiologie sociale selon les propos de certains participants. Cette différence entre l'épidémiologie clinique et l'épidémiologie sociale par rapport au type de données utilisé peut mettre en lumière diverses cultures épistémiques au sein de la discipline de l'épidémiologie. D'après Knorr-Cetina (1999), les cultures épistémiques sont des cultures de création et de justification des connaissances et des amalgames de dispositions de mécanismes et d'éléments liés par affinité, par nécessité et par coïncidence historique qui, dans un domaine donné, constituent la manière dont nous connaissons ce que nous connaissons. La culture épistémique de l'épidémiologie clinique ne serait pas la même que celle de l'épidémiologie sociale en raison de la nature des données utilisées susceptible d'amener des manières différentes de connaître les phénomènes étudiés selon la sous-discipline, reflétant ainsi des univers de recherche distincts au sein de la discipline de l'épidémiologie.

En outre, nous avons constaté que l'épidémiologie traditionnelle utilise des données générées à l'extérieur des systèmes de santé qui n'ont pas été collectées pour effectuer des recherches en santé comme des données de géolocalisation des Affaires municipales (entretien 5Uxxxvi) et des données administratives (entretien 7Mxxi). Ces données ne proviennent pas des systèmes de santé et elles ne sont pas générées pour des recherches en santé. Ces caractéristiques sont importantes dans la définition de la pratique de l'épidémiologie numérique proposée par Salathé (2018) qui affirme qu'il s'agit d'une épidémiologie utilisant des données numériques générées en dehors du système de santé publique, c'est-à-dire des données qui ne sont pas générées dans le but principal de faire de l'épidémiologie. Il a déjà été souligné que quelques entretiens ont fait ressortir des nuances sur la nouveauté de la pratique de l'épidémiologie numérique, car certains participants ont affirmé que l'utilisation de données n'étant pas produites par des systèmes de santé comme des données géographiques est chose courante en épidémiologie depuis longtemps. Ces affirmations pourraient illustrer l'idée que les méthodes de recherches s'appuient souvent sur celles qui les ont précédées (Law *et al.*, 2011 ; Law, 2010) puisqu'il semble que la pratique de l'épidémiologie numérique s'appuie entre autres sur des principes de la discipline de l'épidémiologie en ayant le même but que cette dernière (Salathé, 2018).

« Donc, c'est vraiment des analyses avec des données qu'on collecte sur le terrain et qu'on analyse. » (Chercheur – épidémiologie sociale - entretien 6Bxviii)

« Celle que j'utilise le plus est l'Enquête canadienne sur la mesure de la santé. » (Chercheur – épidémiologie sociale - entretien 9Rxiv)

« Donc pour caractériser les individus, on a souvent recours à des données d'enquêtes qui sont faites par Statistique Canada. Et pour décrire les quartiers, là on utilise différentes façons, ça dépend ce qu'on veut caractériser. On peut utiliser par exemple des données qui nous viennent du ministère des Affaires municipales, ils ont des données sur ce qu'on appelle l'utilisation du sol. Il y a aussi des bases de données qui existent qui répertorient et qui géolocalisent tous les services privés surtout les commerces par exemple. On a d'autres bases de données sur les services publics. La localisation des hôpitaux, des choses ça. Ah il y a d'autres choses que je pense sur les hôpitaux. On utilise aussi des fois, en santé publique ce qu'on appelle les fichiers médicaux administratifs. C'est-à-dire par exemple, les hospitalisations, les naissances, les décès, les registres du cancer, des choses comme ça. Donc il y a beaucoup d'informations qui sont disponibles dans ces bases de données-là aussi. » (Chercheur – épidémiologie sociale - entretien 5Uxxxvi)

« D'abord du recrutement local, c'est-à-dire des bases de données dont on collecte les données nous-mêmes. Et d'autres où vous avez à payer un petit peu pour avoir accès aux données. Donc l'exemple est la UK Bio Bank. Donc c'est la bio banque d'Angleterre où là il y a de l'information pour un demi-million d'anglais au niveau génétique et clinique. » (Chercheur – épidémiologie clinique - entretien 6Fx)

« Most of my research is either using administrative data sets, cohort data, administrative data, and then I work on project related to clinical cohort. » (Chercheur – épidémiologie clinique - entretien 7Mxxi)

D'après Vayena *et al.* (2018), l'inventaire des données courantes en santé et donc, en épidémiologie peut être élargi si des données environnementales y sont ajoutées, notamment sur le climat, la météo, le transport, la pollution, l'énergie de même que des données géospatiales. Ce type de données s'inscrit dans les pratiques de l'épidémiologie environnementale, une sous-discipline de l'épidémiologie. L'épidémiologie environnementale étudie les effets des expositions environnementales sur la santé et les maladies dans les populations en couvrant, en principe, tous les facteurs externes au corps humain susceptibles d'affecter la santé (Brunekreef, 2008). Dans une perspective d'analyse des risques, Aldrich (2000) affirme que l'épidémiologie environnementale étudie les risques pour la santé humaine liés aux dangers environnementaux en s'appuyant en grande partie sur la surveillance de la santé publique et sur des analyses de données portant sur des petites zones et des groupes de population dispersés.

La recherche de terrain apporte des nuances sur les types de données pouvant être utilisés en épidémiologie environnementale. Par exemple un participant effectuant des recherches en épidémiologie environnementale a affirmé utiliser des données qu'il collecte lui-même sur le terrain, des données générées par des organismes comme des programmes de surveillance provinciaux, des données climatologiques, météorologiques et environnementales (entretien 5Cxxvi). Ce chercheur se sert de ces données pour créer ses outils de modélisation. Selon ce chercheur, ces données sont en général bien validées et sont publiquement disponibles sur des sites gouvernementaux ou développées par une équipe de géomatique dans le but d'utiliser ces données pour des modèles de risques.

Un autre participant en épidémiologie environnementale a souligné qu'il utilise les systèmes provinciaux d'information sur les maladies déclarées, c'est-à-dire des bases de données gouvernementales sur la santé (entretien 1Oxxxv). Ce participant a recours aussi à des registres sur le cancer, notamment à celui du cancer du poumon ainsi qu'à des données collectées sur le terrain. Un autre participant de la même sous-discipline a affirmé utiliser des données de sources traditionnelles et des données de sources non traditionnelles (entretien 5Gxxv). Pour ce chercheur, les données de sources traditionnelles comprennent des données des recensements ou celles d'enquêtes à grande échelle ou encore des données collectées sur le terrain. En ce qui concerne les données de source non traditionnelle, ce participant a fait part de son expérience à utiliser des analyses d'images et l'apprentissage automatique (*machine learning*) afin de lier des données avec des mesures de la pollution de l'air prises sur le terrain pour ensuite les combiner avec des données sur le logement et la santé. Ce participant a aussi affirmé que ses travaux de recherche sont souvent fondés sur des combinaisons de données de sources traditionnelles et de sources moins traditionnelles.

« Pour les moustiques eux-mêmes, on va chercher des données sur le terrain, donc on capture les moustiques, on peut aussi avoir recours à des données de captures faites par d'autres organismes, par exemple les programmes de surveillance provinciaux et puis en parallèle à ça, on utilise beaucoup de données climatologiques, météorologiques et environnementales pour créer nos outils de modélisation et puis ces données-là en général ce sont des données assez solidement validées qui sont disponibles soit publiquement sur les sites gouvernementaux, soit des données développées par l'équipe de géomatique auquel je suis rattaché donc c'est plus des produits de développement géomatique si vous voulez qu'on

utilise aussi pour nos modèles de risques. » (Chercheur – épidémiologie environnementale - entretien 5Cxxvi)

« When looking in terms of public health perspective, we will use reported diseases information systems that we have here in the province. These are reported cases that are followed up by the province or the lab or by physicians. So, it is a government health database. The lung cancer data come from a cancer registry. There is also some primary data collection. » (Chercheur – épidémiologie environnementale - entretien 1Oxxxv)

« I use all kinds of data. I guess ... Some cases we are using traditional data sources like the Census or some of these large-scale surveys that are collected...Some of my work is primary data collection, so we do measurements in the field where we measure people's exposure to different pollutants either air pollution or noise pollution... And then for other, thinking about non-traditional sources, I just got out of a call where we are using image-based analysis and machine learning in junction with satellite imagery and measurements that we did at street level air pollution to develop a surface for air pollution in a city in Columbia and that we combine housing and health data. So, you know, it is a combination of measurements using traditional and not so traditional measurements for exposure assessment and then you know obviously primary data collection and Census data and things like that and existing surveys. » (Chercheur – épidémiologie environnementale - entretien 5Gxxv)

En analysant ces résultats sur les sources et les types de données utilisés lors des recherches en épidémiologie environnementale, il est possible d'observer le même questionnement que celui expliqué précédemment en épidémiologie sociale. En effet, il semble que l'épidémiologie environnementale combine des données de sources traditionnelles et non traditionnelles depuis longtemps et qu'une partie de ces données n'a pas été générée pour des recherches en santé ni pour des recherches épidémiologiques. Par exemple, les données des images satellites ne sont pas générées à l'intérieur du système de santé ni pour des recherches en santé et celles-ci sont utilisées en épidémiologie environnementale. En outre, le participant à la recherche de terrain affirmant utiliser ces données ne prétend pas pratiquer l'épidémiologie numérique. Ce dernier a même affirmé qu'il ne connaissait pas l'expression « épidémiologie numérique » avant l'entretien et que ça lui fait plaisir de discuter de cette pratique, mais qu'il n'en connaît pas beaucoup sur ce sujet.

« I have never heard of the term before...I am happy to talk with you, but I don't know much about this. » (Chercheur – épidémiologie environnementale - entretien 5Gxxv)

La définition proposée par Salathé *et al.* (2012) soutient que la pratique de l'épidémiologie numérique s'appuie sur l'analyse de données de sources non traditionnelles dans le cadre de recherches épidémiologiques. De plus, cette définition explique qu'il s'agit d'une épidémiologie utilisant des données numériques générées en dehors du système de santé publique, c'est-à-dire des données n'ayant pas été générées dans le but principal de faire de l'épidémiologie. Ce contexte peut alors apporter certaines confusions sur les différences entre la pratique de l'épidémiologie numérique et celle de l'épidémiologie environnementale, car ces deux sous-disciplines de l'épidémiologie ont recours à des données de sources non traditionnelles dans le cadre de recherches épidémiologiques.

L'une des différences envisageables entre la pratique de l'épidémiologie numérique et celle de l'épidémiologie traditionnelle serait le type de données puisque les données en provenance d'Internet comme celles des médias sociaux sont des données de type non structuré, c'est-à-dire des textes fournis par les utilisateurs. La pratique de l'épidémiologie traditionnelle utiliserait plutôt des données structurées comme des données d'enquêtes ou des tests de laboratoires (Zhang *et al.*, 2020). Cependant, la recherche de Roche *et al.* (2017) fondée sur la pratique de l'épidémiologie numérique combine des données structurées comme des enquêtes sur l'abondance des moustiques et des données non structurées comme les contenus de Tweets à partir de Twitter. Il y a aussi des données non structurées comme des notes et des rapports des médecins sur les patients, c'est-à-dire des données non structurées faisant partie des systèmes de santé (Zhang *et al.*, 2020). Il paraît donc difficile d'attribuer la différence entre la pratique de l'épidémiologie numérique et celle de l'épidémiologie traditionnelle à cette typologie des données.

Les données de sources non traditionnelles utilisées en épidémiologie environnementale proviennent généralement d'organismes reconnus dans les gouvernements ou en milieu académique. Il est possible que les différences entre les données traditionnelles et non traditionnelles reposent sur le secteur d'activité dans lequel les données sont générées (secteur public ou secteur privé) et non sur le fait que les données soient produites à l'extérieur des systèmes de santé pour des motifs autres que des recherches en santé. Les données utilisées en épidémiologie environnementale proviendraient généralement d'organismes publics reconnus comme des ministères et des instituts de recherche. Ces organismes produisent des données pour des motifs

n'étant pas nécessairement liés à des recherches en santé. Les données utilisées lors de la pratique de l'épidémiologie numérique comme les données générées par les médias sociaux peuvent provenir de secteurs autres que les secteurs gouvernementaux ou académiques. Ainsi, à la lumière de ces propos, il est possible que les définitions de la pratique de l'épidémiologie numérique doivent être révisées ou précisées davantage afin de mieux les distinguer de la pratique de l'épidémiologie environnementale notamment. Les analyses suivantes portent sur les données utilisées lors de la pratique de l'épidémiologie numérique.

5.1.2 Les données utilisées dans la pratique de l'épidémiologie numérique

Il a été mentionné précédemment que la pratique de l'épidémiologie numérique impliquerait l'utilisation de données numériques générées en dehors du système de santé publique, c'est-à-dire des données n'ayant pas été générées dans le but principal de faire de l'épidémiologie (Salathé, 2018). Selon certains travaux de recherche comme ceux de Roche *et al.* (2017) expliqués précédemment, les données utilisées dans la pratique de l'épidémiologie numérique peuvent inclure des données entomologiques, des données sur le comportement humain, des données de réseaux routiers, des données de Twitter, des données sur des recensements à l'échelle des populations et des données sur des estimations de cas suspects en provenance de praticiens en santé.

Des recherches dans d'autres domaines comme la science des plantes peuvent aussi utiliser des ensembles de données publics d'images et des techniques d'apprentissage profond (*deep learning*) dans le cadre de la pratique de l'épidémiologie numérique en vue d'identifier facilement certaines caractéristiques de plantes comme des maladies (Mohanty *et al.*, 2016). Ces recherches mettent en lumière une grande diversité dans les données utilisées lors de la pratique de l'épidémiologie numérique.

La recherche de terrain effectuée dans le cadre de cette thèse révèle que plusieurs chercheurs en épidémiologie affirment que la pratique de l'épidémiologie implique des analyses de données d'Internet (entretien 10Txxii, entretien 5Cxxvi) comme celles de médias sociaux (entretien 10Ixxviii, entretien 6Bi, entretien 5Cxxx) et de moteurs de recherche (entretien 10Ixxviii, entretien 3Exxxiii) ainsi que des analyses de données générées par des téléphones intelligents (entretien

9Rxiv, entretien 5Cxxvi, entretien 5Cxxx). Ces données sont mentionnées autant chez les chercheurs n'ayant pas pratiqué l'épidémiologie numérique que ceux l'ayant pratiquée dans le cadre de leurs recherches, et ce, peu importe si les travaux des chercheurs s'inscrivent en épidémiologie sociale, clinique ou environnementale.

« I consider ...you know... mining Twitter to be digital epidemiology or mining Google searches. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

« I assume it means incorporating social media and the larger seas of data that are available to us in more useful way than the standard ways of spoke interviews and surveys and the sort of things that I do now ». (Chercheur - épidémiologie clinique - entretien 6Bi)

« So, I would assume you know data that's coming from the Internet or that is produced by the Internet » (Chercheur - épidémiologie sociale - entretien 10Txxii)

« Ça me fait penser à l'utilisation de nouvelles technologies par exemple, je vois un lien avec l'étude de Katapally qui utilisait beaucoup les applications mobiles sur les téléphones ». (Chercheur - épidémiologie sociale - entretien 9Rxiv)

« Pour moi ça me faisait plus penser à toutes les données qui sont générées allant par Internet, et donc toutes ces connaissances du comportement des gens, le traçage des gens qui sont accessibles via par exemple le suivi de leurs mouvements, avec leur téléphone cellulaire, etc. C'est vrai que c'est plus ces données-là générées par Internet en fait par l'interconnexion des gens par Internet. Voilà. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvi)

« Je pensais à d'autres termes comme big data, les mégadonnées, peut-être les collectes de données par les appareils comme les apps sur les cellulaires ou par des réseaux. » (Chercheur - épidémiologie environnementale - entretien 5Cxxx)

« What I would guess is that digital epidemiology is ... you know how we can observe trends in Googling different terms, and some people say if you look at trends that people google the flu, then you can predict a flu outbreak in a region because a lot of people are searching for flu on Google. That would be my guess. » (Chercheur - épidémiologie sociale - entretien 3Exxxiii)

Un participant ayant pratiqué l'épidémiologie numérique ajoute à ce portrait les données de vidéos (entretien 10Ixxviii). Ce chercheur a utilisé des données générées par des vidéos de piétons dans le cadre d'une recherche sur les relations entre la santé mentale et l'accès à des espaces verts en milieu urbain.

« We used pedestrian videos and extracted green information from pedestrian videos. I see this kind of thing being done with Google Street map data, where you can basically go in and look at the images that we may capture by street view and green space on updating those

images every couple of years so like street view does we can get changes over time and all these kinds of variables... I am an environmental epidemiologist, so I am interested in a topic like this. There are some really really really interesting data sets for green space assessment and in this case, you know we used some of them, but I don't think we do examine them deeply as possible. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

Ces analyses mettent donc en évidence une grande diversité dans les données utilisées dans la pratique de l'épidémiologie numérique ainsi qu'une culture épistémique différente par rapport à celle de l'épidémiologie traditionnelle, la raison étant l'ajout de données non traditionnelles comportant le potentiel d'apporter de nouvelles façons de connaître les phénomènes étudiés. Selon ces constats, les chercheurs ayant l'expérience de la pratique de l'épidémiologie numérique peuvent envisager le potentiel de certains types de données encore moins conventionnels comme des données d'imagerie et des données générées à partir de vidéos pour des recherches épidémiologiques. Ce type de données serait encore sous-exploité dans le cadre de recherches en santé. Cela étant dit, nous analyserons maintenant comment des chercheurs en épidémiologie voient les différences entre les données traditionnelles et les données non traditionnelles pour des recherches en épidémiologie.

5.1.3 Les différences entre les données traditionnelles et les données non traditionnelles

Cette section présentera les différences entre les données traditionnelles et les données non traditionnelles telles que vues par des chercheurs en épidémiologie. Sur cet enjeu, les points de vue varient selon l'expérience de la pratique de l'épidémiologie numérique. Par conséquent, les résultats seront présentés en expliquant les points de vue des chercheurs ne pratiquant pas l'épidémiologie numérique et ceux des chercheurs ayant l'expérience de cette pratique.

5.1.3.1 Les points de vue des chercheurs ne pratiquant pas l'épidémiologie numérique

En principe, les données peuvent agir comme un dispositif médiateur situé entre l'expérience et les phénomènes, c'est-à-dire un intermédiaire entre les phénomènes étudiés lors des recherches et le chercheur. Dans cette perspective, les données représentent les phénomènes du monde réel aux

yeux des chercheurs. Il s'avère primordial pour les chercheurs d'utiliser des données en mesure de les éclairer adéquatement sur les phénomènes étudiés afin de produire des recherches valides. Les données peuvent alors être vues comme un intermédiaire essentiel et aussi un dispositif médiateur parce que c'est à travers les analyses de données que les chercheurs saisissent les caractéristiques des phénomènes étudiés. Il y a donc un enjeu important de confiance envers les données lors des recherches. La confiance envers les données guide les choix du chercheur vis-à-vis de l'utilisation ou de la non-utilisation de ces données lors des recherches.

Pour plusieurs chercheurs en épidémiologie ne pratiquant pas l'épidémiologie numérique, il semble y avoir une plus grande confiance envers les données traditionnelles (comme celles de la surveillance et des enquêtes nationales de Statistique Canada) qu'envers les données non traditionnelles comme les données numériques (y compris les données générées par des médias sociaux, par exemple) selon des résultats de la recherche de terrain. Par exemple, un chercheur en épidémiologie environnementale souligne que même si les données gouvernementales ne sont pas toujours facilement accessibles, au moins le chercheur connaît les caractéristiques de ces données (entretien 1Oxxxiv).

Un chercheur en épidémiologie sociale affirme que les données générées par Internet entraînent plus d'incertitudes sur la source réelle des informations transmises (entretien 6Bxviii). Un autre participant en épidémiologie sociale a communiqué le même point de vue en mentionnant que le chercheur ne sait pas réellement qui fournit les réponses aux questions posées sur un site web ni d'où les données proviennent (entretien 6Fix). De plus, il est impossible de contrôler et d'évaluer la qualité des données non traditionnelles d'après ce chercheur.

Les données traditionnelles seraient donc plus susceptibles de susciter la confiance du chercheur en épidémiologie que les données non traditionnelles. Cette préférence envers les données traditionnelles serait due à une meilleure compréhension de ces données de la part du chercheur en épidémiologie et à une plus grande certitude de la provenance de ces données comparativement aux données numériques.

« I don't say that government data is always easily accessible but at least you know what they are. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxiv)
--

« Si on parle de données qualitatives dans les focus groups par exemple lorsqu'on pose des questions à des gens, il y a une recette qu'on applique et en ligne, on ne sait pas toujours c'est qui qui parle et qui qui va se cacher derrière le téléphone, donc il y a un peu d'incertitude là. Donc ça peut faire en sorte que c'est moins contrôlé donc c'est plus public oui... on connaît pas tout dans le sens... on veut standardiser... et essayer de contrôler les influences externes, c'est quelque chose qu'on n'a pas nécessairement lorsqu'on se sert de données d'Internet parce qu'il faut savoir qui se cache derrière toute cette information-là. Donc il y a un aspect un peu qu'on ne peut pas contrôler nécessairement. » (Chercheur - épidémiologie sociale - entretien 6Bxviii)

« For example, if I want to collect data from PhD students from Montreal, I have a list of who has an e-mail and people respond. But if I pose a random website and I ask a question on a subject, I don't know who is answering, I don't know where the data is coming from, and I have no way of controlling the quality and no way to assign a quality. » (Chercheur - épidémiologie sociale - entretien 6Fix)

En fait, les chercheurs portent leur attention sur ce que les données représentent par rapport à la réalité et ces représentations peuvent prendre la forme de signaux repérables à travers les analyses de données (Knorr-Cetina, 1999). Il semble donc que les signes, les reconstructions et les événements identifiés soient plus difficiles à bien saisir lors des analyses de données non traditionnelles que lors des analyses de données traditionnelles, selon plusieurs participants à la recherche de terrain.

Un autre chercheur en épidémiologie environnementale a partagé un avis similaire en affirmant qu'il est beaucoup plus à l'aise avec les données en provenance du système de santé, car il comprend les biais associés à ces données (entretien 1Oxxxv), notamment avec des enjeux comme les maladies transmises sexuellement où les gens ont tendance à ne pas publier des informations sur ce genre de maladie contrairement à d'autres enjeux moins sensibles comme celui de la grippe. Par conséquent, l'objet de la recherche aurait une de l'influence sur la confiance envers les données non traditionnelles.

« I feel much more comfortable with data from the health care system because I understand the biases, they have a lot more than the social media data. And particularly when it comes with things like sexually transmitted diseases, the information is probably much weaker on social media where people probably not tend to post about their sexually transmitted disease.

The flu, I can kind of understand that. Again, it would probably depend on the issue I am looking at. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxv)

L'un des chercheurs en épidémiologie sociale a fourni une réponse assez précise sur la manière dont il voit les différences entre les données traditionnelles et non traditionnelles (entretien 5Uxxxvi). Donnant l'exemple de la prévalence de la dépression, ce chercheur affirme qu'il existe des échelles et des instruments de mesure traditionnels validés et que leur capacité à bien mesurer une prévalence réelle est connue. Ce chercheur a l'impression que les instruments utilisés lors des analyses de données numériques sont moins connus que ceux utilisés lors des analyses de données traditionnelles. De plus, ce chercheur soutient que les échantillons produits par Statistique Canada sont de très grande qualité. Pour des échantillons extraits à partir d'autres sources comme Facebook, la qualité serait incertaine et il apparaît difficile de savoir qui est réellement représenté dans ces échantillons d'après ce chercheur.

« Mais par exemple, si on veut mesurer... je ne sais pas moi... la prévalence de la dépression dans une population... bon on a des échelles, des instruments de mesure qui ont été validés, qui ont été construits et puis on sait par des analyses psychométriques de ces instruments-là, on sait leurs capacités à bien mesurer une prévalence réelle par exemple. Ça ces choses-là on les connaît moins j'ai l'impression avec le numérique. Euh... mais pour répondre plus précisément à votre question, je pense qu'il faudrait que j'aie lire. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

« C'est que... Statistique Canada va faire des échantillons de populations systématiques ou aléatoires avec une randomisation qui donne des échantillons vraiment de très très très grande qualité alors qu'on a d'autres recherches à l'autre extrême où on y va juste avec des volontaires, on fait des appels sur Facebook et là on ne sait pas trop en fait notre échantillon... il représente qui exactement ? » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

Cependant, tous les chercheurs ne pratiquant pas l'épidémiologie numérique ne démontrent pas nécessairement une grande méfiance vis-à-vis des données non traditionnelles selon les entretiens réalisés lors de la recherche de terrain. L'un des chercheurs en épidémiologie clinique a d'ailleurs un avis différent, car il souligne qu'il y a des données en provenance des systèmes de santé n'ayant pas été collectées pour faire de la recherche comme les données administratives et ces données peuvent s'avérer difficiles à utiliser dans le cadre de recherches épidémiologiques (entretien 6Bi).

Ce chercheur affirme aussi que les données en provenance des systèmes de santé sont souvent désordonnées et que celles-ci peuvent nécessiter beaucoup de travail de la part d'experts avant que le chercheur puisse en extraire les connaissances de manière fiable.

Un chercheur en épidémiologie sociale affirme clairement qu'il ne voit pas de différence entre les données traditionnelles et les données numériques (entretien 3Exxxiii). Sous le même angle, un autre chercheur en épidémiologie sociale a relaté que, selon lui, la qualité des données peut être compromise autant avec les données traditionnelles qu'avec les données non traditionnelles (entretien 5Kxv). D'après ce chercheur, la qualité des données représente un défi pour les deux types de données et les données en provenance de l'extérieur des systèmes de santé peuvent avoir une valeur autant que les données traditionnelles.

Sur une note positive, l'un des chercheurs en épidémiologie sociale soutient que de plus en plus de gens se dirigent vers des données du monde réel (entretien 6Axix). Ce chercheur pense que même si ces données sont désordonnées, celles-ci s'avèrent riches et elles représentent comment les gens communiquent et partagent leurs idées. Ce chercheur affirme aussi qu'une plus grande utilisation doit être faite des données non traditionnelles en épidémiologie. La recherche épidémiologique n'est pas encore arrivée aux méthodes permettant d'utiliser les données non traditionnelles de manière rigoureuse, mais cela vaut la peine de s'y intéresser selon ce chercheur.

« Even if the data are collected in the health care system, they are not necessarily designed for research. So, the same kind of issue applies even if we are talking about health system data often ... it is very messy, and it will require a lot of hands-on expertise before you can reliably extract knowledge from it. I am sure that this is true for Google searches as it is from a health database. » (Chercheur - épidémiologie clinique - entretien 6Bi)

« I guess that there would be something in terms of research design. I personally don't see a difference between digital data vs traditional data. » (Chercheur - épidémiologie sociale - entretien 3Exxxiii)

« Dans les deux cas, la qualité des données de recherche peut être compromise. La qualité des données est un défi dans les deux cas. Ça ne veut pas dire que si des données proviennent à l'extérieur du système de santé, qu'elles n'ont pas de valeur nécessairement. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

« I think that more and more people are looking outside of traditional randomized control trials and other kind of studies and looking at more real-world data and I think that even if

this is messier to deal with from an epidemiological standpoint of view and just scientifically overall it's messier, but it is very rich data and it is how people are communicating and sharing their thoughts and ideas more and more. So, I think that it is looked upon that this is something we need to tap into. We are not there yet with the methods on how to do so in a rigorous way, but this is something that is indeed worthwhile looking into and investing in. »
(Chercheur - épidémiologie sociale - entretien 6Axix)

5.1.3.2 Les points de vue des chercheurs pratiquant l'épidémiologie numérique

Enfin, il est aussi important d'expliquer les points de vue des chercheurs pratiquant l'épidémiologie numérique. Les points de vue de ceux-ci sont différents ou plus précis par rapport aux autres chercheurs n'ayant pas pratiqué l'épidémiologie numérique en raison de leur expérience envers cette nouvelle pratique de recherche.

D'abord, l'un des chercheurs en épidémiologie sociale ayant expérimenté l'épidémiologie numérique affirme qu'il pourrait y avoir plus de biais lors des analyses fondées sur des données non traditionnelles, car celles-ci ne sont pas collectées pour des motifs de soins de santé (entretien 6Axxiii). Selon le participant en épidémiologie environnementale ayant pratiqué l'épidémiologie numérique, les données traditionnelles et non traditionnelles sont différentes à bien des égards (entretien 10Ixxviii). Ce chercheur a insisté sur les biais concernant les groupes d'utilisateurs d'Internet et des médias sociaux. Par exemple, certains jeunes peuvent passer toute la journée sur Internet et ainsi avoir beaucoup plus d'interactions que les personnes plus âgées qui, en général, consultent les médias sociaux moins souvent. Selon ce chercheur, il est évident que les données de sources non traditionnelles comme celles d'Internet sont axées sur les personnes ayant accès à l'Internet. L'autre préoccupation de ce chercheur concerne les intérêts des gens qui varient en ce qui concerne l'utilisation d'Internet et des médias sociaux. Certains communiquent beaucoup sur Internet alors que d'autres restent silencieux. En s'intéressant uniquement aux personnes choisissant de s'engager, surtout avec les médias sociaux, il est possible que le chercheur regarde en réalité une sous-population plus bruyante et non l'ensemble d'une population ciblée dans le cadre de la recherche.

Cette situation peut illustrer la possibilité que les signes fournis par les analyses de données non traditionnelles soient biaisés par rapport à la réalité des phénomènes étudiés. Ces biais seraient

causés par des bruits nuisibles pour des représentations adéquates et des interprétations justes des réalités observées lors des recherches utilisant des données non traditionnelles comme celles d'Internet.

Un autre chercheur en épidémiologie sociale ayant pratiqué l'épidémiologie numérique pense que les données des systèmes de soins de santé sont beaucoup plus rigides et beaucoup plus structurées que les données non traditionnelles (entretien 8Sxx). En revanche, ce chercheur souligne que les liens sont difficiles à établir entre les différentes sources ou les différents organismes dans les systèmes de soins de santé. Il affirme aussi que les données des systèmes de soins de santé se révèlent utiles pour des recherches en santé parce que celles-ci sont fiables et valides.

Toujours selon ce chercheur, les données en provenance d'Internet comportent aussi des avantages à l'égard de la santé. Par exemple, si un chercheur veut comprendre une région du monde où il existe un enjeu de santé important, il est possible de prévenir certains événements. Selon le raisonnement de ce chercheur, utiliser des données provenant de l'extérieur des systèmes de soin de santé (comme des données d'Internet) pour comprendre des phénomènes peut empêcher des gens d'avoir besoin d'entrer dans le système de soins de santé.

En poursuivant ce raisonnement, ce chercheur affirme qu'au départ, les données traditionnelles et non traditionnelles semblent s'exclure mutuellement, mais il fournit un exemple où il est possible de d'entrevoir une intersection entre ces deux types de données susceptible d'aider à résoudre un problème du système de soins de santé en ayant recours aux données de l'Internet ou à celles des téléphones intelligents. Par exemple, en recherche clinique, il existe un moyen de prédire le niveau de stress, d'anxiété et de dépression des personnes à l'aide de données provenant d'outils numériques et de l'Internet. Si les bons algorithmes d'intelligence artificielle sont développés, il est possible de prédire et de prévenir le suicide. Il s'agit d'un avantage pour le système de soins de santé selon les affirmations de ce chercheur. Cet aspect lié à l'intelligence artificielle est sous-développé dans la thèse puisque, comme mentionné précédemment, les enjeux épistémiques sont centrés sur les données, car celles-ci jouent un rôle central en épistémologie et il s'agit d'une phase émergente d'une pratique de recherche dont l'enjeu majeur semble être la source et la nature des données.

Cependant, les propos de ce chercheur amènent une perspective intéressante quant à la distinction entre les données traditionnelles et non traditionnelles. Au lieu de les considérer toujours séparément, il est possible d'examiner des intersections entre ces deux types de données. Les signes repérables lors des analyses de données générées à l'extérieur des systèmes de santé peuvent fournir des informations sur des phénomènes étudiés dans les systèmes de soins de santé. Les algorithmes d'intelligence artificielle pourraient agir comme un dispositif médiateur situé entre l'expérience et les phénomènes. Ce dispositif médiateur enverrait des signaux aux chercheurs pour prédire et prévenir certains événements comme le suicide. Encore faut-il que ces signes soient compris par les chercheurs et que les algorithmes d'intelligences artificielles, c'est-à-dire les dispositifs médiateurs, soient en mesure d'adapter les signaux en leur faisant prendre une forme intéressante selon le regard d'un chercheur en épidémiologie.

« Because it is not collected for the purpose of health care, there could be more biases. »
(Chercheur - épidémiologie sociale- pratiquant de l'ÉN - entretien 6Axxiii)

« Oh, there are all kinds of ways that they are different. First, there is who uses the Internet and how much they use it. So, the problem is that risk, perhaps that is the best way to describe it. You know it's tended towards young people; some young people stands as far as I can tell the entire day on the Internet so they are going to have a lot more interactions than older people who might check in social media once or twice a week or something like that. So, it is going to be mainly biased towards the user group. Obviously, it is going to be biased towards people who have access to the Internet. I think access is becoming more and more universal and it is almost being seen as a human right at this point. So, I think that argument will fade even more over time. I guess the other concern I have is I don't use social media at all. I find it unappealing. And I am sure there is lots of people who don't use these, you know it is not just differentiated by age at that point, it is kind of differentiated by personal factors that may be quite consistent among the people who choose not to engage. So, by looking at the people who choose to engage especially if you are using social media, you're perhaps looking at a noisier sub-section of the population that you would be at the entire population. »
(Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

« I think that health care system data is much more rigid and much more structured, and the linkages are difficult to make in health care systems. And I think that's a big set of challenges, but the health care systems data are reliable and valid right? So, there is a good part to it. When you are thinking about the data that is coming out of the Internet, from a health perspective, there are some advantages if you can pick up to make patterns about how things are evolving... what's happening right now in States if you want to understand where there is a hot spot, you could prevent things. So... outside of the health care system, and you can prevent people from getting into the health care system if you understand data that's coming from the Internet. So, they seem at this moment mutually exclusive, but I will give you an example where you can see an intersection between a health care system problem that could

be solved using data from the Internet or data from smartphones. For example, there is a way to predict people's mood stress, anxiety and depression using data from digital tools and from the Internet. If we develop the right artificial intelligence algorithms, there is a chance to predict and prevent suicide. Now that's a benefit to the health care system, right? The data that is used has nothing to do with the health care system it's coming outside of it. So that's how I would put it. » (Chercheur - épidémiologie sociale- pratiquant de l'ÉN - entretien 8Sxx)

5.1.4 L'échantillonnage populationnel

L'échantillonnage populationnel représente une notion essentielle en épidémiologie parce que cette discipline implique la collecte de données au sein d'un échantillon représentatif d'une population ciblée dans le cadre d'une étude épidémiologique (Ben-Shlomo *et al.*, 2013). De plus, l'épidémiologie utilise des méthodes statistiques pour résumer, examiner les associations et tester des hypothèses spécifiques à partir desquelles elle déduit des conclusions généralisables sur les causes des maladies et l'évaluation des soins de santé dans une population cible (Ben-Shlomo *et al.*, 2013). Les participants à la recherche de terrain étant des chercheurs en épidémiologie, ils ont été nombreux à faire allusion à l'échantillonnage populationnel lors des entretiens. Ce terme a été entendu autant auprès des chercheurs pratiquant l'épidémiologie numérique qu'auprès de ceux n'ayant pas expérimenté cette nouvelle pratique, et ce, peu importe la sous-discipline du chercheur (épidémiologie sociale, épidémiologie clinique et épidémiologie environnementale). Les constats établis lors de cette recherche révèlent deux visions sur l'échantillonnage populationnel dans le cadre de la pratique de l'épidémiologie numérique.

La première vision considère l'échantillonnage populationnel comme problématique dans la pratique de l'épidémiologie numérique. Par exemple, un chercheur en épidémiologie environnementale estime que l'échantillonnage n'est pas adéquat si une étude épidémiologique utilise des données des médias sociaux (entretien 5Gxxix). Un chercheur en épidémiologie sociale a aussi exprimé de l'incertitude sur la qualité des procédures d'échantillonnage des études utilisant des données de Twitter, notamment (entretien 6Pvi). Il n'est pas certain si l'échantillonnage populationnel avec des données de Twitter peut être réalisé selon les mêmes standards que ceux en épidémiologie traditionnelle. Nous avons aussi entendu un chercheur en épidémiologie sociale affirmer qu'il a une plus grande confiance en la qualité des échantillons populationnels effectués

par Statistique Canada que des échantillons populationnels fondés sur des appels volontaires sur Facebook notamment (entretien 5Uxxxvi). Ce chercheur considère qu'il est difficile de savoir ce que représentent exactement des échantillons populationnels extraits à partir de Facebook.

« I think that if I was to see the title, I would be very skeptical. Because I don't think you have a proper sample. You know, everything in epidemiology is all about sampling populations. » (Chercheur - épidémiologie environnementale - entretien 5Gxxix)

« One the challenges is that because we often use sampling procedures, if we can't get the full population, and I think it might be a little challenging ... if you have a whole world of Twitter ... I guess you can sample certain aspects of it... applying our basic statistical techniques that we learned really relies on sampling and I think you have to make sure that you can do that kind of sample in order to make sure that you get reliable answers and that I am not sure if that has been thought through yet. » (Chercheur - épidémiologie sociale - entretien 6Pvi)

« C'est que... Statistique Canada va faire des échantillons de populations systématiques ou aléatoires avec une randomisation qui donne des échantillons vraiment de très très très grande qualité alors qu'on a d'autres recherches à l'autre extrême où on y va juste avec des volontaires, on fait des appels sur Facebook et là on ne sait pas trop en fait notre échantillon... il représente qui exactement ? » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

La deuxième vision considère qu'il existerait un potentiel pour élargir les échantillons populationnels lors des études épidémiologiques avec la pratique de l'épidémiologie numérique. Quelques participants ont ainsi souligné ce potentiel pour les données numériques. Par exemple, un chercheur en épidémiologie sociale n'ayant pas expérimenté la pratique de l'épidémiologie numérique voit un grand potentiel dans l'élargissement des échantillons populationnels avec la pratique de l'épidémiologie numérique (entretien 5Uxxxvi). Pour lui, il s'agit d'un avantage majeur du numérique parce que plus l'échantillon est grand, plus la marge d'erreur est petite, ce qui amène alors plus de précision dans les résultats des études épidémiologiques.

Deux participants pratiquant l'épidémiologie numérique ont affirmé de façon claire que les biais de sélection se réduisent considérablement avec la pratique de l'épidémiologie numérique. Le chercheur en épidémiologie sociale a exprimé cette perspective en soulignant que 90 % de la population canadienne possède un téléphone intelligent (entretien 8Sxx). Ce chercheur estime qu'il

est inutile de recruter seulement 50-60 % d'une population lorsqu'on peut recruter des milliers de gens et collecter les données en temps réel. Il entrevoit une grande valeur ajoutée en épidémiologie avec la pratique de l'épidémiologie numérique et il considère qu'il s'avère important de comprendre le pouvoir des outils numériques dans le cadre de recherches épidémiologiques.

Dans le même sens, mais en distinguant la source de données, le chercheur en épidémiologie environnementale pratiquant l'épidémiologie numérique perçoit les données générées par *Google Trends* comme encore meilleures que celles générées par des médias sociaux (entretien 10Ixxviii). Ce chercheur pense que *Google Trends* peut amener vers des échantillons plus représentatifs de la population puisque Google est utilisé de façon assez universelle pour bien saisir comment les gens pensent. S'il est possible pour le chercheur de traiter ces données de manière adéquate, il peut forcément en tirer quelque chose de très utile d'après les propos de ce participant.

« Par contre, on aurait un grand potentiel d'aller chercher un échantillon beaucoup plus grand, puis ça je ne sais pas dans quelle mesure vous êtes familière avec les statistiques. Plus notre échantillon est grand, plus notre marge d'erreur est petite et plus notre précision est importante. Ça c'est un avantage majeur du numérique, c'est-à-dire, en fait on peut dire qu'on s'éloigne beaucoup de la notion d'échantillon pour tendre davantage vers celle de population directement. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

« So, the big thing that they are asking is what about selection bias? Well, I am like if you could recruit the whole population, where is the selection bias of that? And they compensate that by asking how do you do that? Because 90% of Canadians have a smartphone. That doesn't mean that everyone is going to appear in the data but theoretically, you need to use these tools to understand what's going on. So why do I want to do a study based on a research question by recruiting just 50 or 60% of a population when I can recruit thousands of people and take data in real time? I see that as a big value for epidemiology and we have to understand the power of digital tools. » (Chercheur - épidémiologie sociale- pratiquant de l'ÉN - entretien 8Sxx)

« I actually see Google Trends as stronger than social media. I think it probably captures more representative samples of the population. You know I am pretty aware of Google and where its strength limitations are. So, it is not full of proof for anything, but I think that Google is fairly universally used to capture what is on people's minds. And if you can deal with these data carefully, you can probably get something quite useful out of it. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

Ces analyses révèlent que pour avoir confiance envers les données non traditionnelles, les chercheurs en épidémiologie exigent que l'échantillonnage populationnel des recherches fondées sur ce type de données soit représentatif des standards traditionnels de l'épidémiologie. Ce phénomène signifie que des chercheurs peuvent être incertains de la fiabilité des signes et des apparences perceptibles à travers les analyses des données générées par les données non traditionnelles comme celles des médias sociaux. Cependant, certains chercheurs en épidémiologie peuvent regarder les choses de manière différente parce que selon eux, les données non traditionnelles pourraient offrir une meilleure compréhension des phénomènes étudiés en raison d'une meilleure représentativité des échantillons populationnels devenue possible avec l'utilisation massive d'Internet.

En somme, nous avons constaté des différences dans les données utilisées selon la sous-discipline de l'épidémiologie lors de notre recherche. La différence entre l'épidémiologie clinique et l'épidémiologie sociale par rapport au type de données utilisées peut montrer diverses cultures épistémiques au sein de la discipline de l'épidémiologie. En ce sens, la culture épistémique de l'épidémiologie clinique n'est pas la même que celle de l'épidémiologie sociale en raison de la nature des données utilisées. Cette dernière amènerait des manières différentes de connaître les phénomènes étudiés selon la sous-discipline, créant ainsi des univers de recherche distincts au sein de la discipline de l'épidémiologie. Ces analyses révèlent aussi une grande diversité dans les données utilisées lors de la pratique de l'épidémiologie numérique ainsi qu'une culture épistémique différente par rapport à celle de l'épidémiologie traditionnelle étant donné l'ajout de données non traditionnelles susceptibles d'apporter de nouvelles façons de connaître les phénomènes étudiés.

De plus, notre recherche montre que les chercheurs en épidémiologie n'ayant pas l'expérience de la pratique de l'épidémiologie numérique peuvent avoir une plus grande confiance envers les données traditionnelles comme les données en provenance des systèmes de santé et d'institutions officielles de la statistique comparativement aux sources de données numériques. Sur la base des constats de notre recherche, les données non traditionnelles comme celles générées par les médias sociaux seraient plutôt associées à des problèmes de représentativité populationnelle, car l'exclusion de certaines populations comme les personnes âgées peut engendrer une fausse image de la réalité des enjeux de santé touchant l'ensemble des populations. Les réserves des chercheurs

en épidémiologie envers les données non traditionnelles seraient aussi dues au manque de connaissances sur les biais liés à ces données et à une moins bonne compréhension des apports, des limites, de la validité et de la rigueur des sources de données non traditionnelles comparativement aux sources de données traditionnelles. Il peut y avoir plus de biais lors des recherches fondées sur cette nouvelle pratique que celles basées sur l'épidémiologie traditionnelle d'après le discours de certains participants pratiquant l'épidémiologie numérique. Ces biais s'expliqueraient par le fait que les données d'Internet ne sont pas conçues pour effectuer des recherches en santé et que des populations comme les personnes âgées ou les enfants peuvent être sous-représentées lors des analyses de données parce que celles-ci n'utilisent pas ou très peu l'Internet.

Les données non traditionnelles agiraient comme un dispositif médiateur envoyant des signaux incompris par les chercheurs. De plus, les signaux émis par ce dispositif médiateur pourraient ne pas représenter la réalité du phénomène étudié, diminuant ainsi la confiance des chercheurs en épidémiologie envers les données non traditionnelles. Il semble que les signes, les reconstructions et les événements identifiés soient plus difficiles à bien saisir lors des analyses de données non traditionnelles que lors des analyses de données traditionnelles selon des chercheurs en épidémiologie.

L'un de nos participants pratiquant l'épidémiologie numérique voit de façon différente les distinctions entre les données traditionnelles et non traditionnelles. Au lieu de considérer ces deux types de données comme mutuellement exclusifs, ce chercheur perçoit une intersection entre les données traditionnelles et non traditionnelles. Cette intersection serait utile pour étudier des enjeux de santé comme le stress, l'anxiété et la dépression des personnes à l'aide de données provenant d'outils numériques et de l'Internet. Si les bons algorithmes d'intelligence artificielle sont développés, il devient possible de prédire et de prévenir le suicide, minimisant ainsi le besoin d'entrer dans le système de santé. Dans ce contexte, les algorithmes d'intelligence artificielle pourraient représenter un dispositif médiateur situé entre l'expérience et les phénomènes. Ce dispositif médiateur enverrait des signaux aux chercheurs pour prédire et prévenir certains événements comme le suicide. Cependant, il importe de souligner que ces signes devront être compris par les chercheurs et que les algorithmes d'intelligences artificielles, c'est-à-dire les

dispositifs médiateurs, devront être en mesure d'adapter les signaux en leur faisant prendre une forme intéressante conformément au regard d'un chercheur en épidémiologie.

Dans le cadre de notre recherche, nous avons aussi identifié deux visions au sujet de l'échantillonnage populationnel en lien avec la pratique de l'épidémiologie numérique selon les propos des participants à la recherche de terrain. La première vision considère l'échantillonnage populationnel comme problématique avec la pratique de l'épidémiologie numérique en raison des incertitudes sur la représentativité des échantillons extraits à partir des données numériques. Ce phénomène signifie que des chercheurs peuvent être incertains de la fiabilité des signes et des apparences perceptibles à travers les analyses des données générées par les données non traditionnelles comme celles des médias sociaux.

La deuxième vision considère qu'il y aurait du potentiel pour élargir les échantillons populationnels lors des études épidémiologiques avec la pratique de l'épidémiologie numérique étant donné que les usagers d'Internet sont nombreux. Cette caractéristique de l'Internet peut rendre accessibles des échantillons plus grands avec l'utilisation des données numériques comparativement à l'utilisation de données traditionnelles. Dans une telle logique, les données non traditionnelles sont susceptibles d'offrir une meilleure compréhension des phénomènes étudiés grâce à l'amélioration de la représentativité des échantillons populationnels devenue possible avec l'utilisation massive d'Internet. Dans la prochaine section de ce chapitre, la confiance de chercheurs en épidémiologie envers des dispositifs méthodologiques sera analysée.

5.2 La confiance envers les données numériques comme dispositifs méthodologiques

Dans cette section, les analyses seront concentrées sur la confiance envers les données numériques comme dispositifs méthodologiques utilisés dans la pratique de l'épidémiologie numérique. Nous avons choisi les moteurs de recherche Internet, les médias sociaux et les téléphones cellulaires comme dispositifs méthodologiques puisque ces technologies figurent parmi les plus connues en ce qui concerne la pratique de l'épidémiologie numérique. Rappelons que les dispositifs méthodologiques sont vus comme des mélanges de matériel et de social prenant la forme d'arrangements (Law et Ruppert, 2013). Les dispositifs sont matériels, car ils peuvent comprendre,

et comprennent souvent, des pièces d'équipement. Or, ils sont aussi hétérogènes et sociaux étant donné qu'ils incluent également des personnes comme des techniciens, des utilisateurs et des membres d'une communauté (Law et Ruppert, 2013). Ces arrangements comprennent notamment des textes, des inscriptions, des représentations ou des symboles comme des statistiques sur les échantillons et des graphiques (Law et Ruppert, 2013).

En fait, les moteurs de recherche Internet, les médias sociaux et les téléphones cellulaires ne représentent pas des dispositifs méthodologiques en soi, car ils n'ont pas été conçus pour faire de la recherche. Des chercheurs comme Salathé (2018) et Salathé *et al.* (2012) ont commencé à s'intéresser aux données générées par ces technologies pour effectuer la recherche épidémiologique. La montée de l'intérêt pour ce type de données lors de recherches a incité certains chercheurs en STS à conceptualiser les technologies et les données numériques comme des dispositifs méthodologiques. Il est important de comprendre la confiance de chercheurs en épidémiologie envers ce type de données afin de développer un sens critique par rapport à la pratique de l'épidémiologie numérique.

5.2.1 Les données générées par des moteurs de recherche Internet

Les données générées par les moteurs de recherche Internet tels que *Google Trends* sont utilisées en recherche afin d'étudier entre autres les comportements des gens sur Internet (Mavragani et Ochoa, 2019) et les préoccupations (Basteris *et al.*, 2020). Selon Mavragani *et al.* (2028), *Google Trends* utilise les données sur les fréquences des termes de recherches employés par les usagers. Mavragani *et al.* (2018) soulignent aussi que *Google Trends* montre l'évolution de l'intérêt en ligne pour des séries chronologiques relatives à un terme donné, dans n'importe quel pays ou région, au cours d'une période choisie par le chercheur.

Google Flu Trends (GFT) a aussi été utilisé entre 2008 et 2015 comme un outil en ligne pour la surveillance en temps réel des épidémies de maladie (Carneiro et Mylonakis, 2009). Comme expliqué lors de la revue de la littérature sur l'épidémiologie numérique, GFT a constitué l'objet de certaines controverses engendrées par des incertitudes sur la validité de ses prédictions de la grippe (Pollett *et al.*, 2017b ; Salathé, 2018 ; Duclos, 2019). Malgré sa discontinuité en 2015, il

semble que d'autres recherches plus récentes dans la littérature examinée aient démontré le potentiel de GFT par l'utilisation de méthodologies améliorées dans les modélisations (Clemente *et al.*, 2019) et d'un nouvel algorithme (Guo *et al.*, 2021).

Afin de poser un regard critique sur ce genre de dispositif méthodologique, il est important d'observer les enjeux épistémiques liés aux données générées par les moteurs de recherche Internet à travers les points de vue de chercheurs en épidémiologie. Pour ce faire, nous effectuerons une analyse du niveau de confiance et de connaissance des données générées par les moteurs de recherche Internet en tant que dispositif méthodologique auprès de chercheurs en épidémiologie. Suivra l'utilité de ce dispositif méthodologique et ensuite les difficultés en lien avec l'utilisation de ce type de données pour des recherches épidémiologiques. Nous poursuivrons avec la confiance à l'égard des données générées par les moteurs de recherche Internet en analysant la confiance au sein de l'épidémiologie environnementale, de l'épidémiologie clinique et des chercheurs pratiquant l'épidémiologie numérique. Les propos des chercheurs en épidémiologie sociale de même que ceux des chercheurs ne pratiquant pas l'épidémiologie numérique ne démontrent pas de tendance particulière à l'égard de la confiance envers ce dispositif méthodologique puisque les points de vue exprimés par ces chercheurs peuvent être observés aussi auprès des autres catégories de participants. Cette section sera alors consacrée à l'épidémiologie environnementale, à l'épidémiologie clinique et aux chercheurs pratiquant l'épidémiologie numérique lors des analyses de la confiance à l'égard de ce dispositif méthodologique.

5.2.1.1 Le niveau de confiance et de connaissance du dispositif méthodologique

Selon les résultats de la recherche de terrain, la confiance varie chez les chercheurs en épidémiologie vis-à-vis des moteurs de recherche Internet comme dispositif méthodologique dans le cadre de recherches épidémiologiques. Certains chercheurs en épidémiologie ne connaissent pas ce type de dispositif méthodologique, tandis que d'autres chercheurs voient du potentiel dans les données générées par ce genre de dispositif méthodologique lors de certaines situations comme la surveillance syndromique ou pour étudier les préoccupations au sein des populations sur des enjeux de santé ou encore pour des recherches exploratoires afin d'aider à générer des hypothèses. Enfin,

d'autres chercheurs ne croient pas en la validité des données générées par les moteurs de recherche de Google.

Pour plusieurs participants à cette recherche, les propos ont indiqué qu'ils ne connaissent pas les moteurs de recherche Internet en tant que dispositif méthodologique dans le cadre de recherches épidémiologiques. Par exemple, un chercheur en épidémiologie sociale a affirmé connaître le fonctionnement du moteur de recherche Google, mais il ne peut pas se prononcer sur sa confiance envers un tel dispositif (entretien 6Biv). En outre, un chercheur de la même sous-discipline de l'épidémiologie a affirmé qu'il n'est pas familier du tout avec les approches de recherche utilisant ce genre d'outil et qu'il ne pourrait pas évaluer la qualité des données générées par cet outil (entretien 9Rxiv). Un autre participant en épidémiologie sociale estime pour sa part que GFT a un potentiel différent par rapport aux approches courantes (entretien 6Bxviii). Ce chercheur pense aussi que GFT a une niche, mais qu'il n'est pas certain de la manière dont celui-ci peut être utilisé pour des recherches épidémiologiques. Pour ce chercheur, GFT est difficile à évaluer, car il ne connaît pas l'algorithme derrière cette technologie.

Ce constat révèle que des chercheurs en épidémiologie connaissent le moteur de recherche Google, mais qu'ils ne l'ont jamais considéré comme une ressource pour s'adonner à de la recherche épidémiologique. Le moteur de recherche Google est connu par ces chercheurs, mais ils ne le connaissent pas comme dispositif méthodologique. Il est donc difficile de déterminer le niveau de confiance des chercheurs en épidémiologie envers les données générées par les moteurs de recherche Internet comme dispositif méthodologique. Les propos d'un des participants à la recherche de terrain renvoient à l'idée qu'un algorithme puisse agir comme un dispositif médiateur tel qu'analysé précédemment. Nous avons déjà souligné que les algorithmes d'intelligence artificielle pourraient représenter un dispositif médiateur situé entre l'expérience et les phénomènes. Dans ce cas-ci, le dispositif médiateur serait le moteur de recherche Google et celui-ci enverrait des signaux aux chercheurs permettant d'étudier la grippe. Encore faut-il que ces signes soient compris par les chercheurs et que les algorithmes d'intelligences artificielles, c'est-à-dire les dispositifs médiateurs, soient en mesure d'adapter les signaux en leur faisant prendre une forme intéressante selon le regard d'un chercheur en épidémiologie.

« I think I don't really know. I don't know what I don't know. I do think that the technology for counting these kinds of searches like frequency and so on I think that this is there. I don't think that lacks rigor but ... I couldn't tell you... I know that the way that Google work if I enter one search and you enter one search, you will get different results. » (Chercheur - épidémiologie sociale - entretien 6Biv)

« Honnêtement je dois vous dire que je ne suis pas si familier avec cet outil-là. Donc c'est à prendre avec un grain de sel. Ce que je veux voir mettons que j'évalue un manuscrit pour une revue scientifique, une des choses que je veux voire c'est les données qui sont liées à la fiabilité et à la validité de la mesure. Dans ce cas-là, je ne le sais pas. Donc je ne dis pas que ce n'est pas fidèle pis que ce n'est pas valide. Je ne le sais juste pas. » (Chercheur - épidémiologie sociale - entretien 9Rxiv)

« Je ne connais pas Google Flu Trends, mais selon ce que tu m'en parles, ça a un potentiel différent de ce qu'on utilise présentement. Donc je pense que ça a une niche et comment on peut utiliser ça dans nos recherches c'est une autre question aussi parce que c'est dur à savoir parce que je ne connais pas c'est basé sur quel algorithme. » (Chercheur - épidémiologie sociale - entretien 6Bxviii)

5.2.1.2 L'utilité du dispositif méthodologique

Selon certains participants, les moteurs de recherche Internet comme celui de GFT peuvent servir de signal numérique pour la surveillance syndromique. L'un des chercheurs en épidémiologie environnementale a mentionné que le recours à ce genre de technologie en épidémiologie implique des aspects syndromiques, car il s'agit de recherches fondées sur les symptômes reportés par les gens et qui comportent un aspect de volontariat puisque les gens reportent les symptômes sur une base volontaire (entretien 5Cxxvii). Ce chercheur a aussi fait remarquer qu'il faut comprendre les limites des informations véhiculées par ces données. Un chercheur en épidémiologie sociale associe pour sa part les recherches utilisant ce type de données à des signaux pointant vers certaines directions capables d'aider à générer des hypothèses (entretien 8Hxvi). Ces signaux doivent être vérifiés afin d'assurer leur validité, selon ce chercheur. Un autre chercheur en épidémiologie sociale associe les moteurs de recherche Internet comme celui de GFT à la surveillance numérique (entretien 6Bii). Ce participant soutient qu'il s'agit d'une bonne méthode de surveillance utilisant les données des recherches des gens sur Internet à propos de symptômes, mais ce genre de méthode ne fournit pas l'image complète sur un enjeu de santé et il faut déterminer comment faire en sorte que ce genre d'approche fournisse des informations plus fiables.

« Encore là, au niveau de confiance, on va parler d'indicateurs. On a un programme ici à l'Agence de santé publique du Canada qui n'est pas vraiment différent de ça aussi. Il demande aux gens d'aller sur une plateforme et de se porter volontaire pour décrire un symptôme de grippe par exemple, ça nous permet d'avoir une indication de comment est-ce que ça progresse dans la saison de grippe au travers ça. Donc il y a un aspect syndromique et un aspect de volontariat qui commencent à être beaucoup plus explorés et utilisés dans différentes méthodes. Je pense qu'il faut bien comprendre les limites de ces informations-là. Donc je pense que ça se rapporte beaucoup à la méthodologie de la surveillance syndromique et à l'utilisation d'indicateurs. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvii)

« These searches are signals. As a signal it points out to some directions. I think that there is more work to be done to see how actually accurate it is but yes, it would definitively have a signal that would help generate some hypothesis. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« One of the permissible methods of surveillance is digital surveillance. So, for example, it says we have issues with avian influenza or with Covid-19, it's totally acceptable to think, if you see a ton of people googling flu symptoms or wanting to obtain information on a certain drug at pharmacist, yes absolutely that's a good method of surveillance but it is not the whole picture. You still need to sort of look and see what other things you can do to make it more accurate. But it's quite I believe it would be quite important tool to use, yes. » (Chercheur - épidémiologie sociale - entretien 6Bii)

Plusieurs participants soutiennent que les signaux des moteurs de recherche Internet comme celui de GFT devraient être comparés à ceux provenant de sources différentes comme les données traditionnelles de surveillance pour déterminer la validité de ce type de signal (entretien 9Rxiv, entretien 10Txxii, entretien 6Bi, entretien 10Ixxviii). Selon ces chercheurs, les résultats fondés sur ce genre de dispositif méthodologique pourraient être valides s'ils sont les mêmes que ceux d'autres sources traditionnelles comme la surveillance des médecins, des données administratives ou des données de cohorte. L'un de ces chercheurs dont une partie du travail se situe dans la sphère politique a souligné que les signaux numériques doivent correspondre aux signaux conventionnels pour que ceux-ci puissent être utilisés lors des décisions politiques (entretien 10Ixxviii).

« Je pense que ça vaudrait la peine de regarder plus en détail peut-être utiliser certaines autres sources de données pour trianguler si on veut cette source d'information-là. Peut-être que des données seraient disponibles justement dans le réseau de la santé qui permettraient de

voir que si on obtient le même genre de résultat qu'avec les recherches que les gens font sur Internet. » (Chercheur - épidémiologie sociale - entretien 9Rxiv)

« If there has been a study to cross validate, so to go back to that flu example, if there is been very good research showing that the increase back increase changing flow of the flu data coming from Google predicts with , consistently predicts trends that we observe in other validated surveillance systems like sentinel, physicians surveillance or administrative data or cohort data, then I would build confidence in it, and I would say 'great we validated a source that might be more timely, it might be cheaper, more efficient and then I would feel good about it. But in the absence of that, I would be, at the start of it distrustful. » (Chercheur - épidémiologie sociale - entretien 10Txxii)

« You need to validate against other measures, and I don't know how far they've got in that way. Really something that needs to be explored. » (Chercheur - épidémiologie clinique-entretien 6Bi)

« I see policy limitations. So, in the kind of work I do we often move research policy quite quickly, I would never be comfortable doing that based on digital signals alone, but if digital signals were consistent with conventional signals, I think they can strengthen the policy argument. Digital signals alone I don't think they are adequate to make policy decisions on. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

5.2.1.3 Les difficultés liées au dispositif méthodologique

D'après les propos de certains participants à la recherche de terrain, les difficultés liées au dispositif méthodologique des moteurs de recherche Internet comme GFT semblent liées à la représentativité des populations analysée précédemment dans la section sur l'échantillonnage populationnel. En effet, les données générées par cette source de données peuvent exclure certaines populations qui n'utilisent pas Internet (entretien 6Pvi). Un autre chercheur en épidémiologie environnementale a affirmé qu'il peut y avoir de faux résultats positifs, car si les gens effectuent des recherches sur Internet à propos des symptômes de la grippe par exemple, cela ne signifie pas nécessairement que ces gens ont la grippe (entretien 10xxxv). Donc, selon ce chercheur, il peut y avoir une surestimation de la prévalence de la grippe si les chercheurs se fient uniquement à ces données.

« I think that you have to be very careful about potential biases. I just think about my parents, and I would never use any of that type of posting of anything on the Internet. I think you might exclude the older populations from a lot of studies which could be potentially problematic depending on ... the spread might be more harmful on older populations, and it could be a major health event for older people ... I think you have to think carefully about

who is using these systems and what and who do these data represent.» (Chercheur - épidémiologie sociale - entretien 6Pvi)

« Probably very so so... I know with the Flu Watch I haven't look to it so closely. I remember seeing with the flu watch that they overestimated the flu activity and what was going on with flu. Again because of that false positive that comes with that because people are just googling their symptoms and a lot of things cause fever and all that kind of things, and it might not be the real cause. As an emerging area, this field needs to go through a lot more rigor for me personally to feel more confident. I think this as a very rough idea of that might be happening. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxv)

5.2.1.4 La confiance au sein de l'épidémiologie environnementale

Les chercheurs en épidémiologie environnementale ayant participé à la recherche de terrain ont une confiance et des connaissances limitées envers les données générées par les moteurs de recherche Internet. Certains d'entre eux pensent que des recherches fondées sur des données générées par des moteurs de recherche Internet ne sont pas crédibles pour le domaine de la santé alors que d'autres ne sont pas intéressés par ce type de données et en connaissent très peu sur leur utilisation dans le cadre de recherches épidémiologiques.

Par exemple, l'un des chercheurs en épidémiologie environnementale a affirmé de manière catégorique que les données générées par Google ne sont pas crédibles pour des motifs de santé (entretien 5Gxxix). Ce chercheur soutient que ce type de données convient pour le marketing et non pour le domaine de la santé. Un autre chercheur dans la même sous-discipline de l'épidémiologie tend à se montrer plutôt sceptique envers les recherches fondées sur ce type de données (entretien 5Gxxv). Ce chercheur souligne tout de même qu'il y a toujours des exceptions et que pour tout type de recherche, il faut être honnête sur les limites des sources de données.

Un autre chercheur a communiqué un avis similaire en soulignant que la qualité de ce genre de recherche dépend largement du média et du contexte (entretien 5Cxxx). Ce chercheur pense que la représentativité peut être bonne, car tout le monde effectue actuellement des recherches sur la pandémie de COVID-19. Ce chercheur estime aussi que les analyses de ce type de données sont appropriées pour étudier la grippe puisque ce type d'analyse peut indiquer le début de la saison de

cette maladie. Cependant, ce chercheur n'est pas intéressé par ce type de données, car il s'intéresse plutôt aux mesures sur le terrain comme les mesures sur le niveau d'un virus dans l'eau. Ce chercheur a une plus grande confiance en ce type de mesure puisque, pour lui, les données générées par les moteurs de recherche Internet portent sur des réactions du public. Un autre chercheur en épidémiologie environnementale n'est pas attiré par ce domaine de recherche et il en connaît peu sur ce type de données (entretien 5Cxxvi). Il soupçonne qu'il existe des études comparant les signaux de Google avec ceux provenant du terrain, mais il ne connaît pas les résultats de ces comparaisons.

« Honestly, not credible. You could use Google for marketing purposes. That's what basically it is useful but for health purposes? No. » (Chercheur - épidémiologie environnementale - entretien 5Gxxix)

« Again, I think it depends a little bit on the research question. It certainly depends on the research question. I think I tend to be more skeptical of those kinds of studies. That being said, I think that there is always the exception that has done it. You know... ask a really good research question, use the data appropriately to answer that research question and just be really honest about the limitations of that data source and of the study but it still you know making a really interesting contribution. Unfortunately, I think that probably is the exception, I think most studies are not doing that, I think there is a shift but more often than that I am..... more unfortunate... there is some really big problems with that study and it's too bad that they didn't really take the time to address that. » (Chercheur - épidémiologie environnementale - entretien 5Gxxv)

« Peut-être que pour moi je vais avoir beaucoup de préoccupations basées sur les résultats de ce type de recherche. Je ne veux pas dire qu'il n'y a aucun avantage aucune contribution, c'est juste que ça dépend vraiment du média, du contexte, comme maintenant tout le monde recherche sur la pandémie tout le temps et ça ne représente rien. Je pense que c'est possible avec quelque chose comme la grippe, ça va signaler le début de la saison de la grippe, mais je pense que l'exemple de grippe peut-être que ce n'est pas une bonne utilisation parce qu'on a déjà savoir beaucoup de choses sur la grippe. Moi je suis très intéressée sur le type de projet comme la mesure de... il y avait une étude qui a été lancée à Montréal qui veut mesurer le niveau de virus dans l'eau pour savoir si on peut identifier la deuxième vague, quelque chose qui est plus vraiment basé sur quelque chose mesurable, je mets plus de confiance parce que sinon, c'est vraiment une réaction du public. » (Chercheur - épidémiologie environnementale - entretien 5Cxxx)

« Ce n'est pas vraiment un domaine de recherche qui m'attire, je sais que ça se fait, personnellement, ça ne m'attire pas trop, je vois bien qu'il y a des biens par ce genre d'études, je pense que c'est des signaux qui doivent être valides, mais j'avoue que je ne connais pas assez ce genre d'étude-là et puis j'imagine qu'il doit y avoir des études comparatives qui comparent les signaux Google versus le terrain. Et puis qui doivent donner une espèce

d'analyse de ce genre de signaux-là, mais je ne connais pas les résultats. Donc je ne peux pas vous dire si j'ai confiance ou pas en les résultats. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvi)

5.2.1.5 La confiance au sein de l'épidémiologie clinique

En ce qui concerne les participants en épidémiologie clinique, leur confiance serait assez limitée envers les données générées par les moteurs de recherche Internet dans les contextes de recherches en santé. Plusieurs chercheurs en épidémiologie clinique établissent une distinction claire entre des recherches sur des opinions du public et des recherches ayant comme but d'aider dans les prises de décisions sur les soins de santé lorsqu'ils parlent de leur niveau de confiance envers les données produites par les moteurs de recherche Internet. Presque tous les chercheurs en épidémiologie clinique ayant pris part à la recherche de terrain soutiennent que les données générées par les moteurs de recherche Internet peuvent être utiles pour des recherches sur les opinions et préoccupations des populations. Cependant, ces chercheurs affirment clairement que ce type de données n'est pas utile dans le cadre des recherches sur la santé des populations ni pour la surveillance des maladies.

Par exemple, l'un des chercheurs en épidémiologie clinique soutient que les données générées par les moteurs de recherche Internet peuvent s'avérer très efficaces en marketing, mais il est très sceptique envers ce type de données pour établir des diagnostics sur la santé (entretien 7Mxxi). Il semble que ce chercheur associe l'utilité des données générées par les moteurs de recherche Internet au domaine du marketing et non à celui de la santé. Ces propos semblent indiquer que la pratique de l'épidémiologie numérique est vue comme moins appropriée en épidémiologie clinique que dans d'autres sous-disciplines de l'épidémiologie en raison de la nature des données et des informations utilisées lors des recherches cliniques. La culture épistémique de la pratique de l'épidémiologie numérique serait différente par rapport à la culture épistémique de l'épidémiologie clinique. Selon ce chercheur, il faut d'autres recherches comparatives pour faire confiance aux données générées par les moteurs de recherche Internet dans un contexte où le chercheur doit établir un diagnostic de santé. Des points de vue similaires ont été constatés auprès de cinq autres chercheurs en épidémiologie clinique.

L'un des problèmes soulevés est les difficultés à suivre le nombre de visiteurs uniques utilisant le moteur de recherche Internet (entretien 6Axix). Un individu peut chercher une information une centaine de fois. Pour obtenir des données épidémiologiques valides, il faudrait identifier les utilisateurs uniques et connaître la proportion de personnes ayant accès à la technologie et la population étudiée selon les propos de ce chercheur.

« I heard of some situation where personal conditions have been predicted very accurately. And that was in the marketing context. That maybe a very powerful path in marketing and they might have more resources than health care systems. So, from that standpoint, I would think that it could be powerful. I have a skepticism for health diagnosis that are made in that way. I wouldn't want someone to be treated. In order to really have trust, I would have to see comparative research, I guess. I think that there may be some potential, but I would want to see a comparison of that diagnosis versus if someone was going to be using that to looking at epidemiological trends of a particular condition. » (Chercheur - épidémiologie clinique - entretien 7Mxxi)

« It would depend on the question. Even when we do measure ourselves, we can't track the number of unique visitors that click on any of our tools. We can only track the number of visits. So, it is possible that someone search information on something 100 times. We would need to identify the unique users in order to do that to get some valid epidemiological data and also, we would need to know what proportion of people even have access to technology and population. And for the population, we need to know what the denominator is. So, I think we could at the very least look at trends on what people are looking at, we can see exactly what people are looking for so we can do some very simple descriptive epidemiology without being able to do any kind of surveillance with those numbers. » (Chercheur - épidémiologie clinique - entretien 6Axix)

« Ouf ! You know I have a certain degree of confidence and then I have a certain degree of that I would want to sort of a sense double-checking it out. Like run some algorithms would give me some more confidence in the data ... I guess I wouldn't necessarily accept the data without running some algorithms to make sure at a simplistic level that they are providing the data. » (Chercheur - épidémiologie clinique - entretien 6Biii)

« I can't really comment but the first thing that comes in my mind is that I am doing research on vaccines responses in older adults mostly to determine vaccine responses but one thing that I know by doing searches on that area is that understanding why people do not get vaccinated is a very large area of research. There is a lot of questions sort of why people are not getting vaccinated and what are the causes of it. So, I could see this type of database being very useful for people looking at social trends. There may be a lot of chat on the Internet about influenza and that might relate to spikes of vaccination or how people see vaccination. So, I could see it being useful from a social network research type of area. How it would apply to biomedical research, that I hesitate to comment, I am not really sure. » (Chercheur - épidémiologie clinique - entretien 6Fxi)

« I think you're touching one of the issues, you need to validate against other measures, and I don't know how far they've got in that way. Really something that needs to be explored. The potential benefit of real mass of the data that they have access to does seem to lend itself to interesting questions based on those kinds of searches and in that sector, there is a lot more questions that could be addressed using similar methodology and so. To me it is a matter of again how well the question is matched to the data and whatever there is validity checks that are going into it and how convincing those are in general. My only knowledge of Google Flu Trends is that this is kind of an interesting demonstration, but it is not something necessary guiding healthcare decision in an important way. It is a very interesting demonstration that needs to be validated. » (Chercheur - épidémiologie clinique - entretien 6Bi)

« Having not use them, I would expect that they would be reliable, and the validity would depend on probably more public health questions and their inference-based kind of questions like biomedical, clinical questions certainly. I can't imagine a use case for those. I think it would be for public health concerns. I think their validity is specific to time and place and influence and that it doesn't necessarily represent a yearly or a temporal phenomenon that is going to continue right? Something seasonal etc. or for instance people's concerns around vaccination. In this current state, oil is probably a change rapidly or in the next couple of years especially at this current prices it will probably change that dynamic. People expressing concerns and opinions last year and this year on those issues might change rapidly so any analysis of people's concerns is always over time. So public health and I am concerned with temporal changes. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

Parmi les participants à la recherche de terrain, un seul chercheur en épidémiologie clinique envisage que les données générées par les moteurs de recherche Internet puissent se révéler utiles pour des recherches épidémiologiques comme tout autre ensemble de données (entretien 6Biii). Cependant, ce chercheur souligne qu'il faut avoir une bonne connaissance du fonctionnement de l'algorithme derrière le moteur de recherche afin de bien interpréter et classer les données. Selon lui, il faut toujours expliquer les forces et les faiblesses dans toute recherche. La force des moteurs de recherche Internet s'avère être la grande quantité de données disponible. La faiblesse des données générées par ces technologies réside dans l'incertitude sur la cohérence entre la logique de l'algorithme derrière le moteur de recherche et les types de recherche.

« Je ne suis pas très familier... qu'est-ce que c'est *Google Trends* ? Je vais avoir besoin de plus d'explications. C'est les trucs des big data... j'ai vu quelque chose là-dessus, j'ai vu des schémas avec des bulles de différentes tailles. En fait je dirais qu'il faut comprendre pour savoir, faire de la bonne recherche avec ça il faut comprendre comment le moteur de recherche fonctionne par exemple, si vous mettez votre nom sur Google, il va y avoir des sites qui vont apparaître en premier, mais il y a une logique derrière, donc il faut comprendre

comment le moteur de recherche priorise et classe l'information pour... parce que si la personne n'en est pas consciente, elle va faire de la mauvaise interprétation des data parce que les résultats qu'elle va voir seront basés sur l'algorithme utilisé pour la classification des données. Donc je crois qu'à partir du moment où un chercheur comprend comment le moteur de recherche fonctionne bien comme toute approche en science, il faut toujours parler des forces et des faiblesses de la recherche donc je crois qu'il y a des forces. La force c'est les tonnes de data disponibles qui sont invraisemblables. La faiblesse c'est qu'il y a l'algorithme qui utilise une certaine logique et que du coup peut-être que la logique de l'algorithme ne va pas matcher avec tous les types de recherche forcément.» (Chercheur - épidémiologie clinique - entretien 6Fx)

Nous avons souligné dans le chapitre précédent que les propos de ce chercheur révèlent une attitude de réserve envers la pratique de l'épidémiologie numérique. Cette attention sur l'algorithme derrière le moteur de recherche Google illustre l'idée que les données et les mesures sont contingentes à l'appareil de mesure et que celles-ci sont des représentations de cet appareil plutôt que du monde réel. Par conséquent, le moteur de recherche Google (ou le détecteur) doit être correctement compris par le chercheur parce que c'est à travers cette technologie qu'il pourra analyser des tendances dans les recherches effectuées par les gens sur Internet. Ainsi, le moteur de recherche Google peut être conceptualisé comme un détecteur agissant de la même façon qu'un dispositif médiateur interposé entre les chercheurs et les phénomènes. D'après cette conception, données et les mesures dépendent de l'appareil de mesure et constituent des représentations de cet appareil (Knorr-Cetina, 1999).

Selon ce raisonnement, la manière dont est construite une technologie numérique comme celle du moteur de recherche Google peut influencer les analyses du chercheur en fonction de la conception de l'algorithme derrière ce moteur de recherche. Le moteur de recherche Google et son algorithme deviennent alors un dispositif médiateur entre le chercheur et les phénomènes lors de la pratique de l'épidémiologie numérique. Cette nouvelle pratique peut alors constituer une culture épistémique différente comparée à celle de l'épidémiologie puisque l'algorithme semble jouer un rôle dans les prises de décision et dans le déroulement des recherches lors d'une telle pratique de recherche.

5.2.1.6 La confiance au sein des chercheurs pratiquant l'épidémiologie numérique

Enfin, il est important d'analyser les propos sur la confiance envers les données générées par les moteurs de recherche Internet chez les chercheurs en épidémiologie pratiquant l'épidémiologie numérique, car ils peuvent avoir des perspectives différentes comparativement à ceux des autres chercheurs. Les chercheurs en épidémiologie ayant l'expérience de cette nouvelle pratique expriment une vision assez positive du potentiel des moteurs de recherche Internet comme celui de Google pour des recherches épidémiologiques. L'un des chercheurs en épidémiologie sociale pratiquant l'épidémiologie numérique affirme notamment qu'il pense que le potentiel de Google est très élevé pour des recherches épidémiologiques, mais il insiste sur les défis liés à la propriété des données et sur les difficultés à faire confiance à Google (entretien 8Sxx). Ce chercheur a souligné qu'il sera toujours sceptique vis-à-vis des données générées par Google. Ce à quoi ce chercheur ferait davantage confiance dans les données de Google serait les données générées par ses propres citoyens engagés directement avec lui dans ses recherches.

Il semble que la réponse fournie par ce chercheur sur sa confiance envers les données générées par les moteurs de recherche Internet implique des enjeux sociopolitiques montrant ainsi la symétrie de la coproduction des connaissances (Jasanoff, 2004), c'est-à-dire que les connaissances produites lors des recherches impliquent des enjeux épistémiques, mais aussi des enjeux sociaux et politiques concernant entre autres la propriété des données et la confiance envers certains acteurs détenteurs des données. De tels enjeux sont complexes et nécessitent des analyses approfondies effectuées dans le prochain chapitre portant sur les enjeux sociopolitiques de la pratique de l'épidémiologie numérique.

Le chercheur en épidémiologie environnementale pratiquant l'épidémiologie numérique estime que Google peut inclure des échantillons représentatifs de populations (entretien 10Ixxviii). Ce chercheur souligne aussi qu'il connaît assez bien les limites et les forces de ce dispositif. Selon ce chercheur, les analyses des données de Google ne constituent pas des preuves en soi, mais elles sont utilisées de manière assez universelle pour saisir ce que les gens ont en tête. En outre, si ces analyses de données sont traitées de manière adéquate, il est fort possible d'en tirer quelque chose d'assez utile d'après les propos de ce chercheur. Le point de vue de l'autre chercheur en épidémiologie sociale ayant l'expérience de la pratique de l'épidémiologie numérique va dans le

même sens puisque ce chercheur pense que les données générées par Google peuvent être une bonne chose en épidémiologie (entretien 6Axxiii). Cependant, ce chercheur mentionne que l'utilisation que ces populations font de ces technologies est différente, car il y a des utilisateurs de ces technologies, mais aussi ceux ne voulant pas les utiliser.

« I think the potential is very high, the difficulty is how much do you trust Google? That is a difficult thing because they own the data. To me that is the difficult issue there in understanding what is happening. I find that I will always be a bit skeptical, right? And that's a good thing. What I would trust more in Google data is data coming from my own citizens engaging with us directly. So that would be my answer to that. I am not saying I don't have confidence in it, but I would be a bit skeptical just based on the fact that these are coming from big tech. » (Chercheur - épidémiologie sociale- pratiquant de l'ÉN - entretien 8Sxx)

« I actually see that as stronger than social media. I think it probably captures more representative samples of the population. You know I am pretty aware of Google and where his strength limitations are. So, it is not full of proof for anything, but I think that Google is fairly universal used to capture what are on people's minds. And if you can deal with these data carefully, you can probably get something quite useful out of it. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

« I think it would be nice. There is going to be people who are probably different from people who respond to this kind of survey or whatever on those pages versus people who don't. » (Chercheur - épidémiologie sociale- pratiquant de l'ÉN - entretien 6Axxiii)

En somme, cette analyse des points de vue exprimés par les participants à notre recherche quant aux analyses de données générées par les moteurs de recherche Internet fait ressortir les enjeux épistémiques suivants :

- Les moteurs de recherche Internet comme ceux de Google semblent peu connus comme dispositif méthodologique ;
- La confiance des chercheurs en épidémiologie envers les données générées par les moteurs de recherche Internet varierait en fonction de leur expérience de la pratique de l'épidémiologie numérique et selon leurs pratiques de recherche ;
- Il y aurait du potentiel envers les données générées par les moteurs de recherche Internet lors de la surveillance syndromique ou pour étudier les préoccupations au sein des populations sur des enjeux de santé ou encore dans le cadre de recherches exploratoires afin

- d'aider à générer des hypothèses à la condition que les résultats fondés sur ce type de données soient comparables à ceux obtenus avec d'autres sources de données ;
- Certains chercheurs ne croient pas en la validité des données générées par les moteurs de recherche sur Internet pour des recherches épidémiologiques ;
 - Les réserves vis-à-vis des données générées par les moteurs de recherche Internet seraient liées à la représentativité des populations et aux difficultés à suivre le nombre de visiteurs uniques utilisant le moteur de recherche ;
 - Les chercheurs en épidémiologie environnementale n'ayant pas pratiqué l'épidémiologie numérique et ceux en épidémiologie clinique semblent avoir une confiance assez limitée envers des recherches utilisant des données générées par des moteurs de recherche Internet ;
 - Plusieurs chercheurs en épidémiologie clinique établissent une distinction claire entre des recherches portant sur des opinions du public et des recherches portant sur les soins de santé lorsqu'ils forment leur opinion sur les données générées par des moteurs de recherche Internet. Certains envisagent l'utilité de ce type de données pour des recherches d'opinion et aucun de ces chercheurs ne voit la pertinence de ce type de données lors des recherches sur les soins de santé ;
 - Les chercheurs pratiquant l'épidémiologie numérique auraient une confiance assez bonne envers les données générées par les moteurs de recherche Internet pour des recherches épidémiologiques à la condition que ces données soient utilisées adéquatement en fonction du contexte et de leurs limites.

D'après nos analyses, une technologie peut être connue par les chercheurs d'une discipline, mais celle-ci n'est pas nécessairement considérée comme un dispositif méthodologique utile pour des recherches par tous les chercheurs de cette discipline. L'algorithme derrière un moteur de recherche Internet peut agir comme un dispositif médiateur, c'est-à-dire un intermédiaire entre les phénomènes étudiés lors des recherches et le chercheur (Knorr-Cetina, 1999). Par exemple, un dispositif médiateur comme Google enverrait des signaux aux chercheurs pour étudier la grippe. Encore faut-il que ces signes soient compris par les chercheurs et que les algorithmes d'intelligences artificielles, c'est-à-dire les dispositifs médiateurs, soient en mesure d'adapter les signaux en leur faisant prendre une forme intéressante selon le regard d'un chercheur en épidémiologie.

La pratique de l'épidémiologie numérique serait vue comme moins appropriée en épidémiologie clinique que dans d'autres sous-disciplines de l'épidémiologie en raison de la nature des données et des informations utilisées lors des recherches cliniques d'après nos analyses sur les données générées par les moteurs de recherche Internet en tant que dispositif méthodologique. Il semble donc que la culture épistémique de la pratique de l'épidémiologie numérique serait très différente par rapport à la culture épistémique de l'épidémiologie clinique parce que la première se concentrerait sur des recherches d'opinions et des tendances sociales alors que la seconde porterait sur des recherches cliniques basées sur des données médicales des patients.

5.2.2 Les données générées par des médias sociaux

Les données en provenance des médias sociaux comme celles de Twitter peuvent aussi être utilisées comme dispositif méthodologique en vue d'effectuer des recherches épidémiologiques. Par exemple, Roche *et al.* (2017) ont combiné des données de Twitter et des données traditionnelles lors d'une recherche épidémiologique visant à identifier les principaux moteurs de la dynamique temporelle et spatio-temporelle de l'épidémie de la maladie infectieuse Chikungunya en Martinique. Nous analyserons la confiance à l'égard des données générées par les médias sociaux pour des recherches épidémiologiques à travers les points de vue d'ordre épistémique de chercheurs en épidémiologie envers ce type de données. Contrairement aux données générées par les moteurs de recherche Internet expliquées précédemment, nos analyses n'indiquent pas de tendance particulière selon la sous-discipline (épidémiologie environnementale, épidémiologie sociale, épidémiologie clinique) ni en fonction de l'expérience de la pratique de l'épidémiologie numérique. Les paragraphes suivants seront donc organisés selon les enjeux retrouvés dans les points de vue exprimés par les participants à cette recherche. Cette analyse laisse entrevoir deux enjeux principaux dans les propos des participants, à savoir les réserves et l'utilité des données générées par les médias sociaux pour des recherches sociales.

5.2.2.1 Les réserves

De manière générale, la plupart des chercheurs en épidémiologie auraient peu confiance envers les données générées par les médias sociaux pour des recherches épidémiologiques selon les résultats de la recherche de terrain. Certains d'entre eux ont affirmé clairement qu'ils sont sceptiques envers ce type de données. Un chercheur en épidémiologie sociale soutient qu'il est difficilement concevable à ce que la qualité des données générées par les médias sociaux soit suffisante pour répondre à des questions de recherche (entretien 6Dxii) et les propos d'un autre chercheur en épidémiologie environnementale indiquent qu'il estime que l'échantillon obtenu à partir de ces données n'est pas approprié (entretien 5Gxxix).

« I have trouble believing that the data would be of sufficient quality to answer questions. I'm afraid that I am rather skeptical of the success that we are going to see because of the biases that I mentioned previously. » (Chercheur - épidémiologie sociale - entretien 6Dxii)

« I think that if I was to see the title, I would be very skeptical. Because I don't think you have a proper sample. You know, everything in epidemiology is all about sampling populations. » (Chercheur - épidémiologie environnementale - entretien 5Gxxix)

Certaines réserves liées aux données générées par les médias sociaux ont souvent été observées dans les discours des participants à la recherche de terrain. Ces réserves concernent l'échantillonnage des populations (comme discuté précédemment) et les incertitudes sur la validité des informations publiées dans les médias sociaux.

Les réserves liées à l'échantillonnage résident principalement dans la représentativité des populations étudiées. D'après les propos de plusieurs chercheurs en épidémiologie, la représentativité des populations peut être faussée par les données générées par les médias sociaux, car certaines populations comme les enfants ou les personnes âgées ne les utilisent pas, occasionnant ainsi des problèmes lors de la généralisation des résultats à l'ensemble d'une population (entretien 6Fxi, entretien 5Kxv, entretien 1Oxxxiv, entretien 1Oxxxv, entretien 5Cxxx). Selon plusieurs chercheurs, l'utilisation des médias sociaux peut dépendre du groupe d'âge et en conséquence, certaines populations peuvent se retrouver exclues de l'étude épidémiologique utilisant des données générées à partir des médias sociaux (entretien 1Oxxxiv), tandis que d'autres

peuvent être surreprésentées dans ces données, comme les jeunes utilisant beaucoup les médias sociaux (entretien 5Cxxx) ou des individus très vocaux sur les plateformes (entretien 5Kxv). L'un des participants a aussi souligné que tout dépend de ce que les auteurs de la recherche essaient vraiment de dire sur la portée des résultats obtenus avec les données générées par les médias sociaux (entretien 1Oxxxv).

« The second part of it is that the sampling is very uncertain. So in a proper study design for epidemiological surveys you are going to be trying to sample a representative set of the population that you are trying to understand better and when you just provide a survey out to the masses, even if they are directed as far as you and e-mails invitation that there is something online, you still don't know who exactly is answering and even if this will become less and less of a problem as we are moving forward, some of this does not have access to Twitter so these persons would not answer the questionnaire and similarly for people who don't have access to the Internet which is becoming less and less frequent (but these persons would not answer either). So really your representation of your subset is going to be limited and you can never be surely confident of the approach. » (Chercheur - épidémiologie clinique - entretien 6Fxi)

« Les gens utilisant Twitter ne représentent pas tout le monde. Ce n'est pas tout le monde qui est sur la plateforme Twitter donc ça ne représente pas tout le monde et ceux qui sont les plus vocaux sur cette plateforme, ça ne représente pas non plus tous les utilisateurs de la plateforme. C'est sûr qu'il y a plusieurs limites et c'est juste qu'il faut en prendre conscience. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

« Since you are not from epidemiology, the most important is representativeness, who are included in the population, if you want to generalize with a nation's population, not all the people use Facebook, but I think still the use is qualified to a specific age group. If you want to do a study in a Quebec population, on Facebook it might be a bias because some people don't use Facebook at all. If you target a specific age group for the study, it is fine. But if you want to generalize it, I wouldn't recommend the use of Facebook or Twitter. I really don't know... I don't like it; I think it is too much... Facebook could be interesting for an age group. It is good to approach people who you don't know, if you question the bias to use Facebook, I think it is fine. But I stand that Facebook doesn't represent all the population. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxiv)

« It would depend on how far they are going with the results. It depends on how much they are trying to generalize from it. Again, it is understood that the population is key information. For example, among these groups of people that would typically use Twitter, these are the trends we see, or these are the patterns, or these are the issues. That can't be necessarily applied to the general population. There is a large of subset of the population who do not use Twitter. Again, it depends on what really the authors of the research are trying to say what respect to how far they can take their results. I don't say that there couldn't be very useful information that would come out of that particularly from a qualitative standpoint. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxv)

« Je peux imaginer comme les réseaux sociaux comme Facebook et Instagram ça peut nous aider parce que c'est un réseau utilisé de plus en plus par les jeunes, mais le problème est qu'il faut nous demander ce que les réseaux sociaux en ligne représentent vraiment. Est-ce que ça représente des connexions fortes, est-ce que ça représente la personne que tu vas contacter pour un trouble de santé mentale, probablement non parce que c'est comme un portrait que tu vas donner à l'extérieur et je pense qu'il y a des limites à utiliser ça. Mais je ne veux pas dire que ce n'est pas important, parce que ça va remplacer la connexion en faisant le lien avec la Covid-19, la pandémie, mais je pense que c'est quand même au début de ce qu'il faut faire en recherche, ce n'est vraiment pas mon domaine de recherche. Il faut vraiment s'assurer que ce qu'on mesure ça représente quelque chose qui était le but, je pense. » (Chercheur - épidémiologie environnementale - entretien 5Cxxx)

De plus, nos participants ont souvent affirmé que les problèmes concernant la représentativité des populations peuvent être liés aussi à des minorités bruyantes sur les médias sociaux s'exprimant plus souvent et de manière plus intense que d'autres groupes (entretien 6Fviii, entretien 10Ixxviii, entretien 8Hxvi). Des biais peuvent alors se produire lors des recherches si le contexte de celles-ci n'est pas pris en compte (entretien 8Hxvi).

« A small number of individuals dominate conversations in social media not just a short number of Tweets for instance which is definitely the case but on their influence on other Tweets, etc. That it may not be at all based on the entire population necessarily. So yes, representativeness, what the question is would be a short answer. It would depend entirely on validity. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

« Let's say that there is something about sentiment vaccination, I don't think that the sentiment towards vaccination is reflective of the sentiment towards vaccination in the general population. There is a very noisy minority on Twitter, trying to disrupt to how people feel towards this important health intervention. In that case, unless I saw other information about vaccine sentiment based on a less biased way, I am not going to put a lot of confidence in the material. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

« I am giving you a lot of 'depend' because there are so many contingencies in all these things. It will depend on what you are looking at. And the problem with something like Twitter and something like Facebook or whatever might be (I don't use any of them. I am ignorant of the use of it) is that it can be biased by a few loud people. everybody can go out there and say things that may or not may not be accurate. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

Les incertitudes sur la qualité et la validité des informations véhiculées sur les médias sociaux constituent une autre réserve mentionnée à quelques reprises par des participants à la recherche de terrain. Par exemple, un chercheur en épidémiologie clinique a affirmé qu'il est très difficile de savoir qui sont réellement les gens divulguant des informations sur les médias sociaux et si ces informations sont communiquées de bonne foi (entretien 6Fixi). Cependant, certaines informations obtenues à partir des médias sociaux peuvent se révéler utiles comme le nombre de comptes par individu et les distances géographiques entre différents comptes interconnectés, mais certaines de ces informations ne seraient pas fiables (entretien 6Fix).

« The Twitter is not perfect because respondents can reply twice, you don't know who they are. You only trust that they are who they are saying or hope that they are. There is really no screening process. Survey Monkey approach is very similar. I mean you are able to walk responses and you are able to gather more information using the Survey Monkey approach so that you can at least be able to identify who that person is but again you don't necessarily know who they are and whether or not they are answering in good faith. » (Chercheur - épidémiologie clinique - entretien 6Fixi)

« Not so much for the reason I have given. You don't know who is providing the information, you don't know why they are providing that information. But there are metrics which are reliable: you can measure how many accounts a person is connected to (note: I am saying how many accounts and not how many persons are actually connected), you can measure the geographic distance between different accounts connected together, this is useful information that can be extracted from Twitter. But some of this information is not reliable at all. » (Chercheur - épidémiologie sociale - entretien 6Fix)

Il est possible que ces réserves liées à la représentativité des échantillons populationnels ainsi qu'aux incertitudes sur la qualité et la validité des informations signifient que les signes et les apparences perceptibles à travers les analyses des données générées par les médias sociaux peuvent conduire vers des résultats erronés et ainsi, vers des perceptions déformées par rapport à la réalité des phénomènes étudiés. Il est aussi possible que ce dispositif méthodologique amène le chercheur à voir ou à ne pas voir certains phénomènes étant donné la façon dont ce dispositif est conçu (Law et Ruppert, 2013). Les caractéristiques du dispositif méthodologique constitué par les données générées à partir des médias sociaux semblent dictées par l'inclusion et l'exclusion de certaines populations en fonction de leur utilisation des médias sociaux ainsi que de l'information fournie de la part des usagers des médias sociaux.

5.2.2.2 L'utilité des données générées par les médias sociaux pour des recherches sociales

Enfin, plusieurs participants à la recherche de terrain ont affirmé qu'ils auraient une certaine confiance envers des recherches fondées sur des données générées par les médias sociaux si celles-ci portent sur des opinions, des tendances, des attitudes et des enjeux sociaux (entretien 7Mxxi, entretien 5Kxy, entretien 6Axix, entretien 6Fviii, entretien 6Pvi). Plusieurs chercheurs ont exprimé peu de confiance envers des recherches épidémiologiques utilisant des données générées par des médias sociaux portant sur les soins de santé, les diagnostics et la prévalence des maladies d'après les propos communiqués lors des entretiens (entretien 6Axix, entretien 6Pvi).

« It depends a little bit on the research question, if it's a question about people's interests, I would have a lot of confidence. Something like Twitter probably amplifies certain trends that would depend on the interest of people who would do that a lot. Sometimes a subject is not discussed because it takes someone to first tweet about something in order for that to ... so is that question or tweet being captured of anything of interest, I don't know. » (Chercheur - épidémiologie clinique - entretien 7Mxxi)

« C'est de l'information qui est intéressante, car ça nous parle de tendances, l'information est populaire, mais il faut prendre ça avec les limites qui sont associées à ces données-là. » (Chercheur - épidémiologie sociale - entretien 5Kxy)

« It depends on what the research question is. If the research is just mainly on a quantitative side by looking generally at what people think, share then I would have a lot of confidence because that kind of research just try to understand a point of view understanding that it is not all points of view. If it was on trying to determine numbers that we are looking at incidents such as disease frequency and things like that. That's different. And those types of questions. I would have less confidence in that. But I don't think that those would be the kind of questions that would be asked when using these methods. » (Chercheur - épidémiologie clinique - entretien 6Axix)

« It would depend entirely on what the question was. If the question was around perceptions of something, then Twitter would be maybe fine. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

« I think that if you are asking questions about medical diagnostic, I would say that Twitter is probably not good because you are looking at self-reported stuff. But if you are asking things about self-perceptions on some sort of treatment that people have gone and maybe that would be OK. I think it really depends on the question. » (Chercheur - épidémiologie sociale - entretien 6Pvi)

Seuls quelques participants ont exprimé une vision généralement positive de l'utilisation de données générées par les médias sociaux comme dispositif méthodologique. Par exemple, l'un de ces participants en épidémiologie sociale a affirmé que sa confiance envers les données générées par les médias sociaux s'avère similaire à celle de n'importe quelle autre source de données en tenant compte de la méthodologie employée (entretien 5Uxxxvi). Un autre chercheur de la même sous-discipline affirme que l'information est intéressante parce que celle-ci indique des tendances populaires (entretien 5Kxv). Twitter représente actuellement une bonne fraction de la population mondiale ayant accès à ce genre de bases de données (entretien 5Cxxvi) selon un des participants en épidémiologie environnementale. Ce chercheur souligne tout de même qu'il se poserait des questions sur la comparabilité des résultats entre la population source et la population échantillonnée.

Un autre chercheur en épidémiologie environnementale a souligné que les données générées par les médias sociaux peuvent être utiles afin de recruter des participants aux recherches et pour étudier des attitudes (entretien 1Oxxxiv). Ce type de données serait aussi pertinent pour générer des hypothèses d'après un autre des participants en épidémiologie environnementale (entretien 5Gxxv). À la lumière de ces propos, il semble que les aspects positifs associés aux données des médias sociaux soient l'accessibilité à ces données ainsi que leur potentiel pour des analyses de tendances, pour recruter des participants aux recherches et pour générer des hypothèses.

« Euh... pas moins (de confiance) que n'importe quelle autre source... dans le sens où j'irais voir comment elle a été faite, comment les données ont été recueillies. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

« C'est de l'information qui est intéressante, car ça nous parle de tendances, l'information est populaire, mais il faut prendre ça avec les limites qui sont associées à ces données-là. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

« Hum... je pense c'est... Twitter ça représente quand même une bonne fraction de la population mondiale actuellement qui a accès à ce genre de bases de données, donc c'est sûr que je me poserais la question sur la population source versus la population échantillonnée, la comparabilité, et puis voilà essentiellement, mon questionnement se centrerait là-dessus. Et puis, oui, voilà. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvi)

« I think the social network is good for recruitment... It might be useful to study attitudes. So, I see it as an open source. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxiv)

« From sort of a hypothesis generating standpoint, it seems quite interesting. » (Chercheur - épidémiologie environnementale - entretien 5Gxxv)

Ces analyses sur l'utilité des données générées par les médias sociaux révèlent que des recherches utilisant ce dispositif méthodologique seraient vues comme appropriées si celles-ci portent sur des opinions, des tendances, des attitudes et des enjeux sociaux. Il y aurait peu confiance envers des recherches épidémiologiques utilisant des données générées par des médias sociaux pour étudier des enjeux cliniques comme les soins de santé, les diagnostics et la prévalence des maladies d'après les propos communiqués lors des entretiens. Comme expliqué précédemment, il semble que la culture épistémique de la pratique de l'épidémiologie numérique puisse être vue comme très différente par rapport à la culture épistémique de l'épidémiologie clinique parce que la première se concentrerait sur des recherches d'opinions et des tendances sociales alors que la seconde serait axée sur des recherches cliniques basées sur des données médicales des patients.

Nos analyses sur la confiance de nos participants envers les données générées par les médias sociaux en tant que dispositif méthodologique font ressortir les enjeux épistémiques suivants :

- Les chercheurs en épidémiologie semblent avoir peu de confiance envers les médias sociaux comme dispositif méthodologique ;
- Les problèmes les plus souvent observés lors de la recherche de terrain concernent l'échantillonnage des populations et les incertitudes sur la validité des informations publiées dans les médias sociaux ;
- Il y aurait une certaine confiance envers des recherches fondées sur des données générées par les médias sociaux si celles-ci portent sur des opinions, des tendances, des attitudes et des enjeux sociaux ;
- Il semble y avoir peu de confiance envers des recherches épidémiologiques utilisant des données générées par des médias sociaux si celles-ci étudient les soins de santé et la prévalence des maladies et si leur but consiste à établir des diagnostics ;
- La facilité d'accès aux données des médias sociaux ainsi que leur potentiel dans le recrutement des participants aux recherches et à générer des hypothèses sont les éléments

positifs mentionnés par quelques participants à la recherche de terrain à l'égard de ce type de données.

Les réserves liées à l'utilisation de données générées par les médias sociaux comme dispositif méthodologique peuvent montrer que les signes et les apparences perceptibles à travers les analyses de ces données sont susceptibles de conduire à des résultats erronés et ainsi, à des perceptions déformées par rapport à la réalité des phénomènes étudiés. Le tout en raison des problèmes de la représentativité des échantillons populationnels et des incertitudes sur la qualité et la validité des informations véhiculées sur les plateformes. Il est aussi possible que ce dispositif méthodologique amène le chercheur à déceler ou non certains phénomènes étant donné la façon dont ce dispositif est conçu (Law et Ruppert, 2013). Les caractéristiques du dispositif méthodologique constitué par les données générées à partir des médias sociaux semblent dictées par la conception humaine des médias sociaux, c'est-à-dire l'inclusion et l'exclusion de certaines populations en fonction de leur utilisation des médias sociaux ainsi que de l'information fournie de la part des usagers des médias sociaux. De plus, la culture épistémique de la pratique de l'épidémiologie numérique serait vue comme très différente par rapport à la culture épistémique de l'épidémiologie clinique parce que la première se concentrerait sur des recherches d'opinions et des tendances sociales alors que la seconde se concentrerait sur des recherches cliniques basées sur des données médicales des patients.

5.2.3 Les données générées par des téléphones cellulaires

Selon la revue de la littérature sur l'épidémiologie numérique, les données générées par les téléphones cellulaires désignés aussi comme technologies mobiles ou téléphones intelligents peuvent être utilisées dans le cadre de recherches en épidémiologie (Mohanty *et al.*, 2016 ; Lazzari *et al.*, 2018 ; Cot *et al.*, 2021 ; Vigfusson *et al.*, 2021). Par exemple, Vigfusson *et al.* (2021) ont mesuré les changements de comportement reflétés dans les enregistrements des appels sur les téléphones portables collectés en temps réel en utilisant un ensemble de données d'utilisateurs anonymes de téléphones portables et leur date de diagnostic de syndrome grippal pendant la pandémie H1N1 de 2009. Cette recherche aurait démontré que l'utilisation du téléphone portable pendant la maladie diffère de manière mesurable du comportement habituel. Selon les auteurs, les résultats suggèrent que les enregistrements des appels sur les téléphones portables et les données

de santé liées anonymement peuvent être suffisamment granulaires pour augmenter les efforts de surveillance épidémique et de modélisation des maladies infectieuses. Par conséquent, les téléphones cellulaires peuvent être utilisés comme dispositif méthodologique pour des recherches épidémiologiques. Il devient alors intéressant de comprendre la confiance envers ce dispositif méthodologique selon les points de vue d'ordre épistémique des chercheurs en épidémiologie.

Nous allons maintenant analyser la confiance envers les données générées par les téléphones cellulaires comme dispositif méthodologique parce que la source de ces données est différente par rapport aux données générées par Google et les médias sociaux. Il est donc possible que les points de vue d'ordre épistémique des chercheurs en épidémiologie envers les données générées par les téléphones cellulaires ne soient pas les mêmes que les dispositifs méthodologiques analysés précédemment.

Nous constatons que la confiance varie aussi vis-à-vis des données générées par les téléphones cellulaires dans le cadre de recherches épidémiologiques. En effet, certains ont une confiance assez élevée envers ce type de données alors que d'autres ont des réserves. De plus, nous retrouvons un niveau de confiance élevé et un niveau de confiance assez bas auprès de chaque catégorie de participants : les pratiquants et les non-pratiquants de l'épidémiologie numérique ainsi que les trois sous-disciplines de l'épidémiologie, à savoir l'épidémiologie environnementale, l'épidémiologie clinique et l'épidémiologie sociale. Nous présenterons donc nos constats en commençant avec les perceptions et nous poursuivrons avec les biais liés au dispositif méthodologique des données produites par les téléphones cellulaires.

5.2.3.1 Les perceptions

Les chercheurs ayant l'expérience de la pratique de l'épidémiologie numérique auraient tendance à avoir confiance envers les données des téléphones cellulaires pour effectuer des recherches épidémiologiques. L'un des chercheurs en épidémiologie sociale affirme que ce type de données inspire confiance étant donné qu'une grande partie des populations canadiennes possède un téléphone intelligent (90 %). Cette caractéristique élargit les échantillons populationnels, réduisant ainsi les biais de sélection lors des études des phénomènes (entretien 8Sxx). Le chercheur en

épidémiologie environnementale estime que les données de géolocalisation générées à partir des téléphones cellulaires sont de bonne qualité (entretien 10Ixxviii). Ce chercheur affirme aussi que le potentiel de validité des données générées par les téléphones cellulaires est le plus élevé comparativement à celui des données générées par les médias sociaux qui inspire le plus bas niveau de confiance et à celui des données générées par les moteurs de recherche Internet qui suscite un niveau de confiance moyen.

« So, the big thing that they are asking is what about selection bias? Well, if you could recruit the whole population, where is the selection bias of that? Because 90% of Canadians have a smartphone. That doesn't mean that everyone is going to appear in the data but theoretically, you need to use these tools to understand what's going on. So why do I want to do a study based on a research question by recruiting just 50 or 60% of a population when I can recruit thousands of people and take data in real time? I see that as a big value for epidemiology and we have to understand the power of digital tools. » (Chercheur - épidémiologie sociale-pratiquant de l'ÉN - entretien 8Sxx)

« That I would have a lot of confidence in. I am not sure how much comfort I have with somebody who would appreciate my privacy, but I do think those in terms of potential data validity, I would say that social media is the lowest, Google somewhere in the middle, mobility data from cell phones is actually pretty high. I know there is geo-location and I think that there are pretty solid. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

Cette vision positive des données générées par les téléphones cellulaires semble présente aussi chez les chercheurs ne pratiquant pas l'épidémiologie numérique. Par exemple, un chercheur en épidémiologie environnementale estime que les données générées par les téléphones cellulaires représentent une des meilleures sources hormis les données biologiques puisque tout le monde est toujours connecté avec son appareil (entretien 5Cxxx). Ce type de données serait utile pour effectuer le traçage des déplacements des gens en contexte de la pandémie de la COVID-19. Cependant, ce chercheur souligne que ce type de données pourrait exclure certaines populations plus âgées. Un autre chercheur en épidémiologie environnementale a exprimé un avis similaire en affirmant avoir une bonne expérience de l'utilisation des données générées par les téléphones cellulaires (entretien 5Cxxvii). Il a dit qu'il connaît bien les limites liées à ce type de données. Selon ce chercheur, les limites résideraient plutôt dans l'interprétation des données et moins du côté des limites technologiques. Ce chercheur soutient aussi que la méthodologie de recherche liée

à ce type de données est rigoureuse et qu'il a confiance envers des recherches utilisant ce genre de données.

Les informations géographiques et les mouvements percevables à travers les données générées par les téléphones cellulaires sont fantastiques et assez précis d'après un chercheur en épidémiologie clinique (entretien 6Fx). Toutefois, ce chercheur souligne qu'il faut bien connaître les limites liées à ce type de données comme les problèmes pour localiser les individus ayant fermé leur téléphone cellulaire puisque ce genre de problème pourrait biaiser les études épidémiologiques. Ce chercheur soutient que, peu importe la nouveauté ou l'ancienneté de la technologie utilisée, l'important est de bien comprendre les limites et les biais potentiels des données générées par ces technologies.

Un chercheur en épidémiologie sociale est assez confiant envers les données générées par les téléphones cellulaires dans le cadre de recherches en épidémiologie (entretien 8Hxvi). D'après l'expérience de ce chercheur avec les applications mobiles, le participant à l'étude reçoit un court rappel s'il ne répond pas à l'heure prévue. Il pense que les problèmes se posent lorsque l'on utilise le système de traçage du téléphone et lorsque les gens oublient leur téléphone, mais les gens apportent fréquemment leur téléphone avec eux. Donc, ce chercheur estime avoir un degré de confiance assez élevé pour ce type de dispositif en supposant que les questions de recherche et l'instrument utilisé soient valides.

« À part de la mesure biologique, probablement ça c'est l'une des meilleures mesures, qu'on peut utiliser maintenant, j'imagine, parce qu'on est comme toujours connectés avec nos appareils, mais pour notre âge, et peut-être pas pour les personnes âgées et ceux de 50 ans et 60 ans, c'est l'une des meilleures mesures pour collecter, mais quand pour avoir le questionnaire rempli, pour vraiment « tracker » les déplacements des gens, je pense que ça c'est l'une des meilleures mesures. Je ne sais pas si tu veux lier avec le traçage des personnes qui ont diagnostiqué avec la Covid-19, donc j'imagine que ça c'est l'une des meilleures mesures pour s'assurer qu'une personne reste dans la résidence. » (Chercheur - épidémiologie environnementale - entretien 5Cxxx)

« Un très bon niveau de confiance par rapport à ça. Encore là ça s'est adonné que je me suis mis le nez là-dedans très dernièrement avec un très haut niveau de détails pour étudier ce qui se passait par rapport à la Covid-19, je comprends bien moi personnellement les limites technologiques de cette approche-là, et les limites technologiques ne sont pas incroyablement importantes dans le fond. Il y en a des limites. Donc encore là tout est dans l'interprétation, mais je viens de faire confiance de l'utilisation qu'on peut en faire et évidemment le design de l'étude est d'autant plus rigoureux en termes de représentativité et

des choses comme ça. Ce que j'ai constaté par rapport aux études que j'ai vues récemment est le fait que c'est une méthodologie très solide et j'ai confiance envers l'utilisation de ces données. » (Chercheur - épidémiologie environnementale - 5Cxxvii)

« Je pense que les choses géographiques et les mouvements sont fantastiques... oui j'ai lu sur ça. Je pense que c'est quand même assez précis... c'est comme un GPS un petit peu le cell phone donc... là encore il faut premièrement analyser comment dire quelles sont les limitations par exemple, on ne peut localiser que des téléphones qui sont j'imagine ouverts, comment ça marche, est-ce que ça marche même si le téléphone est fermé ou... par exemple, quand les gens font comme ça du recueil de données... si ça ne marche pas et que les gens éteignent leur cell phone, ben ça va biaiser l'étude parce que peut-être les gens qui éteignent leur cell phone ils sont plus vieux ou des choses comme ça... Donc je pense que dans toute approche épidémiologique, qu'on utilise des nouvelles ou d'anciennes technologies, il faut toujours essayer de comprendre quelles sont les failles, les limitations de notre design expérimental et je pense qu'il y a des limitations. Le danger en science c'est quand les gens ne prennent pas le temps de réfléchir aux biais potentiels dans leur étude. » (Chercheur - épidémiologie clinique - entretien 6Fx)

« I am relatively confident in it based on the things that you are looking at. I don't think people are going to lie on it too much. From my experience in working on these mobile apps, when the question comes up, if they are not answered at a regular time when they are supposed to, they will get a reminder. And these are really short and sweet, so you are not putting a burden on the person. I think the problems are when you are using the tracking in the phone, the GPS ... the problems come when people forget their phone... but I think that people keep their phone with them pretty frequently. So, I think I have a pretty high degree of confidence from this type of system. Assuming that the questions are valid, that the instrument that they are using ... is also validated. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

Cependant, quelques participants ont affirmé qu'ils ne sont pas en mesure d'expliquer la confiance qu'ils ressentent envers les données générées par les téléphones cellulaires, car ils n'ont jamais utilisé ce type de données pour leurs recherches (entretien 6Biii, entretien 5Cxxvi). L'un de ces chercheurs a souligné qu'une partie de la population est incluse dans les données générées par les téléphones cellulaires comme c'est le cas avec les données générées par Google (entretien 5Cxxvi). Ce chercheur estime que le signal obtenu avec ces données devrait avoir une certaine sensibilité, mais il éprouve des difficultés à exprimer son niveau de confiance envers ces données. Un autre chercheur en épidémiologie environnementale affirme que sa confiance envers les données générées par les téléphones cellulaires dépend du contexte. Ce chercheur aurait une certaine confiance envers les données sur les emplacements et les expositions des individus. Or, il aurait

peu de confiance si les téléphones cellulaires sont utilisés pour des enquêtes générales, car la population est inconnue (entretien 5Gxxix).

« I don't know how confident I am. I don't know whether that data is considered reliable or not. I have no idea. » (Chercheur - épidémiologie clinique - 6Biii)

« Je pense que c'est un petit peu les mêmes genres de questionnement que pour les données Google. Donc un ça représente une partie de la population, je pense que certainement que le signal obtenu de ce genre de données-là doit avoir quand même une certaine sensibilité, mais je n'ai aucune idée de la sensibilité du signal donc j'ai du mal à exprimer mon niveau de confiance dans ce genre d'étude-là. Je ne sais pas. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvi)

« It all depends on the context. We often give to our participants instruments that they can measure where they are. You can use their cell phone as well; you know if they allow to. But I've done many studies to see you know where people are and what they might be exposed to. So that I find having some kind of reliability given the accuracy of the satellites and how accurate it can be with that stuff but if that is just a general survey, again, you don't know who the population is. So, I wouldn't give it much credibility. » (Chercheur - épidémiologie environnementale - 5Gxxix)

5.2.3.2 Les biais liés au dispositif méthodologique

D'après nos analyses, les biais liés au dispositif méthodologique des téléphones cellulaires comprennent l'échantillonnage populationnel et les comportements des gens avec leur téléphone cellulaire. Par exemple, un chercheur en épidémiologie clinique a souligné que les données des téléphones cellulaires peuvent être une bonne chose pour étudier des populations pourvues d'un téléphone cellulaire, mais pas du tout pour d'autres populations comme les enfants n'ayant pas accès à cette technologie (entretien 5Cxxxii).

Dans le même sens, un autre participant en épidémiologie clinique soutient que le chercheur doit s'assurer que les données utilisées pour son étude couvrent entièrement la population ciblée pour obtenir une bonne représentativité populationnelle (entretien 6Axix). Ce chercheur souligne aussi que les données générées par les téléphones cellulaires peuvent être sujettes à interprétation, car il peut y avoir des incertitudes sur ce que signifient réellement les textes écrits par les gens. Cette situation peut créer des défis pour bien interpréter les données et pour les rendre fiables en contexte

de recherche d'après ce chercheur. Un chercheur en épidémiologie sociale a insisté sur la façon dont les données sont utilisées lorsqu'il a expliqué sa confiance envers les données générées par les téléphones cellulaires (entretien 9Rxiv). D'après ce chercheur, ces données présentent un bon potentiel pour des recherches exploratoires comme avec les données générées par les moteurs de recherche Google.

« J'imagine que c'est une bonne façon de faire quand on veut aller essayer d'aller chercher les gens qui ont des cellulaires parce que tu sais encore là, qui c'est qui se promène avec un cellulaire ? Ce n'est pas les enfants généralement, moi je travaille beaucoup en pédiatrie aussi donc ce n'est certainement pas un outil qui est intéressant pour moi, tout est relatif à la question j'imagine là... ça l'est sûrement pour un adolescent qui ne vit pas sans son cellulaire tu sais. » (Chercheur - épidémiologie clinique - entretien 5Cxxxii)

« Depends again on what the question is. If we are trying to determine confirmatory type research unless we capture everybody in a particular (what is our target population and what we are trying to understand), unless we capture all of that (including how many people are responding) and is that what is the patient representative of the target population, then I would be confident. And the quality of data again: our people answering or speaking about the same things. Do they know what they are supposed to. It's subject to interpretation especially on how people can write a message in a text, and we don't even spell words out anymore. So, there would be issues around interpreting the data and making it reliable that way. » (Chercheur - épidémiologie clinique - entretien 6Axix)

« Ça dépend toujours comment les données sont utilisées. Je pense qu'en général il y a quand même une certaine similarité avec ce qu'on vient de discuter plus tôt avec *Google Trends*. Je pense que pour la recherche exploratoire il y a vraiment un potentiel là à saisir qui serait avantageux pour les recherches pour avoir accès à de grands nombres de participants. Je pense que ça serait un avantage potentiel. » (Chercheur - épidémiologie sociale - entretien 9Rxiv)

Quant aux biais liés aux comportements des gens avec leur téléphone cellulaire, l'un des participants étudiant l'activité physique affirme qu'il aurait peu de confiance concernant des recherches fondées sur des données générées par des téléphones cellulaires (entretien 6Bxviii). Ce participant a mentionné que de nouvelles études portant sur des millions de personnes ont observé l'activité physique auprès des populations en se basant sur les données fournies par Apple et des compagnies de cellulaire. Les études de ce genre peuvent être facilement trompeuses selon ce chercheur, car les gens n'ont pas toujours leur téléphone cellulaire sur eux. Ce comportement déforme la réalité sur les activités physiques des populations étudiées. Ce chercheur préfère des

études fondées sur des échantillons populationnels plus petits avec des mesures plus robustes que des études fondées sur des millions de gens avec des mesures non validées. Un autre chercheur en épidémiologie clinique partage le même avis parce qu'il a souligné qu'il est possible de localiser seulement les téléphones ouverts et que si les téléphones sont fermés, les études basées sur les données générées par les téléphones cellulaires seront biaisées.

« J'aurais peu confiance aussi en général. Dépendamment dans quel domaine. Ça dépend toujours de quel genre de données. Je sais que ça ne fait pas partie directement de ce qu'on parle, mais dans mon domaine en activité physique, il y a certaines nouvelles études avec des millions de personnes qui ont regardé l'activité physique chez des gens avec leur téléphone cellulaire et ils se servent de ce que Apple et les compagnies de cellulaire donnent et puis on sait qu'il y a beaucoup de téléphones cellulaires chez les femmes qui les portent dans leur sacoches par exemple, donc elles ne l'ont pas sur elles et ce n'est pas nécessairement un gadget qu'on porte toujours donc je pense que ce n'est pas nécessairement, ça n'a pas été validé dans le sens pour l'activité physique donc même s'il y a un échantillon d'un million de personnes, j'ai plus confiance avec un échantillon beaucoup plus faible avec des mesures qui sont plus robustes qu'un très gros échantillon de données avec des mesures qui ne sont pas validées nécessairement. Donc je pense que ça dépend un peu... il faudrait que je voie quelle recherche et puis je pourrais dire, mais en général de ce que j'ai vu présentement, j'ai peu confiance tant que ce n'est pas démontré qu'il y a eu une validation et que les données peuvent être fiables. Mais si ça a été démontré oui je vais le croire. » (Chercheur - épidémiologie sociale - entretien 6Bxviii)

« Là encore il faut premièrement analyser comment dire quelles sont les limitations par exemple, on ne peut localiser que des téléphones qui sont, j'imagine, ouverts, comment ça marche, est-ce que ça marche même si le téléphone est fermé ou... par exemple, quand les gens font comme ça du recueil de données... si ça ne marche pas et que les gens éteignent leur cell phone, ben ça va biaiser l'étude parce que peut-être les gens qui éteignent leur cell phone ils sont plus vieux ou des choses comme ça... » (Chercheur - épidémiologie clinique - entretien 6Fx)

Notre étude sur la confiance des participants envers les données générées par les téléphones cellulaires pour des recherches épidémiologiques relève les enjeux épistémiques suivants :

- Le niveau de confiance vis-à-vis ce type de données dépend du contexte et du type de recherche;

- Il y aurait un niveau de confiance assez élevé envers les outils de localisation, les données sur les mouvements et les informations sur les expositions disponibles à travers les données générées par les téléphones cellulaires ;
- Certains participants n'ont pas pu exprimer leur niveau de confiance en raison de leur manque d'expérience à utiliser ce type de données ;
- Les réserves seraient importantes sur le plan de la représentativité des populations incluses dans les données générées par les téléphones cellulaires puisque certaines populations comme les enfants n'ont pas de téléphones cellulaires ;
- Il semble que les études utilisant des données générées par des téléphones cellulaires peuvent être biaisées si les gens oublient d'apporter leur téléphone cellulaire avec eux et si le téléphone cellulaire n'est pas utilisé.

Ces analyses sur la confiance à l'égard des données générées par les téléphones cellulaires en tant que dispositif méthodologique amènent à réfléchir sur les signes et les apparences secondaires des instruments de mesure comme proposé par Knorr-Cetina (1999). En effet, les téléphones cellulaires peuvent constituer des instruments de mesure à travers les données qu'ils génèrent en émettant des signes aux chercheurs pour interpréter les phénomènes étudiés. Certains ont confiance envers les signes émis par ce type de données alors que d'autres sont réticents envers ces signes pour des études épidémiologiques.

Comme tout appareil de mesure, les données générées par les téléphones cellulaires peuvent causer des complications lors de recherches. En effet, il est possible que les signes marquants des événements intéressants soient étouffés et brouillés par les signes d'autres événements du dispositif méthodologique. Ces effets liés aux données générées par les téléphones cellulaires constituent des événements sous-jacents qui, dans le cas de ce dispositif méthodologique, seraient la sous-représentativité des populations n'ayant pas de téléphone cellulaire et les biais causés par les gens oubliant d'apporter ou d'allumer leur téléphone cellulaire. Ces phénomènes constituent une menace pour la capacité des chercheurs à reconnaître les événements intéressants et ils peuvent falsifier la signature des événements, déformer leur caractère, ou mettre en péril leur identification (Knorr-Cetina, 1999).

En somme, nos analyses sur la confiance envers les données générées par les moteurs de recherche Internet, les médias sociaux et les téléphones cellulaires en tant que dispositifs méthodologiques révèlent une confiance plus élevée à l'égard des données générées par les téléphones cellulaires suivis par les données générées par les moteurs de recherche Internet. Les chercheurs ayant participé à la recherche de terrain ont exprimé un niveau de confiance généralement plus bas envers les données générées par les médias sociaux.

Pour les moteurs de recherche Internet, nous avons constaté qu'une technologie peut être connue par les chercheurs d'une discipline, mais celle-ci n'est pas nécessairement vue comme un dispositif méthodologique utile pour des recherches par tous les chercheurs de cette discipline. Ces constats révèlent aussi que l'algorithme derrière un moteur de recherche Internet peut représenter un dispositif médiateur, c'est-à-dire un intermédiaire entre les phénomènes étudiés et le chercheur. Par exemple, un dispositif médiateur comme Google enverrait des signaux aux chercheurs lors d'études sur la grippe. Encore faut-il que ces signes soient compris par les chercheurs et que les algorithmes d'intelligences artificielles, c'est-à-dire les dispositifs médiateurs, soient en mesure d'adapter les signaux en leur faisant prendre une forme intéressante selon le regard d'un chercheur en épidémiologie.

Pour les données générées par les médias sociaux, les problèmes de la représentativité des échantillons populationnels et des incertitudes sur la qualité et la validité des informations véhiculées sur les plateformes influenceraient les signes et les apparences perceptibles à travers ce dispositif méthodologique en déformant les perceptions par rapport à la réalité des phénomènes étudiés. Ce dispositif méthodologique peut amener le chercheur à repérer ou non certains phénomènes étant donné la façon dont ce dispositif est conçu (Law et Ruppert, 2013). Les caractéristiques du dispositif méthodologique constitué par les données générées à partir des médias sociaux seraient dictées par la conception humaine des médias sociaux, c'est-à-dire l'inclusion et l'exclusion de certaines populations en fonction de leur utilisation des médias sociaux et de l'information fournie de la part des usagers.

Enfin, pour les données générées par les téléphones cellulaires, nous avons interprété nos analyses sur la confiance envers ce dispositif méthodologique en utilisant la conception des signes et des

apparences secondaires des instruments de mesure proposée par Knorr-Cetina (1999). Il est possible que les données générées par les téléphones cellulaires émettent des signes influençant l'interprétation des phénomènes étudiés. Certains ont confiance envers les signes émis par ce dispositif méthodologique alors que d'autres sont réticents envers ces signes dans le cadre d'études épidémiologiques. Comme tout appareil de mesure, ce dispositif méthodologique peut causer des complications lors des recherches parce que les signes marquant des événements intéressants peuvent être étouffés et brouillés par les signes d'autres événements du dispositif. Ces effets liés aux données générées par les téléphones cellulaires constituent des événements sous-jacents qui, dans le cas de ce dispositif méthodologique, seraient la sous-représentativité des populations n'ayant pas de téléphone cellulaire et les biais causés par les gens oubliant d'apporter ou d'allumer leur téléphone cellulaire. Ces phénomènes constituent une menace pour la capacité des chercheurs à reconnaître les événements intéressants et ils peuvent falsifier la signature des événements, déformer leur caractère, ou mettre en péril leur identification (Knorr-Cetina, 1999).

5.3 La conclusion sur les enjeux épistémiques

Dans ce chapitre, l'objectif est de saisir les enjeux épistémiques qui viennent avec l'émergence de la pratique de l'épidémiologie numérique. Les enjeux épistémiques observés lors de nos analyses sont la pertinence des données envers les questions de recherche épidémiologique, la signification des données et la confiance envers des dispositifs méthodologiques utilisés dans la pratique de l'épidémiologie numérique. Nous avons analysé comment les chercheurs en épidémiologie voient la signification des données dans le cadre de leurs pratiques courantes de recherche et dans le contexte de la pratique de l'épidémiologie numérique. Nous avons aussi analysé la confiance des chercheurs en épidémiologie envers des dispositifs méthodologiques utilisés lors de la pratique de l'épidémiologie numérique.

Il existerait des différences significatives dans les données utilisées selon la sous-discipline de l'épidémiologie d'après les constats de notre recherche. La différence entre l'épidémiologie clinique et l'épidémiologie sociale par rapport au type de données utilisées réside dans les diverses cultures épistémiques au sein de la discipline de l'épidémiologie. Il semble donc que la culture épistémique de l'épidémiologie clinique ne soit pas la même que celle de l'épidémiologie sociale

et de l'épidémiologie environnementale en raison de la nature des données utilisées. Le type de données utilisé amènerait des manières différentes de connaître les phénomènes étudiés selon la sous-discipline, créant ainsi des univers de recherche distincts au sein de la discipline de l'épidémiologie. Nos analyses révèlent aussi une grande diversité dans les données utilisées lors de la pratique de l'épidémiologie numérique ainsi qu'une culture épistémique différente par rapport à celle de l'épidémiologie traditionnelle en raison de l'ajout de données de sources non traditionnelles susceptible d'apporter de nouvelles façons de connaître les phénomènes étudiés en épidémiologie.

Sur la base des résultats de notre recherche, une part assez importante de chercheurs en épidémiologie ne pratiquant pas l'épidémiologie numérique peut éprouver une plus grande confiance envers les données de sources traditionnelles comme les systèmes de santé et d'institutions officielles de la statistique qu'envers les données de sources non traditionnelles dans le cadre de recherches épidémiologiques. Les réserves des chercheurs en épidémiologie envers les données de sources non traditionnelles seraient liées entre autres au manque de connaissances sur les biais liés à ces données et à une moins bonne compréhension des apports, des limites, de la validité et de la rigueur des sources de données non traditionnelles comparativement aux sources de données traditionnelles.

À la lumière de ces constats, les données de sources non traditionnelles constitueraient des dispositifs médiateurs envoyant des signaux dont les chercheurs ne comprennent pas. Ces signaux pourraient aussi ne pas représenter adéquatement la réalité du phénomène étudié, diminuant ainsi la confiance des chercheurs en épidémiologie envers les données de sources non traditionnelles. Ainsi, il serait plus difficile de bien saisir les signes, les reconstructions et les événements lors des analyses de données de sources non traditionnelles que lors des analyses de données de sources traditionnelles.

De plus, il existerait deux visions de l'échantillonnage populationnel en lien avec la pratique de l'épidémiologie numérique selon les constats établis lors de notre recherche. La première vision considère l'échantillonnage statistique comme problématique avec la pratique de l'épidémiologie numérique étant donné les incertitudes sur la représentativité des échantillons extraits à partir des

données numériques. Ce phénomène signifie que des chercheurs craignent que les signes et les apparences perceptibles à travers les analyses des données générées par les données de sources non traditionnelles comme celles des médias sociaux conduisent à des résultats erronés et, ainsi, vers des perceptions déformées par rapport à la réalité des phénomènes étudiés.

La deuxième vision envisage du potentiel dans la pratique de l'épidémiologie numérique pour élargir les échantillons populationnels lors des études épidémiologiques parce que les usagers d'Internet sont nombreux. Cette caractéristique de l'Internet peut rendre accessibles des échantillons plus grands grâce à l'utilisation des données numériques comparativement à l'utilisation de données traditionnelles. Certains chercheurs envisagent donc que les données de sources non traditionnelles soient en mesure d'apporter une meilleure compréhension des phénomènes étudiés en raison d'une plus grande représentativité des échantillons populationnels devenue possible avec l'utilisation massive d'Internet.

Nous avons aussi analysé la confiance de chercheurs en épidémiologie envers des dispositifs méthodologiques utilisés dans la pratique de l'épidémiologie numérique. Les données générées par les moteurs de recherche Internet, les médias sociaux et les téléphones cellulaires ont été choisies comme dispositifs méthodologiques puisque ces technologies sont parmi les plus connues en ce qui concerne la pratique de l'épidémiologie numérique.

Notre recherche montre que la confiance serait relativement bonne envers les données générées par les moteurs de recherche Internet si les recherches étudient des opinions, des tendances, des attitudes ou des enjeux sociaux. Toutefois, il semble y avoir peu de confiance envers des recherches épidémiologiques utilisant des données générées par Internet portant sur les soins de santé, les diagnostics et la prévalence des maladies. Toujours selon les constats établis lors de notre recherche, les réticences à l'égard des données générées par les moteurs de recherche Internet semblent liées aux risques d'exclure ou de sous-représenter certaines populations, aux difficultés à suivre le nombre de visiteurs uniques et aux incertitudes quant à la validité des informations publiées sur Internet. La facilité d'accès à ce type de données ainsi que leur potentiel pour la surveillance syndromique (à la condition que les signaux en provenance d'Internet soient comparables à ceux de sources de données traditionnelles) pour recruter des participants aux recherches et pour générer

des hypothèses constituent les éléments positifs identifiés lors de nos analyses au sujet de la confiance envers ce dispositif méthodologique.

Sur la base des résultats obtenus lors de notre recherche, le niveau de confiance serait généralement plus bas envers les données générées par les médias sociaux. Les signes et les apparences influencés par les problèmes de la représentativité des échantillons populationnels et les incertitudes sur la qualité et la validité des informations véhiculées sur les plateformes des médias sociaux entraîneraient des résultats erronés et des perceptions déformées de la réalité des phénomènes étudiés. Ce dispositif méthodologique amènerait le chercheur à voir ou à ne pas voir certains phénomènes étant donné la façon dont ce dispositif est conçu. Les caractéristiques du dispositif méthodologique constitué par les données générées à partir des médias sociaux semblent dictées par la conception humaine des médias sociaux, c'est-à-dire l'inclusion et l'exclusion de certaines populations en fonction de leur utilisation des médias sociaux ainsi que l'information fournie de la part des usagers des médias sociaux.

Pour les données générées par les téléphones cellulaires, le niveau de confiance serait assez élevé envers les outils de localisation, les données sur les mouvements et les informations sur les expositions disponibles avec les téléphones cellulaires selon les discours de certains participants à notre recherche. Plusieurs participants ont exprimé des réserves par rapport à ce type de données en raison de l'exclusion de certaines populations ne possédant pas de téléphone cellulaire et du fait que parfois les gens se séparent de leur téléphone cellulaire, ce qui engendre ainsi des biais dans les études épidémiologiques.

Le fait que les chercheurs expriment des réserves montre qu'ils sont conscients des limites de ces dispositifs méthodologiques n'étant pas à la fine pointe pour être largement déployés lors des recherches. Ceci constitue une barrière à la diffusion de la pratique de l'épidémiologie numérique. À ce stade d'émergence de cette nouvelle pratique, il serait intéressant d'observer comment ces dispositifs méthodologiques s'améliorent et comment ils évoluent d'une expérience à l'autre. En effet, le cas de GFT montre que ces dispositifs méthodologiques peuvent évoluer à la suite d'améliorations (Pollett *et al.*, 2017b ; Salathé, 2018 ; Duclos, 2019 ; Clemente *et al.*, 2019). Les processus d'amélioration de ces dispositifs méthodologiques, les porteurs et les mécanismes

mettant en place les améliorations ainsi que la manière dont évoluent les perceptions des chercheurs envers ces dispositifs restent une question ouverte à ce stade de l'émergence de la pratique de l'épidémiologie numérique.

Selon les constats établis lors de notre recherche, il semble que les données générées par les moteurs de recherche Internet soulèvent des questions en tant que dispositif méthodologique au niveau technologique alors que les données générées par les médias sociaux soulèveraient des questions au niveau humain, conditionnées par les signes et les apparences secondaires de ce dispositif méthodologique. Il est donc possible que les complications lors des recherches diffèrent selon le dispositif méthodologique utilisé.

Notre recherche montre aussi que les données générées par les moteurs de recherche Internet semblent susciter des questions plutôt d'ordre technique auprès des chercheurs en épidémiologie et plus particulièrement auprès de ceux n'ayant pas pratiqué l'épidémiologie numérique. Ces questions renvoient à la façon dont est conçu le moteur de recherche Internet. La conception technique des moteurs de recherche Internet apparaît comme obscure auprès de chercheurs en épidémiologie, amenant ainsi une attitude de réserve ou prudente vis-à-vis des données produites par ces technologies dans le cadre de recherches épidémiologiques. Certains chercheurs en épidémiologie estiment que la manière dont est construite une technologie numérique comme celle du moteur de recherche Google peut influencer les analyses du chercheur en fonction de la conception de l'algorithme derrière ce moteur de recherche. Il semble que l'algorithme puisse être vu comme un dispositif médiateur, c'est-à-dire un intermédiaire entre le chercheur et les phénomènes étudiés capable d'influencer les décisions du chercheur lors des recherches.

Pour les données générées par les médias sociaux, les questionnements semblent se situer sur le plan des signes et des apparences secondaires du dispositif selon les discours de plusieurs participants à notre recherche. En utilisant des données produites par des médias sociaux pour des recherches épidémiologiques, le chercheur serait plutôt confronté à des biais, à des bruits et à des distorsions causés par des comportements humains susceptibles d'émettre des signes inappropriés et d'amener des apparences trompeuses. Certains chercheurs ont souligné que les signes et les apparences visibles à travers les contenus des médias sociaux sont incertains parce que des

populations peuvent être surreprésentées sur les plateformes, les contenus véhiculés ne sont pas nécessairement vrais, les utilisateurs des plateformes ne sont pas toujours connus et des populations comme les enfants et les personnes âgées peuvent être exclues si elles n'utilisent pas les médias sociaux. Nous avons aussi entendu ces enjeux lors de quelques discussions sur les données générées par des moteurs de recherche Internet, mais ils n'ont pas été mentionnés aussi souvent que lors des discussions sur les médias sociaux. Il est donc possible que les questionnements des chercheurs à l'égard des données générées par les médias sociaux soient liés à la conception humaine de ce dispositif méthodologique alors que les questionnements à propos des données produites par les moteurs de recherche Internet seraient davantage liés à la conception technologique de ce dispositif méthodologique.

Ce chapitre a fait ressortir les enjeux épistémiques de la pratique de l'épidémiologie numérique tels que vus par des chercheurs canadiens en épidémiologie afin d'examiner cette pratique émergente de manière critique. Comme toute discipline, l'épidémiologie est aussi influencée par des enjeux sociopolitiques liés à la production de ses connaissances et non pas seulement par les enjeux épistémiques. Le prochain chapitre sera alors consacré aux enjeux sociopolitiques de la pratique de l'épidémiologie numérique.

CHAPITRE VI

LES ENJEUX SOCIOPOLITQUES DE LA PRATIQUE DE L'ÉPIDÉMIOLOGIE NUMÉRIQUE

Dans ce chapitre, l'objectif est de connaître les enjeux sociopolitiques du contexte d'émergence de la pratique de l'épidémiologie numérique. Nous voulons plus précisément faire ressortir les dimensions sociales et politiques de la pratique de l'épidémiologie numérique à travers les points de vue de chercheurs en épidémiologie afin de compléter de manière critique les observations du chapitre précédent sur les enjeux épistémiques de cette nouvelle pratique. Pour ce faire, nous situerons la pratique de l'épidémiologie numérique par rapport à des questions plus larges, dont les enjeux éthiques et l'interdisciplinarité.

Nos analyses sont inspirées par la coproduction des connaissances (Jasanoff, 2004), les pratiques organisationnelles dans les sciences (Knorr-Cetina, 1999), les travaux de recherche sur la vie sociale des méthodes (Law et Ruppert, 2013) et les enjeux politiques des données (Ruppert *et al.*, 2017) expliqués précédemment dans le cadre théorique de la thèse. Notre recherche révèle notamment que les enjeux sociopolitiques de la pratique de l'épidémiologie numérique sont particulièrement marqués par l'expertise et la légitimité.

6.1 Les enjeux d'expertise liés à la pratique de l'épidémiologie numérique

Pour analyser les enjeux sociopolitiques de la pratique de l'épidémiologie numérique, nous commencerons avec les questions relatives à l'expertise. Nos analyses montrent que les préoccupations vis-à-vis de l'expertise en lien avec la pratique de l'épidémiologie numérique concernent les nouvelles compétences et les spécialités envisagées, la formation des chercheurs en épidémiologie et l'interdisciplinarité.

6.1.1 Les nouvelles compétences et les spécialités envisagées

Une bonne partie des participants à la recherche de terrain estiment que la pratique de l'épidémiologie numérique nécessite de nouvelles compétences et l'expertise de diverses spécialités. Nos analyses font effectivement ressortir la nécessité de mise à jour en épidémiologie, pour les compétences en informatique et en science des données ainsi que pour les compétences non techniques. Les résultats seront présentés selon les points de vue des participants à la recherche de terrain parce que nous n'avons pas constaté de tendance particulière en fonction de la sous-discipline du participant ni selon l'expérience de la pratique de l'épidémiologie numérique.

6.1.1.1 Mise à jour de la discipline de l'épidémiologie

Plusieurs participants ont fait allusion à la nécessité de se mettre à jour afin de mieux profiter des nouvelles données disponibles pour des recherches en épidémiologie. Par exemple, l'un des participants en épidémiologie sociale a affirmé que la discipline devra s'actualiser en acquérant les nouvelles méthodes requises pour nettoyer les nouvelles bases de données (entretien 5Kxv). Ce chercheur estime que ces nouvelles méthodes ne changeront pas énormément les approches d'analyse en épidémiologie, mais qu'il faudra disposer des moyens pour gérer toutes ces données. Ce chercheur envisage aussi que ces méthodes puissent s'avérer pertinentes pour les futures générations d'épidémiologistes afin qu'elles soient formées sur les méthodes spécifiques à la pratique de l'épidémiologie numérique.

« Je ne pense pas que ce soit des nouvelles tâches, mais des nouvelles compétences. Par exemple, moi, lorsque j'étais aux études en épidémiologie, ce ne sont pas des choses qu'on a apprises, mais je pense qu'aujourd'hui, c'est essentiel de se mettre à jour sur ces méthodes-là pour pouvoir profiter des données qui sont disponibles. La discipline devra se mettre à jour avec les nouvelles méthodes requises pour nettoyer ces grosses bases de données-là. Pour les analyser, ça ne change pas énormément nos approches, mais il faudra avoir des moyens pour pouvoir gérer toutes ces données. Je pense que ça serait pertinent que les futures générations d'épidémiologiste soient formées sur les méthodes spécifiques à l'épidémiologie numérique. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

6.1.1.2 Les compétences en informatique et en science des données

Tous les participants à la recherche de terrain ont fait allusion au développement de compétences en informatique lorsqu'ils ont discuté des nouvelles compétences nécessaires à la pratique de l'épidémiologie numérique. Cet enjeu d'expertise a été mentionné autant chez les chercheurs en épidémiologie sociale, environnementale et clinique que chez les chercheurs ayant pratiqué l'épidémiologie numérique et ceux ne l'ayant pas pratiquée.

Des chercheurs ont souligné qu'il importera de mieux comprendre la validité et la fiabilité des données numériques ainsi que l'apprentissage automatique (*machine learning*) et l'intelligence artificielle dans le cadre de la pratique de l'épidémiologie numérique. Par exemple, l'un des chercheurs en épidémiologie sociale souligne que l'épidémiologie traditionnelle devra souvent déterminer les informations valables et fiables lors des enquêtes ou des collectes de données (entretien 8Hxvi). D'après ce chercheur, un tout nouveau monde est en train de s'ouvrir dans lequel il faudra peser le pour et le contre en matière de validité et de fiabilité des données numériques parce que si ces données ne sont pas valides et ne constituent pas une bonne source d'information, les recherches ne feront qu'embrouiller le problème. Ce chercheur affirme aussi que c'est la même chose que la médecine fondée sur les preuves nécessitant de disposer de bonnes sources de données et de technologues formés pour bien analyser ces données. Ces technologues devront donc apprendre comment accéder aux données, comment les évaluer et comment les utiliser dans leurs études selon les propos de ce chercheur.

Quant aux analyses, ce chercheur en épidémiologie sociale souligne qu'il faudra aller au-delà du simple calcul des incidents et des risques relatifs. En effet, le processus d'analyse va au-delà des statistiques du fichier de base selon ce chercheur, car ce nouveau monde nécessite l'apprentissage automatique et l'intelligence artificielle. Ce même participant affirme que le chercheur en épidémiologie devra développer une compréhension de base sur ces choses et sur la manière dont celles-ci peuvent être utilisées pour des recherches. Il estime que cette situation s'appliquera à beaucoup de gens et particulièrement aux épidémiologistes quantitatifs. De plus, ce chercheur considère que les besoins de faire des analyses, de l'apprentissage automatique, de l'intelligence artificielle et d'autres choses de ce genre vont se généraliser, peu importe si cela plaît aux gens ou

non. Il est fort probable que les grandes quantités de données disponibles constituent un moteur important vers la numérisation en épidémiologie d'après les propos de ce chercheur.

Dans le même sens, un chercheur en épidémiologie environnementale a affirmé que des informaticiens sont désormais engagés au département d'épidémiologie de son université (entretien 5Gxxv). Les propos de ce chercheur indiquent qu'il y a maintenant des besoins de compétences en programmation dans le cadre de certaines recherches en épidémiologie. Le discours d'un chercheur en épidémiologie clinique exprime un avis similaire. En effet, il affirme que la pratique de l'épidémiologie numérique amène obligatoirement le développement de compétences en informatique et en gestion de bases de données (entretien 5Cxxxii). Un autre participant en épidémiologie clinique complète ce portrait en soulignant que les chercheurs auront besoin de plus de compétences en informatique, dans le partage des données et sur les technologies liées à cette nouvelle pratique (entretien 6Biii). De plus, un autre chercheur en épidémiologie clinique a affirmé qu'il faudra développer des capacités à analyser des bases de données de grande échelle (entretien 6Bi).

Nos analyses révèlent aussi qu'un chercheur en épidémiologie sociale suppose entre autres que des compétences en exploration de données (*data mining*) peuvent être requises pour pratiquer l'épidémiologie numérique (entretien 5Uxxxvi). Dans le même ordre d'idées, un chercheur en épidémiologie environnementale souligne que cette nouvelle pratique peut nécessiter des compétences en analyse de texte, notamment avec les données des médias sociaux nécessitant d'analyser tous les textes publiés et de comprendre ce qu'ils disent (1Oxxxv).

« First, it takes a better understanding of the validity and reliability of that data. Traditional epidemiology will take a lot of time to figure out what is valid and reliable information from surveys and data collection. This is going to be a whole new world and figure out the pros and cons in terms of validity and reliability and the old term of garbage in garbage out ... you can cut a lot of information but if it's not valid and it is not a good source of information, then you are just going to confuse the problem. It is the same thing as evidence-based medicine ... you need to have good data sources and technologists trained in that ... they will have to learn how to access the data, how to evaluate it, and how to use it in their studies. That's one part of it. The second part is the actual analysis. As I just mentioned, moving beyond just calculating incidents and relative risks. Now the analysis process goes beyond the basis file stat. In terms of machine learning and artificial intelligence, and I think that these are going to be required ... at least having a basic understanding of what these

things are and how they can be used and for many people and especially quantitative epidemiologists will have tasks and training on this. I think in general more and more people are going to require doing analysis, machine learning and artificial intelligence and things like that whatever if people like it or not these things are there, and they are going to be required. It is a necessary thing that is going to happen with all the data that is available. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« It changed a bit whom we hire. So, I think we started to see... you know... in epidemiology departments there are computer scientists that I think we're not there before, because we need people that can do some of the programming that we may or may not have expertise to be able to do. So, I think that the hiring has changed a little bit and just thinking about the set-up like you often need to have a much more powerful server than you did previously because you have a massive amount of data storage that's required. » (Chercheur - épidémiologie environnementale - entretien 5Gxxv)

« C'est sûr que ça amène un développement de compétences aussi comme on en parlait un peu tantôt, tu sais en informatique, en gestion de bases de données parce que là ça va être assez gigantesque tu sais, de l'utilisation non seulement de logiciels d'analyse de données, mais aussi de gestion de données. Je pense que ça amène un développement obligatoire de compétences par rapport à ça c'est sûr. » (Chercheur - épidémiologie clinique - entretien 5Cxxxii)

« I think that they would probably need more of the grounding in computer science, in data sharing and technologies that are related to that. » (Chercheur - épidémiologie clinique - entretien 6Biii)

« I mean large-scale database analysis is clearly an important set of skills. » (Chercheur - épidémiologie clinique - entretien 6Bi)

« Écoutez j'imagine que ça prend des compétences en programmation ne serait-ce que pour savoir comment utiliser par exemple des API, faire du *data mining* j'imagine. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

« I would say that we need some good computing skills to extract that data out because it is quite challenging. Other skills of text-based analysis are also needed for that. Usually, our regular data went through a data cleansing process. With social media, you need to parse all that text that comes in and try to learn what it says. » (Chercheur - épidémiologie environnementale - entretien 1Oxxxv)

De plus, l'un des chercheurs en épidémiologie pratiquant l'épidémiologie numérique a souligné la croissance du domaine de la science des données (entretien 10Ixxviii).

« Next, assuming that you have the data, you need people who are competent in dealing with the data and those people are often not epidemiologists or people with an MPH degree or

people who come from our field. They are often computer scientists, sometimes statisticians and more and more as we see data science grow as a field. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

Selon l'article de Subrahmanya *et al.* (2021), la science des données peut se définir comme un domaine interdisciplinaire visant à extraire des connaissances et des idées à partir de nombreuses données structurées et non structurées, en utilisant des méthodes scientifiques, des techniques d'exploration de données, des algorithmes d'apprentissage automatique et le big data. La définition de la science des données proposée par ces auteurs se lit comme ceci :

Data science is an interdisciplinary field that extracts knowledge and insights from many structural and unstructured data, using scientific methods, data mining techniques, machine-learning algorithms, and big data. (Subrahmanya *et al.*, 2021, p. 1473)

Avec tout l'accent mis sur la valeur des données récemment, l'expertise de ce domaine serait en grande demande. De plus, certains travaux en STS soulignent la présence d'une guerre de territoire disciplinaire impliquant l'informatique, les statistiques, les mathématiques, les sciences de l'information ainsi que diverses collaborations avec les secteurs scientifiques, étatiques et industriels, chacun revendiquant l'expertise, le financement et la sphère réglementaire de la science des données (Ribes, 2019). Il y a donc de fortes chances que ce chercheur fasse allusion à la science des données étant donné la place grandissante qu'elle occupe dans différents secteurs d'activités comme la recherche. D'ailleurs, certains chercheurs en STS comme Kitchin (2014) soutiennent que la science des données (désignée cette fois comme le big data et les nouvelles analyses de données) mène à des innovations perturbatrices capables de reconfigurer dans de nombreux cas les façons de mener les recherches :

Big Data and new data analytics are disruptive innovations which are reconfiguring in many instances how research is conducted. (Kitchin, 2014, p. 1)

6.1.1.3 Les compétences non techniques

Concernant les compétences non techniques, l'un des chercheurs en épidémiologie sociale pratiquant l'épidémiologie numérique (entretien 8Sxx) et un chercheur en épidémiologie sociale n'ayant pas expérimenté cette nouvelle pratique (entretien 5Cxxxii) pensent que celle-ci nécessite de comprendre la gestion des données, c'est-à-dire de saisir d'où viennent les données et comment on doit les acquérir (entretien 8Sxx). Le chercheur en épidémiologie sociale ayant l'expérience de cette nouvelle pratique précise aussi que les capacités d'établir des collaborations interdisciplinaires avec l'informatique notamment représentent des compétences essentielles pour pratiquer l'épidémiologie numérique.

Le chercheur en épidémiologie environnementale pratiquant l'épidémiologie numérique apporte quant à lui d'autres précisions à propos des compétences essentielles en lien avec cette nouvelle pratique. D'après l'expérience de ce chercheur, cette pratique requiert que des responsables de la protection de la vie privée, soit des avocats et des personnes compétentes en négociation d'accords, partagent de l'information lorsque cela s'avère nécessaire (entretien 10Ixxviii). Ce chercheur donne l'exemple des données des téléphones cellulaires parce que celles-ci peuvent provenir d'un fournisseur particulier, nécessitant ainsi de savoir prendre les dispositions requises pour accéder à ces données.

« I am not a coder, I am not a programmer, but I understand coding, I understand programming. And I came from a highly bio-statistical background, as we do in epidemiology. So once having that background we have the understanding what needs to be done. So that was step one. Step 2 is understanding data management. What I mean by data management is not about database management is understanding where the data is coming from, where is going, how do we need to acquire the data. So that's the second step. The third step is interdisciplinary collaboration. So, computer scientists know to code, how to program but they don't know what we want. When I say 'we' I say epidemiologists or health professionals. So, without interdisciplinary collaboration, we will be working in silos instead of working outside of silos. So, I established collaboration with computer scientists and at the moment in our university, we have a collaboration between my lab and another computer science lab. And we actually collaborate, and we work together. So, for example, we share students together. So, we have students in computer science who are working on my projects. So, this is how our research is conducted. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

« I think honestly one important thing is data management, being able to manage large quantities of data which is often not thought adequately like there could be like a small course on data management but honestly if you talk to anyone that stands as a computational epidemiologist will find that digital epidemiology has 80% of the time to do with data management like setting up the data to make it workable. So, I think that's a huge thing. » (Chercheur - épidémiologie sociale - entretien 5Cxxxii)

« It is not the easiest. We certainly come across. So, the first thing is getting access to the data. So, in some cases the access is relatively easy to get. In some cases, it is more difficult to get such as the cell phone data, if you are looking at data from a certain provider. My experience of that has been you need your privacy officers, and you lawyers, and all of the people with those skills sets to negotiate those information-sharing agreements if they are necessary. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

D'autres compétences non techniques sont ressorties lors des analyses des propos des participants n'ayant pas pratiqué l'épidémiologie numérique. Un participant en épidémiologie clinique a mentionné l'importance de posséder des compétences en consultation auprès de compagnies afin d'accéder aux données numériques, alors qu'un autre participant dans la même sous-discipline a mentionné l'importance des compétences en analyse de bases de données à grande échelle (entretien 6Fviii). Lors d'une discussion avec un autre chercheur en épidémiologie sociale, celui-ci a déclaré que la pratique de l'épidémiologie numérique peut constituer un moteur considérable pour aller chercher de la nouvelle expertise spécialisée afin de bien manipuler et de bien comprendre les données numériques qui sont plus difficiles à utiliser comme celles des téléphones cellulaires et les données de satellites (entretien 5Cxxvii).

« Consulting skills as a soft skill because you would have to work with companies I imagine. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

« La plupart de ces données-là sont ancrées dans des domaines d'expertise très spécialisés. Les données cellulaires ne sont pas des données faciles à analyser parce que ça prend des gens qui sont avec une spécialité dans l'utilisation d'analyse des ondes et puis ça ce n'est pas donné à tout le monde. Les données satellitaires sont des données qui sont aussi... et même si elles sont acceptées en termes de "voici la donnée", ce ne sont pas des données faciles à manipuler puis à comprendre. Donc je pense que c'est un moteur très fort d'aller chercher de la nouvelle expertise spécialisée, c'est une demande très forte d'aller chercher de l'expertise spécialisée qu'on doit avoir par rapport à l'utilisation de ces données-là. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvii)

En somme, rappelons que nos analyses sur les nouvelles compétences et les spécialités envisagées avec la pratique de l'épidémiologie numérique mettent en évidence la nécessité de mise à jour en épidémiologie, pour les compétences en informatique et en science des données ainsi que pour les compétences non techniques. Deux chercheurs parmi les trois pratiquants de l'épidémiologie numérique ont été clairs sur l'intersection de plusieurs types de compétences dans leurs discours. Selon ces chercheurs, la pratique de l'épidémiologie numérique implique à la fois des compétences en informatique et en gestion de données de même que des compétences ne relevant pas nécessairement de la technique comme la collaboration interdisciplinaire, la protection de la vie privée et la négociation des accords de partage de l'information. Nous avons constaté que quelques participants envisagent des besoins de compétences pour traiter les données des médias sociaux, les données des téléphones cellulaires, les données des satellites et pour l'exploration de données (*data mining*).

Cette analyse montre aussi que la pratique de l'épidémiologie numérique implique une vie sociale au même titre que les méthodes de recherches. Les méthodes de recherches liées à la pratique de l'épidémiologie numérique ne sont pas des boîtes à outils neutres parce que celles-ci peuvent influencer des pratiques et susciter de nouveaux besoins en matière de compétences en épidémiologie. Certains chercheurs en épidémiologie intègrent déjà cette nouvelle pratique dans le cadre de leurs recherches alors que d'autres comprennent le besoin de développer des compétences, notamment en informatique pour être en mesure de pratiquer l'épidémiologie numérique. Il est aussi possible d'entrevoir la symétrie de la coproduction des connaissances, c'est-à-dire que les connaissances produites lors des recherches impliquent des enjeux épistémiques et technologiques, mais aussi des enjeux sociaux et politiques (Jasanoff, 2004). En effet, la pratique de l'épidémiologie numérique implique des enjeux techniques nécessitant des compétences en informatique et en science des données et des enjeux sociopolitiques amenant des besoins de compétences non techniques comme la collaboration interdisciplinaire, la protection de la vie privée et la négociation des accords de partage de l'information.

6.1.2 La formation des chercheurs en épidémiologie

La formation des chercheurs en épidémiologie constitue un autre enjeu d'expertise assez souvent mentionné par des participants à la recherche de terrain. Pour cet enjeu, les résultats seront présentés en fonction de l'expérience de la pratique de l'épidémiologie numérique du participant puisque des tendances sont différentes en fonction de cette caractéristique. Nous commencerons par analyser les points de vue des chercheurs en épidémiologie n'ayant pas l'expérience de cette pratique et ensuite, nous enchaînerons avec les points de vue des chercheurs en épidémiologie pratiquant l'épidémiologie numérique.

6.1.2.1 Les points de vue des chercheurs ne pratiquant pas l'épidémiologie numérique

La plus grande partie des participants soutient que les chercheurs en épidémiologie doivent suivre une formation pour pratiquer l'épidémiologie numérique parce qu'ils ne l'ont jamais expérimentée lors de leurs travaux et que leur programme d'études n'avait pas intégré cette pratique au cours de leur cheminement d'apprentissage de la discipline. Un chercheur en épidémiologie sociale pratiquant l'épidémiologie depuis longtemps a souligné qu'il n'est pas vraiment familier avec la pratique de l'épidémiologie numérique (entretien 8Hxvi). Ce chercheur estime qu'il aurait besoin d'une formation sur les statistiques, sur les hypothèses ainsi que sur les avantages et inconvénients de cette nouvelle pratique.

Dans le même sens, un chercheur en épidémiologie clinique estime que les chercheurs en épidémiologie ont besoin de guides sur la manière de pratiquer l'épidémiologie numérique afin d'assurer de la cohérence lors de cette nouvelle pratique (entretien 6Axix). D'après ce chercheur, la moindre des choses serait que les chercheurs sachent entre autres comment obtenir des indicateurs à partir des différents sites Web. Ce chercheur estime aussi qu'il faut comprendre ce qu'est un visiteur sur le Web, ce qu'est un utilisateur unique, d'où viennent ces personnes, leur situation géographique et ce que l'on peut déduire de toutes ces données lors de la pratique de l'épidémiologie numérique.

Un autre des participants en épidémiologie sociale ajoute à ce portrait que la plupart des chercheurs ne possèdent pas l'expertise et qu'il faut plusieurs experts pour pratiquer l'épidémiologie numérique (entretien 6Bxviii). Il affirme aussi que les chercheurs doivent recevoir de l'entraînement pour apprendre cette nouvelle pratique et qu'il faut faire en sorte qu'une telle pratique devienne mieux connue en épidémiologie. De plus, ce chercheur souligne qu'il faudra démontrer comment cette pratique peut bien fonctionner et la manière dont celle-ci peut être intégrée aux pratiques courantes en épidémiologie de manière concrète. Ce participant est ouvert à la pratique de l'épidémiologie numérique, mais il faut du soutien pour les chercheurs n'étant pas familiers avec cette nouvelle pratique.

« I've been in epidemiology in quite some time, and I am not super familiar with digital epidemiology, so I would need to take a crash course on the statistics about it, hypothesis, pros, cons and things like that. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« We have a lot of reporting guidelines and have guidelines for doing different kinds of research and I think that something like this would need to be put in place (need guidelines to be put in place or could be useful to help us to be consistent on how we use digital epidemiology). That way at least we could compare the results. They would be comparable for at least doing things consistently. Some sort of guideline on how to do this kind of work. The very least, researchers need to know how to get metrics from the various websites or those kinds of things, understanding what a visitor is, what is a unique user and what they can infer from this pile of data, where these people are coming from, what geographical location. » (Chercheur - épidémiologie clinique - entretien 6Axix)

« Ça prend plusieurs experts dans ce domaine-là. Je pense que la plupart des chercheurs n'ont pas cette expertise-là. Il faut qu'on soit entraîné. Je pense que c'est sûr que quelqu'un comme moi qui n'a jamais utilisé ça même que le terme, je ne l'avais jamais entendu avant, donc je n'étais pas sûr de bien le comprendre, il faut que ça soit plus véhiculé et démontré que ça peut bien fonctionner et comment on peut l'intégrer. Donc être plus concret aussi. Je suis ouvert à ça. C'est juste que je ne sais pas comment l'utiliser. Je n'ai pas d'expertise pour ça. Peut-être que si j'avais des assistants de recherche qui connaissent ça et qui me montreraient comment je pourrais utiliser ça et intégrer à ma recherche je serais vraiment ouvert à ça. C'est juste que je n'ai pas l'expertise moi-même et je ne l'ai jamais utilisée. » (Chercheur - épidémiologie sociale - entretien 6Bxviii)

De plus, quelques participants à la recherche de terrain ont exprimé des préoccupations relatives à la formation des futurs chercheurs en épidémiologie avec l'émergence de la pratique de l'épidémiologie numérique. Par exemple, un participant en épidémiologie sociale a déclaré que ce

serait une bonne chose de commencer tôt à enseigner cette nouvelle pratique lors de l'éducation et des formations des futurs chercheurs (entretien 8Hxvi). Ce chercheur pense aussi que les étudiants en épidémiologie devront se familiariser avec l'apprentissage automatique et l'intelligence artificielle parce que, selon lui, ces choses-là nécessiteront beaucoup de procédures d'analyses différentes. Toujours selon ce chercheur, cette nouvelle pratique est difficile et tout le monde ne deviendra pas un expert de la pratique de l'épidémiologie numérique. Cependant, il pense qu'il faut mettre en place des mécanismes pour familiariser les étudiants en épidémiologie avec cette nouvelle pratique et offrir la possibilité d'un apprentissage plus avancé pour ceux qui sont intéressés à approfondir la pratique de l'épidémiologie numérique.

Le discours d'un autre participant en épidémiologie environnementale va dans le même sens. Ce chercheur souligne que les nouvelles sources de données numériques contribuent à l'apprentissage de l'épidémiologie (entretien 5Gxxv). Ces nouvelles sources de données peuvent se révéler difficiles à utiliser et celles-ci peuvent inciter les étudiants à aller vers des ressources à l'extérieur de la discipline de l'épidémiologie pour compléter leur apprentissage selon ce chercheur. Cette démarche peut s'avérer intéressante pour les étudiants, mais celle-ci peut rendre le processus d'apprentissage plus compliqué pour devenir chercheur en épidémiologie d'après ce chercheur.

« I think that starting early would be good in education and training. And human resources that are going to be required. I am also thinking from an analysis perspective students do need to become more familiar with some of these machine learning and artificial intelligence because I think that these are going to take a lot more of different analysis procedures. This is difficult and not everybody is going to become an expert, but I think that some familiarity with them at least an opportunity for more advanced learning in it has to be put in place, I think. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« Right now, just thinking about student training... I think that all of these new data sets are really exciting right? But it is also just ... when you think about it from a student training perspective, so in addition to making sure that your students are trained as epidemiologists, so that they have all the credentials, they understand good study design, they understand what a good measurement means, they know how to do advanced statistical analysis, in addition to that like epidemiology is just harder now in a sense that there are these massive data sources that can be challenging to use, it forces them to step outside of epidemiology and talk with other people, I think that those are good things but it certainly complicates the process of them trying to learn how to be an epidemiologist. » (Chercheur - épidémiologie environnementale - entretien 5Gxxv)

En revanche, deux autres participants ont un point de vue différent. L'un de ces participants en épidémiologie clinique a affirmé que la pratique de l'épidémiologie numérique ne nécessite pas de connaissances additionnelles (entretien 5Cxxx). En effet, ce chercheur soutient que la discipline de l'épidémiologie possède les fondements requis pour pratiquer l'épidémiologie numérique à l'exception de situations exceptionnelles où la recherche pourrait exiger des habiletés spécialisées et de la collaboration avec d'autres chercheurs. Dans la même perspective, un autre participant en épidémiologie sociale estime que l'approche du chercheur en épidémiologie serait d'en apprendre un peu plus sur les sources de données et les développements des technologies, mais il considère que le chercheur en épidémiologie utilisera toujours son raisonnement déductif de base et que ce raisonnement le mènera loin dans le cadre de la pratique de l'épidémiologie numérique (entretien 6Pvi).

« Je ne pense pas qu'on ait besoin vraiment de nouvelles connaissances basées sur l'utilisation de l'épidémiologie numérique, sauf si cette une étude très spécifique, très spécialisée et qu'on besoin de collaboration d'un autre chercheur dans un autre milieu pour nous guider, mais je ne pense pas qu'avec l'épidémiologie de base qu'on a vraiment besoin de nouvelles compétences. » (Chercheur - épidémiologie environnementale - entretien 5Cxxx)

« I think things for epidemiologists our approach would be to learn a little bit more about the data sources to make sure to know on what we are getting into but in terms of the methodologies that we use, there might be some new things developed but I think we will still use our core deducting reasoning and I think that would take us far. » (Chercheur - épidémiologie sociale - entretien 6Pvi)

6.1.2.2 Les points de vue des chercheurs ayant l'expérience de l'épidémiologie numérique

Il est important de considérer les points de vue de chercheurs pratiquant l'épidémiologie numérique. Leurs avis peuvent différer par rapport à ceux expliqués précédemment. En effet, l'un des participants en épidémiologie environnementale ayant l'expérience de la pratique de l'épidémiologie numérique a affirmé clairement que les habiletés requises pour cette nouvelle pratique ne font pas nécessairement partie des habiletés courantes en épidémiologie (entretien 10Ixxviii). De plus, selon ce chercheur, la plus grande limite dans le domaine est la difficulté à établir les liens entre les spécialistes des données et les épidémiologistes afin que ces derniers puissent avoir la possibilité d'exploiter ces données s'ils les jugent utiles. En fait, la formation dans

le contexte de la pratique de l'épidémiologie numérique doit se faire aussi auprès des informaticiens adoptant cette nouvelle pratique afin qu'ils puissent interpréter les données épidémiologiques correctement d'après les propos de ce chercheur. Ce chercheur affirme aussi qu'il préférerait enseigner aux informaticiens comment faire de l'épidémiologie plutôt qu'enseigner l'informatique à des épidémiologistes. La question de la formation est donc essentielle dans le cadre de la pratique de l'épidémiologie numérique selon ce chercheur et l'apprentissage n'est pas uniquement nécessaire pour le chercheur en épidémiologie, mais aussi pour les informaticiens contribuant à des projets de recherche épidémiologique.

Ce chercheur ayant l'expérience de la pratique de l'épidémiologie numérique souligne aussi qu'il aimerait voir l'apparition de programmes permettant à chaque personne intéressée par l'épidémiologie d'acquérir de solides compétences en science des données, puis de procéder à l'analyse et à l'interprétation. En illustrant ses propos avec les données géographiques de Google, ce chercheur a expliqué aussi que l'interprétation n'est pas nécessairement simple et qu'il faut souvent des interactions entre les personnes responsables de traiter l'information et le chercheur afin de comprendre ce que ces données signifient réellement.

« Very computational skills, I would say. I am as much as a data scientist as I am an epidemiologist. So, I've been working with large data sets and manipulate them and do what I want them, but I know that it is not necessarily common skills set in epidemiology and that would be I guess the biggest limitation in the field is making those connections between data scientists and epidemiologists so that epidemiologists could have the opportunity to take advantage of those data if they feel that they would be useful. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

« You just can't turn a data scientist on digital data from which you are trying to draw epidemiological conclusions. They really do need some founding in our science. So, there is often some extra training involved and well known for saying I'd rather teach computer scientists to epidemiology than teaching epidemiologists computer science. But they need a foundation our skills set. So, one of the things I would love to see is programs where each people who are interested in epidemiology with solid data science skills and then you get there, and you need to do your analysis and interpret. The interpretation is not necessarily straight forward and often takes some back and forth with people who are dealing with the information with you to understand really what these data mean, we had a lot of this around the Google location data they're putting out information about, how much time people are spending at home, how much time they are spending in grocery stores, how that compares with other periods and once you start to interpret those data, you realize what are they doing

here, what is this algorithm? So, this is back and forth about that. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

Les deux autres participants en épidémiologie sociale pratiquant l'épidémiologie numérique insistent moins sur la nécessité de la formation des chercheurs en épidémiologie. L'un de ces participants pense qu'il faut une compréhension globale de la pratique de l'épidémiologie numérique si un chercheur en épidémiologie adopte cette nouvelle pratique (entretien 6Axxiii). Ce participant estime également qu'il n'est pas nécessaire pour le chercheur en épidémiologie de faire beaucoup de programmation ou des choses de ce genre parce que les chercheurs collaborent avec gens en informatique s'occupant de la programmation. Le point de vue du troisième participant pratiquant l'épidémiologie numérique est similaire à celui-ci, mais il diffère à propos des habiletés techniques comme celle du codage (entretien 8Sxx). Ce chercheur soutient qu'il faut développer des capacités à apprendre le codage et à utiliser les logiciels afin de ne pas être entièrement dépendant de ressources externes.

« I think just having a general understanding of digital epidemiology like a very broad and high-level working would be helpful. I think that we don't have to expect people to do a lot of programming and that type of things. People collaborate with the people who are doing that. But at least a general understanding of the principles of digital epidemiology and understanding the usefulness and how you can use it in your research. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 6Axxiii)

« I think the biggest one is interdisciplinary engagement. Second is building some capacity in learning how to code, learning how to use some of the software programs so that you are not completely rely on external persons. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

À la lumière de ces résultats, les perspectives des chercheurs au sujet de la formation sur la pratique de l'épidémiologie numérique varient selon la façon dont les chercheurs travaillent, c'est-à-dire la culture épistémique de ces chercheurs. Le participant en épidémiologie environnementale pratiquant l'épidémiologie numérique effectue le travail de traitement et d'analyse des données numériques lui-même alors que les deux autres participants travaillent en collaboration pour les tâches de leurs projets de recherche relevant de l'informatique. Il semble donc logique que le participant en épidémiologie environnementale soit d'avis que la formation des chercheurs

constitue un enjeu important étant donné qu'il accomplit la majeure partie du travail lui-même. Cette façon de travailler nécessite d'autres habiletés que celles incluses dans la formation traditionnelle en épidémiologie. Les deux autres participants travaillant en collaboration mettent moins l'accent sur la formation des chercheurs en épidémiologie, car ils vont chercher des ressources externes pour pratiquer l'épidémiologie numérique.

En somme, les résultats de nos analyses témoignent que la formation des chercheurs et des futurs chercheurs constitue un enjeu important dans le cadre de la pratique de l'épidémiologie numérique parce que celle-ci nécessiterait des apprentissages au sein de la discipline de l'épidémiologie. Il est donc possible que la discipline de l'épidémiologie doive s'adapter pour équiper ses chercheurs s'ils veulent pratiquer l'épidémiologie numérique. Certains participants soutiennent qu'il faudra réfléchir sur la façon de mettre en œuvre cette adaptation pour l'apprentissage de cette nouvelle pratique.

D'autres proposent de former les chercheurs et les futurs chercheurs de manière à leur fournir une compréhension générale de la pratique de l'épidémiologie numérique sans aller trop en détail, car souvent les projets de recherche fondés sur cette nouvelle pratique se font en collaboration interdisciplinaire, notamment avec l'informatique pour les aspects techniques de cette pratique. Une formation en épidémiologie peut s'avérer nécessaire pour les informaticiens contribuant à des projets de recherche épidémiologique afin d'aider à bien interpréter les données numériques d'après certains propos entendus lors de la recherche de terrain. Par conséquent, la formation des chercheurs en épidémiologie sur la pratique de l'épidémiologie numérique serait essentielle. Il se pourrait aussi qu'il soit important d'enseigner l'épidémiologie aux chercheurs d'autres disciplines comme celle de l'informatique afin d'établir une connexion entre les spécialistes des données et les chercheurs en épidémiologie lors des recherches épidémiologiques fondées sur cette nouvelle pratique.

Ces constats constituent une autre illustration de la perspective soutenant que les méthodes de recherches liées à la pratique de l'épidémiologie numérique ne sont pas des boîtes à outils neutres parce que celles-ci peuvent susciter de nouveaux besoins pour la formation des chercheurs en épidémiologie. Les chercheurs et les futurs chercheurs n'étant pas nécessairement bien préparés

pour pratiquer l'épidémiologie numérique, il est possible que l'adoption de cette nouvelle pratique en épidémiologie demeure rare au sein de la discipline.

6.1.3 L'interdisciplinarité

L'interdisciplinarité constitue un enjeu important étudié en long et en large en STS et dans d'autres domaines comme les sciences humaines et sociales. Il semble que l'interdisciplinarité soit apparue comme « une voie nouvelle dans le développement de l'enseignement et de la recherche universitaires » il y a quelques décennies (Darbellay, 2011). Gibbons *et al.* (1994) soutiennent que les découvertes sont construites à l'extérieur des frontières d'une discipline dans le nouveau mode de la production des connaissances (Gibbons *et al.*, 1994). Certains chercheurs interprètent cette perspective comme une thèse soutenant « la dissolution des frontières disciplinaires » (Louvel, 2015, p. 76). Cette thèse est critiquée, notamment par Heilbron et Gingras (2015) qui expliquent leur perspective s'appuyant sur les travaux empiriques de Jerry Jacobs. Les travaux de ce chercheur concluent que les disciplines ne sont pas des silos isolés, mais plutôt des points de rencontre reliés par un vaste réseau de programmes, de centres et d'instituts de recherche interdisciplinaire (Jacobs, 2013). Peu importe si l'interdisciplinarité est vue comme une nouveauté ou une continuité, elle représente un enjeu majeur dans les études sociales sur les sciences et les technologies (Louvel, 2015 ; Smith-Doerr, 2017).

L'interdisciplinarité constitue un des enjeux d'expertise le plus souvent mentionné par les participants à la recherche de terrain lors des discours sur la pratique de l'épidémiologie numérique, et ce, peu importe l'expérience de la pratique de l'épidémiologie numérique et de la sous-discipline du chercheur. D'ailleurs, l'interdisciplinarité serait reflétée dans la validation par les pairs des travaux fondés sur la pratique de l'épidémiologie numérique parce que ces travaux sont acceptés par des comités de lecture pour des revues interdisciplinaires comme la revue *Frontiers in Public Health* (Gianfredi *et al.*, 2018 ; Müller et Salathé, 2019 ; Siddiqui *et al.*, 2022). Notre recherche laisse entrevoir deux thèmes principaux quant à l'interdisciplinarité lors de la pratique de l'épidémiologie numérique, à savoir la discipline de l'informatique et les autres disciplines.

6.1.3.1 La discipline de l'informatique

Les collaborations interdisciplinaires envisagées pour la pratique de l'épidémiologie numérique seraient principalement fondées sur le besoin de l'expertise informatique. Cet avis est revenu très souvent dans les discours des participants à la recherche de terrain. Par exemple, un chercheur en épidémiologie environnementale a souligné qu'il voit des équipes composées de chercheurs issus de plusieurs disciplines avec une présence importante de l'informatique (entretien 5Cxxvi). Ce chercheur pense aussi que la profession d'épidémiologiste a beaucoup changé récemment, car il faut soit savoir coder et effectuer d'autres tâches informatiques, soit intégrer une équipe comprenant des informaticiens.

« Moi ce que je vois se dessiner à l'horizon ce sont des équipes de plus en plus multidisciplinaires avec une grosse composante d'informaticiens finalement parce que je pense que ça change beaucoup l'approche, la profession d'épidémiologiste par rapport à la façon il y a juste encore 30 ans quoi. Ça a plus grand-chose à voir et puis, je pense, que ça s'en va en s'accroissant. Les épidémiologistes maintenant, il faut qu'ils sachent coder, etc. ou alors ils sont dans des équipes avec des informaticiens. Donc c'est le profil des équipes je pense qui va être de plus en plus multidisciplinaire. Ça va être une nécessité. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvi)

Les chercheurs en épidémiologie se voient donc collaborer avec des informaticiens lors de la pratique de l'épidémiologie numérique. Ce phénomène peut illustrer l'influence des technologies dans la formation des groupes d'acteurs lors des activités de recherche comme suggéré par Knorr-Cetina (1999). La pratique de l'épidémiologie numérique peut nécessiter d'utiliser des logiciels d'analyse et de traitement de données différents de ceux utilisés couramment en épidémiologie. Ainsi, les technologies sont susceptibles d'influencer les façons dont les chercheurs collaborent dans le cadre de la pratique de l'épidémiologie numérique.

6.1.3.2 Les autres disciplines

Cependant, les aspects technologiques ne dictent pas tout selon des chercheurs en épidémiologie. La plupart des participants à la recherche de terrain affirment que l'expertise informatique est de plus en plus nécessaire pour traiter les nouvelles sources de données non traditionnelles, mais

l'expertise des autres disciplines comme celle de l'épidémiologie sont aussi importantes dans le cadre des activités de recherche. La plupart des participants estiment qu'il faut à la fois des gens en informatique et des gens issus d'autres disciplines afin d'être en mesure de bien conduire des recherches épidémiologiques fondées sur la pratique de l'épidémiologie numérique.

Par exemple, l'un des participants en épidémiologie environnementale voit le chercheur en épidémiologie comme un chef d'orchestre travaillant avec d'autres chercheurs en provenance de différentes disciplines puisque les données numériques font partie du paysage de domaines d'expertise spécialisée (entretien 5Cxxvii). D'après ce chercheur, certaines données comme celles des images radar sont difficiles à comprendre, nécessitant ainsi d'établir des partenariats avec des ingénieurs et des mathématiciens.

Or, un participant en épidémiologie sociale s'est exprimé encore plus clairement dans ce sens. Ce chercheur a souligné que la pratique de l'épidémiologie numérique va accentuer les besoins pour l'interdisciplinarité et pour travailler en équipe selon diverses perspectives (entretien 8Hxvi). D'après ce chercheur, le chercheur en épidémiologie sera efficace pour les analyses, mais il faut aussi des informaticiens et des statisticiens pour s'assurer que les données sont traitées correctement. De plus, ce chercheur soutient que l'expertise de praticiens en santé s'avère aussi importante, car les études sur les populations nécessitent d'être capable d'expliquer les résultats. Il est possible de trouver beaucoup d'associations dans les analyses de données sophistiquées, mais encore faut-il que ces associations signifient quelque chose. Donc, selon les propos de ce chercheur, il faut l'expertise de l'épidémiologiste, du praticien en santé, de l'informaticien et des autres disciplines pertinentes afin d'assurer de la rigueur dans les études épidémiologiques.

« Les épidémiologistes, on va être un peu des chefs d'orchestre avec une lunette multidisciplinaire, la plupart de ces données-là sont ancrées dans des domaines d'expertise très spécialisés... Donc ça prend des gens avec... par exemple, particulièrement pour les images radar qui sont des données vraiment complexes à comprendre, donc ça oblige les épidémiologistes à faire des partenariats avec des disciplines vraiment non traditionnelles dans différents aspects... donc on parle beaucoup aussi d'utilisation de nouvelles méthodes de l'intelligence artificielle pour essayer de faire des analyses là-dedans et puis ça aussi ça va porter l'accès à de nouvelles associations avec des ingénieurs et puis des mathématiciens et des choses comme ça. » (Chercheur - épidémiologie environnementale - entretien 5Cxxvii)

« I think you have to have both. I think digital epidemiology is just increasing our need for interdisciplinary and being able to work together from different perspectives. Epidemiologists are going to be very strong in the analysis, for example machine learning, but you need a computer scientist, you need a statistician to help to make sure it is being done appropriately. On top of that play, I also that you need the practitioner ... the work with the study of the population also involves being able to explain the results. this goes back to my comment about data digging and data mining. If you've got some scales and some stats and some computing science, you can run things and just come up with some associations but is it important, does it mean anything ... have you evaluated it in some sort of scientific rigor from a perspective of health science, so you need the epidemiologists, practitioners, computer scientists working all together with this. You still need them all. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

En plus d'illustrer l'importance des technologies dans les formations des équipes de recherche comme expliquée précédemment, cette vision illustre la gestion des recherches orientées vers l'objet d'études comme proposé par Knorr-Cetina (1999) parce que les chercheurs s'organisent aussi autour d'un objet de recherche et non pas seulement selon les frontières de leur discipline. Un objet de recherche peut souvent impliquer l'expertise de diverses disciplines et ainsi, inciter les chercheurs à former des équipes interdisciplinaires en fonction du projet de recherche.

Donc, en principe, la gestion des recherches lors de la pratique de l'épidémiologie numérique serait basée sur les objets des recherches ou sur les tâches à effectuer. Gibbons *et al.* (1994) ont aussi introduit cette idée en affirmant que le nouveau mode de la production des connaissances serait hétérogène et diversifié sur les plans organisationnels ainsi qu'au niveau des habiletés et des expériences. En outre, les propos de plusieurs chercheurs ayant participé à la recherche de terrain vont encore plus loin en soutenant que toutes les expertises sont importantes lors des recherches en santé et qu'aucune ne prend le dessus sur l'autre. Il faut toutes les expertises pour mener à bien les projets de recherches et non pas seulement celle de l'informatique.

En somme, la pratique de l'épidémiologie numérique constituerait un moteur considérable intensifiant les collaborations interdisciplinaires et entre diverses spécialités dans le cadre des recherches selon les constats établis lors de notre recherche. Tous les participants pensent que la pratique de l'épidémiologie numérique nécessite de collaborer avec des chercheurs en informatique en raison des besoins de compétences. La technologie semble ainsi jouer un rôle important dans

les formations des équipes de recherches, mais ce n'est pas la seule. L'expertise des autres disciplines comme celle de l'épidémiologie ou de la statistique sont aussi importantes dans le cadre des activités de recherche. Certains participants ont souligné aussi l'importance de collaborer avec des experts de divers domaines spécialisés dans le but de comprendre ou de traiter correctement certains types de données comme les données générées par les téléphones cellulaires ou celles en provenance des satellites. La plupart des participants pensent qu'il faut à la fois des gens en informatique et des gens provenant d'autres disciplines pour bien conduire des recherches épidémiologiques fondées sur la pratique de l'épidémiologie numérique. Donc, en plus des technologies, les objets des recherches et les tâches à effectuer semblent influencer la formation des équipes dans la pratique de l'épidémiologie numérique.

Pour résumer, notre recherche montre que les enjeux d'expertise liés à la pratique de l'épidémiologie numérique sont les nouvelles compétences et les spécialités envisagées, la formation des chercheurs en épidémiologie et l'interdisciplinarité. Rappelons que nos analyses sur les nouvelles compétences et les spécialités envisagées en lien avec cette nouvelle pratique mettent en évidence la nécessité de mise à jour en épidémiologie, l'importance des compétences en informatique et en science des données ainsi que des compétences non techniques comprenant la collaboration interdisciplinaire, la protection de la vie privée et la négociation des accords de partage de l'information.

Nos analyses montrent également que la pratique de l'épidémiologie numérique a une vie sociale au même titre que les méthodes de recherches. Les méthodes de recherches liées à la pratique de l'épidémiologie numérique ne sont pas des boîtes à outils neutres parce que celles-ci peuvent influencer des pratiques et susciter de nouveaux besoins en matière de compétences en épidémiologie. La pratique de l'épidémiologie numérique implique des enjeux techniques nécessitant des compétences en informatique et en science des données et des enjeux sociopolitiques exigeant des compétences non techniques comme la collaboration interdisciplinaire, la protection de la vie privée et la négociation des accords de partage de l'information.

Les chercheurs et les futurs chercheurs en épidémiologie peuvent avoir besoin de formation pour pratiquer l'épidémiologie numérique selon les discours de certains participants à notre recherche. Il est donc possible que la discipline de l'épidémiologie doive s'adapter pour équiper ses chercheurs intéressés à pratiquer l'épidémiologie numérique. Il faudra réfléchir sur la façon de mettre en œuvre cette adaptation pour l'apprentissage de cette nouvelle pratique. Ces constats constituent une autre illustration que les méthodes de recherches liées à la pratique de l'épidémiologie numérique ne sont pas des boîtes à outils neutres parce que celles-ci peuvent susciter de nouveaux besoins pour la formation des chercheurs en épidémiologie. Les chercheurs et les futurs chercheurs n'étant pas nécessairement bien préparés pour pratiquer l'épidémiologie numérique, il est possible que l'adoption de cette nouvelle pratique en épidémiologie demeure rare au sein de la discipline.

Quant à l'interdisciplinarité, la pratique de l'épidémiologie numérique représenterait un moteur considérable intensifiant les collaborations interdisciplinaires et entre diverses spécialités dans le cadre des recherches d'après les constats établis lors de notre recherche. La pratique de l'épidémiologie numérique nécessite de collaborer avec des chercheurs en informatique étant donné les besoins de compétences techniques. La technologie semble donc jouer un rôle important dans les formations des équipes de recherches, mais d'autres éléments entrent en jeu. Il faudrait à la fois des gens en informatique et des gens dans d'autres disciplines afin de conduire efficacement des recherches épidémiologiques fondées sur la pratique de l'épidémiologie numérique. Par conséquent, en plus des technologies, les objets des recherches et les tâches à faire influenceraient la formation des équipes lors de la pratique de l'épidémiologie numérique.

6.2 La légitimité de la pratique de l'épidémiologie numérique

Dans cette section, nous nous interrogerons sur la façon dont la légitimité de la pratique de l'épidémiologie numérique est envisagée. Pour ce faire, nous analyserons la reconnaissance institutionnelle, les barrières à la légitimité ainsi que les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique. Les résultats seront présentés selon les points de vue des participants puisqu'il n'y a pas de constats particuliers en fonction de la sous-discipline du chercheur ni selon l'expérience de la pratique de l'épidémiologie numérique.

6.2.1 La reconnaissance institutionnelle de la pratique de l'épidémiologie numérique

La reconnaissance institutionnelle de la pratique de l'épidémiologie numérique peut être analysée en étudiant comment cette nouvelle pratique s'insère dans les recommandations de formation des chercheurs en épidémiologie de même que dans les programmes d'études en épidémiologie. Nous analyserons aussi les liens avec l'informatique lors des enseignements en épidémiologie parce que la discipline de l'informatique joue un rôle important dans le cadre de la pratique de l'épidémiologie numérique comme expliqué précédemment.

6.2.1.1 Les recommandations de formation des chercheurs en épidémiologie

Les propos de la plupart des participants à la recherche de terrain laissent entendre que leur université n'a pas établi de recommandation sur la formation des chercheurs en épidémiologie quant à l'apprentissage de la pratique de l'épidémiologie numérique. Par exemple, un chercheur en épidémiologie sociale a affirmé que les recommandations institutionnelles à l'égard de l'apprentissage des chercheurs sur cette nouvelle pratique sont très limitées (entretien 6Fviii). Ce chercheur souligne que des étudiants apprennent un peu à pratiquer l'épidémiologie numérique et qu'il existe quelques conférences en santé publique sur l'intelligence artificielle. D'après ce chercheur, certains ateliers sont annoncés, mais ceux-ci ne sont pas inclus dans les des cours de base réguliers au sein de son université. Ce chercheur a aussi affirmé qu'il ne sait pas si la pratique de l'épidémiologie numérique fera partie des compétences essentielles en santé publique.

Selon les propos d'un autre chercheur en épidémiologie sociale, son département s'est engagé dans l'internationalisation de ses programmes d'études et d'enseignement. Ce département n'aurait pas planifié de développer ses programmes afin d'y intégrer la pratique de l'épidémiologie numérique (entretien 6Biv). Le chercheur de ce département croit que son université devrait investir davantage dans ce domaine. Il précise que son université offre un cours sur les big data depuis longtemps, mais ce cours explique comment utiliser des ensembles de données administratives et ne se concentre pas sur les données numériques comme celles des médias sociaux ou de Google.

Un autre chercheur en épidémiologie sociale est d'avis que la formation sur la pratique de l'épidémiologie devrait être offerte en informatique et pas nécessairement en épidémiologie (entretien 6Pvi). Ce chercheur affirme que son département est très petit et que personne ne travaille à extraire des informations à partir de Twitter ou d'autres sources de ce genre. Un chercheur en épidémiologie clinique ne sait pas si son université offre de la formation sur la pratique de l'épidémiologie numérique (entretien 6Biii). Il est certain que son université offre des cours sur la manière d'utiliser des données de sources traditionnelles comme celles de Santé Canada, mais il est très incertain en ce qui concerne l'offre de formation de son université sur la façon d'utiliser des données à l'extérieur du système de santé. Ce chercheur estime qu'il y a probablement des besoins pour des cours et de la formation dans ce domaine. Un autre chercheur en épidémiologie sociale a affirmé clairement qu'il n'a pas reçu de recommandation de son université pour pratiquer l'épidémiologie numérique (entretien 8Hxvi). Il a souligné qu'en tant qu'épidémiologiste, il utilise un peu les données numériques, mais il se perçoit comme étant très ignorant sur le sujet. Ce chercheur essaie toujours de comprendre quelles sont les définitions et où se situent les choses dans ce domaine et il n'a reçu aucune directive ou recommandation officielle de son université sur ce sujet.

« Very limited. Students learn about it a bit. We have a few lectures from our master's students in public health on kind of artificial intelligence with examples of digital epidemiology tools that have been used. There are some workshops that I see advertised but not part of the regular core of courses. This is interesting because we are doing some work on core competences in Canada and alignment with what we are teaching the core competencies, but I don't know if that would be part of core competencies in our public health training. » (Chercheur - épidémiologie sociale - entretien 6Fviii)

« The University of Ottawa has committed to the internationalization of its curriculum and approach so there is an interest in digital technologies to facilitate that. But in our department, we haven't start to think about curriculum development in this area ... we probably should...but it is not done yet... we do have a big data course in our epidemiology program (since a long time) that is only focused on how to use administrative data sets rather than the kind of data you are describing like social media or Google data or that kind of things. » (Chercheur - épidémiologie sociale - entretien 6Biv)

« Not in epidemiology right now in terms of public health. I think all of that kind of work has been more computer science driven. Pulling information from Twitter and other sources right now we don't do any of that here. We are a very small public health department. There are only 2.5 epidemiologists in the department and none of them are doing that type of work right now. » (Chercheur - épidémiologie sociale - entretien 6Pvi)

« I don't think... it is a good question...I don't know ... our institution certainly has courses on how to use data, how to get data from let's say Health Canada, etc. etc... but beyond the health sector, I am not sure. So, I think there is probably a need for some courses and training. » (Chercheur - épidémiologie clinique- entretien 6Biii)

« No. I have never been told that I have to or recommending that I have to do some more work on that. As epidemiologists we use it a little bit but again, I still see myself as very ignorant on the topic trying to figure out what the definitions are where certainly things fit in but no, I haven't received any formal guidelines or recommendations. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

6.2.1.2 Les programmes d'études en épidémiologie

Toujours concernant la reconnaissance institutionnelle de la pratique de l'épidémiologie numérique, nous avons aussi examiné les façons dont cette nouvelle pratique peut s'intégrer aux programmes d'études en épidémiologie. Les perspectives varient sur cette question.

Une bonne partie des participants sont d'avis que les programmes d'études des universités en épidémiologie devraient s'adapter pour offrir aux étudiants des mécanismes d'apprentissage de la pratique de l'épidémiologie numérique. Par exemple, l'un des chercheurs en épidémiologie environnementale ayant participé à la recherche de terrain affirme que la pratique de l'épidémiologie numérique doit occuper une place essentielle dans l'apprentissage de l'épidémiologie présentement et pour le futur (entretien 5Cxxvii). Selon ce chercheur, il faut que les épidémiologistes soient formés pour développer une très bonne aisance dans l'utilisation, la manipulation et la compréhension des données numériques innovantes.

L'un des participants en épidémiologie sociale pratiquant l'épidémiologie numérique estime qu'il existe un énorme potentiel de changement dans les programmes d'épidémiologie avec l'émergence de cette nouvelle pratique (entretien 8Sxx). Ce chercheur souligne que les choses ont déjà commencé à changer dans certaines universités comme McGill University qui offre un programme où les étudiants utilisent les big data lors d'études épidémiologiques. Il pense aussi que l'intégration de la pratique de l'épidémiologie numérique aux programmes d'études en épidémiologie deviendra plus courante au cours des dix prochaines années.

Dans une optique similaire, un chercheur en épidémiologie clinique a fait allusion à un programme à l'Université McMaster (entretien 6Fviii). Pour la première fois, ce programme tentera d'introduire des étudiants en épidémiologie aux méthodes d'apprentissage automatique afin qu'ils aient une compréhension conceptuelle de ces méthodes. Ce chercheur affirme également qu'il n'y a pas beaucoup de programmes enseignant ce genre de méthodes et que peu de personnes au Canada sont capables de les enseigner.

Pour sa part, un participant en épidémiologie sociale croit que des programmes de son université offrent des bourses en format stage pour des étudiants au baccalauréat intéressés par les données numériques (entretien 9Rxiv). Ce chercheur est aussi d'avis que l'université devrait soutenir ce domaine. Un autre chercheur en épidémiologie sociale considère que la pratique de l'épidémiologie numérique doit être enseignée tôt aux étudiants en épidémiologie (entretien 8Hxvi). Ce chercheur a précisé que l'enseignement de cette nouvelle pratique doit inclure comment accéder aux différentes sources de données. Il pense également que les avantages et les inconvénients de ces sources de données de même que leur précision, validité et fiabilité doivent être expliqués aux étudiants en épidémiologie. Quant aux analyses, il souligne que les étudiants doivent se familiariser avec des techniques d'apprentissage automatique et d'intelligence artificielle. Ces techniques s'avèrent difficiles d'après ce chercheur. Il estime que tout le monde ne va pas devenir un expert, mais qu'il faut au moins familiariser les étudiants en épidémiologie avec ces techniques et offrir la possibilité d'un apprentissage plus avancé pour ceux intéressés.

« C'est une place absolument essentielle, critique. On ne veut pas former des épidémiologistes pour faire des études cotées moins du passé. Je pense qu'il va falloir former tous nos épidémiologistes avec une très bonne aisance dans l'utilisation, la manipulation et la compréhension des données numériques innovantes. Le domaine de l'épidémiologie numérique est critique et essentiel dans l'apprentissage de l'épidémiologie présentement et pour le futur. Et ça complète d'autres phases de la formation de l'épidémiologie. »
(Chercheur - épidémiologie environnementale - entretien 5Cxxvii)

« So, I think there is huge potential in epidemiological programs not just in our university but epidemiological programs across Canada would change and I think they already changed. I know their programs in McGill where they don't call it digital epidemiology, but people are using big data for epidemiological purposes. I think it will become more mainstream in the next 10 years. So, at this point of time, I would say the potential is high, although it may not

realize its pick point. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

« There is already a lot of interest around growing capacity around machine learning methods and you are starting to see it in traditional epidemiological programs. Not a lot but the example at McMaster, it is the first time this year in our program, they will be testing epidemiology students in machine learning methods from a broad perspective. It doesn't mean that they will do anything, but they will understand it conceptually. We don't have a lot of faculties that are particularly teaching it and it is difficult to teach because the people who could teach it, there is not a lot in Canada. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

« Je pense que c'est un domaine qui devient de plus en plus important. À mon université, je pense que cette année il y avait un appel pour les bourses de stages d'été pour les étudiants au bac, les priorités incluait... je ne sais plus comment ça s'appelle, mais c'est lié aux données numériques. Je pense que c'est quelque chose que l'Université devrait supporter en général. » (Chercheur - épidémiologie sociale - entretien 9Rxiv)

« I think that starting early would be good in education and training. And human resources that are going to be required. This is similar as what I mentioned earlier, I would expect or encourage now more focus on the access to different data sources and some of the advantages and disadvantages of those data sources. from a practical process perspective as well as accuracy and validity and reliability. I think those should be included. It's important and this is happening. So, it should be included. I am also thinking from an analysis perspective students do need to become more familiar with some of these machine learning and artificial intelligence because I think that these are going to take a lot more of different analysis procedures. this is difficult and not everybody is going to become an expert, but I think that some familiarity with them at least and opportunity for more advanced learning in it has to be put in place, I think. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

Cependant, l'un des chercheurs en épidémiologie sociale ayant participé à la recherche de terrain soutient que l'intégration de la pratique de l'épidémiologie numérique dans les programmes d'études en épidémiologie ne devrait pas relever de l'université (entretien 5Uxxxvi). En effet, il affirme que les professeurs du programme devraient décider d'inclure cette pratique dans leurs cours. De plus, ce chercheur estime qu'il n'y aurait pas une clientèle pour un cours complet de 45 heures dédié à la pratique de l'épidémiologie numérique. En outre, il a déclaré qu'il pourrait y avoir des ateliers, c'est-à-dire une séance de cours dans la semaine réservée à la pratique de l'épidémiologie numérique.

« Je pense que ça peut être d'abord en sensibilisant les étudiants, les futurs chercheurs et chercheuses à l'existence de ces données-là, à leur potentiel, aussi peut-être à leurs limites puis leur disponibilité peuvent aussi j'imagine servir à l'enseignement. À mon sens, ça ne devrait pas relever de l'Université comme telle. Je pense que ce sont les enseignants du programme qui devraient inclure ça. À la limite, à l'échelle plus institutionnelle, on pourrait penser à des cours, mais je ne pense pas qu'il y aurait une clientèle pour un cours de 45 heures spécialisé en épidémiologie numérique. Par contre, on peut penser à des ateliers, tu sais dans un cours d'épidémiologie à ce qu'une des semaines de cours soit consacrée à l'épidémiologie numérique, des choses comme ça. C'est pour ça que j'ai l'impression que ça relève davantage des enseignants que de l'institution. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

D'autres chercheurs affirment que la discipline de l'épidémiologie devra se mettre à jour concernant les nouvelles méthodes liées à la pratique de l'épidémiologie numérique. Un participant en épidémiologie sociale est d'avis que cette nouvelle pratique ne change pas beaucoup les approches d'analyse en épidémiologie, mais qu'il faut des moyens pour gérer et nettoyer ce type de données (entretien 5Kxv). Ce participant a mentionné que l'enseignement de la pratique de l'épidémiologie numérique serait pertinent pour les générations futures. Il envisage que cette nouvelle pratique occupe une place relativement importante dans le travail des futurs épidémiologistes.

Un chercheur en épidémiologie environnementale affirme quant à lui que la santé publique devra probablement effectuer du rattrapage puisqu'il existe peu de gens en épidémiologie formés en informatique selon lui (entretien 5Gxxv). Ce chercheur souligne qu'il y a plus de gens que dans le passé avec des habiletés en informatique en épidémiologie, mais qu'il reste encore beaucoup de rattrapage à faire sur ce plan. Un chercheur en épidémiologie clinique soutient que les plateformes numériques doivent être expliquées dans les cours d'épidémiologie afin que ces cours soient à jour (entretien 6Fxi). Ce chercheur pense également que si de nouvelles tâches en épidémiologie apparaissent avec l'émergence de la pratique de l'épidémiologie numérique, celles-ci doivent être incluses dans les cours en épidémiologie afin de les garder à jour.

« La discipline devra se mettre à jour avec les nouvelles méthodes requises pour nettoyer ces grosses bases de données-là. Pour les analyser, ça ne change pas énormément nos approches, mais il faudra avoir des moyens pour pouvoir gérer toutes ces données. Je pense que ça serait pertinent que les futures générations d'épidémiologistes soient formées sur les méthodes spécifiques à l'épidémiologie numérique. L'épidémiologie numérique va occuper une place

assez importante dans le travail des futurs épidémiologistes. Ça devrait commencer à être intégré dans le curriculum assez rapidement. Donc, c'est très important. » (Chercheur - épidémiologie sociale - entretien 5Kxv)

« Public health probably has to catch up a little bit ... I think... we are not training people on having a lot of computers ... I think that there is more than before, but I don't think that you will see a lot of people with computer science background in public health... really on the programming side and leveraging some of these non-traditional data sets, I think we are starting to see that more, but I think probably public health needs to catch up a little bit and figure out people that they can hire for jobs that could leverage better some of those data and information and resources that are being generated by universities. I think that is better than it was 10 years ago and 5 years ago, but I think that there is still a lot of catching to do. » (Chercheur - épidémiologie environnementale - entretien 5Gxxv)

« I think that courses that involve study design, data collection they must certainly if they want to stay modern should at least be mentioning digital epidemiology platforms. If not, entire workshops being offered I imagine that this is available now if not, it should be available. I don't know if an entire course is necessary but there is certainly a need to be for the young epidemiologists that needs to be information available in their training program that let them know about digital platforms, how they are used, what they are used for and how they should be using them. If there are different tasks (as you mentioned before) in the way that you need to approach these data in the process of studies. That needs to be thought and kept up to date. » (Chercheur - épidémiologie clinique - entretien 6Fxi)

6.2.1.3 Les liens avec l'informatique lors des enseignements en épidémiologie

De plus, il semble que le jumelage entre l'informatique (y compris la science des données) et l'épidémiologie devrait s'intensifier dans les programmes d'études en épidémiologie étant donné la place grandissante de la pratique de l'épidémiologie numérique. D'après un participant en épidémiologie sociale, il n'y a pas beaucoup d'écoles de santé publique au Canada donnant des cours orientés vers l'épidémiologie computationnelle, ce qui nécessite souvent d'aller chercher des étudiants d'ailleurs en informatique pour effectuer ce genre d'analyse (entretien 5Cxxxii). Ces étudiants en informatique ne possèdent pas de compétences en épidémiologie et souvent, ils ne tiennent pas compte des biais lors des analyses. Selon ce chercheur, il s'agit d'un contexte de travail avec une approche différente, mais il estime que ce rapprochement entre le côté informatique et le côté épidémiologique s'avère tout à fait nécessaire dans le cadre de la pratique de l'épidémiologie numérique.

L'un des chercheurs en épidémiologie environnementale pratiquant l'épidémiologie numérique affirme que toute personne voulant devenir épidémiologiste à l'ère moderne doit suivre un cours de science des données parce que même si la personne n'effectuera pas ce travail, elle devra communiquer efficacement avec des gens spécialisés dans ce domaine (entretien 10Ixxviii). Ce chercheur donne l'exemple de la statistique. Selon lui, tous les épidémiologistes ne font pas réellement de la statistique, mais ils détiennent les connaissances leur permettant de dialoguer avec des statisticiens ayant le même langage. Il soutient donc que les épidémiologistes des temps modernes devront d'une façon ou d'une autre se familiariser avec la science des données tout comme avec la statistique dans le contexte de l'épidémiologie traditionnelle. Ce chercheur a aussi exprimé sa déception envers le retrait du cours en science des données au sein du programme de maîtrise en santé publique de son département. Il estime que c'est un pas vers la mauvaise direction. Il soutient que des programmes conjoints en épidémiologie, santé publique et science des données sont nécessaires.

« I don't know many schools in Canada that are teaching more like computational epidemiology courses it seems to be quite traditional still and so often we have to go elsewhere to find students that can conduct these types of analysis but again, students let's say in computer science don't have the competency in epidemiology so often, they don't even consider the biases, even like model checking, etc. So, it's quite different in terms of approach. But I find that quite necessary in digital epidemiology to have that paring with more like computational side with epidemiological side. » (Chercheur - épidémiologie sociale - entretien 5Cxxxii)

« I think anybody who wants to become an epidemiologist in the modern age should have to take a data science course or a data science course designed for epidemiologists. Even if it is not that type of work that they end up doing, chances are that they will be working with people who do that type of work, and it is good for them to be able to communicate with each other. It is just like we make them take statistics. A lot of epidemiologists don't ever do real statistics, but they have the background to be able to talk to statisticians when the time comes, and they can have that conversation with a common language. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

« They just dropped the data science course out of the MPH program at UBC and I am very disappointed about that, and I think it was the wrong decision. So that's kind of adverse thing. But then I think we need to look at programs that are joined epidemiology and-or public health and data science program. So, we see these that pop pop all over the place and that's fantastic. In a lot of these cases, people will end up working in epidemiological setting and they are going to end up working with health data and health data or health related data are different from other types of data. So, I would really like to see like a joint master's in public

health and data science and see those types of programs emerging. » (Chercheur - épidémiologie environnementale - pratiquant de l'ÉN - entretien 10Ixxviii)

Enfin, un chercheur en épidémiologie sociale a fait remarquer que l'apprentissage de la pratique de l'épidémiologie numérique ne se fera pas nécessairement en milieu universitaire (entretien 3Exxxiii). Selon lui, les gens possédant les habiletés techniques pour pratiquer l'épidémiologie numérique et intéressés par cette pratique vont s'y mettre sans être liés à une université. Parfois, la pratique de l'épidémiologie numérique ne sera pas possible en raison des contraintes liées à l'éthique de recherches, selon ce participant. Il affirme également que ce contexte va créer des divergences entre les chercheurs universitaires dépendants des évaluations par les pairs et les chercheurs à l'extérieur des universités faisant leur travail avec moins de contraintes que leurs homologues.

« As the barriers for conducting research go down because the people with the technical skills to do it don't have to be based in universities, they don't need the university connections, sometime these practices are held back by universities because of research ethics, etc. but people interested and if they have the technical skills to do it, they will just do it. And that is going to create tensions between universities researchers who are dependent on peer review for their work and not academic researchers will just do what they want because they have to. » (Chercheur - épidémiologie sociale - entretien 3Exxxiii)

En somme, la pratique de l'épidémiologie numérique semble peu adoptée et peu diffusée sur le plan institutionnel selon constats établis lors de notre recherche. Cette nouvelle pratique serait toutefois considérée comme importante et nécessaire en milieu universitaire. Cependant, peu d'actions semblent avoir été entreprises afin d'intégrer cette pratique dans les universités hormis quelques initiatives lancées de manière sporadique dans certains départements d'épidémiologie. Il n'y aurait pas de recommandation institutionnelle ni pour pratiquer ni pour apprendre la pratique de l'épidémiologie numérique. Cette situation semble indiquer que le chercheur en épidémiologie n'est pas particulièrement encouragé à intégrer cette pratique à ses travaux. De plus, les programmes d'études en épidémiologie enseigneraient peu la pratique de l'épidémiologie numérique. Selon des participants à notre recherche, la pratique de l'épidémiologie numérique doit obtenir une plus grande reconnaissance institutionnelle et il doit y avoir plus de jumelage avec l'informatique afin de mieux équiper la discipline de l'épidémiologie pour comprendre, manipuler

et analyser les nouvelles sources de données numériques et ainsi, se garder à jour pour continuer à contribuer à la santé publique.

Cependant, nous avons constaté aussi que certains chercheurs en épidémiologie ont une vision différente. En effet, il est possible que l'apprentissage de cette nouvelle pratique ne se fasse pas nécessairement en milieu universitaire, car les gens intéressés dotés des habiletés techniques requises pour pratiquer l'épidémiologie numérique vont s'y mettre sans qu'ils soient liés à une université. Ce phénomène pourrait créer des divergences entre les chercheurs universitaires et les chercheurs à l'extérieur des milieux académiques.

Ces analyses sur la reconnaissance institutionnelle de la pratique de l'épidémiologie numérique témoignent de l'absence de la pratique de l'épidémiologie numérique dans l'apprentissage de la culture épistémique de l'épidémiologie. Il semble que l'apprentissage de cette nouvelle pratique en milieu universitaire se fasse de manière sporadique et tacite selon les intérêts et les habiletés des étudiants et des professeurs puisque les mécanismes officiels d'apprentissage de l'épidémiologie n'ont pas ou très peu intégré la pratique de l'épidémiologie numérique aux curriculums de la discipline.

De plus, il se peut que l'apprentissage de la pratique de l'épidémiologie soit réalisé en dehors des universités. Cette éventualité a été soulevée seulement une fois lors des entretiens, mais sa portée est importante. En effet, il s'agit d'une illustration d'un des attributs du nouveau mode de la production des connaissances proposé par Gibbons *et al.* (1994), à savoir une augmentation du nombre de sites potentiels où des connaissances peuvent être créées (Gibbons *et al.*, 1994). Les universités et les collèges ne seraient plus les seuls sites de production des connaissances puisque désormais, des instituts non universitaires, des agences gouvernementales et des laboratoires industriels produisent aussi des connaissances. Cette situation peut être vue comme étant problématique, car certains soutiennent que les connaissances produites à l'extérieur des structures légitimes ne peuvent pas être qualifiées de scientifiques (Gibbons *et al.*, 1994). Une opposition à l'égard des structures établies peut alors émerger lorsque des scientifiques agissent de manière différente par rapport aux normes techniques et sociales propres à leur domaine. Par conséquent, la pratique de l'épidémiologie numérique pourrait être marquée par une telle opposition. Des

barrières peuvent aussi freiner la légitimité de cette nouvelle pratique. Ces barrières sont analysées dans ce qui suit.

6.2.2 Les barrières à la légitimité de la pratique de l'épidémiologie numérique

Le but de cette section est de connaître les freins à la légitimité de la pratique de l'épidémiologie numérique. Pour ce faire, nous allons analyser les barrières à la pratique de l'épidémiologie numérique selon les points de vue des participants à la recherche de terrain.

6.2.2.1 Les cadres juridiques liés aux données

La barrière à la légitimité de la pratique de l'épidémiologie numérique la plus souvent mentionnée par les participants à la recherche de terrain est celle concernant les cadres juridiques liés aux données, qui constitue un enjeu complexe à gérer lors des projets de recherche. Par exemple, un chercheur en épidémiologie sociale a expliqué les complications par l'emplacement de serveurs sécuritaires abritant les données (entretien 8Hxvi). Selon lui, la sécurité des serveurs représente un enjeu énorme en raison de la confidentialité. Les pays ne disposent pas tous des mêmes cadres juridiques liés à la confidentialité des données. Cette situation peut poser des défis pour accéder aux données selon l'emplacement des serveurs. D'après ce chercheur, un bon nombre d'universités au Canada entreposent leurs données dans des serveurs aux États-Unis où les mécanismes juridiques ne sont pas les mêmes que ceux au Canada. Ces données sont entreposées dans des serveurs régis par les juridictions et les règles des États-Unis, ce qui fait perdre au Canada certains droits par rapport à la gestion de ses données selon ce chercheur.

De plus, ce chercheur a discuté de l'exemple de la COVID-19 à Taiwan. Il semble que ce pays ait vraiment réussi à contrôler la situation de COVID-19, notamment grâce à une décision exécutive du gouvernement consistant à relier les informations personnelles sur les voyages contenues dans les fichiers électroniques aux informations sur la santé. C'est excellent en ce qui a trait à la santé, mais il faut avoir une discussion éthique avec des experts sur les droits de la personne et la vie privée selon ce chercheur. Il faudra donc faire appel à des experts dans le domaine de l'éthique, car même les dossiers médicaux électroniques posent un grand nombre de problèmes d'après ce

chercheur. Ces préoccupations sur les enjeux éthiques ont été exprimées par tous les participants à la recherche de terrain.

Un chercheur en épidémiologie clinique soutient d'ailleurs que la propriété privée n'est pas très bien conceptualisée lorsqu'il est question de l'utilisation des dossiers médicaux des patients en contexte de recherche (entretien 6Fviii). Ce chercheur affirme aussi que la législation et l'éthique en matière de protection de la vie privée sont très lentes à s'adapter à la nouvelle réalité numérique des collectes et des traitements des données non traditionnelles. Le grand problème réside dans le fait que ces données ont une valeur marchande d'après ce chercheur.

« I think that we are going to have more place and secure server locations and by secure this is a big issue I think because of confidentiality: who is housing these things, security systems around these things ... I know that there are quite number of universities and places in Canada that house all their data in the US and the legal mechanism of that are quite different because the data kept on there is legally under their jurisdiction and under their rules. So, we lose I think some of our rights and a lot of issues around that. Security maintained in Canada I think that there is going to be a lot more discussions about ethics (more discussions are needed on that subject), and legislation and what is appropriate and what can we take and what can we link and cannot link. the Covid example is a good one. In Taiwan, they were really able to control the Covid situation and part of it was the government made executive decision consisting of linking personal travel information from the electronic files with health information. this is excellent from a health perspective and try to contain this, but you have to have that ethical discussion on human rights and privacy and this type of things. These are ethical debates. So, people that are experts in that field are going to be required because it is coming up even with medical health records electronically there is a huge number of issues around those, and people are trying to figure out. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« The reason why is because the private legislation for research use of health records of someone is like is the health record of someone the private ownership of the individuals is not very well conceptualized. The constraints of privacy legislation and ethics which are very slow to catch up with the new digital reality of capture and so forth. The big issue is that these data are sold. » (Chercheur - épidémiologie clinique - entretien 6Fviii)

Ces propos sur les cadres juridiques liés aux données montrent la vivacité des méthodes de recherche et des données étant donné que celles-ci impliquent des enjeux plus larges que des questions techniques. La pratique de l'épidémiologie numérique impliquerait donc des méthodes de recherche et des données qui ne seraient pas de simples boîtes à outils neutres parce que celles-

ci nécessitent des efforts importants pour légiférer les usages et l'accès aux données et que ces efforts suscitent des opinions et des investissements. Comme proposé par Savage (2013), nous suggérons de concevoir les méthodes de recherche et les données de la pratique de l'épidémiologie numérique en tant qu'objets d'études au lieu de les considérer comme quelque chose d'ennuyeux relevant simplement de la technique. Nous inspirant de la vie sociale des méthodes, nous soutenons que les sites de réseaux sociaux, les algorithmes, les moteurs de recherche Internet, les bases de données représentent des dispositifs capables d'influencer les relations sociales et constituent des objets de recherche intéressants. Ainsi, les méthodes de recherche et les données deviennent à la fois l'objet des intérêts et des véhicules indispensables aux chercheurs pour mener à bien leurs travaux de recherche (Savage, 2013).

De plus, l'application de méthodes et l'utilisation de données impliquent des conséquences et créent des associations dictées par des intérêts et des objectifs tels que la sécurité publique, le marketing ciblé, le contrôle des maladies et la génomique des populations (Law *et al.*, 2011). Ce contexte nécessite des réflexions critiques sur les méthodes et les données parce que celles-ci ne reflètent pas nécessairement les préoccupations des milieux universitaires ou des communautés de citoyens. Les défenseurs des méthodes de recherche et des données auraient des intérêts divers et poursuivraient des objectifs différents (Law *et al.*, 2011) que ceux de la santé publique. La pratique de l'épidémiologie numérique peut répondre à des intérêts divers et ainsi, susciter des questions sur la légalité de l'usage et de l'accès aux données numériques.

Les caractéristiques de cette nouvelle pratique révèlent que les connaissances scientifiques et les technologies incorporent et sont aussi incorporées par les pratiques sociales, les identités, les normes, les conventions, les discours, les instruments et les institutions. Dans ce sens, la pratique de l'épidémiologie numérique est conditionnée par des enjeux scientifiques et technologiques, mais aussi par des pratiques sociales comprenant entre autres des normes et des institutions.

6.2.2.2 Les ressources technologiques

Les ressources technologiques ont été mentionnées brièvement à quelques reprises lors des entretiens avec les participants à la recherche de terrain. Le manque de ressources technologiques

peut constituer une barrière à la légitimité de la pratique de l'épidémiologie numérique puisque ces ressources sont essentielles pour collecter, traiter et analyser les données numériques de manière adéquate. Si les ressources technologiques requises pour pratiquer l'épidémiologie numérique ne sont pas suffisantes, il sera difficile de reconnaître cette nouvelle pratique au sein des universités puisque les chercheurs ne seront pas équipés pour mettre en œuvre des projets de recherche fondés sur cette nouvelle pratique.

Selon les propos entendus lors des entretiens, les ressources technologiques désignent les capacités techniques des équipements informatiques. Il est attendu que les capacités techniques requises pour pratiquer l'épidémiologie numérique soient supérieures à celles des équipements courants en épidémiologie. L'un des chercheurs en épidémiologie sociale a notamment souligné que les équipements et les technologies s'améliorent rapidement et qu'il faudra des investissements sur ce plan avec la pratique de l'épidémiologie numérique qui nécessite de traiter de grands volumes de données (entretien 8Hxvi). Pour un chercheur en épidémiologie environnementale, les capacités des serveurs et la vitesse des ordinateurs doivent être suffisantes (entretien 5Gxxv). Ce chercheur souligne aussi que les ordinateurs requis pour travailler avec ce type de données sont dispendieux.

Dans le même sens, un chercheur en épidémiologie sociale pense que les systèmes informatiques typiques ne sont pas adéquats pour pratiquer l'épidémiologie numérique (entretien 6Pvi). Un chercheur en épidémiologie clinique exprime le même point de vue en affirmant que la pratique de l'épidémiologie numérique nécessite des équipes avec beaucoup de connaissances pour assurer le support des technologies de l'information et qu'il faut des serveurs avec des capacités suffisantes (entretien 7Mxxi). Les propos d'un autre chercheur en épidémiologie environnementale vont dans le même sens étant donné que celui-ci fait allusion aux ressources matérielles comprenant des ordinateurs de haute performance, des capacités de stockage et des serveurs (entretien 5Cxxvi).

« Equipment and technology are improving so quickly these things will have to be invested in. » (Chercheur - épidémiologie sociale - entretien 8Hxvi)

« You need sufficient capacity with servers and computers and speeds of computers and all of that. Any of our students who are working with those data they have very expensive computers. » (Chercheur - épidémiologie environnementale - entretien 5Gxxv)

« You might need a bit more intensive computer equipment because if you are pulling out millions of Twitter data you will need more in terms of equipment. I don't think that our typical computer systems here will be able to run that. » (Chercheur - épidémiologie sociale - entretien 6Pvi)

« It would take extremely knowledgeable teams of IT support. Plus, the server's capacity to handle it. » (Chercheur - épidémiologie clinique - entretien 7Mxxi)

« Ben il y a les ressources humaines et les ressources matérielles donc je ne sais pas laquelle - Matérielles - Des ordinateurs haute performance, des capacités de stockages aussi, des serveurs, ce genre de choses-là. (Chercheur - épidémiologie environnementale - entretien 5Cxxvi)

Ces propos sur les ressources technologiques montrent que les barrières à la pratique de l'épidémiologie numérique peuvent comprendre des enjeux technologiques en plus des enjeux sociopolitiques. Ce phénomène peut illustrer la symétrie dans la coproduction des connaissances comme proposé par Jasanoff (2004). Selon Jasanoff (2004), la coproduction serait symétrique puisqu'elle implique des dimensions sociales, épistémiques, technologiques et matérielles. Les enjeux liés aux ressources technologiques rappellent les corrélations matérielles lors de la pratique de l'épidémiologie numérique. Cette nouvelle pratique nécessite des ressources matérielles comme des technologies assez puissantes pour traiter les données numériques en plus des compétences d'analyse des chercheurs pour analyser ces données.

6.2.2.3 Le manque de confiance envers les détenteurs de données et les gouvernements

Une dernière barrière constituant une entrave à la légitimité de la pratique de l'épidémiologie numérique serait le manque de confiance envers les détenteurs de données et les gouvernements. Certains travaux publiés dans la littérature expriment un haut niveau de préoccupation sur la confiance en affirmant que toute donnée sur des sujets humains soulève inévitablement des problèmes de confidentialité, et que les risques réels d'abus de ces données s'avèrent difficiles à quantifier (*Nature*, 2007) :

« Any data on human subjects inevitably raise privacy issues, and the real risks of abuse of such data are difficult to quantify. » (*Nature*, 2007)

D'après la revue *Nature*, il existe des écarts importants entre la législation de la plupart des gouvernements et les phénomènes actuels sur Internet. La confiance du public pourra être gagnée uniquement en redonnant un sentiment de contrôle et de propriété aux personnes fournissant leurs données sur internet (*Nature*, 2007). Et ce n'est qu'à partir de ce moment que les gens se sentiront plus à l'aise à l'idée que des études soient faites sur les détails intimes des résultats de leurs recherches anonymes, leurs courriers électroniques, leurs déplacements et leurs appels téléphoniques (*Nature*, 2007).

Lors de la recherche de terrain, deux participants en épidémiologie sociale (dont un pratiquant de l'épidémiologie numérique) ont discuté de la confiance à l'égard du traçage numérique. D'après ces chercheurs, la confiance des gens est essentielle pour assurer un traçage efficace des contacts sociaux afin d'aider la santé publique (entretien 9Qxxiv, entretien 8Sxx). L'un de ces chercheurs a souligné que la confiance est la monnaie en santé publique (entretien 9Qxxiv). En effet, si une communauté est confiante envers la santé publique ou envers les dirigeants, beaucoup de choses sont possibles selon ce chercheur. Si la confiance est perdue, les mesures prises se révéleront totalement inefficaces, peu importe l'argent disponible. Pour ce chercheur, il est primordial que les gens sachent que le traçage numérique des citoyens se fait parce que c'est important pour la santé et non pas parce que les gouvernements veulent savoir où sont les personnes. S'il y a fausse route, les conséquences négatives seront incontrôlables selon ce chercheur, qui estime que la confiance du public est la chose la plus importante à l'ère numérique.

« A lot of the discussion that we had about using the app for contact tracing was around the fact that in public health, trust is everything. Trust is the currency that we use. If a community trusts you in public health or individuals' decision makers trusted in public health, then many things are possible. But once you lost that trust, it doesn't matter how much money you will have to do what you want to do, it doesn't matter how much power you have, if you don't have trust, then you will be completely ineffective. And this question about 'are you tracking me because it is important to my health or are you tracking me because you want to know where I am, who I am seeing and where am I going?'. If we get that wrong, it is going to be uncontrolled damage and so if I am going to highlight the single most important thing that we have to get right in digital, it will be that. » (Chercheur - épidémiologie sociale - entretien 9Qxxiv)

« A classic example is Covid-19. Now almost every government wants you to be on a contact tracing app but why aren't all people joining the contact tracing app? Because they don't trust

to be on the app but if you do gain their trust if you engage them as a new scientist, they will download it and that's what we are doing at this point of time with the community of health.
» (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

Le manque de confiance envers les détenteurs de données et les gouvernements peut être dû aux perceptions des gens vis-à-vis de leurs intérêts comme expliqué précédemment. Comme proposé par Law *et al.* (2011), les méthodes et les données seraient utilisées pour des intérêts et des objectifs très divers tels que la sécurité publique, le marketing ciblé, le contrôle des maladies et la génomique des populations. Le manque de confiance envers les détenteurs de données et les gouvernements serait lié aux incertitudes sur les intérêts poursuivis par ces acteurs. Par conséquent, l'adoption de la pratique de l'épidémiologie numérique par les chercheurs en épidémiologie et l'engagement des participants aux recherches fondées sur cette nouvelle pratique pourraient être freinés si les intérêts des détenteurs de données et des gouvernements paraissent obscurs aux yeux des citoyens.

En somme, la complexité des cadres juridiques, l'insuffisance des ressources technologiques et le manque de confiance des gens envers les détenteurs de données et les gouvernements constituent les barrières à la légitimité de la pratique de l'épidémiologie numérique identifiées lors des entretiens avec les chercheurs en épidémiologie ayant pris part à la recherche de terrain. Les propos exprimés sur les barrières à la légitimité de la pratique de l'épidémiologie numérique illustrent la vivacité des méthodes de recherche et des données, l'importance d'adopter une perspective critique vis-à-vis de celles-ci et la symétrie de la coproduction des connaissances.

6.2.3 Les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique

Les conséquences individuelles et collectives peuvent être définies comme des réalités collatérales liées à des dispositifs méthodologiques dont les effets viennent en cours de route, et ce, de manière accidentelle lors des recherches (Rubio et Baert, 2012). Law et Ruppert (2013) soulignent d'ailleurs que ce que font les dispositifs méthodologiques n'est pas nécessairement écrit sur leur emballage, c'est-à-dire que ces dispositifs font des choses qui ne sont pas toujours prévues. Et les conséquences de ce que font ou peuvent faire ces choses sont souvent l'objet de contestations sociales et politiques (Law et Ruppert, 2013). Ceci suggère que toute analyse sociale des dispositifs

méthodologiques sera marquée par un certain niveau de susceptibilité critique (Law et Ruppert, 2013). Par exemple, les caméras de sécurité peuvent porter atteinte à la vie privée et en conséquence de ce phénomène, des contestations sociales ou politiques peuvent survenir.

De plus, la pratique de l'épidémiologie numérique implique l'utilisation de données générées par les gens sur Internet comme celles des médias sociaux, des moteurs de recherche et des téléphones cellulaires. L'utilisation de ces données peut entraîner des conséquences comme des risques informationnels auprès des citoyens, c'est-à-dire des conséquences individuelles et collectives importantes à comprendre pour évaluer cette nouvelle pratique de manière critique. Nous analyserons donc les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique telles que vues par des chercheurs en épidémiologie, c'est-à-dire les risques vis-à-vis de la sécurité et de la confidentialité des données numériques et les pouvoirs de grandes entités privées détentrices des données numériques.

6.2.3.1 Les risques vis-à-vis de la sécurité et de la confidentialité des données numériques

Les risques vis-à-vis de la sécurité et de la confidentialité des données numériques ont été très souvent mentionnés par les participants lors des discussions sur les enjeux éthiques liés à la pratique de l'épidémiologie numérique. Selon plusieurs participants, les questions liées à l'éthique se trouvent être un des enjeux les plus urgents à régler afin d'assurer la légitimité de la pratique de l'épidémiologie numérique.

Ce résultat peut indiquer une grande sensibilité au sein de chercheurs en épidémiologie au fait que ce n'est pas parce que des données sont accessibles que leur utilisation est nécessairement éthique (Boyd et Crawford, 2012). Il semble que les implications éthiques en science des données sont très peu connues et que beaucoup de questions demeurent sans réponse quant aux caractéristiques d'un consentement éclairé (Boyd et Crawford, 2012). Selon Boyd et Crawford (2012), il peut être déraisonnable de demander aux chercheurs d'obtenir le consentement de chaque personne publiant un tweet, mais il est problématique pour les chercheurs de justifier leurs actions comme étant éthiques simplement parce que les données sont accessibles. En effet, ce n'est pas parce que le contenu est accessible au public qu'il est destiné à être consommé par n'importe qui. Il n'est donc

pas surprenant de constater qu'une grande partie des participants à notre recherche se montre sensible à la sécurité et à la protection des données dans le cadre de la pratique de l'épidémiologie numérique.

Par exemple, un chercheur en épidémiologie clinique a affirmé que les processus éthiques devront être mis à jour et qu'il ne sait pas comment les données numériques peuvent être entièrement confidentielles (entretien 7Mxxi). Un chercheur en épidémiologie sociale a souligné quant à lui la nécessité d'un consentement explicite des participants dans le cadre des recherches (entretien 5Uxxxvi). Selon ce chercheur, tout dépend de la manière dont on peut accéder aux données et quelles sont les données utilisées. Il peut aussi y avoir un inconfort sur le plan éthique avec l'agrégation de données au niveau populationnel d'après ce chercheur parce que, selon lui, les utilisateurs des données numériques n'ont pas nécessairement obtenu le consentement éclairé pour utiliser les données produites.

De plus, un autre chercheur en épidémiologie sociale affilié à un institut de recherche en humanités médicales critiques et en éthique soutient qu'il faut bâtir de fortes relations éthiques avec les gens lors des recherches et qu'il faut aussi des cadres éthiques (entretien 6Bii). Ce chercheur propose d'inclure des gens des communautés sur les comités éthiques afin d'élaborer une étude ne portant pas atteinte aux populations étudiées. D'un autre côté, des situations très complexes concernant les permissions éthiques peuvent également survenir, selon ce chercheur. Par exemple, les permissions éthiques peuvent rendre les choses très compliquées pour des ingénieurs informaticiens devant obtenir les données de santé dans le but de construire une technologie d'apprentissage automatique capable de prédire les surdosages de drogues en vue de prévenir des décès.

D'après l'expérience d'un autre chercheur en épidémiologie sociale, il existerait peu de différences entre les données traditionnelles et les données numériques outre les problèmes éthiques (entretien 3Exxxiii). Ce participant est d'avis que les chercheurs sont confrontés à des problèmes éthiques plus importants avec les données numériques qu'avec les données traditionnelles et plus particulièrement lorsqu'ils commencent à lier des ensembles de données. D'ailleurs, il y aurait une section sur l'établissement de liens entre les ensembles de données lorsque les chercheurs remplissent une demande de recherche au sein de l'université de ce participant.

En outre, les propos d'un chercheur en épidémiologie sociale ont fait ressortir la nécessité que les gens soient conscients de l'utilisation potentielle des données numériques (entretien 6Pvi). Ce chercheur n'est pas certain que ce soit le cas lorsque les gens donnent leur avis à grande échelle sur Internet. Il soutient donc qu'il faut agir avec prudence en utilisant les données numériques puisque certaines d'entre elles sont susceptibles de divulguer des informations pouvant nuire aux personnes concernées par ces données.

« The most pressing issue would be confidentiality. I think that ethical processes will have to be updated. I don't know how we can make these data fully confidential. » (Chercheur - épidémiologie clinique - entretien 7Mxxi)

« Ben là il y a des enjeux éthiques importants là-dedans. Je pense qu'il faut que le consentement soit explicite. Donc ça dépend de comment on a accès aux données en fait ça dépend de quelles données on parle... Aussi à l'échelle populationnelle, l'agrégation de données, mais là, c'est là où j'ai un petit inconfort avec l'enjeu éthique c'est-à-dire que je ne pense pas que les utilisateurs de données font un consentement éclairé sur l'utilisation des données qu'ils produisent. » (Chercheur - épidémiologie sociale - entretien 5Uxxxvi)

« So, the very first one is ethics. And actually, this is one of the things why I am here temporary in Pennsylvania University, I am affiliated with the Critical Medical Humanities and the ethics institute... Any way my idea is that we need to build very strong ethical relationships with the people... think that ethics is part of that, and I think you can always ask people from a community to come and sit on your ethics board, to help you to design a study which will not harm people or at least to the best of your knowledge... We need the ethics framework. Ethics is very a big issue: we have problems with ethics permissions and on the other side, the worst is happening because it is too complicated for computer engineers to get the health data to build a machine learning technology to predict overdoses... and at the same thing, people are dying from that... it is kind of a problem for me. » (Chercheur - épidémiologie sociale - entretien 6Bii)

« I personally don't see a difference for digital data vs traditional data. However, there is a lot of ethical problems that you have to face with digital data and when you start linking data sets. So, at Dalhousie University for example, when you are filling out a research application, there is a section about linking datasets. » (Chercheur - épidémiologie sociale - entretien 3Exxxiii)

« I think ethical issues. And I didn't really have the chance to do that given that I don't know that this is. But that would be my first thing. Thinking about ethics to make sure that a) people should be aware that this data is potential being used. I am not sure if is the case where people are posing their opinions widely. I think that they should be made aware. And also, being careful in terms of harms of people. You don't want to harm people by pulling some of their data and finding certain things that could potentially be back to harm them. I

think we need to be careful about that in that type of work as well. I would say that people need to be careful about ethical issues. » (Chercheur - épidémiologie sociale - entretien 6Pvi)

6.2.3.2 Les pouvoirs de grandes entités privées détentrices des données numériques

Enfin, nous avons constaté que les rapports de pouvoir en faveur des grandes entités privées détentrices des données numériques constituent des réalités collatéralesconséquences collectives de la pratique généralisée de l'épidémiologie numérique identifiées par des participants à la recherche de terrain. Certains participants envisagent notamment que l'utilisation de données numériques dans le cadre des recherches peut renforcer le pouvoir de grandes instances privées comme Google. Ces instances sont propriétaires des données numériques et elles exercent le contrôle sur la publication des informations sur Internet. Quelques participants à la recherche de terrain ont clairement exprimé le fait qu'ils n'aiment pas l'idée que de grandes instances privées puissent contrôler les données et encore moins en contexte de recherche sur la santé publique. De plus, ces participants éprouvent de la difficulté à faire confiance aux données des grandes instances privées comme celles de Google.

Par exemple, un chercheur en épidémiologie sociale a mentionné que le système de surveillance *Google Flu Trends (GFT)* pourrait s'avérer utile pour la santé publique (entretien 6Jv). Cependant, il n'apprécie pas le fait que de grandes multinationales se retrouvent avec autant de pouvoir et il est incertain sur les motifs de ces instances privées les poussant à contribuer à des recherches. Un autre chercheur en épidémiologie sociale pense que les gouvernements ou les grandes entreprises pourraient orienter des conversations sur Internet en générant des milliers de tweets fictifs amenant vers des mensonges sur la réalité des événements (entretien 3Exxxiii).

En outre, un chercheur en épidémiologie sociale pratiquant l'épidémiologie numérique a expliqué qu'il a de la difficulté à faire confiance aux données de Google, car cette instance privée est propriétaire des données (entretien 8Sxx). Il considère que cette situation peut créer des problèmes pour comprendre le monde réel. Il pense aussi qu'il sera toujours sceptique vis-à-vis de ce type de données. Ce chercheur a également souligné qu'il aurait confiance envers les données de Google générées par les citoyens engagés directement avec le laboratoire de recherche qu'il dirige.

« What a system like that (Google Flu Trends) could do is give you a signal a couple of weeks ahead of time to get ready for an epidemic. I think we have been trying to do such a thing in our hospital - some syndromic surveillance for a while but I hate that idea of a big multinational corporation having so much power. Why also they do it... » (Chercheur - épidémiologie sociale - entretien 6Jv)

« We live in a world where research is perhaps trying its best, but they can't possibly compete with governments and corporations who want to use social media for whatever purpose they want. They could orient conversations by creating thousands of fake individuals Tweets and populating some sort of trends and then you see some stuff like these really niche positions overstated massive way on social medias because some governments want to cause something in some countries and make a bunch of noise on the internet. So, I think that social media information can be very noisy and convey lies. » (Chercheur - épidémiologie sociale - entretien 3Exxxiii)

« I think the potential is very high (with Google Trends), the difficulty is how much do you trust Google? That is a difficult thing because they own the data. To me that is the difficult issue there in understanding what is happening. I find that I will always be a bit skeptical right? And that's a good thing. What I would trust most in Google data is data coming from my own citizens engaging with us directly. So that would be my answer to that. I am not saying I don't have confidence in it, but I would be a bit skeptical just based on the fact that these are coming from big tech. » (Chercheur - épidémiologie sociale - pratiquant de l'ÉN - entretien 8Sxx)

Cette situation de pouvoir entre les mains des grandes instances comme Google concerne la politique des données telle que développée par Ruppert *et al.* (2017). Ces chercheurs s'intéressent notamment à la manière dont les données génèrent de nouvelles formes de relations de pouvoir et de politique à des échelles différentes et interconnectées. Quelques participants à la recherche de terrain ont communiqué leurs préoccupations concernant le pouvoir de grandes instances privées grandement influencé par la propriété des données numériques. Ce résultat révèle que les nouvelles formes de relations de pouvoir et de politique générées par les données peuvent être fondées sur la propriété des données numériques. La pratique de l'épidémiologie numérique pourrait alors engendrer une concentration des pouvoirs auprès des grandes instances propriétaires des données numériques.

En somme, les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique identifiées lors de notre recherche sont les risques vis-à-vis de la sécurité et de la

confidentialité des données numériques ainsi que les rapports de pouvoir en faveur des grandes entités privées détentrices de ces données. Ces constats constituent une preuve que la pratique de l'épidémiologie numérique implique des méthodes de recherche vivantes, c'est-à-dire capables d'agir sur les relations sociales, notamment en portant atteinte à certaines populations par l'utilisation des données sur leur état de santé ou en contribuant à accroître le pouvoir auprès de grandes compagnies privées, lesquelles détiennent les données numériques. En s'inspirant de Jasanoff (2004), cette analyse fait ressortir des dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales constituant la pratique de l'épidémiologie numérique.

De plus, ce travail d'analyse met en lumière les enjeux politiques des données tels que proposés par Ruppert *et al.* (2017). Les données sont des entités capables de reconfigurer les relations entre les gouvernements, les citoyens et les industries. Au lieu de conceptualiser les données comme des représentations inertes, Ruppert *et al.* (2017) soutiennent que les données ont des pouvoirs concrets influençant grandement la vie politique et que ces objets sont investis de certains pouvoirs, influences et rationalités. La pratique de l'épidémiologie numérique est grandement influencée par ce contexte sociopolitique des données. Par exemple, l'accès aux données peut être limité pour les chercheurs parce que celles-ci sont la propriété privée de grandes sociétés comme Google ou de compagnies de téléphones cellulaires. Il peut aussi y avoir des enjeux de confiance vis-à-vis des acteurs détenant les données étant donné que ces grands acteurs contrôlent entièrement l'accès aux données numériques.

Il ne faut pas non plus oublier les incertitudes sur le fonctionnement de certaines technologies telles que des moteurs de recherche Internet dont les données peuvent être utilisées comme dispositifs méthodologiques pour des recherches. Les algorithmes derrière ces dispositifs méthodologiques peuvent sélectionner certaines informations et en exclure d'autres en fonction des intérêts des sociétés propriétaires des données et des technologies numériques. Ce contrôle exercé sur les données peut susciter des doutes envers la légitimité de la pratique de l'épidémiologie numérique au sein de chercheurs en épidémiologie parce qu'ils sont incertains des intérêts poursuivis par de grandes sociétés comme Google. Donc, un dispositif méthodologique comme le moteur de recherche Google est matériel parce qu'il possède des caractéristiques technologiques, mais il

s'avère aussi hétérogène parce qu'il implique également des personnes comme des utilisateurs et des communautés (Law et Ruppert, 2013) ainsi que de grandes sociétés privées pouvant augmenter leurs pouvoirs avec la propriété des données numériques.

Pour résumer nos analyses sur la légitimité de la pratique de l'épidémiologie numérique, rappelons que la pratique de l'épidémiologie numérique semble peu adoptée et peu diffusée sur le plan institutionnel. Or, cette nouvelle pratique serait nécessaire et importante en milieu universitaire. Cependant, peu d'actions seraient entreprises en vue d'intégrer cette pratique dans les universités, outre quelques initiatives lancées de manière sporadique dans certains départements d'épidémiologie.

Notre recherche montre aussi que les cadres juridiques, les ressources technologiques et le manque de confiance des gens envers les détenteurs de données et les gouvernements constituent des barrières à la légitimité de la pratique de l'épidémiologie numérique. Les propos exprimés sur les barrières à la légitimité de la pratique de l'épidémiologie numérique illustrent la vivacité des méthodes de recherche et des données, l'importance d'adopter une perspective critique vis-à-vis de celles-ci et la symétrie de la coproduction des connaissances.

Enfin, les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique comprennent les risques vis-à-vis de la sécurité et de la confidentialité des données numériques ainsi que les rapports de pouvoir en faveur des grandes entités privées détenant ces données. La pratique de l'épidémiologie numérique est grandement influencée par le contexte sociopolitique des données. Par conséquent, la pratique de l'épidémiologie numérique implique des méthodes de recherche vivantes, c'est-à-dire capables d'agir sur les relations sociales, notamment en portant atteinte à certaines populations par l'utilisation des données sur leur état de santé ou en contribuant à accroître le pouvoir auprès de grandes compagnies privées propriétaires des données numériques. Nos analyses font ressortir des dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales (Jasanoff, 2004) constituant la pratique de l'épidémiologie numérique.

En outre, des doutes subsisteraient envers la légitimité de la pratique de l'épidémiologie numérique chez certains chercheurs en épidémiologie étant donné qu'ils sont incertains des intérêts poursuivis par les grands acteurs détenteurs des données. Un dispositif méthodologique comme le moteur de recherche Google serait donc matériel parce qu'il possède des caractéristiques technologiques, mais aussi hétérogènes parce qu'il implique également des personnes comme des utilisateurs et des communautés (Law et Ruppert, 2013) ainsi que de grandes sociétés privées susceptibles d'augmenter leurs pouvoirs avec la propriété des données numériques.

6.3 La conclusion sur les enjeux sociopolitiques

Dans ce chapitre, l'objectif était de connaître les enjeux sociopolitiques liés au contexte d'émergence de la pratique de l'épidémiologie numérique. Cette recherche révèle que les enjeux sociopolitiques de la pratique de l'épidémiologie numérique sont particulièrement marqués par l'expertise et la légitimité. D'après nos constats, les enjeux d'expertise associés à la pratique de l'épidémiologie numérique sont les nouvelles compétences et les spécialités envisagées, la formation des chercheurs en épidémiologie et l'interdisciplinarité. Pour cerner la légitimité de la pratique de l'épidémiologie numérique, nous avons analysé la reconnaissance institutionnelle, les barrières à la légitimité ainsi que les conséquences individuelles et collectives de cette nouvelle pratique.

La recherche effectuée dans le cadre de cette thèse révèle que la pratique de l'épidémiologie numérique comprend plusieurs enjeux sociopolitiques en plus des enjeux épistémiques discutés dans le chapitre précédent. Ces analyses ont fourni un survol des préoccupations de chercheurs en épidémiologie vis-à-vis de cette nouvelle pratique sous l'angle sociopolitique. Pour certains enjeux, il se dégage un certain consensus dans les points de vue des chercheurs en épidémiologie comme dans le cas de l'importance de la discipline de l'informatique pour pratiquer l'épidémiologie numérique. Cependant, pour d'autres enjeux comme la reconnaissance institutionnelle de cette nouvelle pratique et les besoins de formation des chercheurs pour pratiquer l'épidémiologie numérique, les chercheurs en épidémiologie n'envisagent pas nécessairement les mêmes perspectives. Des recherches plus approfondies portant sur des enjeux sociopolitiques particuliers seraient alors pertinentes pour mieux comprendre les différentes facettes des points de vue des

chercheurs en épidémiologie afin de prendre les bonnes décisions sur la pratique de l'épidémiologie numérique. Le prochain chapitre sera consacré à la discussion des résultats obtenus lors de cette recherche doctorale.

CHAPITRE VII

LA DISCUSSION

Ce chapitre de discussion s'appuie sur les analyses des chapitres précédents concernant les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. Le but de ce chapitre est d'expliquer les contextes de la coproduction des connaissances dans le cas de l'épistémologie numérique et, plus largement, à l'ère numérique. Pour ce faire, nous commencerons par un rappel des principaux résultats obtenus à la suite de l'analyse des points de vue des chercheurs en épidémiologie ayant participé à la recherche de terrain. Ensuite, notre attention sera centrée sur l'écosystème de la pratique de l'épidémiologie numérique. Nous poursuivrons en abordant les contextes de la coproduction des connaissances à l'ère numérique, puis nous terminerons en discutant la question de la numérisation des disciplines.

7.1 Les caractéristiques et les enjeux épistémiques et sociopolitiques de la pratique de l'épidémiologie numérique

La recherche effectuée dans le cadre de cette thèse s'appuie sur des entretiens semi-structurés réalisés auprès de 36 chercheurs canadiens en épidémiologie. Les résultats obtenus ont été analysés en fonction de l'expérience de la pratique de l'épidémiologie numérique, de la sous-discipline de l'épidémiologie de ces chercheurs (épidémiologie sociale, épidémiologie clinique et épidémiologie environnementale) et des points de vue exprimés sur les différents aspects traités lors de la recherche de terrain. Rappelons que le but de cette recherche est de documenter et de comprendre les enjeux épistémiques et sociopolitiques qui façonnent l'émergence de la pratique de l'épidémiologie numérique. Pour atteindre ce résultat, nous nous sommes fixé deux objectifs, à savoir 1) comprendre ce que signifie la pratique de l'épidémiologie numérique et 2) identifier les enjeux épistémiques et sociopolitiques qui accompagnent l'émergence de la pratique de l'épidémiologie numérique.

Au niveau de ses caractéristiques, la revue de la littérature sur la pratique de l'épidémiologie numérique montre que celle-ci peut être définie comme une épidémiologie utilisant des données

numériques générées en dehors du système de santé publique. Sa particularité est que les données qu'elle utilise ont été générées dans un but autre que de faire de l'épidémiologie (Salathé, 2018). La recherche effectuée dans le cadre de cette thèse témoigne que l'univers conceptuel de la pratique de l'épidémiologie numérique est marqué par un contexte de nouveauté, où les concepts, outils et usages ne sont pas clairs et peuvent être compris différemment d'une personne à l'autre. Le vocabulaire utilisé pour désigner ces nouvelles pratiques évoque souvent les données et les technologies numériques. Quant au terme « numérique », il peut susciter de la confusion dans la compréhension des nouvelles pratiques de recherche issues de la mise en données de la société. Cependant, l'expérience de ces nouvelles pratiques de recherche peut aider à mieux comprendre le travail associé à ces nouvelles pratiques et aussi à saisir leur raison d'être. Il y aurait aussi des incertitudes sur la nouveauté de la pratique de l'épidémiologie numérique et sur les différences entre cette nouvelle pratique et l'épidémiologie traditionnelle. Ces incertitudes semblent liées au fait que l'utilisation de données en provenance de l'extérieur des systèmes de santé et de données non structurées comme des textes s'observe autant dans la pratique de l'épidémiologie numérique que dans la pratique de l'épidémiologie traditionnelle. De plus, l'épidémiologie environnementale utilise des données générées en dehors des systèmes de santé depuis longtemps, comme des données générées par des satellites lors d'études environnementales.

La pratique de l'épidémiologie numérique en serait à ses débuts dans son processus d'adoption et de diffusion au sein des chercheurs canadiens en épidémiologie selon les constats établis lors de notre recherche. Elle l'est aussi au sein des programmes d'études en épidémiologie au sein des universités canadiennes⁸. Seulement trois chercheurs canadiens en épidémiologie ont fait l'expérience de cette nouvelle pratique parmi les 36 participants. Sur ces trois chercheurs, deux pratiquaient l'épidémiologie sociale, et un l'épidémiologie environnementale. Lors des entrevues, nous avons constaté des réserves de la part des 36 chercheurs envers la pratique de l'épidémiologie numérique. La plupart de ces chercheurs semblaient ouverts à pratiquer l'épidémiologie numérique, mais démontraient une attitude prudente face à cette nouvelle pratique, qu'ils percevaient comme une façon de faire de l'épidémiologie présentant du potentiel, mais aussi des lacunes.

⁸ Sur la base du groupe de chercheurs en épidémiologie ayant participé à la recherche de terrain.

D'après les constats établis lors de la recherche de terrain, sur le plan épistémique, les données générées par les téléphones cellulaires seraient le dispositif méthodologique ayant le plus haut niveau de confiance chez les 36 chercheurs canadiens en épidémiologie, suivi des données générées par les moteurs de recherche Internet et finalement de celles générées par les médias sociaux. De plus, la pratique de l'épidémiologie numérique semble conditionnée par les aspects humains et sociaux des dispositifs méthodologiques utilisés dans des recherches fondées sur cette nouvelle pratique.

Notre recherche montre que les aspects humains des dispositifs méthodologiques utilisés lors de la pratique de l'épidémiologie numérique concernent les interactions entre les utilisateurs des technologies numériques et les technologies numériques elles-mêmes ainsi que la qualité des informations fournies par les gens utilisant ces technologies. Le niveau d'interaction entre les utilisateurs et les technologies numériques comme les médias sociaux influencerait surtout l'échantillonnage populationnel. En effet, les gens n'utilisant pas ces technologies peuvent être exclus d'une population alors que d'autres peuvent être surreprésentés, car ils sont de grands adeptes de ces technologies. De plus, la qualité des informations fournies par les utilisateurs de ces technologies est difficile à vérifier, ce qui suscite des doutes sur le potentiel réel de la pratique de l'épidémiologie numérique. La recherche de terrain montre que ces difficultés se situent au niveau des signes et des apparences secondaires du dispositif. En utilisant des données générées par les médias sociaux, le chercheur serait plutôt confronté à des biais, à des bruits et à des distorsions à cause de comportements humains (comme l'utilisation intensive ou le refus d'utilisation des médias sociaux). Ces données sont susceptibles d'émettre des signes inappropriés et d'amener des apparences trompeuses. Cette perspective se retrouve autant auprès des chercheurs ayant l'expérience de la pratique de l'épidémiologie numérique que de ceux qui n'ont pas l'expérience d'une telle pratique.

En ce qui concerne les aspects technologiques des dispositifs méthodologiques utilisés dans la pratique de l'épidémiologie numérique, certains chercheurs peuvent concevoir ces dispositifs comme des boîtes noires du fait qu'ils ignorent leur fonctionnement sur le plan technique. Sur la base des propos de plusieurs de nos participants, les données générées par le moteur de recherche Google doivent être utilisées avec prudence à cause du manque de connaissances techniques sur ce

genre de dispositif. Aussi, le fonctionnement des algorithmes des technologies numériques utilisées comme dispositifs méthodologiques demeure obscur aux yeux des chercheurs. Or, les algorithmes jouent un rôle important dans la pratique de l'épidémiologie numérique en raison de l'influence qu'ils ont sur les perceptions des phénomènes étudiés, à travers la manière dont ils traitent les informations communiquées par les utilisateurs des technologies numériques.

L'attention portée aux algorithmes illustre l'idée que les données et les mesures sont contingentes à l'appareil de mesure et que celles-ci sont des représentations de l'appareil plutôt que du monde réel. À cet effet, le fonctionnement du moteur de recherche Google (ou le détecteur) doit être correctement compris par le chercheur parce que c'est à travers cette technologie que le chercheur pourra analyser les tendances dans les recherches effectuées par les gens sur Internet. Ainsi, le moteur de recherche Google est un détecteur agissant comme un dispositif médiateur interposé entre les chercheurs et les phénomènes. Par ailleurs, les données et les mesures dépendent de la conception des technologies et sont des représentations de ces technologies (Knorr-Cetina, 1999). Cette réalité a un impact crucial sur l'adaptation de l'appareillage analytique et interprétatif des données dans les recherches dans le domaine de l'épidémiologie numérique. Par qui, de quelle façon, à quel rythme et dans quels lieux sera mis en œuvre ce processus d'adaptation? Selon nous, ce questionnement permet d'associer l'étude des enjeux épistémiques aux enjeux sociopolitiques de la pratique de l'épidémiologie numérique.

En effet, notre recherche révèle que les enjeux sociopolitiques de la pratique de l'épidémiologie numérique sont particulièrement marqués par l'expertise et la légitimité. Nos analyses montrent que les méthodes de recherche liées à la pratique de l'épidémiologie numérique ne sont pas des boîtes à outils neutres. Au contraire, elles peuvent influencer des pratiques et susciter de nouveaux besoins en matière de compétences en épidémiologie. À nouveau, ce constat renvoie au principe de la symétrie dans la coproduction des connaissances (Jasanoff, 2004), c'est-à-dire l'importance de considérer à la fois les enjeux épistémiques et les enjeux sociaux et politiques pour comprendre comment les connaissances sont produites. En effet, la pratique de l'épidémiologie numérique implique d'une part des enjeux techniques nécessitant des compétences en informatique et en science des données, et d'autre part des enjeux sociopolitiques. Les enjeux sociopolitiques créent des besoins pour des compétences non techniques. On peut citer la collaboration interdisciplinaire,

la protection de la vie privée et la négociation des accords de partage de l'information. Ce contexte illustre l'impératif de l'interdisciplinarité qui s'impose sur la pratique et sur la formation de l'épidémiologie numérique. Comment et dans quelles conditions l'interdisciplinarité se met-elle en place?

Nos analyses indiquent que les cadres juridiques, les ressources technologiques et le manque de confiance envers les détenteurs de données et les gouvernements sont les barrières à la légitimité de la pratique de l'épidémiologie numérique. Les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique incluent les risques liés à la sécurité et à la confidentialité des données numériques ainsi que les rapports de pouvoir en faveur des grandes entités privées détenant ces données. Ces constats révèlent que la pratique de l'épidémiologie numérique implique des méthodes de recherche vivantes, c'est-à-dire capables d'agir sur les relations sociales, notamment en portant atteinte à certaines populations par l'utilisation des données sur leur état de santé ou en contribuant à accroître le pouvoir des grandes compagnies privées détentrices des données numériques. En s'inspirant de Jasanoff (2004), ces analyses font ressortir des dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales constituant la pratique de l'épidémiologie numérique. L'écosystème canadien entourant la pratique de l'épidémiologie numérique ne semble pas proche du point de basculement menant à son adoption à grande échelle. Comme nous l'avons constaté, plusieurs enjeux restent encore à régler autant sur le plan épistémique que sur le plan sociopolitique.

En somme, la recherche effectuée dans le cadre de cette thèse a permis d'apporter des éclaircissements sur les enjeux liés à l'émergence de la pratique de l'épidémiologie numérique. Nous avons cerné les caractéristiques et la diffusion de la pratique de l'épidémiologie numérique, contribuant ainsi à une meilleure compréhension de ce que signifie cette nouvelle pratique. Nous avons aussi identifié les enjeux épistémiques et décelé les enjeux sociopolitiques de cette nouvelle pratique à travers les points de vue des chercheurs canadiens en épidémiologie ayant participé à la recherche de terrain, ce qui nous a permis de cerner les aspects technologiques, épistémiques et humains d'une telle pratique. Enfin, notre recherche contribue à poser des questions importantes quant à l'adaptation de l'appareillage analytique et interprétatif des données dans le contexte de recherches épidémiologiques numériques et sur la façon dont l'interdisciplinarité et la

transdisciplinarité se mettent en place dans le contexte d'émergence de la pratique de l'épidémiologie numérique.

7.2 L'écosystème de la pratique de l'épidémiologie numérique

L'écosystème de la pratique de l'épidémiologie numérique est complexe et peut certainement être différent selon les pays. Pour le Canada, la diversité et la taille de la population canadienne, de même que la variabilité géographique (climat et infrastructures de soins de santé) et les disparités dans l'accès aux technologies (la fracture numérique) sont des éléments propres au contexte canadien pouvant être pertinents à mobiliser dans les analyses sur la pratique de l'épidémiologie numérique afin d'enrichir notre compréhension de cette nouvelle pratique. En effet, la population canadienne a dépassé les 40 millions en 2023 en grande partie à cause du solde migratoire international (Statistique Canada, 2023) amenant ainsi des personnes issues de milieux culturels, linguistiques et socio-économiques différents.

De plus, le Canada est le deuxième pays le plus vaste au monde après la Russie (Larousse, 2023) constitué par plusieurs régions géographiques au bord de l'océan Atlantique, Pacifique et Arctique (RNC, 2022) et pourvu d'une infrastructure de santé publique complexe et décentralisée entre les gouvernements fédéral, provinciaux, territoriaux et autochtones (INSPQ, 2023). Il y a aussi les problèmes d'équité en santé auprès des communautés autochtones canadiennes, démontrés par de nombreuses études (Statistique Canada, 2015; Reading et Wien, 2009; Crowshoe *et al.*, 2021) soulignant entre autres les disparités entre les populations autochtones et non-autochtones pour l'accès aux soins de santé. La pratique de l'épidémiologie numérique peut aider à comprendre ces enjeux afin d'adapter les interventions de santé publique.

Cependant, cette nouvelle pratique fait face à des défis, notamment l'accès aux technologies numériques parce que des communautés rurales et autochtones seraient encore touchées par la fracture numérique (Sénat du Canada, 2023). Il est essentiel de comprendre la fracture numérique et son impact sur la collecte de données et l'accès aux soins de santé. Notre recherche révèle que l'écosystème de la pratique de l'épidémiologie inclut des institutions, des acteurs et des technologies, des enjeux de confiance au sein des chercheurs ainsi que des tensions.

7.2.1 Les institutions, les acteurs et les technologies

La littérature révèle que l'écosystème de la pratique de l'épidémiologie numérique comprend des chercheurs de diverses disciplines en milieu universitaire et au sein d'organismes internationaux comme l'OMS et le RKI (Hill-Cawthorne et Sorrell, 2016; Makri, 2017). Cet écosystème inclut aussi des acteurs non experts à travers la science citoyenne (Hamer *et al.*, 2018) et de grandes entités privées comme Google (Roche *et al.*, 2017; Fung *et al.*, 2013; Cheng *et al.*, 2021; Sato *et al.*, 2021) et Twitter (Roche *et al.*, 2017, Fung *et al.*, 2013; Cheng *et al.*, 2021). Cependant, les résultats de notre recherche montrent des différences quant à certaines composantes de l'écosystème de la pratique de l'épidémiologie numérique.

En effet, les chercheurs ayant participé à la recherche de terrain ont mentionné les organismes produisant des données officielles, comme Statistique Canada (entretien 5Gxxix), lorsqu'ils ont expliqué leur confiance envers les données traditionnelles en épidémiologie. Les organismes produisant des données officielles seraient donc des acteurs de confiance pour les chercheurs en épidémiologie. Or, cette catégorie d'acteur était absente des propos des chercheurs qui exprimaient spécifiquement des réserves sur la pratique de l'épidémiologie numérique. Dans les discours sur la pratique de l'épidémiologie numérique, les grandes sociétés privées comme Google et Twitter ont été mentionnées très souvent, et ce, autant chez les chercheurs pratiquant l'épidémiologie numérique que ceux n'ayant pas l'expérience de cette nouvelle pratique.

Ce constat semble indiquer que les organismes produisant des données officielles sont davantage associés aux sources de données traditionnelles en épidémiologie et beaucoup moins à la pratique de l'épidémiologie numérique, probablement à cause du fait que ces organismes existent depuis longtemps et bien avant l'émergence de cette nouvelle pratique. Par ailleurs, nous avons questionné les participants à la recherche de terrain sur les moteurs de recherche Internet, les médias sociaux et les téléphones cellulaires, puisque ces technologies sont parmi les plus connues en ce qui concerne la pratique de l'épidémiologie numérique. L'accent que nous avons mis sur ces technologies lors de la recherche de terrain a fait en sorte que le rôle des organismes produisant des données officielles n'a pas été abordé dans les discours de nos participants sur la pratique de l'épidémiologie numérique. D'autres recherches sont alors nécessaires pour comprendre le rôle de ces organismes en lien avec la pratique de l'épidémiologie numérique. Il s'agit d'un autre enjeu

majeur dont il faudra tenir compte lors de la définition des nouvelles « règles de jeu » de cette pratique émergente.

D'après la littérature, les acteurs non experts, c'est-à-dire les profanes, peuvent avoir une certaine place dans le paysage de la pratique de l'épidémiologie numérique, notamment dans le cadre de projets de science citoyenne (Hamer *et al.*, 2018). Notre recherche semble le confirmer à travers l'exemple d'un chercheur en épidémiologie sociale qui pratique l'épidémiologie numérique et qui utilise l'approche de la science citoyenne pour faire participer des citoyens à la collecte de données en temps réel à travers leurs applications mobiles (entretien 8Sxx). Cependant, la science citoyenne n'a été mentionnée par aucun autre chercheur en épidémiologie ayant participé à la recherche de terrain. D'autres recherches sont nécessaires pour mieux comprendre comment fonctionnent la science citoyenne et les autres approches impliquant des profanes, comme celle de la production participative (*crowdsourcing*)⁹, dans le contexte de la pratique de l'épidémiologie numérique.

Sur le plan technologique, la pratique de l'épidémiologie numérique implique effectivement l'utilisation massive d'entités informatiques, de technologies et d'objets numériques producteurs de données, incluant des médias sociaux (Salathé *et al.*, 2016; Chary *et al.*, 2017; Bar-Lev *et al.*, 2021; Mavragani et Ochoa, 2019; Basteris *et al.*, 2020), des moteurs de recherche Internet (Rovetta, 2022, Adawi *et al.*, 2017; Duclos, 2019), des technologies mobiles (Mohanty *et al.*, 2016; Lazzari *et al.*, 2018; Cot *et al.*, 2021; Vigfusson *et al.*, 2021), des algorithmes (Hassan *et al.*, 2019; Hochreiter et Waldhauser, 2013; Durazzi *et al.*, 2021) ainsi que l'intelligence artificielle et l'apprentissage automatique (*machine learning*) (Sau, 2017; Kaveh-Yazdy et Zareh-Bidoki, 2018; Magumba et Nabende, 2021). Cependant, l'expérience de ces dispositifs dans le cadre de recherches épidémiologiques semble limitée dans le contexte canadien d'après notre recherche de terrain, puisque seulement trois personnes parmi les 36 chercheurs rencontrés pratiquent l'épidémiologie numérique (entretien 8Sxx, entretien 6Axxiii et entretien 10Ixxviii). De plus, les résultats de notre recherche montrent que les chercheurs en épidémiologie doivent se former pour pouvoir pratiquer l'épidémiologie numérique, puisque les compétences nécessaires à cette pratique

⁹ La recherche par production participative permet aux chercheurs d'engager des milliers de personnes des milieux académiques et du public pour fournir des données ou des analyses de données (Ranard *et al.*, 2014).

ne sont pas enseignées dans les programmes en épidémiologie (entretien 6Bxviii). Ces constats soulèvent des questions quant au curriculum et à la formation des chercheurs dans le contexte d'émergence d'une telle pratique de recherche épidémiologique.

Ce manque de formation et d'expérience peut expliquer les réserves que suscite cette nouvelle pratique chez les chercheurs en épidémiologie clinique en particulier. Nous avons constaté qu'aucun des participants à la recherche de terrain en épidémiologie clinique n'avait pratiqué l'épidémiologie numérique au moment des entretiens en 2020. Il semble y avoir un consensus chez les participants à notre recherche à savoir que la pratique de l'épidémiologie numérique serait mieux alignée avec des recherches sociales par comparaison avec les recherches cliniques. Notre recherche montre que plusieurs chercheurs en épidémiologie considèrent la pratique de l'épidémiologie numérique comme étant moins pertinente pour des recherches cliniques (entretien 6Fviii et entretien 5Cxxxi). L'épidémiologie clinique constitue une sous-discipline importante de l'épidémiologie (Giroux, 2012; Frérot *et al.*, 2018). Si la plupart des chercheurs en épidémiologie clinique résistent à pratiquer l'épidémiologie numérique, il y a des chances que la diffusion de cette nouvelle pratique demeure faible en épidémiologie, du moins, par rapport aux technologies présentes. Nul doute que les développements technologiques futurs risquent d'avoir un impact important sur la diffusion de cette pratique.

Il faut souligner que le groupe de chercheurs ayant participé à la recherche de terrain est de petite taille par rapport à l'ensemble des épidémiologistes canadiens, ce qui nous amène à nuancer nos constats sur la diffusion et le niveau d'expérience de cette nouvelle pratique en épidémiologie au Canada. Il reste que les chercheurs ayant participé à la recherche de terrain évoquent très souvent les données générées par des technologies numériques comme les médias sociaux, les moteurs de recherche Internet et les téléphones cellulaires lorsqu'ils expliquent leurs points de vue par rapport à la pratique de l'épidémiologie numérique, et ce, autant sur le plan épistémique (entretien 6Fx) que sur le plan sociopolitique (entretien 8Hxvi).

7.2.2 La confiance au sein des chercheurs

Notre recherche montre que le niveau de confiance au sein des chercheurs varie selon le dispositif méthodologique utilisé dans la pratique de l'épidémiologie numérique. En effet, nous avons

constaté que les données générées par les téléphones cellulaires (entretien 6Fx) ont le niveau de confiance le plus élevé, suivies par les données générées par les moteurs de recherche Internet. Les données générées par les médias sociaux auraient le niveau de confiance le moins élevé (entretien 10Ixxviii). Plus généralement, notre recherche montre que le niveau de confiance des chercheurs est plus élevé envers les données traditionnelles, dont celles générées par les enquêtes officielles de Statistique Canada, qu'envers les données numériques, comme celles générées par les médias sociaux (entretien 5Uxxxvi).

À la lumière de la littérature, la pratique de l'épidémiologie numérique implique l'utilisation d'une grande variété de dispositifs. Or, notre recherche montre que le niveau de confiance n'est pas nécessairement le même vis-à-vis tous ces dispositifs. Comme nous l'avons expliqué précédemment, les dispositifs ne sont pas seulement des formalités dans le cadre des méthodes de recherche (Savage, 2013). En l'occurrence, ils suscitent de nombreux questionnements quant à la confiance des chercheurs vis-à-vis des recherches fondées sur les données générées par les dispositifs numériques. Notre recherche montre aussi que la confiance des chercheurs envers les données numériques n'est pas toujours très forte du point de vue épistémique, en raison d'incertitudes sur la qualité des informations publiées par les usagers d'Internet (entretien 6Fxi) et sur la validité de l'échantillonnage populationnel (entretien 5Gxxix). Ces constats rejoignent ceux de plusieurs chercheurs, comme Salathé *et al.* (2012) et Bakker *et al.* (2016) qui soulignent des disparités dans les usagers d'Internet, ou comme Salathé *et al.* (2013), Towers *et al.* (2015) et Génois et Barrat (2018) qui mentionnent les enjeux liés aux différences entre les informations publiées sur le web et la réalité, dans la pratique de l'épidémiologie numérique.

Sur le plan sociopolitique, la confiance des chercheurs serait compromise par les doutes sur les intérêts des grandes entités privées détentrices des données numériques comme Google les incitant à publier gratuitement des données sur Internet (entretien 8Sxx) et par l'ampleur grandissante des pouvoirs de ces grandes entités à l'ère numérique (entretien 6Jv). Ces constats rejoignent les perspectives développées par plusieurs chercheurs soutenant que les méthodes de recherche et les données sont des phénomènes plus larges que des opérations techniques, car celles-ci peuvent, d'une part, avoir un impact important sur la vie des individus, notamment à cause des risques encourus par la divulgation de données sensibles (Mayer-Schönberger et Cukier, 2013; Kitchin,

2014; Lupton, 2015) et, d'autre part, entraîner une concentration de pouvoir auprès de grands détenteurs de données comme Google et Facebook (Mayer-Schönberger et Cukier, 2013; Zwitter, 2014; Ruppert *et al.*, 2015).

Comme plusieurs travaux étudiant les méthodes de recherche et les données l'ont soulevé, les technologies et les dispositifs utilisés lors de la pratique de l'épidémiologie numérique impliquent des enjeux plus larges que des enjeux purement techniques (Savage, 2013; Law *et al.*, 2011). Les pratiques de recherche impliquent des objets (infrastructures de serveurs, dispositifs et câbles), des langages (code, programmation et algorithmes) et des personnes (scientifiques, entrepreneurs, ingénieurs, informaticiens, concepteurs) susceptibles de créer, ensemble, de nouveaux mondes (Ruppert *et al.*, 2017). Cette perspective fait ressortir l'importance des entités non humaines (comme des instruments de mesure et des données) influençant une partie importante du travail des experts (Knorr-Cetina, 2001) dans le cadre des pratiques de recherche. Les individus et les organisations, de même que les technologies et les dispositifs, forment des réseaux hétérogènes, temporaires et fluides lors des pratiques de recherche (Vinck, 2007). Tous ces acteurs exercent une influence lors des projets de recherche, et ce, autant au niveau épistémique qu'au niveau sociopolitique. L'influence sur le plan épistémique est causée par les apparences secondaires et les signes produits par les dispositifs, les technologies, les mesures et les données (Knorr-Cetina, 1999). Celle du point de vue sociopolitique est produite, quant à elle, en raison des enjeux éthiques liés entre autres à l'utilisation des données sur les gens générées, sans leur consentement, par les technologies numériques (Zwitter, 2014; Ruppert *et al.*, 2015) et au risque d'atteindre à leur vie privée (Dencik et Sanchez-Monedero, 2022). Ces constats font jaillir une conclusion très importante : l'épidémiologie numérique est loin d'être une pratique homogène.

7.2.3 Les tensions

D'un côté, notre recherche révèle que la pratique de l'épidémiologie numérique peut contribuer à la discipline de l'épidémiologie sur le plan épistémique pour la surveillance syndromique à la condition que les signaux en provenance d'Internet soient comparables à ceux de sources de données traditionnelles (10Ixxviii), pour le recrutement des participants aux recherches

(entretien 1Oxxxiv) et pour l'émission d'hypothèses lors de recherches exploratoires (entretien 9Rxiv). D'un autre côté, du point de vue sociopolitique, cette nouvelle pratique suscite des inquiétudes d'une part, sur le plan éthique en lien avec le consentement qui n'est pas toujours explicite dans le contexte de l'utilisation de données numériques (entretien 5Uxxxvi) et d'autre part, au niveau politique à cause des pouvoirs de grandes entités privées détentrices des données numériques (entretien 6Jv, entretien 3Exxxiii et entretien 8Sxx). Donc certaines conséquences individuelles et collectives de la pratique de l'épidémiologie numérique peuvent aller à l'encontre de la justice sociale.

Sur le plan éthique, la pratique généralisée de l'épidémiologie numérique entraîne des conséquences pour les individus parce que cette nouvelle pratique requiert l'utilisation de leurs données. Parfois ces données sont sensibles et leur divulgation peut causer des préjudices aux personnes concernées si les dispositions nécessaires ne sont pas prises pour minimiser les risques d'atteinte à la vie privée.

Il faut aussi que les chercheurs pratiquant l'épidémiologie numérique obtiennent le consentement éclairé des individus en leur communiquant clairement les objectifs de la recherche, la manière dont les données seront utilisées et les risques potentiels encourus. Il semble que beaucoup de questions demeurent sans réponse quant aux caractéristiques d'un consentement éclairé (Boyd et Crawford, 2012) parce qu'il est très difficile d'obtenir le consentement de chaque personne publiant de l'information sur Internet et que les chercheurs ne peuvent pas justifier leurs actions comme étant éthiques simplement parce que les données sont accessibles.

De plus, la pratique de l'épidémiologie numérique s'appuie considérablement sur des données numériques susceptibles d'exclure des individus n'ayant pas accès aux technologies numériques introduisant ainsi des erreurs dans la représentativité et la généralisation des résultats lors des études épidémiologiques. Le contexte canadien est particulièrement marqué par le phénomène de la fracture numérique touchant entre autres les communautés rurales et autochtones. Donc ces communautés peuvent se retrouver exclues lors des études épidémiologiques fondées sur la pratique de l'épidémiologie numérique.

La surveillance des maladies en temps réel rendue possible avec la pratique de l'épidémiologie numérique soulève aussi des enjeux éthiques parce que cette activité nécessite l'identification et le suivi des personnes atteintes de maladies infectieuses entraînant des risques de discrimination envers ces personnes si leurs données ne sont pas bien protégées. Les algorithmes utilisés pour surveiller les maladies sont souvent méconnus par les chercheurs en santé publique. Il y a des incertitudes non seulement sur la façon dont les algorithmes fonctionnent sur le plan technique, mais aussi sur les motivations réelles régissant leur développement. Cependant, la surveillance des maladies est importante pour la santé publique. Il y a donc un conflit important entre la nécessité de protéger les données des individus et les obligations des gouvernements de veiller à la santé publique. D'ailleurs ce conflit entre les intérêts individuels et collectifs est analysé dans certains travaux comme ceux de Mittelstadt *et al.* (2018) questionnant l'obligation morale de la participation aux travaux de l'épidémiologie numérique, un questionnement lié à l'équilibre entre intérêts publics et individuels. Selon ces chercheurs, assurer cet équilibre apparaît comme difficile en pratique parce que l'importance relative des intérêts publics et individuels ne peuvent être décidées qu'au cas par cas, en tenant dûment compte des intérêts et des risques locaux.

Cette réflexion sur la surveillance nous amène à souligner des enjeux au niveau politique contribuant à renforcer cette tension centrale liée à la pratique de l'épidémiologie numérique. En effet, la confiance du public envers les instances effectuant la surveillance des maladies est essentielle pour assurer un traçage efficace des contacts sociaux afin d'aider la santé publique (entretien 9Qxxiv, entretien 8Sxx). Il est indispensable que les gens sachent que le traçage numérique des citoyens se fait parce que c'est important pour la santé et non pas parce que les gouvernements visent à s'approprier les informations sur leur vie privée.

La confiance est aussi une question importante en ce qui concerne Google en tant que source de données pour des recherches en santé. La pratique de l'épidémiologie numérique est grandement influencée par la confiance des chercheurs vis-à-vis de Google (entretien 8Sxx). Notre recherche montre d'une part que des chercheurs pensent que cette nouvelle pratique présente un haut potentiel pour des recherches épidémiologiques et d'autre part, les chercheurs ont de la difficulté à faire confiance à Google puisque cette grande société privée exerce le plein contrôle sur les données

généérées par son moteur de recherche Internet et manque de transparence sur la manière dont les algorithmes fonctionnent.

Les enjeux de pouvoir liés aux développements de l'intelligence artificielle sont énormes. Certains chercheurs comme Martineau et Folco (2023) soulignent qu'il faut développer une théorie critique des algorithmes pour comprendre l'impact de leur usage généralisé et expliquer les multiples dimensions institutionnelles, politiques, économiques, culturelles et idéologiques de la logique algorithmique afin de mettre en lumière les rapports de domination et les durs contrecoups de ces innovations. Ces chercheurs ainsi que Zuboff (2019) établissent clairement la relation entre le capitalisme et les développements de l'intelligence artificielle. Dans la même perspective, le travail de Rouvroy (2018) vise à démontrer comment les gens sont pris pour des objets d'observation et de classification dans la gouvernance actuelle fondée sur les algorithmes en examinant comment ces machines cartographient et traduisent le monde, les trajectoires, les comportements, les actions, les choix, les préférences et les attitudes. Nous serions désormais face à un nouvel ordre économique où l'expérience humaine sert de matériel brut à des fins cachées de commercialisation, prédictions et de ventes (Zuboff, 2019).

Des tensions sont alors perceptibles en ce qui concerne la pratique de l'épidémiologie numérique puisque nous envisageons un dilemme entre le potentiel de cette nouvelle pratique pour faire avancer les connaissances d'une discipline en utilisant des méthodes innovantes de l'ère numérique et les risques encourus sur le plan éthique et au niveau politique. Dans un tel contexte, la tension centrale ne se situerait pas au niveau du dilemme entre la nouveauté et la continuité des traditions d'une discipline (D'Agostino, 2008), mais entre l'innovation et la protection des données et des citoyens.

En somme, la pratique de l'épidémiologie numérique implique une grande diversité d'acteurs et de technologies interagissant sous forme de réseaux fluides et hétérogènes lors des projets de recherche. Notre recherche a mis en évidence une tension centrale entre les apports épistémiques de cette nouvelle pratique et ses enjeux éthiques et sociopolitiques. Cette analyse aboutit au constat que la pratique de l'épidémiologie numérique soulève des enjeux épistémiques et sociopolitiques

ayant trait à l'interdisciplinarité et à la transdisciplinarité qui seront discutées dans la prochaine section qui fera l'analyse des contextes de la coproduction des connaissances à l'ère numérique.

7.3 Les contextes de la coproduction des connaissances à l'ère numérique

Les enjeux identifiés quant à la pratique de l'épidémiologie numérique permettent de faire des hypothèses sur les pratiques de recherche utilisant des données numériques dans différents domaines. Cette section de la discussion a comme objectif d'expliquer les contextes de la coproduction des connaissances à l'ère numérique. Soulignons que dans la discussion, le terme « coproduction » est utilisé et non le terme « production » des connaissances, car le cadre théorique de cette thèse se situe entre le déterminisme technologique et naturel et le déterminisme social. Nous soutenons que les pratiques de recherche impliquent des aspects technologiques, scientifiques, conceptuels et sociaux. Cette perspective STS est alignée de manière générale avec le concept de la coproduction des connaissances défini par Jasanoff (2004). Notre recherche sur la pratique de l'épidémiologie numérique témoigne de la combinaison de différents modes de la coproduction des connaissances, de l'ampleur grandissante des recherches fondées sur les connaissances algorithmiques et de la numérisation des disciplines dans les contextes de l'ère numérique.

7.3.1 Les modes de la coproduction des connaissances

S'inspirant des modes et des attributs de la production des connaissances développés par Gibbons *et al.* (1994), notre recherche sur la pratique de l'épidémiologie numérique montre que les nouvelles pratiques de recherche issues de l'ère numérique peuvent combiner différents modes lors de la coproduction des connaissances. Rappelons que Gibbons *et al.* (1994) proposent de conceptualiser la production des connaissances en deux modes, c'est-à-dire le Mode 1 et le Mode 2 afin d'expliquer les transformations en recherche. Gibbons *et al.* (1994) soutiennent que ces transformations marquent un tournant vers un nouveau mode de production des connaissances susceptible de remplacer ou de réformer les institutions, les disciplines, les pratiques et les politiques établies (Gibbons *et al.*, 1994). Le Mode 1 désigne le mode traditionnel de produire des connaissances en contexte disciplinaire et fondamentalement cognitif tandis que le Mode 2 désigne

le nouveau mode de produire des connaissances en contextes sociaux et économiques plus larges, transdisciplinaires et appliqués (Gibbons *et al.*, 1994).

D'après nos constats, la pratique de l'épidémiologie numérique aurait les mêmes objectifs et objets de recherche que ceux de la discipline de l'épidémiologie (Salathé, 2018) dont les activités sont gouvernées en grande partie par la science académique (Gibbons *et al.*, 1994). Cette caractéristique porte à croire que la pratique de l'épidémiologie numérique peut résoudre les problèmes en suivant les codes de pratique et en utilisant des concepts et des théories de la discipline de l'épidémiologie. De plus, les chercheurs en épidémiologie continueraient à diriger les recherches lors de la pratique de l'épidémiologie numérique comme chefs d'orchestre même si celle-ci nécessite plusieurs autres expertises spécialisées (entretien 5Cxxvii). Nous soutenons que ces caractéristiques illustrent l'attribut disciplinaire du Mode 1 lors de la coproduction des connaissances dans le contexte de la pratique de l'épidémiologie numérique. Ces caractéristiques illustrent aussi l'interdisciplinarité. Les moteurs de l'interdisciplinarité lors de cette pratique seraient les technologies, les expertises exigées lors des projets de recherche et le recours à de nouvelles sources de données générées en dehors des systèmes de santé (entretien 5Cxxvi, entretien 8Sxx et entretien 5Cxxvii). L'épidémiologie et l'épidémiologie numérique partagent les mêmes objectifs, mais leurs écosystèmes (humains et non humains) ne sont pas les mêmes.

Nous avons aussi constaté que la pratique de l'épidémiologie numérique s'organiserait autour d'une application particulière telle que la surveillance des maladies pour informer des décisions de santé publique (Clemente, 2019; entretien 6Fix). Cette caractéristique montre que la coproduction des connaissances à l'ère numérique peut aussi être caractérisée par le Mode 2, c'est-à-dire un nouveau mode appliqué se voulant utile pour la société en général et les gouvernements (Gibbons *et al.*, 1994). De plus, selon nos analyses, la pratique de l'épidémiologie numérique ne serait pas dirigée par un seul organisme central coordonnant toutes les activités. Les expertises des équipes sont très diversifiées et la composition de ces équipes change en fonction des problèmes à résoudre et des projets de recherche. Telles qu'expliquées par Gibbons *et al.* (1994), ces caractéristiques s'illustrent par le nombre moyen d'auteurs par article sur la pratique de l'épidémiologie numérique, par la diversité des spécialités et disciplines impliquées lors des recherches (notamment des médecins et des informaticiens) et par la pluralité des institutions et organisations d'appartenance

des chercheurs travaillant sur un projet. De telles caractéristiques s'appliquent au Mode 2 de la coproduction des connaissances et celles-ci représentent l'hétérogénéité et la diversité des acteurs impliqués lors des nouvelles pratiques de recherche à l'ère numérique.

La transdisciplinarité constitue une autre caractéristique du Mode 2 de la coproduction des connaissances selon les travaux de Gibbons *et al.* (1994). Rappelons que selon Popa *et al.* (2015), la transdisciplinarité est une approche de recherche critique et autoréflexive dont l'objectif est de contribuer au progrès sociétal et scientifique. La transdisciplinarité met en relation des problèmes sociétaux et scientifiques en produisant de nouvelles connaissances fondées sur différentes perspectives scientifiques et extrascientifiques (Popa *et al.*, 2015). Cette approche de recherche était absente des propos de la plupart des chercheurs ayant participé à la recherche de terrain effectuée dans le cadre de cette thèse. Un seul chercheur en épidémiologie sociale pratiquant l'épidémiologie numérique dans l'ensemble de ses projets a affirmé qu'il utilise une approche fondée sur la science citoyenne en engageant des citoyens dans la collecte de données en temps réel à travers leurs applications mobiles (entretien 8Sxx).

Ailleurs dans le monde, le laboratoire suisse *Digital Epidemiology Lab* (DEL, 2021) serait transdisciplinaire, car il utilise des approches de production participative (AICrowd, 2023) dans le but de résoudre des problèmes de santé du monde réel. Ce laboratoire réunit les efforts et l'expertise de spécialistes et de passionnés de l'apprentissage machine et de la science des données qui s'efforcent en collaborant de trouver les solutions les plus précises, les plus efficaces et les plus performantes. Ces constats incitent à conclure que l'épidémiologie numérique se pratique dans des contextes transdisciplinaires. Donc, il semble que de nouvelles pratiques de recherche à l'ère numérique peuvent être transdisciplinaires, c'est-à-dire pourvues d'un autre attribut du Mode 2 de la coproduction des connaissances d'après la littérature sur la pratique de l'épidémiologie numérique. Cependant, d'autres recherches portant sur la transdisciplinarité à l'ère numérique seront nécessaires pour tirer des conclusions quant à la façon dont cette approche s'intègre réellement dans le travail des chercheurs.

En fait, nos analyses montrent que les contextes des nouvelles pratiques de recherche à l'ère numérique se caractérisent par une combinaison du Mode 1 et du Mode 2 parce que ces nouvelles

pratiques ont des attributs appartenant à chacun de ces modes de la production de connaissances. Ces nouvelles pratiques semblent avoir un plus grand nombre d'attributs caractérisés par le Mode 2 de la production des connaissances que par le Mode 1. Cependant, le Mode 1 est important, car ces nouvelles pratiques semblent s'appuyer considérablement sur des théories, des concepts et des expertises des disciplines selon notre étude de cas sur la pratique de l'épidémiologie numérique. Les nouvelles pratiques de recherche à l'ère numérique seraient disciplinaires, mais aussi appliquées, hétérogènes et axées sur l'informatique et la modélisation de données (cet attribut étant visible tout au long de cette thèse). La transdisciplinarité serait aussi présente dans certains contextes des nouvelles pratiques de recherche à l'ère numérique de même que la diversification des sites de la création des connaissances. Là encore, d'autres recherches seront nécessaires pour comprendre comment la transdisciplinarité et la diversification des sites de la création des connaissances entrent en jeu dans ces nouvelles pratiques de recherche, notamment en ce qui concerne la division du travail entre les chercheurs, experts professionnels et profanes.

7.3.2 La complémentarité entre l'épidémiologie numérique et l'épidémiologie traditionnelle

D'après la littérature, les technologies utilisées en épidémiologie numérique amènent des changements importants dans les approches de recherche en passant d'une production fondée sur les connaissances des experts vers une production fondée sur les connaissances des algorithmes (Eckmanns *et al.*, 2019). De plus, il semble que les connaissances sur le monde physique et sur les maladies infectieuses soient extraites de plus en plus sous forme de signaux et de signes projetés sous l'emblème du numérique et de moins en moins à partir des processus statistiques fondés sur les analyses humaines (Eckmanns *et al.*, 2019). Dans un tel contexte, les connaissances et les informations générées par les algorithmes dépasseraient les traditions humaines d'évaluation, d'analyse, d'hypothèse et d'essais cliniques essentielles aux calculs statistiques utilisés dans les recherches épidémiologiques traditionnelles (Eckmanns *et al.*, 2019). Cette perspective mise de l'avant par Eckmanns *et al.* (2019) sous-entend que la pratique de l'épidémiologie numérique serait transformationnelle par rapport aux pratiques courantes en épidémiologie. Est-ce vraiment le cas? Nos analyses à partir de notre recherche sur la pratique de l'épidémiologie numérique au Canada montrent que cette nouvelle pratique serait complémentaire vis-à-vis de l'épidémiologie traditionnelle.

Selon plusieurs chercheurs interviewés, la pratique de l'épidémiologie numérique va compléter ce qui se fait présentement en épidémiologie (entretien 6Bxviii). Cette nouvelle pratique pourrait aider notamment à avoir des échantillons populationnels plus larges. Elle permet de tester des relations entre les phénomènes observés dans les données des médias sociaux par exemple et ceux observés dans les données des systèmes de surveillance. Elle permet aussi de mesurer des choses impossibles à mesurer autrement comme des normes et des valeurs au sein de certains groupes à l'aide de données extraites des discours sur Twitter notamment et à accélérer les processus de surveillance des maladies (Chapitre IV). Cette complémentarité est aussi soulignée dans la littérature sur l'épidémiologie numérique, notamment dans le cadre de la pharmacovigilance combinant des données de sources traditionnelles et des données générées par les patients contribuant ainsi à renforcer les systèmes de santé publique (Salathé, 2016).

Une autre illustration de la complémentarité entre la pratique de l'épidémiologie numérique et l'épidémiologie traditionnelle est le fait que plusieurs chercheurs pensent que les signaux générés par les dispositifs numériques comme *Google Trends* doivent correspondre aux signaux générés par les systèmes de surveillance traditionnels des maladies. S'il y a des différences entre les signaux émis par des sources traditionnelles et ceux en provenance des dispositifs numériques, les chercheurs n'auront pas confiance envers les signaux générés par les dispositifs numériques (entretien 10Ixxviii). Donc, les chercheurs en épidémiologie ne sont pas prêts à faire confiance uniquement aux signaux générés par les dispositifs numériques selon les constats établis lors de notre recherche. À la lumière de ce constat, il est difficile d'envisager que les connaissances en épidémiologie soient extraites uniquement sous forme de signaux générés par les dispositifs numériques et que les techniques algorithmiques dépassent les traditions essentielles aux calculs statistiques utilisés dans les recherches épidémiologiques traditionnelles (Eckmanns *et al.*, 2019).

Cependant, l'émergence et la reconnaissance de l'épidémiologie computationnelle soulèvent des questions (Li Vigni, 2021). Par l'application de méthodes de simulation basées sur la science des systèmes fondés sur des modèles mathématiques d'épidémies, cette spécialité apporterait des contributions importantes à l'évaluation des risques de contagion, des populations vulnérables et des effets des mesures d'intervention visant à arrêter la propagation des épidémies (Marathé et Vullikanti, 2013 cités dans Li Vigni, 2021). L'expertise offerte par l'épidémiologie

computationnelle ferait désormais partie des mécanismes de préparation sanitaire de nombreux gouvernements depuis les années 2000 et encore plus récemment depuis la pandémie de la COVID-19 malgré de vives critiques de la part de médecins, d'infectiologues, de biostatisticiens et de journalistes scientifiques (Li Vigni, 2021).

La discipline maîtresse de l'épidémiologie computationnelle serait la physique statistique et les chercheurs de cette spécialité auraient en commun la maîtrise de capacités mathématiques et informatiques poussées (Li Vigni, 2021). Les pratiquants de l'épidémiologie computationnelle doivent d'abord acquérir des connaissances en épidémiologie par la lecture de la littérature et des interactions avec les professionnels de la santé publique et ensuite, utiliser leurs connaissances en physique, en informatique, en mathématiques, en écologie et en géographie (Li Vigni, 2021). Lors des projets de recherche, l'informatique fournirait les modèles alors que l'expertise de l'épidémiologie serait utilisée « pour nommer et interpréter les variables, caractériser les équations et borner la dynamique du phénomène modélisé » (Li Vigni, 2021, p. 29). À la lumière de ces derniers constats, l'épidémiologie constitue une composante aussi importante que celle de l'informatique lors de la pratique de l'épidémiologie computationnelle et de l'épidémiologie numérique montrant ainsi la complémentarité entre la discipline de l'épidémiologie et les autres disciplines à l'ère numérique.

De plus, les données traditionnelles des systèmes de santé (comme les résultats de tests médicaux et des données collectées sur le terrain par les chercheurs en santé) et les données numériques (comme les données numériques générées par Internet et les applications mobiles des individus) ne seraient pas nécessairement exclusives et pourraient même se compléter pour résoudre des problèmes de santé (entretien 8Sxx). Par exemple, les données générées par Internet ou par des téléphones cellulaires peuvent prédire le stress, l'anxiété et la dépression si les bons algorithmes d'intelligence artificielle sont développés. On peut alors faire en sorte que les gens n'aient pas besoin d'aller vers les systèmes de santé. Ce type de données n'a rien à voir avec le système de soins de santé puisque celles-ci proviennent de l'extérieur de ce système, mais elles peuvent être bénéfiques pour les soins de santé.

La complémentarité entre ces deux types de sources de données montre que les connaissances des domaines sociaux et des domaines médicaux peuvent être intégrées pour comprendre des enjeux de santé en vue d'agir de manière efficace pour aider les patients (Duclos et Carle, 2009). De plus, les connaissances générées à l'extérieur du domaine médical peuvent enrichir les recherches cliniques avec les bonnes méthodologies et les outils appropriés. Cette complémentarité devra être expliquée à l'ensemble des chercheurs en épidémiologie et peut-être encore plus aux chercheurs en épidémiologie clinique afin de démontrer la valeur ajoutée de la pratique de l'épidémiologie numérique pour ainsi ajouter d'autres ressources à l'expertise de cette discipline en vue de mieux saisir la complexité des enjeux de santé et de prendre les bonnes décisions sur les soins de santé.

7.4 Vers la numérisation des disciplines?

Pour conclure cette discussion, il est important de souligner que les enjeux épistémiques et sociopolitiques liés aux méthodes et aux données numériques se posent pour d'autres disciplines. En effet, plusieurs autres domaines et disciplines commencent à prendre le tournant numérique, dont le domaine des sciences humaines et sociales. En sciences humaines, l'influence de l'ère numérique s'observe par l'émergence des humanités numériques (Vinck, 2016) qui quantifient les phénomènes culturels et sociaux (comme des textes, des images et des films) à travers les données numériques générées à l'aide de l'expertise informatique. Cependant, la numérisation en histoire serait encore imparfaite à cause des difficultés à saisir la complexité et la singularité des événements historiques (Vinck, 2016) avec les technologies informatiques. Il semble que des historiens utilisent tout de même des technologies numériques non pas pour quantifier les phénomènes observés, mais pour faciliter l'accès à un grand nombre de documents rassemblés dans des bases de données afin de pouvoir les rapprocher et les comparer facilement (Vinck, 2016).

En sociologie, des chercheurs s'interrogent sur l'avenir de la recherche sociologique et de la théorisation à l'ère numérique (Lupton, 2015). Certains chercheurs étudient de manière approfondie les méthodes de recherches sociales à l'ère numérique. Par exemple, Boullier (2015) conceptualise la mutation numérique comme une troisième génération pour les sciences sociales représentant une époque particulière de réflexivité sur les façons de concevoir les recherches sociales. La matière première des époques précédentes à l'ère numérique serait les personnes, les

identités et les communautés (Boullier, 2015). Les traces sur les plateformes constitueraient la matière première des recherches sociales à l'ère numérique selon Boullier (2015). Donc, l'ère numérique peut constituer une mesure incitative à exploiter les traces sur les plateformes afin d'étudier la société ou les opinions. Toujours selon Boullier (2015), la troisième génération des sciences sociales ne parlerait plus de personnes, mais de comptes, plus de communautés, mais de clusters, plus de sociabilité, mais de connectivité.

D'autres travaux soulignent les défis pour différencier les études de la dynamique des médias et les études sur les dynamiques sociales lors des analyses des données générées par Internet dans le cadre de recherches sociales (Marres et Weltevrede, 2013; Marres et Gerlitz, 2016). Ces défis semblent liés au fait que des tendances observées à partir de données numériques et des médias sociaux ne signifient pas nécessairement que de telles tendances sont observées dans la vie réelle. Des enjeux techniques liés aux données du web peuvent aussi renforcer les incertitudes face aux méthodes numériques de recherche. De tels enjeux incluent la durée limitée des services web et l'instabilité des flux de données (Rogers, 2015). Ces perspectives rejoignent certaines réserves exprimées par des chercheurs en épidémiologie vis-à-vis de la pratique de l'épidémiologie numérique. Comme nous l'avons expliqué précédemment, la confiance des chercheurs en épidémiologie envers les données numériques n'est pas toujours très grande sur le plan épistémique à cause des incertitudes sur la qualité des informations publiées par les usagers d'Internet (entretien 6Fxi) et de la validité de l'échantillonnage populationnel (entretien 5Gxxix). Donc, il y a des incertitudes sur l'écart entre la vie réelle et ce qui est observable dans la dynamique des médias selon des chercheurs.

Certains chercheurs soutiennent que les sociologues devraient développer d'autres façons de faire de la sociologie en réponse à l'ère numérique en particulier pour conserver leur position prééminente d'experts en recherche sociale (Lupton, 2015). Certains chercheurs en sociologie empirique soutiennent que les nouvelles générations de sociologues empiriques devront analyser des données différentes par rapport à celles habituellement analysées en sociologie, c'est-à-dire des données transactionnelles « sociales » collectées, traitées et analysées de façon routinière par un large éventail d'institutions privées et publiques (Savage et Burrows, 2007).

Cette perspective ne ferait pas l'unanimité au sein des chercheurs en épidémiologie selon le groupe de participants à notre recherche sur la pratique de l'épidémiologie numérique. Nos constats montrent que plusieurs chercheurs soutiennent que l'ère numérique est en train d'influencer les pratiques en épidémiologie, mais il n'y a pas ce questionnement sur l'avenir de la discipline de l'épidémiologie au même niveau que certains chercheurs en sociologie. Plusieurs participants à notre recherche ont tout de même souligné que la discipline de l'épidémiologie doit se mettre à jour (entretien 5Kxv) et que les chercheurs en épidémiologie doivent être formés sur la pratique de l'épidémiologie numérique (entretien 8Hxvi).

Les nouvelles pratiques de recherche de l'ère numérique complètent ce qui se fait présentement en épidémiologie sans pour autant la remplacer (entretien 6Bxviii) selon plusieurs chercheurs en épidémiologie. De plus, certains chercheurs en épidémiologie se voient comme des chefs d'orchestre lors de la pratique de l'épidémiologie numérique, puisque selon eux l'expertise de l'épidémiologie reste indispensable pour mener à bien les recherches épidémiologiques à l'ère numérique (entretien 5Cxxvii).

Les chercheurs en épidémiologie semblent conscients de la nécessité des engagements interdisciplinaires afin d'utiliser d'autres expertises spécialisées (entretien 5Cxxvii), surtout celles de l'informatique (entretien 5Cxxvi) et de la science des données (entretien 10Ixxviii) avec l'émergence des nouvelles pratiques de recherche de l'ère numérique. Par contre, pour eux, l'expertise de l'épidémiologie demeure essentielle pour interpréter correctement les données numériques lors des projets de recherche. Donc, pour le moment, les épidémiologistes ne semblent pas avoir de crainte vis-à-vis de leur place d'experts en épidémiologie à l'ère numérique comme il est possible de le constater dans les propos de certains chercheurs en sociologie (Lupton, 2015; Savage et Burrows, 2007). Cette attitude exprimée par des chercheurs en épidémiologie rejoint la perspective expliquée par Kitchin (2014) affirmant que même si la science des données mène vers des innovations perturbatrices capables de reconfigurer dans de nombreux cas les façons de mener les recherches, les experts spécialistes sont plus que jamais nécessaires lors des recherches à l'ère numérique afin d'évaluer et d'expliquer les résultats des travaux de la science des données (Porway, 2013 cité dans Kitchin, 2014). Donc, il se peut que l'ère numérique exige encore plus de connaissances spécialisées lors des nouvelles pratiques de recherches fondées sur les analyses de

données numériques afin de donner un sens à ces analyses pour saisir la complexité des phénomènes étudiés et prendre les décisions optimales lors de la gestion des enjeux sociétaux.

En somme, les nouvelles pratiques de recherche à l'ère numérique seraient disciplinaires et interdisciplinaires, mais aussi appliquées, hétérogènes et axées sur l'informatique et la modélisation de données sur la base des résultats de notre recherche. Cependant, d'autres recherches sont nécessaires afin de comprendre comment la transdisciplinarité et la diversification des sites de la création des connaissances entrent en jeu dans ces nouvelles pratiques de recherche, notamment en ce qui concerne la division du travail entre les chercheurs, experts professionnels et profanes. Nous avons aussi constaté que d'autres domaines et disciplines commencent à prendre le tournant numérique, dont les sciences humaines et la discipline de la sociologie. Il semble que l'ère numérique suscite plusieurs questionnements sur l'avenir de la sociologie. D'après notre recherche, il y aurait moins d'inquiétude quant à l'avenir de l'épidémiologie à l'ère numérique, puisque plusieurs chercheurs en épidémiologie soutiennent que l'expertise des spécialistes est indispensable lors des recherches à l'ère numérique afin d'évaluer et d'expliquer les résultats des travaux de la science des données.

CONCLUSION

Nous débuterons ce chapitre par un retour réflexif sur l'ensemble de la démarche de thèse. Ensuite, nous présenterons les contributions sur le plan empirique et théorique, ainsi que les principaux éléments de la discussion. Nous terminerons en soulignant les limites de notre recherche et en proposant des pistes pour des recherches futures.

La démarche de la thèse

L'épidémiologie numérique est un objet d'étude fort intéressant dans le cadre d'une thèse en STS parce que celle-ci constitue nouvelle façon de faire des recherches en épidémiologie qui émerge dans le contexte historique actuel en plein effervescence marqué par la diffusion massive des technologies numériques ainsi que les développements de l'intelligence artificielle et l'utilisation de nouvelles sources de données pour faire des recherches. Il s'agit donc d'une pratique qui suscite la curiosité. Cependant, les étapes d'une recherche sur la pratique de l'épidémiologie numérique comportent plusieurs défis.

D'abord, l'épidémiologie numérique est pratiquée de manière sporadique au sein de nombreuses institutions dans des domaines très variés ce qui amène des difficultés pour choisir le terrain de la recherche. Étant donné que cette nouvelle pratique est proposée pour l'épidémiologie, nous avons décidé d'étudier les perspectives de chercheurs canadiens en épidémiologie envers une telle pratique. Cette approche nous a permis de recueillir des informations importantes sur les façons de voir cette nouvelle pratique, mais comme précisé plus loin, nous n'avons pas pu connaître les perspectives de chercheurs d'autres disciplines ni de différentes régions du monde. Les défis sur le plan théorique ont aussi été importants parce que la pratique de l'épidémiologie numérique implique de nombreux aspects incluant entre autres des enjeux techniques relevant de l'informatique, les questionnements éthiques et épistémiques, l'interdisciplinarité et les compétences requises lors d'une telle pratique.

Notre recherche de terrain montre que les perspectives de chercheurs canadiens en épidémiologie comportent des enjeux épistémiques et sociopolitiques. Les cultures épistémiques développées par Knorr-Cetina (1999) se sont avérées très pertinentes pour analyser les enjeux épistémiques de la pratique de l'épidémiologie numérique parce que ce cadre théorique fait ressortir l'influence de la conception des instruments (c'est-à-dire les sources de données et des algorithmes pour le cas de l'épidémiologie numérique) sur les mesures et les signes repérables lors des recherches. Pour expliquer les enjeux sociopolitiques, nous avons utilisé des travaux théoriques sur les enjeux sociétaux des méthodes de recherche et des données parce que les cultures épistémiques (Knorr-Cetina, 1999) n'incluent pas de réflexion sur ces enjeux. Ces travaux sur les méthodes de recherche et les données sont intéressants pour étudier la pratique de l'épidémiologie numérique parce qu'ils apportent de l'originalité à la thèse qui examine cette nouvelle pratique sous l'angle des méthodes de recherche. Dans cette perspective, l'épidémiologie numérique est vue comme une pratique régies par des méthodes de recherche ayant des impacts d'ordre sociétal et non seulement d'ordre technologique.

Le point fort de notre recherche au niveau méthodologique est d'avoir pu avoir la participation de 36 chercheurs. Le point faible est le manque de participants ayant l'expérience de la pratique de l'épidémiologie numérique. Seulement trois chercheurs ont pratiqué l'épidémiologie numérique parmi le groupe de participant à notre recherche. Ce résultat est plutôt surprenant étant donné que le Canada fait partie des pays ayant publié le plus grand nombre de publications. Il est possible que les chercheurs pratiquant l'épidémiologie numérique soient très sollicités pour leur expertise ce qui peut faire en sorte que ces chercheurs soient moins disponibles pour un entretien avec un membre de la population étudiante ne faisant pas partie de leurs réseaux de connaissance. Néanmoins, les propos des participants étaient riches et très pertinents pour apporter un regard critique envers cette nouvelle pratique. Comme précisé plus loin, des études de laboratoire auprès de chercheurs pratiquant l'épidémiologie numérique seront très utiles pour des recherches futures afin de saisir cette nouvelle pratique en elle-même.

Les entretiens avec les participants ont eu lieu entre février et juin 2020, c'est-à-dire pendant la période confinement entraînée par la pandémie de la COVID-19. Cette période étant particulièrement perturbatrice, quelques chercheurs n'ont pas pu participer à l'entretien malgré l'intérêt pour notre objet de recherche. Cependant dans d'autres situations, ce contexte s'est avéré

avantageux pour s'entretenir avec des chercheurs de différentes régions canadiennes parce que l'utilisation des médias de communications en ligne comme celui de Zoom s'est répandue très rapidement au cours de cette période.

Sur le plan empirique, les analyses sur les enjeux épistémiques (Chapitre V) et sur les caractéristiques de la pratique de l'épidémiologie numérique (Chapitre IV) ont atteint leur objectif. Les travaux sur les cultures épistémiques de Knorr-Cetina (1999) ont réellement aidé à analyser en détail les aspects d'ordre épistémiques de la pratique de l'épidémiologie numérique et la comparaison entre cette nouvelle pratique et l'épidémiologie traditionnelle en vue d'illustrer des différences qui caractérisent les pratiques comparées.

Nous avons fait face à plus de défis lors des analyses des enjeux sociopolitiques de la pratique de l'épidémiologie numérique (Chapitre VI). Les aspects sociopolitiques sont nombreux en ce qui concerne les enjeux numériques et les recherches abondent sur ces questions. Les propos des participants à notre recherche étaient moins approfondis sur les aspects sociopolitiques de la pratique de l'épidémiologie numérique. Il est possible que ces enjeux auraient été expliqués plus en profondeur auprès d'autres chercheurs pratiquant l'épidémiologie numérique puisque ce type d'enjeu a été mieux élaboré auprès des participants ayant l'expérience de cette nouvelle pratique. Les autres participants n'ayant pas cette expérience ont communiqué un discours plus détaillé sur les enjeux épistémiques de cette nouvelle pratique. Il se peut aussi qu'un travail plus élaboré sur le contexte et les enjeux de l'épidémiologie aurait permis de mieux cerner les enjeux sociopolitiques de la pratique de l'épidémiologie numérique.

Le Chapitre VI permet de bien comprendre les enjeux d'expertise de la pratique de l'épidémiologie numérique parce que ce chapitre informe sur les nouvelles compétences et spécialités requises pour pratiquer l'épidémiologie numérique : il faut une mise à jour de la discipline de l'épidémiologie ainsi que des compétences en informatique et en science des données, mais aussi des compétences non techniques telles que la gestion des données, la collaboration interdisciplinaire, la protection de la vie privée et la négociation des accords de partage de l'information.

Les contributions sur le plan empirique

La recherche doctorale menée dans le cadre de cette thèse révèle que les enjeux épistémiques de la pratique de l'épidémiologie numérique renvoient à la signification des données et à la confiance envers les dispositifs méthodologiques utilisés dans le cadre de cette nouvelle pratique. Les enjeux sociopolitiques concernent surtout l'expertise en lien avec la pratique de l'épidémiologie numérique et la légitimité de cette pratique émergente. Les chercheurs en épidémiologie expriment un point de vue critique envers la pratique de l'épidémiologie numérique à cause des problèmes liés à la représentativité des échantillons populationnels fondés sur les données numériques ainsi qu'aux incertitudes sur la validité des informations communiquées sur Internet et sur la qualité des données numériques. De plus, les chercheurs en épidémiologie sont grandement préoccupés par les enjeux de formation et de compétences des chercheurs de leur discipline et des futurs chercheurs, les collaborations avec l'informatique, les questions liées à l'éthique et à la protection de la vie privée ainsi que par l'accès aux ressources technologiques requises pour pratiquer l'épidémiologie numérique. Ces enjeux soulevés dans les point de vue critique auprès du groupe de participants à notre recherche montrent que les « prouesses de l'épidémiologie numérique » peuvent impliquer des problèmes au niveau épistémique et entraîner des conséquences individuelles et collectives sur le plan sociopolitique.

Ces enjeux sont importants parce que les méthodes de recherche fondées sur la pratique de l'épidémiologie numérique doivent susciter la confiance auprès des chercheurs pour assurer leur pérennité. Les chercheurs auront confiance envers cette nouvelle pratique si les données sont de qualité suffisante pour arriver à des conclusions valides lors des recherches et si leur utilisation est conforme aux normes éthiques de leur discipline. Cependant, il faut effectuer des recherches pour comprendre les rapports de force entre les disciplines. En effet, la pratique de l'épidémiologie numérique est grandement marquée par l'interdisciplinarité (notamment en ce qui a trait aux collaborations avec l'informatique). Le groupe de participants à notre recherche n'ont pas fait allusion à des rapports de force entre les disciplines lors de la pratique de l'épidémiologie numérique. Au contraire, certains pensent que l'épidémiologie agit comme un chef d'orchestre dans le contexte de cette nouvelle pratique. Rappelons que seulement trois chercheurs pratiquent l'épidémiologie numérique au sein de notre groupe de participants. Cette question sur les rapports

de force doit être analysée lors d'une recherche effectuée auprès d'un nombre plus grand de chercheurs pratiquant l'épidémiologie numérique.

De plus, les résultats de nos analyses montrent que la pratique de l'épidémiologie numérique ne serait pas très présente dans la culture épistémique de la discipline de l'épidémiologie. La raison étant qu'une bonne partie des participants étaient incertains par rapport aux caractéristiques de cette nouvelle pratique et que celle-ci semble peu connue et peu expérimentée par les communautés de chercheurs canadiens en épidémiologie. À ce jour, l'écosystème canadien de la pratique de l'épidémiologie numérique ne semble pas prêt à entamer le point de basculement menant vers son adoption à grande échelle parce que, tel que nous l'avons constaté, plusieurs enjeux restent encore à régler autant sur le plan épistémique que sociopolitique.

Notre recherche montre que sur le plan épistémique, la pratique de l'épidémiologie numérique est crédible pour des recherches exploratoires et des recherches sociales analysant des opinions, mais cette nouvelle pratique est moins appropriée pour des recherches cliniques pour déterminer l'incidence réelle des problèmes de santé et pour assurer la surveillance des maladies. La culture épistémique de la pratique de l'épidémiologie numérique serait très différente par rapport à la culture épistémique de l'épidémiologie clinique parce que la première se concentrerait sur des recherches d'opinions et des tendances sociales alors que la seconde porterait sur des recherches cliniques basées sur des données médicales des patients.

Nos constats montrent que les chercheurs en épidémiologie n'ayant pas l'expérience de la pratique de l'épidémiologie numérique expriment une plus grande confiance envers les données traditionnelles comme celles en provenance des systèmes de santé et d'institutions officielles de la statistique comparativement aux sources de données numériques. Pour les données numériques, nos analyses révèlent une confiance plus élevée à l'égard des données générées par les téléphones cellulaires, suivies par les données générées par les moteurs de recherche Internet. Les chercheurs ayant participé à la recherche de terrain ont exprimé un niveau de confiance généralement plus bas envers les données générées par les médias sociaux.

En effet, les signes et les apparences perceptibles à travers les analyses des données numériques peuvent conduire vers des résultats erronés, et ainsi vers des perceptions déformées des

phénomènes étudiés. De plus, il se peut que ce type d'analyse amène le chercheur à voir ou à ne pas voir certains phénomènes à cause de la façon dont ces dispositifs sont conçus. Les incertitudes du dispositif méthodologique des médias sociaux seraient liées à des aspects humains alors que les incertitudes liées à d'autres dispositifs méthodologiques comme celui des moteurs de recherche Internet seraient davantage liées aux aspects technologiques.

Notre recherche témoigne aussi que les données non traditionnelles comme celles générées par les médias sociaux sont plutôt associées à des problèmes de représentativité populationnelle, car l'exclusion de certaines populations comme les personnes âgées peut engendrer une fausse image de la réalité des enjeux de santé touchant l'ensemble des populations. Les questionnements d'ordre épistémique semblent se situer au niveau des signes et des apparences secondaires du dispositif susceptibles de confronter le chercheur à des biais, à des bruits et à des distorsions à cause de comportements humains comme l'utilisation intensive ou le refus d'utilisation des médias sociaux. Ces comportements pourraient émettre des signes inappropriés et amener des apparences trompeuses lors des analyses de données.

Les algorithmes derrière les moteurs de recherche Internet constituent des dispositifs médiateurs, c'est-à-dire des dispositifs agissant comme intermédiaires entre les phénomènes étudiés lors des recherches et le chercheur selon les constats établis lors de notre recherche. Les mesures et les données sont contingentes à l'appareil de mesure et celles-ci sont des représentations de cet appareil plutôt que du monde réel (Knorr-Cetina, 1999). Il s'agit d'un enjeu épistémique majeur pour l'adaptation de l'appareillage analytique et interprétatif des données lors des recherches épidémiologiques numériques. Diverses stratégies et expertises provenant de différents secteurs seront nécessaires à ce processus d'adaptation démontrant ainsi l'intérêt d'associer l'étude des enjeux épistémiques aux enjeux sociopolitiques de la pratique de l'épidémiologie numérique.

Notre recherche révèle que sur le plan sociopolitique, la pratique de l'épidémiologie numérique est très marquée par des enjeux d'expertise et de légitimité. Les enjeux d'expertise liés à la pratique de l'épidémiologie numérique sont les nouvelles compétences et les spécialités envisagées, la formation des chercheurs en épidémiologie et l'interdisciplinarité selon les constats établis lors de notre recherche. Sur le plan institutionnel, cette nouvelle pratique serait considérée comme

importante et nécessaire. Cependant, il semble y avoir peu d'action entreprise afin d'intégrer cette pratique dans les universités à part quelques initiatives lancées de manière sporadique dans certains départements d'épidémiologie.

Les méthodes de recherche et les données liées à la pratique de l'épidémiologie numérique peuvent influencer les pratiques de recherche, susciter de nouveaux besoins en matière de compétences en épidémiologie et influencer les dynamiques sociales d'après les discours entendus auprès des participants à notre recherche. Les méthodes de recherche et les données ont une vie sociale parce que celles-ci peuvent agir sur les relations sociales (Savage, 2013), notamment en portant atteinte à certaines populations par l'utilisation des données sur leur état de santé ou en contribuant à accroître le pouvoir auprès de grandes compagnies privées détentrices de données numériques.

En s'inspirant de Jasanoff (2004), nos analyses font ressortir des dimensions d'éthique, de valeurs, de légalité et de pouvoir au sein des formations épistémiques, matérielles et sociales constituant la pratique de l'épidémiologie numérique. Ces résultats illustrent aussi les enjeux politiques des données puisqu'il est possible que les données soient des entités capables de reconfigurer les relations entre les gouvernements, les citoyens et les industries (Ruppert *et al.*, 2017). Par exemple, l'accès aux données peut être limité pour les chercheurs parce qu'elles sont la propriété privée de grandes sociétés comme Google ou de compagnies de téléphones cellulaires. Le contrôle d'accès détenu par les propriétaires de données peut alors entraîner des enjeux de confiance dans le contexte de la pratique de l'épidémiologie numérique.

Enfin, ces résultats signifient aussi que le contrôle de grandes sociétés comme Google sur les données peut susciter des doutes envers la légitimité de la pratique de l'épidémiologie numérique au sein des chercheurs en épidémiologie incertains à propos des intérêts poursuivis par ces sociétés. Un tel phénomène illustre l'idée qu'un dispositif méthodologique comme le moteur de recherche Google est matériel parce que celui-ci a des caractéristiques technologiques. Il est aussi hétérogène parce qu'il implique des personnes comme des utilisateurs et des communautés (Law et Ruppert, 2013) en plus des grandes sociétés privées susceptibles d'augmenter leurs pouvoirs avec la propriété des données numériques.

Selon notre recherche, les cadres juridiques, les ressources technologiques et le manque de confiance des gens envers les détenteurs de données et les gouvernements sont des barrières importantes à la légitimité de la pratique de l'épidémiologie numérique. Les conséquences individuelles et collectives de la pratique de l'épidémiologie numérique incluent les risques liés à la sécurité et à la confidentialité des données numériques ainsi que les rapports de pouvoir en faveur des grandes entités privées détenant ces données.

Il ressort alors de cela qu'un tel contexte implique des enjeux techniques nécessitant, d'une part, des compétences en informatique et en science des données et, d'autre part, des enjeux sociopolitiques créant des besoins pour des compétences non techniques telles que la collaboration interdisciplinaire, la protection de la vie privée et la négociation des accords de partage de l'information. L'interdisciplinarité et la transdisciplinarité deviennent alors essentielles lors de la pratique de l'épidémiologie numérique et lors de la formation dans cette discipline. Cependant, des questions demeurent sans réponse quant à la mise en place de l'interdisciplinarité et de la transdisciplinarité dans le contexte de cette pratique.

Les contributions sur le plan théorique

Nous soutenons que la pratique de l'épidémiologie numérique constitue une culture épistémique. La raison étant que cette nouvelle pratique peut faire émerger des façons différentes de prendre connaissance des phénomènes étudiés dans la discipline de l'épidémiologie à cause de l'utilisation de sources de données non traditionnelles et aussi de nouvelles technologies pour faire de l'épidémiologie. Les cultures épistémiques développées par Knorr-Cetina demeurent incontournables pour étudier les nouvelles pratiques de recherche à l'ère numérique comme la pratique de l'épidémiologie numérique. Les cultures épistémiques sont importantes parce qu'elles reflètent différentes façons de faire de la science. Elles permettent aussi de cerner les influences des instruments et des mesures sur les perceptions des phénomènes étudiés lors de la conduite des recherches allant jusqu'à opérationnaliser les interactions entre les chercheurs eux-mêmes et entre les chercheurs et les instruments qu'ils utilisent. De plus, les cultures épistémiques considèrent les particularités de la pratique de l'épidémiologie numérique en matière de méthodes de recherche et d'expertise nécessitant de la distinguer par rapport à d'autres pratiques de recherche.

Cependant, les cultures épistémiques développées par Knorr-Cetina (1999) ne peuvent pas expliquer les enjeux sociopolitiques ni les contextes de la coproduction des connaissances liés à la pratique de l'épidémiologie numérique. Comme nous l'avons mentionné précédemment, les dimensions sociopolitiques de cette nouvelle pratique sont aussi importantes que ses dimensions épistémiques parce que les méthodes et les données ne sont pas seulement des moyens techniques pour faire de la recherche (Savage, 2013; Law *et al.*, 2011; Law, 2010; Lury *et al.*, 2020). En effet, une telle pratique nécessite l'utilisation de méthodes et de données susceptibles d'avoir des impacts significatifs sur la vie des gens lorsqu'ils divulguent leurs informations personnelles sur Internet et de renforcer les pouvoirs de certains acteurs détenteurs des données numériques comme Google et Twitter (Beer, 2009; Mayer-Schönberger et Cukier, 2013; Kitchin, 2014; Zwitter, 2014; Ruppert *et al.*, 2015; Ruckenstein et Granroth, 2020). De plus, les enjeux sociétaux des méthodes de recherche et des données considèrent les pratiques comme faisant partie d'un assemblage ou d'une écologie plus large que la sphère de la recherche (Lury *et al.*, 2020). Ce genre de travaux s'intéresse aussi aux données qu'ils perçoivent comme des objets investis de certains pouvoirs, influences et rationalités faisant ainsi ressortir les enjeux politiques des données (Ruppert *et al.*, 2017) incluant les questions éthiques et légales liées aux données numériques.

Les travaux théoriques s'intéressant aux enjeux sociétaux des méthodes de recherche et des données mettent de l'avant des idées afin de comprendre le contexte global des pratiques de recherche, mais ces travaux ne vont pas assez en profondeur pour opérationnaliser les pratiques de recherche comme le fait Knorr-Cetina (1999) avec les cultures épistémiques. Il faut maintenant opérationnaliser ces cadres théoriques dans le but d'expliquer comment et pourquoi les réseaux se forment lors des recherches ainsi que les relations et interactions entre les acteurs (chercheurs, professionnels, citoyens et profanes), les technologies, les données et les algorithmes. Malgré certaines critiques (Heilbron et Gingras, 2015; Nowotny *et al.*, 2003), les travaux de Gibbons *et al.* (1994) constituent un point de départ intéressant vers l'opérationnalisation des cadres théoriques sur les enjeux sociétaux des méthodes de recherche et des données. Les attributs de la production des connaissances développés par Gibbons *et al.* (1994) aident à identifier les caractéristiques des contextes liés aux nouvelles pratiques de recherche issues de la mise en données de la société. Donc, la complémentarité entre diverses théories et différents concepts constitue un enjeu théorique

fondamental pour l'étude de l'émergence des nouvelles pratiques de recherche à l'ère numérique comme celle de la pratique de l'épidémiologie numérique. Ces nouvelles pratiques sont des phénomènes complexes devant être analysés en tenant compte de plusieurs dimensions comme les disciplines, les pratiques de recherche, les contextes de la coproduction des connaissances, l'interdisciplinarité et la mise en données de la société.

Les principaux éléments de la discussion

Les enjeux identifiés quant à la pratique de l'épidémiologie numérique permettent de faire des hypothèses sur les pratiques de recherche utilisant des données numériques dans différents domaines. À travers les discussions dans le cadre de cette thèse, nous avons affirmé que l'écosystème des nouvelles pratiques de recherche à l'ère numérique est très hétérogène puisqu'il inclut une multitude d'acteurs, des technologies numériques, des données et des algorithmes interagissant sous forme de réseaux fluides et hétérogènes lors des projets de recherche. Cet écosystème est influencé par une tension centrale entre les apports épistémiques de ces nouvelles pratiques et leurs enjeux sociopolitiques. Cette tension est caractérisée, d'une part, par la volonté d'utiliser de nouvelles méthodes de recherche en vue de faire avancer les connaissances des disciplines et, d'autre part, par les préoccupations suscitées par la concentration des pouvoirs des grandes sociétés privées détentrices des données numériques et aux risques liés à la divulgation de certaines données sur les citoyens et utilisateurs des technologies numériques.

Les nouvelles pratiques de recherche à l'ère numérique sont disciplinaires et interdisciplinaires, mais aussi appliquées, hétérogènes et axées sur l'informatique et la modélisation de données. La transdisciplinarité semble présente lors de ces nouvelles pratiques, mais celle-ci doit être étudiée davantage afin d'arriver à des conclusions plus sûres quant à la façon dont cette approche s'intègre réellement dans le travail des chercheurs. Les connaissances produites en épidémiologie numérique complètent celles produites en épidémiologie traditionnelle. En plus de l'épidémiologie, d'autres domaines et disciplines commencent à prendre le tournant numérique, dont le domaine des sciences humaines et la discipline de la sociologie. L'ère numérique en recherche entraînerait des questionnements sur l'avenir de la sociologie. Cependant, ces questionnements semblent peu

présents auprès des chercheurs en épidémiologie quant à l'avenir de leur discipline. Ils considèrent leur expertise comme indispensable lors des recherches à l'ère numérique afin d'évaluer et d'expliquer les résultats des travaux de la science des données.

Les limites de cette recherche et des pistes pour des recherches futures

Nous avons analysé les enjeux épistémiques et sociopolitiques à l'égard de la pratique de l'épidémiologie numérique à travers les points de vue de 36 chercheurs canadiens en épidémiologie. Ce groupe de participants ne représente pas tous les chercheurs canadiens en épidémiologie ni l'ensemble des chercheurs en épidémiologie dans le monde.

Afin de poursuivre les réflexions, il serait intéressant d'effectuer des études comparatives avec d'autres communautés de chercheurs en épidémiologie situées dans divers pays. D'ailleurs, les équipes de recherche dépassent souvent les frontières des pays puisque celles-ci sont souvent composées par des membres en provenance de différentes régions du monde. Il s'avère aussi intéressant d'analyser le contexte organisationnel de l'épidémiologie au Canada et dans le monde pour mieux situer les discours des participants aux recherches. Cependant, ce travail constituerait une recherche en soi puisque le contexte organisationnel en santé publique est large et complexe.

De plus, cette recherche s'est concentrée sur les points de vue de chercheurs en épidémiologie. D'autres intervenants sont aussi importants lors des nouvelles pratiques de recherche de l'ère numérique comme les organismes produisant des données officielles, les informaticiens, les professionnels des politiques publiques et les chercheurs d'autres disciplines. Dans le but d'approfondir cette analyse, nous proposons d'étudier les enjeux épistémiques et sociopolitiques auprès de différents intervenants. Cette démarche permettrait de saisir les impacts et les perceptions des nouvelles pratiques de recherche à l'ère numérique auprès de diverses communautés.

D'autres recherches doivent être effectuées quant au rôle des organismes produisant des données officielles dans le contexte des nouvelles pratiques de recherche à l'ère numérique. Il faut aussi examiner la transdisciplinarité et la diversification des sites de la création des connaissances lors de ces nouvelles pratiques de recherche, notamment en ce qui concerne la division du travail entre

les chercheurs, experts professionnels et profanes. Les études sur les nouvelles pratiques de recherche à l'ère numérique sont très prometteuses parce qu'elles sont conditionnées par un écosystème complexe d'acteurs, de technologies, de données et d'algorithmes suscitant ainsi de nombreuses questions autant sur les aspects épistémiques et sociétaux à cause de la diversité des compétences, que sur les objectifs et les intérêts auprès des intervenants concernés par ces nouvelles pratiques de recherche.

Sous un autre angle, nous proposons aussi des études de laboratoire auprès d'entités dédiées à la pratique de l'épidémiologie numérique en vue de comprendre, de manière approfondie, comment se fait la pratique de l'épidémiologie numérique et de cerner ses contributions et ses défis au quotidien dans le cadre des recherches en santé publique. Ces efforts permettraient d'approfondir les connaissances sur les façons de pratiquer l'épidémiologie numérique, sur les apports de cette nouvelle pratique dans les activités quotidiennes des chercheurs et sur les besoins qu'elle suscite en matière de ressources technologiques, d'habiletés et de compétences. De telles initiatives permettraient aussi de comprendre les réseaux de collaboration entre les chercheurs et les acteurs à l'extérieur des organismes de recherche lors de la pratique de l'épidémiologie numérique, notamment dans les contextes transdisciplinaires de la science citoyenne et de la production participative (*crowdsourcing*) impliquant des acteurs experts et profanes de différents secteurs.

À ce jour, cette nouvelle pratique présente des faiblesses quant à l'échantillonnage populationnel, aux incertitudes sur l'interprétation des tendances visibles à travers les données numériques, à l'éthique, aux risques d'atteinte à la vie privée des gens et à l'augmentation du pouvoir des grandes sociétés détentrices des données numériques. Toutefois, certains travaux analysant les utilisations de nouvelles sources de données à des fins de recherche dans le domaine de la santé montrent que cette pratique détient un grand potentiel pour compléter ou améliorer les pratiques courantes en épidémiologie. Par exemple Bour *et al.* (2021) expliquent que les données générées par les médias sociaux peuvent être collectées et étudiées pour améliorer la surveillance des maladies et de la santé publique notamment lors des études sur les maladies transmissibles comme celle de la grippe. Ainsi, la pratique de l'épidémiologie numérique peut amener la création de connaissances nouvelles nécessaires à la gestion efficace des enjeux actuels en santé des populations et en environnement.

ANNEXE A
LE GUIDE D'ENTRETIEN

Renée Carpentier, étudiante-chercheuse
Université du Québec à Montréal
Programme de doctorat en Science, technologie et société

Projet de thèse intitulé

Transformations numériques et expertise : le cas de l'épidémiologie numérique

Thèmes/questions principales	Questions complémentaires
<p>Accueil</p>	<p>Objectif de la recherche :</p> <p>Connaître le point de vue de chercheurs épidémiologistes au sujet de l'épidémiologie numérique afin de comprendre comment les technologies numériques affectent la pratique de l'épidémiologie.</p> <p>Modalités de participation :</p> <p>Les participants sont demandés d'expliquer leur opinion sur l'épidémiologie numérique en répondant oralement aux questions formulées dans le cadre de cet entretien.</p> <p>Consentement :</p> <p>Lecture du formulaire de consentement (anonymat, confidentialité, possibilité de se retirer de la recherche en tout temps)</p> <p>Les questions du participant</p> <p>Signature de deux copies du formulaire de consentement. Pour les entretiens à distance, les formulaires signés seront numérisés et envoyés par courrier électronique à chaque participant et à</p>

Thèmes/questions principales	Questions complémentaires
	l'étudiante chercheuse au carpentier.renee@courrier.uqam.ca
<p>Introduction</p> <p>A. J'aimerais que vous me parliez de votre travail en épidémiologie</p>	<p>A.1 Quelle formation et expérience avez-vous dans le domaine de l'épidémiologie ?</p> <p>A.2 Sur quelles maladies ou enjeu de santé publique vos recherches portent-elles récemment ?</p> <p>A.3 Est-ce que vous travaillez sur des recherches portant sur les systèmes informatiques de santé ?</p> <p>A.4 Est-ce que vous travaillez sur des recherches portant sur les méthodes statistiques ?</p> <p>A.5 Est-ce vous travaillez sur des recherches portant sur l'intelligence artificielle ?</p> <p>A.6 Quelles banques de données utilisez-vous dans vos travaux de recherches ?</p> <p>A.7 Est-ce que vous avez déjà utilisé l'épidémiologie numérique dans le cadre de vos recherches ?</p> <p>A.8 Est-ce que vous prévoyez contribuer à des projets en épidémiologie numérique au cours des prochaines années ?</p> <p>A.9 Quels sont vos projets en cours ?</p> <p>A.10 Quels seraient vos projets futurs ?</p>
<p>Thème 1 : Les pratiques de recherche</p> <p>1.1 Comment percevez-vous l'impact des technologies numériques sur la recherche en épidémiologie ?</p>	<p>1.1.1 Quels changements (positifs et négatifs) les technologies numériques ont-elles apporté pour les recherches en épidémiologie ?</p> <p>1.1.2 Quelles technologies numériques peuvent influencer les pratiques de recherche en épidémiologie ?</p>

Thèmes/questions principales	Questions complémentaires
<p>1.2 Quel est votre point de vue sur l'utilisation de l'épidémiologie numérique dans les pratiques de recherches épidémiologiques ?</p>	<p>1.1.3 Pour quelles activités de recherche les technologies numériques peuvent-elles avoir une influence dans le domaine de l'épidémiologie ?</p> <p>1.1.4 Qu'est-ce que les technologies numériques peuvent apporter pour les recherches en épidémiologie ?</p> <p>1.1.5 Quelle place occupent les technologies numériques par rapport aux autres technologies dans vos recherches ?</p> <p>1.1.6 Quelles précautions doit-on prendre vis-à-vis des technologies numériques pour les recherches en épidémiologie ?</p> <p>1.2.1 De quelle façon l'épidémiologie numérique pourrait changer des pratiques de recherche en épidémiologie ?</p> <p>1.2.2 Comment voyez-vous l'utilité de l'épidémiologie numérique pour les recherches épidémiologiques ?</p> <p>1.2.3 Qu'en est-il de la fiabilité de l'épidémiologie numérique pour les recherches épidémiologiques ?</p> <p>1.2.4 Pour quel(s) objet(s) de recherche épidémiologique il serait souhaitable d'utiliser l'épidémiologie numérique ?</p> <p>1.2.5 Pour quel(s) objet(s) de recherche épidémiologique il vaudrait mieux éviter d'utiliser l'épidémiologie numérique ?</p> <p>1.2.6 Quels avantages offre l'épidémiologie numérique selon vous ?</p> <p>1.2.7 Quels problèmes implique l'épidémiologie numérique selon vous ?</p>

Thèmes/questions principales	Questions complémentaires
1.5 Quels éléments d'information pourraient être ajoutés pour ce thème sur les pratiques ?	1.5.1 Quels éléments d'information avons-nous oublié de mentionner sur les pratiques de recherche ?
<p data-bbox="199 453 618 516">Thème 2 : Les perceptions sur l'épidémiologie numérique</p> <p data-bbox="199 562 618 653">2.1 Que représente pour vous l'épidémiologie numérique dans le cadre de votre travail ?</p> <p data-bbox="199 1304 618 1451">2.2 Selon vos connaissances, quels sont les mécanismes de diffusion des recherches en épidémiologie ?</p> <p data-bbox="199 1759 618 1881">2.3 Quelle place pouvez-vous envisager pour l'épidémiologie numérique dans les programmes d'études visant à</p>	<p data-bbox="638 453 1243 516">2.1.1 Quelle importance accordez-vous à l'épidémiologie numérique pour votre travail ?</p> <p data-bbox="638 562 1252 625">2.1.2 Quelle importance accorde votre institut à l'épidémiologie numérique selon vous ?</p> <p data-bbox="638 672 1341 762">2.1.3 Quelles conséquences pouvez-vous entrevoir sur le travail scientifique de votre discipline avec l'émergence de l'épidémiologie numérique ?</p> <p data-bbox="638 808 1349 905">2.1.4 Quel intérêt voyez-vous à joindre un réseau de collaboration pour contribuer à des recherches utilisant l'épidémiologie numérique ?</p> <p data-bbox="638 951 1349 1083">2.1.5 Quelle place est accordée à l'épidémiologie numérique dans les recommandations d'apprentissage et de formation de votre institut en tant que membre du personnel ?</p> <p data-bbox="638 1129 1349 1226">2.1.6 Comment voyez-vous la place de l'épidémiologie numérique vis-à-vis des approches habituelles de recherche reconnues par votre institut ?</p> <p data-bbox="638 1304 1317 1367">2.2.1 Quelles revues publient les recherches utilisant l'épidémiologie numérique ?</p> <p data-bbox="638 1413 1260 1509">2.2.2 Quel est le facteur d'impact des revues qui publient des travaux utilisant l'épidémiologie numérique selon vous ?</p> <p data-bbox="638 1556 1317 1688">2.2.3 Quels forums de discussions et de transmission d'information (sites web, conférences d'experts ou autres) sont disponibles sur l'épidémiologie numérique ?</p> <p data-bbox="638 1766 1325 1862">2.3.1 De quelle façon l'épidémiologie numérique pourrait influencer certains contenus des programmes d'études en épidémiologie de votre institut ?</p>

Thèmes/questions principales	Questions complémentaires
<p>développer les compétences des futurs experts en épidémiologie ?</p> <p>2.4 Comment peut-on définir l'épidémiologie numérique ?</p> <p>2.5 Quels éléments d'information pourraient être ajoutés pour ce thème sur les perceptions vis-à-vis de l'épidémiologie numérique ?</p>	<p>2.3.2 Selon vous, quelle position devrait adopter votre institut quant à l'adaptation des programmes existants pour y inclure l'épidémiologie numérique ?</p> <p>2.3.3 Quel niveau d'intérêt pouvez-vous déduire auprès des étudiants actuels pour apprendre sur l'épidémiologie numérique ?</p> <p>2.4.1 Selon vos perceptions et les discussions précédentes, quelle est votre définition de l'épidémiologie numérique ?</p> <p>2.5.1 Quels éléments d'information avons-nous oublié de mentionner sur les perceptions vis-à-vis de l'épidémiologie numérique ?</p>
<p>Thème 3 : Le cas pratique de <i>Google Flu Trends</i> et de <i>FoodRepo</i></p> <p>3.1 Qu'est-ce que vous pensez de <i>Google Flu Trends</i> et de <i>FoodRepo</i> en tant que technologie pour des recherches en épidémiologie ?</p> <p>3.2 Quels éléments d'information pourraient être ajoutés pour ce thème sur <i>Google Flu Trends</i> et <i>FoodRepo</i> ?</p>	<p>3.1.1 Comment voyez-vous le potentiel de <i>Google Flu Trends</i> pour contribuer à des études valides sur la grippe ?</p> <p>3.1.2 Comment voyez-vous le potentiel de <i>Food Repo</i> pour contribuer à des études valides sur des enjeux de santé liés à la nutrition ?</p> <p>3.1.3 Quels motifs guident votre perspective sur <i>Google Flu Trends</i> ?</p> <p>3.1.4 Quels motifs guident votre perspective sur <i>FoodRepo</i> ?</p> <p>3.1.5 Comment jugez-vous la qualité de la contribution d'une de ces technologies par rapport à l'autre pour des recherches épidémiologiques ?</p> <p>3.2.1 Quels éléments d'information avons-nous oublié de mentionner sur <i>Google Flu Trends</i> et <i>FoodRepo</i> ?</p> <p>3.2.2 Quels sont les autres cas pratiques qui seraient intéressants à examiner ?</p>

Thèmes/questions principales	Questions complémentaires
<p>Thème 4 : Le futur de l'épidémiologie numérique</p> <p>4.1 Quel avenir voyez-vous pour l'épidémiologie numérique en tant que pratique de recherche ?</p> <p>4.2 Quels éléments d'information pourraient être ajoutés pour ce thème sur le futur de l'épidémiologie numérique ?</p>	<p>4.1.1 Quel(s) intérêt(s) les chercheurs en épidémiologie ont-ils pour les travaux de recherches portant sur l'épidémiologie numérique ?</p> <p>4.1.2 Quelles connaissances et compétences devraient selon vous développer les chercheurs épidémiologistes pour travailler sur des projets de recherche impliquant l'épidémiologie numérique ?</p> <p>4.1.3 Quelles stratégies seraient nécessaires pour mieux bénéficier de l'épidémiologie numérique ?</p> <p>4.1.4 Quels seraient les enjeux les plus importants pour l'épidémiologie avec l'émergence de l'épidémiologie numérique ?</p> <p>4.1.5 Quelle crédibilité scientifique pouvez-vous envisager pour l'épidémiologie numérique ?</p> <p>4.2.1 Quels éléments d'information avons-nous oublié de mentionner sur le futur de l'épidémiologie numérique ?</p>

Interview Guide

Renée Carpentier, Student-Researcher

Université du Québec à Montréal

Doctoral Program in Science, Technology and Society

Thesis Project Entitled

Digital transformations and expertise: the case of digital epidemiology

Topic/Main Questions	Complementary Questions
Welcome Remarks	<p>Objective of the research:</p> <p>To gain insight from epidemiological researchers about digital epidemiology in order to understand how digital technologies affect the practice of epidemiology.</p> <p>Terms and conditions of participation:</p> <p>Participants are asked to explain their opinion on digital epidemiology by answering orally to the questions formulated in this interview.</p> <p>Consent:</p> <p>Reading the consent form (anonymity, confidentiality, ability to withdraw from the research at any time)</p> <p>Participant's questions</p> <p>Signature of two copies of the consent form. For distance interviews, the signed forms will be scanned and sent by e-mail to each participant and to the student-researcher at carpentier.renee@courrier.uqam.ca.</p>
<p>Introduction</p> <p>A. I would like you to tell me about your work in epidemiology</p>	<p>A.1 What training and experience do you have in the field of epidemiology?</p> <p>A.2 What diseases or public health issues have you been researching recently?</p> <p>A.3 Are you working on research related to health informatics systems?</p> <p>A.4 Are you working on research related to statistical methods?</p> <p>A.5 Are you working on research related to artificial intelligence?</p> <p>A.6 Are you working on research related to the operation of data banks?</p>

Topic/Main Questions	Complementary Questions
	<p>A.7 Have you ever used digital epidemiology in your research?</p> <p>A.8 Do you plan to use digital epidemiology in any of your work in the coming years?</p> <p>A.9 What are your current projects?</p> <p>A.10 What are your future plans?</p>
<p>Topic 1: Research Practices</p> <p>1.1 How do you perceive the impact of digital technologies on epidemiological research?</p> <p>1.2 What are your views on the use of digital epidemiology in epidemiological research practices?</p>	<p>1.1.1 What level of impact can digital technologies have on epidemiological research?</p> <p>1.1.2 What digital technologies can influence epidemiological research practices?</p> <p>1.1.3 For which research activities can digital technologies influence epidemiological research?</p> <p>1.1.4 What can digital technologies bring to epidemiological research?</p> <p>1.1.5 How do digital technologies compare to other technologies in your research?</p> <p>1.1.6 What precautions should be taken regarding digital technologies for epidemiological research?</p> <p>1.2.1 How might digital epidemiology change epidemiology research practices?</p> <p>1.2.2 How do you see the usefulness of digital epidemiology for epidemiological research?</p> <p>1.2.3 What about the reliability of this approach for epidemiological research?</p>

Topic/Main Questions	Complementary Questions
<p>1.3 How do you see the quality of the digital data?</p> <p>1.4 How do you see the workflow in research using digital epidemiology?</p>	<p>1.2.4 For which object(s) of study would it be desirable to use digital epidemiology?</p> <p>1.2.5 For which object(s) of study it would be better to avoid using numerical epidemiology?</p> <p>1.2.6 What advantages do you think digital epidemiology offers?</p> <p>1.2.7 What problems do you think digital epidemiology implies?</p> <p>1.3.1 How are numerical data different from data commonly used for epidemiological research?</p> <p>1.3.2 How much confidence do you place in epidemiological research that has used data from social networks such as Twitter?</p> <p>1.3.3 How confident are you in epidemiological research that has used data from Internet search engines such as Google Trends?</p> <p>1.3.4 How confident are you in epidemiological research that has used data from mobile devices such as cell phones?</p> <p>1.4.1 What changes would digital epidemiology bring to the way you do your work?</p> <p>1.4.2 What form would partnerships take in such research projects?</p> <p>1.4.3 What would be the role of the research epidemiologist in such research?</p> <p>1.4.4 How do you see the financial support to carry out such research?</p>

Topic/Main Questions	Complementary Questions
<p>1.5 What information could be added for this topic related to research practices?</p>	<p>1.4.5 What resource issues are important to consider in conducting research using digital epidemiology?</p> <p>1.4.6 Who would make the decisions in such research?</p> <p>1.4.7 How would decisions be made in conducting such research?</p> <p>1.5.1 What information did we forget to mention about research practices?</p>
<p>Topic 2: Perceptions of digital epidemiology</p> <p>2.1 What does digital epidemiology mean to you in your work?</p> <p>2.2 To the best of your knowledge, what are the mechanisms for</p>	<p>2.1.1 How important is digital epidemiology to your work?</p> <p>2.1.2 How important do you think digital epidemiology is to your institute?</p> <p>2.1.3 What consequences can you foresee for the scientific work of your discipline with the emergence of digital epidemiology?</p> <p>2.1.4 What interest do you see in joining a collaborative network to contribute to research using digital epidemiology?</p> <p>2.1.5 How does digital epidemiology figure in the learning and training recommendations of your institute as a staff member?</p> <p>2.1.6 How do you see the place of digital epidemiology in relation to the usual research approaches recognized by your institute?</p> <p>2.2.1 What journals publish research using digital epidemiology?</p>

Topic/Main Questions	Complementary Questions
<p>disseminating epidemiological research?</p> <p>2.3 What place can you envisage for digital epidemiology in the curricula aimed at developing the skills of future epidemiology experts?</p> <p>2.4 How can digital epidemiology be defined?</p> <p>2.5 What information could be added for this topic related to perceptions of digital epidemiology?</p>	<p>2.2.2 What is the impact factor for journals that publish work using digital epidemiology in your opinion?</p> <p>2.2.3 What forums for discussion and information transmission (websites, expert conferences, etc.) are available on digital epidemiology?</p> <p>2.3.1 How might digital epidemiology influence some of the content of your institute's epidemiology curricula?</p> <p>2.3.2 In your opinion, what position should your institute take regarding the adaptation of existing curricula to include digital epidemiology?</p> <p>2.3.3 What level of interest can you infer from current students in learning about digital epidemiology?</p> <p>2.4.1 Based on your perceptions and previous discussions, what is your definition of digital epidemiology?</p> <p>2.5.1 What information did we forget to mention about perceptions of digital epidemiology?</p>
<p>Topic 3: The case study of Google Flu Trends and FoodRepo</p> <p>3.1 What do you think of Google Flu Trends and FoodRepo as a technology for epidemiological research?</p>	<p>3.1.1 How do you see the potential of Google Flu Trends to contribute to valid flu studies?</p> <p>3.1.2 How do you see the potential of Food Repo to contribute to valid studies on nutrition-related health issues?</p> <p>3.1.3 What reasons guide your perspective on Google Flu Trends?</p> <p>3.1.4 What reasons guide your perspective on Food Repo?</p>

Topic/Main Questions	Complementary Questions
<p>3.2 What information could be added for this topic related to Google Flu Trends and FoodRepo?</p>	<p>3.1.5 How do you judge the quality of the contribution of one of these technologies compared to the other for epidemiological research?</p> <p>3.2.1 What information did we forget to mention on Google Flu Trends and FoodRepo?</p> <p>3.2.2 What other case studies would be interesting to consider?</p>
<p>Topic 4: The future of digital epidemiology</p> <p>4.1 What future do you see for digital epidemiology as a research practice?</p> <p>4.2 What information could be added for this topic related to the future of digital epidemiology?</p>	<p>4.1.1 How could digital epidemiology become part of standard epidemiological research practice?</p> <p>4.1.2 What knowledge and skills do you think epidemiological researchers should develop to use digital epidemiology?</p> <p>4.1.3 What strategies would be needed to better benefit from digital epidemiology?</p> <p>4.1.4 What would be the most important issues for epidemiology with the emergence of digital epidemiology?</p> <p>4.1.5 What scientific credibility can you envisage for digital epidemiology?</p> <p>4.2.1 What information did we forget to mention about the future of digital epidemiology?</p>

ANNEXE B
LA CODIFICATION

La codification des propos des participants à la recherche de terrain à partir du logiciel Nvivo

Catégorisation et codification	Nombre de participants	Nombre de références
ADOPTION-RÉSISTANCE ENVERS L'ÉPIDÉMIOLOGIE NUMÉRIQUE (ÉN)	0	0
Crédibilité scientifique de l'ÉN	22	29
Comprendre les limites des données	5	5
Critères d'acceptabilité	16	22
Difficultés	4	5
Place occupée dans le travail épidémiologique	4	7
Point tournant-Facteurs décisifs	3	5
Potentiel	17	32
Prise de décisions	1	1
Prudence	7	9
Type de problèmes résolus	1	1
Fiabilité des résultats en ÉN	7	14
Aisance du côté éthique	1	1
Biais	15	22
Comparaison aux données traditionnelles	9	14
Connaissances des limites	3	9
Contrôle de qualité	3	3
Dangers	5	8
Doit être démontrée	7	19
Évolution de la fiabilité	3	5
Fiabilité variable	13	14
Participation aux recherches	1	1
Qualité de la source	5	6
Validation de corrélations	3	3
Niveau de confiance Google Trends	26	44
Aspect syndromique	12	16
Capacité de capter les perceptions (<i>people's mind</i>)	2	3
Comparaison à d'autres sources	5	7
Comprendre le fonctionnement	1	2
Connaissances sur les populations	3	4
Controverses	7	8
Étudier les comportements	1	1
Indicateurs	4	5

Catégorisation et codification	Nombre de participants	Nombre de références
Informer les populations	1	1
Scepticisme	17	28
Universalité	4	4
Utilisateur unique	1	1
Niveau de confiance réseaux sociaux	29	40
Contextualisation de la recherche	11	15
Devoirs méthodologiques	8	9
Minorités bruyantes, bruits	6	6
Paramètres épidémiologiques	13	15
Perceptions du public	8	9
Plateformes	1	1
Tendances	4	4
Niveau de confiance tel cellulaires	32	47
Expositions et résultats	7	7
Mobilité et distance	13	13
CARACTÉRISTIQUES DE L'ÉN	0	0
Acteurs	1	4
Citoyens	2	3
Compagnies de téléphone cellulaire	1	1
Google	4	5
Gouvernement fédéral	2	4
Gouvernements provinciaux	1	1
Instances privées	3	5
Laboratoires	3	4
Paires (<i>peer review</i>)	1	2
Réseaux	3	3
Santé publique	2	3
Twitter	1	1
Définitions proposées de l'ÉN	28	36
Accéder à des BD encodées numériquement et de diverses sources	3	3
Alignement santé des populations et politiques	2	3
Application de technologies	2	2
Au-delà des soins et systèmes de santé	5	7
Big data, prédictions, modélisations	10	13
Chaos, ambiguïtés, désordre	1	1
Diverses sources	3	4
Données en temps réel	4	6
Exclusions de la définition	2	2
Génétique récréative	1	3

Catégorisation et codification	Nombre de participants	Nombre de références
Intégrer des données de divers secteurs et disciplines	1	2
Pertinence du concept	1	7
Santé numérique	2	2
Utilisation de SIG, données géographiques, données env., satellites	5	6
Utilisation d'outils électroniques	1	1
Utilisation d'outils et données numériques en épidémiologie	15	17
Étiquette 'Épidémiologie Numérique'	15	26
Limites et problèmes de l'ÉN	6	13
Aspects statistiques	7	9
Clarté et documentation des limites	4	4
Convergence résultats EN avec épidémiologie traditionnelle	3	6
Données manquantes	2	2
Identifier les cas de contrôle	1	1
Importance du contact humain	1	2
Information au niveau de l'individu vs écologique	2	4
Limites en recherche	13	14
Limites politiques	1	1
Objet de recherche	5	6
Précision de la donnée	4	8
Quantité vs. Qualité	1	1
Représentativité populationnelle	25	44
Problématiques et contexte de santé (voir publications ÉN)	2	13
Activité physique	4	6
Capacités cardiovasculaires	1	1
Changements climatiques	2	2
Contamination environnementale	1	1
Contexte économique, social et infrastructures	3	5
COVID-19	17	37
Cyberintimidation	1	1
Éclotions de maladies	1	1
Espaces verts	3	3
Expositions aux maladies	7	8
Maladie de Lyme	1	1
Maladies chroniques	1	1
Maladies et bactéries transmises par l'environnement	1	2
Maladies et santé mentales	3	4
Maladies respiratoires	2	5

Catégorisation et codification	Nombre de participants	Nombre de références
Maladies transmises par les vecteurs	3	8
Maladies transmises sexuellement	1	3
Pandémie	1	3
Pharmacovigilance	3	5
Politiques de santé	2	6
Populations vulnérables	1	2
Recherche sur les expositions et Recherche sur les résultats	1	2
Risques	7	9
Synthèse de connaissances	3	3
Taux de participation aux recherches	3	3
Vaccination	1	1
Victimisation	1	1
Vieillesse	1	1
Virus du Nil	1	1
CHANGEMENTS	0	0
Changements pratiques recherche	9	16
Augmentation de l'usage tech, données et outils numériques	4	6
Utilisation de différents outils	4	5
Cadres de travail	3	4
Au-delà des méthodes traditionnelles	1	2
Collaboration interdisciplinaire	4	4
Collection de données	7	10
Conflits numériques	1	3
Exploration	4	4
Confusion	2	2
Études pilotes	4	5
Innovation	5	11
Nouvelles hypothèses	2	2
Progression des explorations	3	4
Recherche exploratoire	8	9
Nouvelle méthode de mesure	1	1
Nouvelles approches de recherche	3	3
Paradigme (incluant <i>data-driven paradigm</i>)	2	3
Partage de données	2	2
Partenariats privé-public	2	2
Quantité des travaux de recherche	1	1
Suivis de participants, collecte de données	2	3
Contenu des travaux de recherche	14	18

Catégorisation et codification	Nombre de participants	Nombre de références
Aller vers la complexité	4	6
Causalité	2	2
Différentes questions de recherche	2	2
Exactitude des données	1	2
Intégration de la dimension spatiale	2	2
Niveau de connaissances des populations étudiées	1	1
Nouvelles hypothèses	4	4
Opportunités	12	23
Utilisation de grands volumes de données	1	1
Vitesse de collecte des données	1	2
CONTINUITÉS	11	28
Analyses et Déterminants	4	7
Analyses temporelles et saisonnières	1	2
Éclotions de maladies	1	1
Facteur de risque	4	4
Facteurs environnementaux	3	3
Indicateurs	3	3
Nombre d'appels aux lignes de santé	1	1
Nombre et raisons des visites aux urgences	1	1
Nouveaux sons de cloche	3	4
Sentiments	1	1
Changement versus continuité	2	2
<i>Digital divide</i>	1	1
Emphase sur question de recherche	25	80
Épidémiologie de santé publique versus épidémiologie clinique	7	15
Épidémiologie environnementale	5	17
Apport complémentaire	3	5
Côté populationnel	3	4
Évaluation et analyses des risques	6	7
Expérience de l'approche numérique	2	6
Granularité des données	1	1
Information géographique	4	6
Maladies vectorielles	1	3
Pollution de l'air	3	4
Méthodes de recherche en santé	3	3
Methodologies	6	16
Protocoles de recherche	5	7
Rôle de l'épidémiologiste en ÉN	4	6
Caractérisation des limites des données (rigueur)	9	14

Catégorisation et codification	Nombre de participants	Nombre de références
Comprendre la causalité	3	5
Connaître les méthodes	6	7
Dégager le sens, conception, représentation, réflexion	4	13
Élaborer les questions et directions de recherche	10	15
Rôle d'intégration et interprétation des données	8	13
Valider	10	15
EXPERTISE	0	0
Compétences dans les équipes	3	4
Enjeux d'expertise en ÉN	6	16
Capacités et compréhension de la mesure	5	7
Chef d'orchestre	8	9
Connaissance de l'ÉN	25	52
Aucune connaissance	21	29
Entendu parler, vu dans des études	10	15
Expérience concrète	6	6
Expertise spécialisée	19	33
Gestion de données	7	11
Habilités de consultation	1	1
Habilités en communication	2	2
Habilités en informatique	14	21
Informatique	24	51
Jugement	6	6
Langage	2	2
Nouvelles compétences	7	11
Spécialiste en science des données	3	4
Épidémiologie nutritionnelle (nutrition)	6	12
FoodRepo	14	20
Problématiques de nutrition	3	6
Interprétation des données	13	19
Science citoyenne	1	2
Signaux en santé publique	2	3
Signaux conventionnels	1	2
Signaux numériques	6	9
Signaux syndromiques	2	4
Utilité de l'ÉN	7	13
Assistance ou intervention écologique momentanée	2	3
Atteindre des populations plus larges ou difficile à atteindre	7	10
Comprendre l'environnement social	1	1
Comprendre les comportements, traçage	4	6

Catégorisation et codification	Nombre de participants	Nombre de références
Comprendre les prédictions	2	2
Connaître les opinions, discours	3	3
Contextualiser des résultats	2	3
Contrôle, surveillance des maladies	3	3
Davantage d'objectivité	1	1
Davantage explicite	2	2
Données empiriques	12	13
Efficienc e, efficacité	15	24
Évaluer la qualité de l'air	1	1
Évaluer l'impact de politiques	1	1
Imagerie	1	2
Indicateurs précoces	2	3
Infrastructures de santé limitées	1	1
Lecture automatisée	1	2
Recrutement des participants aux recherches	3	4
Repérer les changements saisonniers et temporels	1	1
Répondre à des questions impossibles à répondre avant	12	20
Ressembler à la vérité	4	4
Santé mentale	2	2
Spatialisation des résultats	4	7
Utilisation avec épidémiologie traditionnelle	2	2
LÉGITIMITÉ DE L'ÉN	0	0
Accessibilité à l'ÉN (aussi pour l'expertise)	5	16
Accès aux données et disponibilité	14	22
Accès aux technologies	5	8
Barrières	13	29
Barrières organisationnelles	6	9
Cadres légaux liés aux données	4	7
Complexité, liens entre données	10	11
Dépendance de la qualité des données	13	25
Financement, ressources	13	18
Habilités techniques	7	11
Nettoyage, traitement des données	3	5
Propriété des données	9	10
Routines	6	9
Apprentissage sur les nouvelles données et approches	6	14
Confiance envers les données	1	3
Confiance relative au public	2	5
Contextualisation de la recherche	12	20

Catégorisation et codification	Nombre de participants	Nombre de références
Caractéristiques de la source, des données, acteurs impliqués	7	7
Contexte local des recherches	3	5
Effets possibles au public	1	1
Transparence sur les limites et méthodes	7	8
Enjeux éthiques	22	40
Anonymisation	1	1
Assemblage de données	1	1
Cadres politiques et légaux	2	3
Confidentialité, vie privée	7	7
Consentement	3	7
Dispositions requises	3	3
Éducation	4	4
Intégrité	2	2
Possession des données	4	5
Processus	2	5
Formation, connaissances en ÉN	14	24
Interdisciplinarité (lunette interdisciplinaire)	17	28
Compétences du domaine	4	6
Engagements interdisciplinaires	7	9
Travail d'équipe	8	11
Partenariats	22	37
Qualité du design de l'étude	8	16
Situations et agents facilitant l'accès (<i>thinking outside the box</i>)	6	9
Validation des données numériques	5	6
Connaître les réalités sur les données	2	2
Devoirs par rapport à de nouvelles données	6	7
PERCEPTIONS ENVERS L'ÉN	4	10
Complémentarité approches trad et numériques	11	24
Degré d'importance	30	55
Données traditionnelles vs données numériques	7	12
Collecte pour objectifs de recherche	4	4
Désinformation	7	12
Données produites par les citoyens	1	1
Données structurées, contrôlées versus désordonnées, bruyantes	12	16
Frontière données trad vs numériques	3	4
Importance de la source de données	3	7
Rigueur et validité	19	42

Catégorisation et codification	Nombre de participants	Nombre de références
Utilisation	5	6
Importance dans les programmes d'études (aussi dans Expertise)	19	37
Perspective générationnelle	8	15
Présence de l'ÉN maintenue en recherche	2	3
Progrès scientifique versus inertie	9	21
Promesses, attentes	9	15
Qualité (parfaite) des données	5	6
Révolution en épidémiologie	7	15
Pressing Issues for DE	25	25
RESSOURCES EN ÉPIDÉMIOLOGIE	0	0
Sources de données	3	5
<i>Crowd sourcing</i>	1	1
Données de laboratoires	1	1
Données ouvertes	2	2
Données primaires	9	11
Données secondaires	6	6
Dossiers électroniques de santé	3	3
Organismes détenteurs de données	4	6
Programme Sentinelle	1	1
Registres	2	2
Systèmes de rapports	1	1
Technologies	0	0
Applications de traçage	4	4
Capacités techniques	1	2
Capteurs	2	4
Logiciels d'analyse de données	8	9
Outils mobiles	2	3
Plateformes	1	1
Satellites	4	4
Systèmes d'information géographique (SIG)	2	2
Tableau	1	1
Technologies répandues	1	3
Téléphones Intelligents	4	10
Types de données	1	8
BD d'études de cohortes	5	6
Données administratives	7	7
Données cellulaires	1	1
Données cliniques (cohortes)	1	1

Catégorisation et codification	Nombre de participants	Nombre de références
Données de surveillance (provinces, fédéral, laboratoires)	5	7
Données écologiques	7	12
Données environnementales	2	2
Données générées à partir du Web	1	1
Données géographiques	7	7
Données populationnelles et facteur de risque	1	1
Données satellitaires	3	3
Médias sociaux	6	9
Sondages, Recensements, Enquêtes	7	8
SCEPTICISME	12	23
TRANSFORMATIONS NUMÉRIQUES	3	7
Adaptation	9	10
Avancements	2	2
Expertise technique	1	1
Nouvelles générations	1	1
Problèmes rencontrés	10	20
Types de transformation numérique	0	0
Analyse d'imagerie	1	1
Applications Cloud	3	3
Automatisation	4	6
Capteurs de mouvements, vélo mètres	2	2
Conditions de pérennité	2	6
Dispositifs médicaux	2	3
Grands ensembles de données	1	2
Intelligence artificielle (algorithmes)	6	9
<i>Machine learning</i>	9	14
Marqueurs de mesure	1	1
Plateformes pour transférer les données	2	2
Site web	1	3
Sondages, questionnaires en ligne	8	9
Techniques de modélisations	1	1
Technologies numériques de communication	4	5
Utilité des transformations numériques	2	2
Amélioration de la mesure	3	4
Amélioration des processus versus contenu des recherches	1	1
Analyse de données géographiques interactives	1	1
Collecte des données	3	5
Utilité pour essais cliniques	1	2

BIBLIOGRAPHIE

- Adawi, M., Bragazzi, N. L., Watad, A., Sharif, K., Amital, H. et Mahroum, N. (2017). Discrepancies Between Classic and Digital Epidemiology in Searching for the Mayaro Virus: Preliminary Qualitative and Quantitative Analysis of Google Trends. *JMIR public health and surveillance*, 3(4), e93. <https://doi.org/10.2196/publichealth.9136>
- Aicrowd. (2023). *Crowdsourcing AI to Solve Real-World Problems*. <https://www.aicrowd.com/>. Consulté le 1^{er} avril 2023
- Aldrich T. E. (2000). Environmental epidemiology forward. *Chemosphere*, 41(1-2), 59–67. [https://doi.org/10.1016/s0045-6535\(99\)00390-2](https://doi.org/10.1016/s0045-6535(99)00390-2)
- Angius, Nicola, Giuseppe Primiero et Raymond Turner. (2024). "The Philosophy of Computer Science", *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), forthcoming URL = <https://plato.stanford.edu/archives/sum2024/entries/computer-science/>
- Aramaki, E., Maskawa, S. et Morita, M. (2011). Twitter catches the flu: Detecting influenza epidemics using twitter. Paper presented at the EMNLP 2011 - *Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1568-1576. <https://aclanthology.org/D11-1145>
- Armstrong, M. D, Maupomé, D. et Meurs, M.-J. (2021). Topic Modeling in Embedding Spaces for Depression Assessment. *Proceedings of the Canadian Conference on Artificial Intelligence*. <https://doi.org/10.21428/594757db.9e67a9f0>
- Association of Atlantic Universities, Council of Ontario Universities, Individual institutions, and Bureau de coopération interuniversitaire (2022). *Enrolment by university*. Universities Canada. <https://www.univcan.ca/universities/facts-and-stats/enrolment-by-university/>. Consulté le 4 novembre 2023
- Bakker, K. M., Martinez-Bakker, M. E., Helm, B. et Stevenson, T. J. (2016). Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24), 6689–6694. <https://doi.org/10.1073/pnas.1523941113>
- Bar-Lev, S., Reichman, S. et Barnett-Itzhaki, Z. (2021). Prediction of vaccine hesitancy based on social media traffic among israeli parents using machine learning strategies. *Israel Journal of Health Policy Research*, 10(1). <https://doi.org/10.1186/s13584-021-00486-6>
- Basteris, A., Mansourvar, M. et Kock Wiil, U. (2020). Google Trends and Seasonal Effects in Infodemiology: A Use Case About Obesity. *Studies in health technology and informatics*, 272, 245–248. <https://doi.org/10.3233/SHTI200540>

Bastin, G. et Francony, J. (2016). L'inscription, le masque et la donnée: Datafication du web et conflits d'interprétation autour des données dans un laboratoire invisible des sciences sociales. *Revue d'anthropologie des connaissances*, 10,4, 505-530. <https://doi.org/10.3917/rac.033.0505>

Bates M. (2017). Tracking Disease: Digital Epidemiology Offers New Promise in Predicting Outbreaks. *Institute of Electrical and Electronics Engineers (IEEE) pulse*, 8(1), 18–22. <https://doi.org/10.1109/MPUL.2016.2627238>

Beer, D. (2016). Circulation. Dans D. Beer (dir.), *Metric Power* (p. 77-125). UK : Palgrave. Macmillan. <https://link.springer.com/book/10.1057/978-1-137-55649-3>

----- (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, 11(6), 985–1002. <https://doi.org/10.1177/1461444809336551>

----- (2017). The data analytics industry and the promises of real-time knowing: perpetuating and deploying a rationality of speed. *Journal of Cultural Economy*, 10(1), 21–33. <https://doi.org/10.1080/17530350.2016.1230771>

Ben-Shlomo, Y., Brookes, S. et Hickman, M. (2013). *Epidemiology, evidence-based medicine and public health*. John Wiley & Sons. <https://www.wiley.com/en-us/Epidemiology%2C+Evidence+based+Medicine+and+Public+Health%2C+6th+Edition-p-9781118463567>

Benyei, P., Pardo-de-Santayana, M., Aceituno-Mata, L., Calvet-Mir, L., Carrascosa-García María, Rivera-Ferre, M., Perdomo-Molina, A., et Reyes-García Victoria. (2021). Participation in citizen science: insights from the conect-e case study. *Science, Technology, & Human Values*, 46(4), 755–788. <https://doi.org/10.1177/0162243920948110>

Beraldo, D. et Milan, S. (2019). From data politics to the contentious politics of data. *Big Data & Society*. <https://doi.org/10.1177/2053951719885967>

Bonneuil, C. et Joly, P. (2013). I. Sciences, innovations et économie : des sciences néolibérales ?. Dans: Christophe Bonneuil (dir.), *Sciences, techniques et Société* (p. 21-36). Paris: La Découverte. <https://www-cairn-info.proxy.bibliotheques.uqam.ca/sciences-techniques-et-societe-9782707150974.htm>

Boullier, D. (2015). Les sciences sociales face aux traces du big data: Société, opinion ou vibrations? *Revue Française De Science Politique*, 65(5-6), 805–828. <https://doi.org/10.3917/rfsp.655.0805>

Bour, C., Ahne, A., Schmitz, S., Perchoux, C., Dessenne, C. et Fagherazzi, G. (2021). The use of social media for health research purposes: Scoping review. *Journal of Medical Internet Research*, 23(5). e25736. <https://doi.org/10.2196/25736>

Bourdieu, P. (2001). Science de la science et réflexivité : cours du Collège de France (2000-2001) (Ser. Cours et travaux). Raisons d'agir. <https://www.raisonsdagir-editions.org/catalogue/science-de-la-science-et-reflexivite/>

- Boyd, Dahah et Kate Crawford. (2012). Critical Questions for big data. *Information, Communication & Society*, 15:5, 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Brunekreef, B. (2008). Environmental epidemiology and risk assessment. *Toxicology Letters*, 180(2), 118–122. <https://doi.org/10.1016/j.toxlet.2008.05.012>
- Carneiro, H. A. et Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 49(10), 1557–1564. <https://doi.org/10.1086/630200>
- Castells, Manuel. (2001). La révolution des technologies de l'information. Dans *La société en réseaux. L'ère de l'information* (p. 53-108). Nouvelle édition. Paris : Fayard. <https://www.fayard.fr/sciences-humaines/la-societe-en-reseaux-tome-1-lere-de-linformation-9782213608457>
- Centre for Research on Socio-Cultural Change (CRSCC). (2022). *Social Life of Methods (SLOM)*. <https://www.cresc.ac.uk/research/social-life-of-methods/>. Consulté le 5 avril 2023
- Cervellin, G., Comelli, I. et Lippi, G. (2017). Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *Journal of epidemiology and global health*, 7(3), 185–189. <https://doi.org/10.1016/j.jegh.2017.06.001>
- Chary, M., Genes, N., Giraud-Carrier, C., Hanson, C., Nelson, L. S. et Manini, A. F. (2017). Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. *Journal of medical toxicology : official journal of the American College of Medical Toxicology*, 13(4), 278–286. <https://doi.org/10.1007/s13181-017-0625-5>
- Chen, J. et Wang, Y. (2021). Social media use for health purposes: Systematic review. *Journal of Medical Internet Research*, 23(5), 17917. <https://doi.org/10.2196/17917>
- Cheng, I. K., Heyl, J., Lad, N., Facini, G. et Grout, Z. (2021). Evaluation of twitter data for an emerging crisis: An application to the first wave of COVID-19 in the UK. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-98396-9>
- Clemente, L., Lu, F. et Santillana, M. (2019). Improved Real-Time Influenza Surveillance: Using Internet Search Data in Eight Latin American Countries. *JMIR public health and surveillance*, 5(2), e12214. <https://doi.org/10.2196/12214>
- Cot, C., Cacciapaglia, G. et Sannino, F. (2021). Mining google and apple mobility data: Temporal anatomy for COVID-19 social distancing. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-83441-4>
- Crowshoe LL, Sehgal A, Montesanti S, Barnabe C, Kennedy A, Murry A, et al. (2021). The Indigenous primary health care and policy research network: Guiding innovation within primary health care with Indigenous peoples in Alberta. *Health Policy*, 125(6):725–31. <https://doi.org/10.1016/j.healthpol.2021.02.007>

- Cruz, T. M. (2022). The social life of biomedical data : capturing, obscuring, and envisioning care in the digital safety-net. *Social Science & Medicine*, 294, 114670–114670. <https://doi.org/10.1016/j.socscimed.2021.114670>
- Cutcliffe, S. (2001). Book Review: Epistemic Cultures: How the Sciences Make Knowledge. *Science, Technology, & Human Values*, 26(3), 390–393. <https://doi.org/10.1177/016224390102600308>
- D’Agostino, F. (2008). Naturalizing the essential tension. *Synthese : An International Journal for Epistemology, Methodology and Philosophy of Science*, 162(2), 275–308. <https://doi.org/10.1007/s11229-007-9192-7>
- Darbellay, F. (2011). Vers une théorie de l’interdisciplinarité? Entre unité et diversité. *Nouvelles perspectives en sciences sociales*, 7(1), 65–87. <https://doi.org/10.7202/1007082ar>
- Dencik, L. et Sanchez-Monedero, J. (2022). Data justice. *Internet Policy Review*, 11(1). <https://doi.org/10.14763/2022.1.1615>
- Denecke, K. (2017). An ethical assessment model for digital disease detection technologies. *Life sciences, society and policy*, 13(1), 16. <https://doi.org/10.1186/s40504-017-0062-x>
- Dias da Silva, P., Heaton, L. et Millerand, F. (2017). Une revue de littérature sur la « science citoyenne » : la production de connaissances naturalistes à l’ère numérique. *Natures Sciences Sociétés*, 25, 370-380. <https://doi.org/10.1051/nss/2018004>
- Dicker RC, Coronado F, Koo D et Parrish RG. (2006). Introduction to Epidemiology. Dans *Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics*. (3^e éd.). Centers for Disease Control and Prevention. <https://www.cdc.gov/csels/dsepd/ss1978/index.html>
- Digital Epidemiology Lab (DEL). (2021). *Digital Epidemiology Lab - Shaping the digital transformation of healthcare systems*. <https://www.digitalepidemiologylab.org/>. Consulté le 1^{er} avril 2023
- Duclos, V. (2019). Algorithmic futures. *Medicine Anthropology Theory*, 6(3). <https://doi.org/10.17157/mat.6.3.660>
- Duclos, V. et Carle, M.-E.(dir.). (2009). Présentation : penser les frontières du social et du médical. *Altérités : revue d’anthropologie du contemporain*, 6(1), 1–8. Récupéré le 29 mars 2023 de <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/20990>
- Durazzi, F., Müller, M., Salathé, M. et Remondini, D. (2021). Clusters of science and health related twitter users become more isolated during the COVID-19 pandemic. *Scientific Reports*, 11(1), 19655. <https://doi.org/10.1038/s41598-021-99301-0>
- Dussault G. (1995). Epidemiology and health services management. *Epidemiological bulletin*, 16(2), 1–5. https://www3.paho.org/english/sha/epibul_95-98/be952services.htm.

Eckmanns, T., Füller, H. et Roberts, S. L. (2019). Digital epidemiology and global health security; an interdisciplinary conversation. *Life sciences, society and policy*, 15(1), 2. <https://doi.org/10.1186/s40504-019-0091-8>

Elsevier B.V. (2023). *Scopus*. <https://www-scopus-com.proxy.bibliotheques.uqam.ca/search/form.uri?display=basic#basic>. Consulté le 1^{er} avril 2023

Encyclopédie canadienne (EC). (2023). *Entomologie*. <https://www.thecanadianencyclopedia.ca/fr/article/entomologie>. Consulté le 6 avril 2023

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. et Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492LSF>

Fontanet, A. (2019). L'épidémiologie ou la science de l'estimation du risque en santé publique : Leçon inaugurale prononcée au Collège de France le jeudi 31 janvier 2019. Dans *L'épidémiologie ou la science de l'estimation du risque en santé publique : Leçon inaugurale prononcée le jeudi 31 janvier 2019*. Paris : Collège de France. <https://books.openedition.org/cdf/8273>

Frérot, M., Lefebvre, A., Aho, S., Callier, P., Astruc, K. et Aho Glélé, L. S. (2018). What is epidemiology? Changing definitions of epidemiology 1978-2017. *PloS one*, 13(12), e0208442. <https://doi.org/10.1371/journal.pone.0208442>

Fung, I. C., Fu, K. W., Ying, Y., Schaible, B., Hao, Y., Chan, C. H. et Tse, Z. T. (2013). Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. *Infectious diseases of poverty*, 2(1), 31. <https://doi.org/10.1186/2049-9957-2-31>

Galway, L., Bell, N., Sae, A. S., Hagopian, A., Burnham, G., Flaxman, A., Weiss, W. M., Rajaratnam, J. et Takaro, T. K. (2012). A two-stage cluster sampling method using gridded population data, a GIS, and Google Earth(TM) imagery in a population-based mortality survey in Iraq. *International journal of health geographics*, 11, 12. <https://doi.org/10.1186/1476-072X-11-12>

Génois, M. et Barrat, A. (2018). Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1), 11. <https://doi.org/10.1140/epjds/s13688-018-0140-1>

Gianfredi, V., Bragazzi, N. L., Nucci, D., Martini, M., Rosselli, R., Minelli, L. et Moretti, M. (2018). Harnessing Big Data for Communicable Tropical and Sub-Tropical Disorders: Implications From a Systematic Review of the Literature. *Frontiers in public health*, 6, 90. <https://doi.org/10.3389/fpubh.2018.00090>

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. et Trow, M. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. Sage Publications. <https://uk.sagepub.com/en-gb/eur/the-new-production-of-knowledge/book204307>

- Gingras, Y. (1991). L'institutionnalisation de la recherche en milieu universitaire et ses effets. *Sociologie Et Sociétés*, 23(1), 41–54. <https://doi.org/10.7202/001297ar>
- Ginsberg, J., Mohebbi, M., Patel, R. *et al.* (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012–1014. <https://doi.org/10.1038/nature07634>
- Giordano, A., Vabanesi, M., Dalla Costa, G., Cerri, F., Comi, G., Martinelli, V. et Fazio, R. (2019). Assessing seasonal dynamics of Guillain-Barré syndrome with search engine query data. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 40(5), 1015–1018. <https://doi.org/10.1007/s10072-019-03757-y>
- Giroux, É. (2012). De l'épidémiologie de santé publique à l'épidémiologie clinique. Quelques réflexions sur la relation entre épidémiologie et clinique (1920-1980). *Bulletin d'histoire et d'épistémologie des sciences de la vie*, 19, 21-43. <https://doi.org/10.3917/bhesv.191.0021>
- Gopichandran, V. et Elango, V. (2018). Data Ethics in Epidemiology: Autonomy, Privacy, Confidentiality and Justice. Dans Mishra, A., Subbiah, K. (eds) *Ethics in Public Health Practice in India*. Springer, Singapore. https://doi.org/10.1007/978-981-13-2450-5_7
- Gruebner, O., Lowe, S. R., Sykora, M., Shankardass, K., Subramanian, S. V. et Galea, S. (2018). Spatio-Temporal Distribution of Negative Emotions in New York City After a Natural Disaster as Seen in Social Media. *International journal of environmental research and public health*, 15(10), 2275. <https://doi.org/10.3390/ijerph15102275>
- Guernier, V., Milinovich, G. J., Bezerra Santos, M. A., Haworth, M., Coleman, G. et Soares Magalhaes, R. J. (2016). Use of big data in the surveillance of veterinary diseases: early detection of tick paralysis in companion animals. *Parasites & vectors*, 9(1), 303. <https://doi.org/10.1186/s13071-016-1590-6>
- Guo, S., Fang, F., Zhou, T., Zhang, W., Guo, Q., Zeng, R., Chen, X., Liu, J., & Lu, X. (2021). Improving google flu trends for covid-19 estimates using weibo posts. *Data Science and Management*, 3, 13–21. <https://doi.org/10.1016/j.dsm.2021.07.001>
- Hamer, S. A., Curtis-Robles, R. et Hamer, G. L. (2018). Contributions of citizen scientists to arthropod vector data in the age of digital epidemiology. *Current opinion in insect science*, 28, 98–104. <https://doi.org/10.1016/j.cois.2018.05.005>
- Harezlak, J. et Di, C. (2019). Editorial for the Special Issue "Medical Device Data: Challenges, Statistical Methods and Applications". *Statistics in biosciences*, 11(2), 207–209. <https://doi.org/10.1007/s12561-019-09247-1>
- Hartley, D. M., Nelson, N. P., Arthur, R. R., Barboza, P., Collier, N., Lightfoot, N., Linge, J. P., van der Goot, E., Mawudeku, A., Madoff, L. C., Vaillant, L., Walters, R., Yangarber, R., Mantero, J., Corley, C. D. et Brownstein, J. S. (2013). An overview of internet biosurveillance. *Clinical microbiology and infection : the official publication of the European Society of Clinical*

Microbiology and Infectious Diseases, 19(11), 1006–1013. <https://doi.org/10.1111/1469-0691.12273>

Hasan, S. M. S., Fox, E. A., Bisset, K. et Marathe, M. V. (2017). EpiK: A Knowledge Base for Epidemiological Modeling and Analytics of Infectious Diseases. *Journal of healthcare informatics research*, 1(2), 260–303. <https://doi.org/10.1007/s41666-017-0010-9>

Hassan Zadeh, A., Zolbanin, H. M., Sharda, R. et Delen, D. (2019). Social media for nowcasting flu activity: spatio-temporal big data analysis. *Information Systems Frontiers : A Journal of Research and Innovation*, 21(4), 743–760. <https://doi.org/10.1007/s10796-018-9893-0>

He, Z., Zhang, C. J. P., Huang, J., Zhai, J., Zhou, S., Chiu, J. W., Sheng, J., Tsang, W., Akinwunmi, B. O. et Ming, W. K. (2020). A New Era of Epidemiology: Digital Epidemiology for Investigating the COVID-19 Outbreak in China. *Journal of medical Internet research*, 22(9), e21685. <https://doi.org/10.2196/21685>

Heilbron, J. et Gingras, Y. (2015). La résilience des disciplines. *Actes de la recherche en sciences sociales*, 210, 4-9. <https://doi.org/10.3917/arss.210.0004>

Henman, P. (2022). Digital Social Policy: Past, Present, Future. *Journal of Social Policy*, 51(3), 535-550. <https://doi.org/10.1017/S0047279422000162>

Higgins, T. S., Wu, A. W., Sharma, D., Illing, E. A., Rubel, K., Ting, J. Y. et Snot Force Alliance (2020). Correlations of Online Search Engine Trends With Coronavirus Disease (COVID-19) Incidence: Infodemiology Study. *JMIR public health and surveillance*, 6(2), e19702. <https://doi.org/10.2196/19702>

Hill, R. K. (2016). What an Algorithm Is. *Philosophy & Technology*, 29(1), 35-59. <https://doi.org/10.1007/s13347-014-0184-5>

Hill-Cawthorne, G. A. et Sorrell, T. C. (2016). Future directions for public health research in emerging infectious diseases. *Public health research & practice*, 26(5), 2651655. <https://doi.org/10.17061/phrp2651655>

Ho, J., Stange, C., Suhrborg, R., Wurzbacher, C., Drewes, J. E. et Tiehm, A. (2022). SARS-CoV-2 wastewater surveillance in Germany: Long-term RT-digital droplet PCR monitoring, suitability of primer/probe combinations and biomarker stability. *Water research*, 210, 117977. <https://doi.org/10.1016/j.watres.2021.117977>

Hochreiter R. et Waldhauser C. (2013). A Stochastic Simulation of the Decision to Retweet. Dans Perny P., Pirlot M., Tsoukiàs A. (eds) *Algorithmic Decision Theory*. ADT 2013. Lecture Notes in Computer Science, vol 8176. Springer, Berlin, Heidelberg

Hoeyer K. (2019). Data as promise: Reconfiguring Danish public health through personalized medicine. *Social studies of science*, 49(4), 531–555. <https://doi.org/10.1177/0306312719858697>

Höhle M. (2017). A statistician's perspective on digital epidemiology. *Life sciences, society and policy*, 13(1), 17. <https://doi.org/10.1186/s40504-017-0063-9>

Hockin J. (1991). The Field Epidemiology Training Program at the LCDC. *The Canadian journal of infectious diseases = Journal canadien des maladies infectieuses*, 2(3), 129–130. <https://doi.org/10.1155/1991/620450>

Hors-Fraile, S., Malwade, S., Spachos, D., Fernandez-Luque, L., Su, C. T., Jeng, W. L., Syed-Abdul, S., Bamidis, P. et Li, Y. J. (2018). A recommender system to quit smoking with mobile motivational messages: study protocol for a randomized controlled trial. *Trials*, 19(1), 618. <https://doi.org/10.1186/s13063-018-3000-1>

Institut de recherche pour le développement (IRD). (2023). *Fontenille Didier*. http://www.red.ird.fr/annuaire/membres_de_droit/fontenille.htm. Consulté le 6 avril 2023

Institut national de santé publique du Québec (INSPQ). (2023). *Infrastructure de santé publique*. <https://ccnpps-ncchpp.ca/fr/infrastructure-de-sante-publique/>. Consulté le 30 octobre 2023

Jacobs, J. A. (2013). *In defense of disciplines: interdisciplinarity and specialization in the research university*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/I/bo16468807.html>

Jasanoff, Sheila (2004). The idiom of co-production. Dans *The Idiom of Coproduction. States of Knowledge: The Coproduction of Science and the Social Order*. Sheila Jasanoff (dir.), Routledge. <https://www.routledge.com/States-of-Knowledge-The-Coproduction-of-Science-and-the-Social-Order/Jasanoff/p/book/9780415403290>

Karas, M., Bai, J., Strączkiewicz, M., Harezlak, J., Glynn, N. W., Harris, T., Zipunnikov, V., Crainiceanu, C. et Urbanek, J. K. (2019). Accelerometry data in health research: challenges and opportunities. *Statistics in biosciences*, 11(2), 210–237. <https://doi.org/10.1007/s12561-018-9227-2>

Katapally, T.R. (2023). *Vision*. <https://tarunkatapally.com/vision/>. Consulté le 1^{er} avril 2023

Katapally, T.R., Hammami, N. et Chu, L. M. (2021). A randomized community trial to advance digital epidemiological and mHealth citizen scientist compliance: A smart platform study. *PloS one*, 16(11), e0259486. <https://doi.org/10.1371/journal.pone.0259486>

Kaveh-Yazdy, F. et Zareh-Bidoki, A. M. (2018). Search engines, news wires and digital epidemiology: Presumptions and facts. *International journal of medical informatics*, 115, 53–63. <https://doi.org/10.1016/j.ijmedinf.2018.03.017>

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <https://doi.org/10.1177/2053951714528481>

----- (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications. <https://methods.sagepub.com/book/the-data-revolution>

Knipe, D., Gunnell, D., Evans, H., John, A. et Fancourt, D. (2021). Is Google Trends a useful tool for tracking mental and social distress during a public health emergency? A time-series analysis. *Journal of affective disorders*, 294, 737–744. <https://doi.org/10.1016/j.jad.2021.06.086>

Knorr-Cetina, K. et Reichmann, Werner. (2015). Epistemic Cultures, Editor(s): James D. Wright, *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition), Elsevier, Pages 873-880, ISBN 9780080970875, <https://doi.org/10.1016/B978-0-08-097086-8.10454-4>

Knorr-Cetina, K. (2001). Objectual Practice. Dans Schatzki, T. R., Knorr-Cetina, K. et Savigny, E. von. (2001). *The practice turn in contemporary theory*. Routledge. <https://ebookcentral.proquest.com/lib/uqam/detail.action?docID=235322>

----- (1999). *Epistemic cultures : How the sciences make knowledge*. Harvard University Press. <https://ebookcentral.proquest.com/lib/uqam/detail.action?docID=6131400#>

Kostkova P. (2018). Disease surveillance data sharing for public health: the next ethical frontiers. *Life sciences, society and policy*, 14(1), 16. <https://doi.org/10.1186/s40504-018-0078-x>

Larousse. (2023). *Canada : géographie physique*. https://www.larousse.fr/encyclopedie/divers/Canada_g%C3%A9ographie_physique/186049. Consulté le 30 octobre 2023

Lau, E. H., Zheng, J., Tsang, T. K., Liao, Q., Lewis, B., Brownstein, J. S., Sanders, S., Wong, J. Y., Mekaru, S. R., Rivers, C., Wu, P., Jiang, H., Li, Y., Yu, J., Zhang, Q., Chang, Z., Liu, F., Peng, Z., Leung, G. M., Feng, L., ... et Yu, H. (2014). Accuracy of epidemiological inferences based on publicly available information: retrospective comparative analysis of line lists of human cases infected with influenza A(H7N9) in China. *BMC medicine*, 12, 88. <https://doi.org/10.1186/1741-7015-12-88>

Law, J. (2016). 1 STS as Method. Dans Felt, U., Fouché, R., Miller, C. A. et Smith-Doerr, L. (eds.). (2017). *The handbook of science and technology studies*, fourth edition. Mit Press. <https://ebookcentral.proquest.com/lib/uqam/detail.action?docID=5052910>

----- (2010). *The Double Social Life of Methods*. Heterogeneities.net. <http://heterogeneities.net/publications/Law2010DoubleSocialLifeofMethod5.pdf>. Consulté le 29 mars 2023

Law, J. et Ruppert, E. (2013). The social life of methods: devices. *Journal of Cultural Economy*, 6(3), 229–240. <https://doi.org/10.1080/17530350.2013.812042>

Law, J. , Ruppert, E. et Savage, M. (2011). The Double Social Life of Methods. *CRESC Working Paper Series, Working Paper No. 95*. <http://www.open.ac.uk/researchprojects/iccm/files/iccm/Law%20Savage%20Ruppert.pdf>. Consulté le 8 avril 2023

- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: traps in big data analysis. *Science*, 343(6176), 1203–1205
- Lazzari, G., Jaquet, Y., Kebaili, D. J., Symul, L. et Salathé, M. (2018). FoodRepo: An Open Food Repository of Barcoded Food Products. *Frontiers in nutrition*, 5, 57. <https://doi.org/10.3389/fnut.2018.00057>
- Leal-Neto, O., Egger, T., Schlegel, M., Flury, D., Sumer, J., Albrich, W., Babouee Flury, B., Kuster, S., Vernazza, P., Kahlert, C. et Kohler, P. (2021). Digital SARS-CoV-2 Detection Among Hospital Employees: Participatory Surveillance Study. *JMIR public health and surveillance*, 7(11), e33576. <https://doi.org/10.2196/33576>
- Lee, E. C., Asher, J. M., Goldlust, S., Kraemer, J. D., Lawson, A. B. et Bansal, S. (2016). Mind the Scales: Harnessing Spatial Big Data for Infectious Disease Surveillance and Inference. *The Journal of infectious diseases*, 214(suppl_4), S409–S413. <https://doi.org/10.1093/infdis/jiw344>
- Li Vigni, F. (2021). L'épidémiologie computationnelle à l'ère de la COVID-19: Enjeux disciplinaires et politiques d'une spécialité fondée sur l'étude des réseaux. *Réseaux*, 228, 23-60. <https://doi.org/10.3917/res.228.0023>
- Lippi, G. et Cervellin, G. (2019). Is digital epidemiology reliable?-insight from updated cancer statistics. *Annals of translational medicine*, 7(1), 15. <https://doi.org/10.21037/atm.2018.11.55>
- Liu, Y., Feng, G., Tsui, K.-L. et Sun, S. (2021). Forecasting influenza epidemics in hong kong using google search queries data: a new integrated approach. *Expert Systems with Applications*, 185. <https://doi.org/10.1016/j.eswa.2021.115604>
- Loi, M. (2019). The digital phenotype: a philosophical and ethical exploration. *Philosophy & Technology*, 32(1), 155–171. <https://doi.org/10.1007/s13347-018-0319-1>
- (2021). How to fairly incentivize digital contact tracing. *Journal of Medical Ethics*, 47(12), E76
- Louvel, S. (2015). Ce que l'interdisciplinarité fait aux disciplines: Une enquête sur la nanomédecine en France et en Californie. *Revue française de sociologie*, 56, 75-103. <https://doi.org/10.3917/rfs.561.0075>
- Lupton, D. (2016). Digital health technologies and digital data : new ways of monitoring, measuring and commodifying human bodies. Dans *Research Handbook on Digital Transformations*. Edited by F. Xavier Olleros, and Majlinda Zhegu, Edward Elgar Publishing Limited. <https://ebookcentral.proquest.com/lib/uqam/detail.action?docID=4697091>
- (2015). *Digital sociology*. Routledge, Taylor & Francis Group. <https://www.routledge.com/Digital-Sociology/Lupton/p/book/9781138022775>
- Lury, C., Tironi, M. et Bernasconi, R. (2020). The Social Life of Methods as Epistemic Objects: Interview with Celia Lury. *Revista Diseña*, (16), 32-55.

https://www.researchgate.net/publication/339855469_The_Social_Life_of_Methods_as_Epistemic_Objects_Interview_with_Celia_Lury

Magumba, M. A. et Nabende, P. (2021). Evaluation of different machine learning approaches and input text representations for multilingual classification of tweets for disease surveillance in the social web. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00528-5>

Makri A. (2017). Robert Koch Institut: towards digital epidemiology. *Lancet* (London, England), 390(10097), 833. [https://doi.org/10.1016/S0140-6736\(17\)32278-X](https://doi.org/10.1016/S0140-6736(17)32278-X)

Markovsky, B. (2000). Book review: epistemic cultures: how the sciences make knowledge. *Contemporary Sociology*, 29(3), 556–557

Marres, N. et Gerlitz, C. (2016). Interface methods: renegotiating relations between digital social research, sts and sociology. *The Sociological Review*, 64(1), 21–46. <https://doi.org/10.1111/1467-954X.12314>

Marres, N. et Weltevrede, E. (2013). Scraping the social?: issues in live social research. *Journal of Cultural Economy*, 6(3), 313–335. <https://doi.org/10.1080/17530350.2013.772070>

Marres, N. (2012). The redistribution of methods: on intervention in digital social research, broadly conceived. *The Sociological Review*, 60: 139-165. <https://doi.org/10.1111/j.1467-954X.2012.02121.x>

Martineau, J., & Durand Folco, J. (2023). Thèse 1. Nous avons besoin d'une théorie critique des algorithmes. Dans *Le capital algorithmique : accumulation, pouvoir et résistance à l'ère de l'intelligence artificielle* (Ser. Théorie, th 13). Écosociété

Mattelart, Armand. (2009). L'émergence du paradigme techno-informationnel. Dans *Histoire de la société de l'information*, 4^e édition, Paris, La Découverte, p. 30-46. <https://www.cairn.info/histoire-de-la-societe-de-l-information--9782707157980.htm>.

Mattern, Shannon. (2013). Methodolatry and the Art of Measure. *Places Journal*, November. <https://doi.org/10.22269/131105>

Mavragani, A. et Ochoa, G. (2019). Google Trends in Infodemiology and Infeveillance: Methodology Framework. *JMIR public health and surveillance*, 5(2), e13439. <https://doi.org/10.2196/13439>

Mavragani, A, Ochoa, G. et Tsagarakis, K. P. (2018). Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. *Journal of medical Internet research*, 20(11), e270. <https://doi.org/10.2196/jmir.9366>

Mayer-Schönberger Viktor et Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McGill University (MGU). (2023). *Dual Degree Programs in Public Health Data Science*. <https://www.mcgill.ca/epi-biostat-occh/education/grad/dual-degree-programs-public-health-data-science>. Consulté le 1^{er} avril 2023

- Merianos, A. et Peiris, M. (2005). International Health Regulations (2005). *The Lancet*, 366(9493), 12491251. [https://doi.org/10.1016/S0140-6736\(05\)67508-3](https://doi.org/10.1016/S0140-6736(05)67508-3)
- Miettinen, O. S. (2010). Up from ‘Clinical Epidemiology’ & EBM. Dans *Up from Clinical Epidemiology & EBM* (p. 15-20). Springer, Dordrecht. https://doi.org/10.1007/978-90-481-9501-5_5
- Mittelstadt, B., Benzler, J., Engelmann, L., Prainsack, B. et Vayena, E. (2018). Is there a duty to participate in digital epidemiology? *Life Sciences, Society and Policy*, 14(1), 1–24. <https://doi.org/10.1186/s40504-018-0074-1>
- Mohanty, S. P., Hughes, D. P. et Salathé, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in plant science*, 7, 1419. <https://doi.org/10.3389/fpls.2016.01419>
- Morabia, A. (2014). History of Epidemiological Methods and Concepts. Dans Ahrens, W., Pigeot, I. (eds) *Handbook of Epidemiology*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-09834-0_52
- Müller, M. M. et Salathé, M. (2019). Crowdbreaks: Tracking Health Trends Using Public Social Media Data and Crowdsourcing. *Frontiers in public health*, 7, 81. <https://doi.org/10.3389/fpubh.2019.00081>
- National Research Council (NRC) (US) Committee on Environmental Epidemiology; National Research Council (US) Commission on Life Sciences. (1997). *Environmental Epidemiology: Volume 2: Use of the Gray Literature and Other Data in Environmental Epidemiology*. Washington (DC): National Academies Press (US); 1997. 1, *Environmental Epidemiology: The Context*. <https://www.ncbi.nlm.nih.gov/books/NBK233640/>
- Nature (2007). A matter of trust. *Nature*, 449(7163), 637–638. <https://doi.org/10.1038/449637b>
- Nowotny, Helga, Peter Scott et Michael Gibbons. (2003). Introduction: ‘Mode 2’ Revisited: The New Production of Knowledge. *Minerva*, 41: 179–94. <https://doi.org/10.1023/A:1025505528250>
- (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press. <https://www.wiley.com/en-us/Re+Thinking+Science%3A+Knowledge+and+the+Public+in+an+Age+of+Uncertainty-p-9780745626079>
- O’Brien, D. et Lockley, P. (2015). The Social Life of Cultural Value. Dans MacDowall, L., Badham, M., Blomkamp, E., Dunphy, K. (eds) *Making Culture Count. New Directions in Cultural Policy Research*. Palgrave Macmillan, London. https://doi.org/10.1007/978-1-137-46458-3_7
- Olleros, F. X. et Zhegu, M. (Eds.). (2016). Digital transformations : an introduction. Dans *Research handbook on digital transformations*. Edited by F. Xavier Olleros, and Majlinda Zhegu. Edward Elgar Publishing Limited. <https://ebookcentral.proquest.com/lib/uqam/detail.action?docID=4697091#>

Organisation mondiale de la santé (OMS). (2020). (2 mars 2020). *Maladies à transmission vectorielle*. <https://www.who.int/fr/news-room/fact-sheets/detail/vector-borne-diseases>. Consulté le 1^{er} avril 2023

----- (2016). *International Health Regulations (2005)*, 3^e éd. World Health Organization. <https://apps.who.int/iris/handle/10665/246107>. Consulté le 30 mars 2023

Osborne, J. (2014). Teaching scientific practices: meeting the challenge of change. *Journal of Science Teacher Education : The Official Journal of the Association for Science Teacher Education*, 25(2), 177–196. <https://doi.org/10.1007/s10972-014-9384-1>

Paillé, P. et Mucchielli, A. (2016). Chapitre 11. L'analyse thématique. Dans P. Paillé et A. Mucchielli (dir), *L'analyse qualitative en sciences humaines et sociales* (p. 235-312). Paris: Armand Colin. <https://doi-org.proxy.bibliotheques.uqam.ca/10.3917/arco.paill.2016.01.0235>

Paillé P. (1994). L'analyse par théorisation ancrée. *Cahiers De Recherche Sociologique*, 23(23), 147–181. <https://doi.org/10.7202/1002253ar>

Perra, N. et Gonçalves, B. (2015). Modeling and Predicting Human Infectious Diseases. *Social Phenomena: From Data Analysis to Models*, 59–83. https://doi.org/10.1007/978-3-319-14011-7_4

Pinch, T. (2000). Epistemic cultures: how the sciences make knowledge by karin knorr-cetina. *American Journal of Sociology*, 105, 1472–1474

Pollett, S., Althouse, B. M., Forshey, B., Rutherford, G. W. et Jarman, R. G. (2017a). Internet-based biosurveillance methods for vector-borne diseases: Are they novel public health tools or just novelties?. *PLoS neglected tropical diseases*, 11(11), e0005871. <https://doi.org/10.1371/journal.pntd.0005871>

Pollett, S., Boscardin, W. J., Azziz-Baumgartner, E., Tinoco, Y. O., Soto, G., Romero, C., Kok, J., Biggerstaff, M., Viboud, C. et Rutherford, G. W. (2017b). Evaluating Google Flu Trends in Latin America: Important Lessons for the Next Phase of Digital Disease Detection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 64(1), 34–41. <https://doi.org/10.1093/cid/ciw657>

Popa, F., Guillermin, M. et Dedeurwaerdere, T. (2015). A pragmatist approach to transdisciplinarity in sustainability research: from complex systems theory to reflexive science. *Futures*, 65, 45–56. <https://doi.org/10.1016/j.futures.2014.02.002>

Porta, M. (2014). Digital Epidemiology. Dans *A Dictionary of Epidemiology*: Oxford University Press. <https://www-oxfordreference-com.proxy.bibliotheques.uqam.ca/view/10.1093/acref/9780199976720.001.0001/acref-9780199976720-e-2118>. Consulté le 30 Mars 2023

----- (2008). *A Dictionary of Epidemiology*: Oxford University Press. [https://www-oxfordreference-](https://www-oxfordreference-com.proxy.bibliotheques.uqam.ca/view/10.1093/acref/9780199976720.001.0001/acref-9780199976720-e-2118)

com.proxy.bibliotheques.uqam.ca/view/10.1093/acref/9780195314496.001.0001/acref-978019531449. Consulté le 30 mars 2023

Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., Seymour, A. K. et Merchant, R. M. (2014). Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *Journal of General Internal Medicine*, 29(1), 187–203. <https://doi.org/10.1007/s11606-013-2536-8>

Ressources naturelles Canada (RNC). (2022). *Cartes de référence*. https://ressources-naturelles.canada.ca/sciences-terre/geographie/atlas-canada/explorez-nos-cartes/cartes-de-r%C3%A9f%C3%A9rence/16847?_gl=1*1hbuqnv*_ga*NDM2OTk3OTMwLjE2OTg2Njk0ODE.*_ga_C2N57Y7DX5*MTY5ODY3MjY1NS4yLjEuMTY5ODY3MjY2OS4wLjAuMA... Consulté le 30 octobre 2023

Ribes, D. (2019). STS, meet data science, once again. *Science, Technology, & Human Values*, 44(3), 514–539. <https://doi.org/10.1177/0162243918798899>

Riordan, B. C., Merrill, J. E., Ward, R. M. et Raubenheimer, J. (2022). When are alcohol-related blackout tweets written in the united states?. *Addictive Behaviors*, 124, 107110–107110. <https://doi.org/10.1016/j.addbeh.2021.107110>

Robert Koch Institute (RKI). (10 janvier 2023a). *Mission Statement*. https://www.rki.de/EN/Content/Institute/Mission_Statement/Mission_Statement_node.html. Consulté le 1er avril 2023

----- (10 janvier 2023b). *Project Group P4 : Computational Epidemiology*. https://www.rki.de/EN/Content/Institute/DepartmentsUnits/ProjectGroups/P4/project_group_4.html. Consulté le 1^{er} avril 2023

Roche, B., Gaillard, B., Léger, L., Pélagie-Moutenda, R., Sochacki, T., Cazelles, B., Ledrans, M., Bateau, A., Fontenille, D., Etienne, M., Simard, F., Salathé, M. et Yébakima, A. (2017). An ecological and digital epidemiology analysis on the role of human behavior on the 2014 chikungunya outbreak in martinique. *Scientific Reports*, 7(1), 5967–5967. <https://doi.org/10.1038/s41598-017-05957-y>

Rogers, R. (2015). Digital Methods for Web Research. Dans *Emerging Trends in the Social and Behavioral Sciences* (eds R.A. Scott and S.M. Kosslyn). <https://doi.org/10.1002/9781118900772.etrds0076>

Rogers, Everett M. (2003). *Diffusion of Innovations*, 5th Edition. Free Press. <https://uqam-bib.on.worldcat.org/search/detail/491078853?queryString=Diffusion%20of%20Innovations%20Rogers&databaseList=3784%2C283%2C3539%2C638%2C3536%2C3155&stickyFacetsChecked=false&clusterResults=false&groupVariantRecords=false>

Rothman, KJ. (2002). *Epidemiology, An Introduction*. New York: Oxford University Press. <https://global.oup.com/academic/product/epidemiology-9780199754557?cc=us&lang=en&>

Rouvroy, A. (2018). Mapping as governance in an Age of Autonomic Computing: Technology, Virtuality and Utopia. Dans P. Bargués-Pedreny, D. Chandler, & E. Simon (Eds.), *Mapping and Politics in the Digital Age* (pp. 118-134). Routledge

Rovetta A. (2022). Google Trends as a Predictive Tool for COVID-19 Vaccinations in Italy: Retrospective Infodemiological Analysis. *JMIRx Med*, 3(2), e35356.
<https://doi.org/10.2196/35356>

Rowbotham, S., McKinnon, M., Leach, J., Lamberts, R. et Hawe, P. (2019). Does citizen science have the capacity to transform population health science? *Critical Public Health*, 29(1), 118–128.
<https://doi.org/10.1080/09581596.2017.1395393>

Rubio, F.D. et Baert, P. (2012). The Politics of Knowledge. An introduction. Dans F. Domínguez Rubio et P. Baert. (eds.) *The Politics of Knowledge*, Routledge: London, p. 1-10.
<https://www.routledge.com/The-Politics-of-Knowledge/Baert-Rubio/p/book/9780415704755>

Ruckenstein, M. et Granroth, J. (2020). Algorithms, advertising and the intimacy of surveillance. *Journal of Cultural Economy*, 13(1), 12–24. <https://doi.org/10.1080/17530350.2019.1574866>

Ruppert, E., Isin, E. et Bigo, D. (2017). Data politics. *Big Data & Society*, 4(2).
<https://doi.org/10.1177/2053951717717749>

Ruppert, E. *et al.* (2015). *A Social Framework for Big Data*.
https://www2.warwick.ac.uk/fac/cross_fac/cim/research/archive/socialising-big-data/sfbd.pdf.
Consulté le 29 mars 2023

Ruppert, E., Law, J. et Savage, M. (2013). Reassembling social science methods: the challenge of digital devices. *Theory, Culture & Society*, 30(4), 22–46.
<https://doi.org/10.1177/0263276413484941>

Rutty, Christopher et Sullivan, Sue C. (2010). *La santé publique : Une histoire canadienne*.
https://www.cpha.ca/sites/default/files/assets/history/book/history-book-print_all_f.pdf

Sadowski, J. (2019). When data is capital: datafication, accumulation, and extraction. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951718820549>

----- (2018). Digital epidemiology: what is it, and where is it going?. *Life sciences, society and policy*, 14(1), 1. <https://doi.org/10.1186/s40504-017-0065-7>

----- (2016). Digital Pharmacovigilance and Disease Surveillance: Combining Traditional and Big-Data Systems for Better Public Health. *The Journal of infectious diseases*, 214(suppl_4), S399–S403. <https://doi.org/10.1093/infdis/jiw281>

----- (2015). Digital epidemiology: tracking diseases in the mobile age. *The Conversation*.
<https://theconversation.com/digital-epidemiology-tracking-diseases-in-the-mobile-age-37741>.
Consulté le 30 octobre, 2023

Salathé, M., Freifeld, C. C., Mearu, S. R., Tomasulo, A. F. et Brownstein, J. S. (2013). Influenza A (H7N9) and the importance of digital epidemiology. *The New England journal of medicine*, 369(5), 401–404. <https://doi.org/10.1056/NEJMp1307752>

Salathé, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., Campbell, E. M., Cattuto, C., Khandelwal, S., Mabry, P. L. et Vespignani, A. (2012). Digital epidemiology. *PLoS computational biology*, 8(7), e1002616. <https://doi.org/10.1371/journal.pcbi.1002616>

Sato, K., Mano, T., Iwata, A. et Toda, T. (2021). Need of care in interpreting Google Trends-based COVID-19 infodemiological study results: potential risk of false-positivity. *BMC medical research methodology*, 21(1), 147. <https://doi.org/10.1186/s12874-021-01338-2>

Sau A. (2017). A Simulation Study on Hypothetical Ebola Virus Transmission in India Using Spatiotemporal Epidemiological Modeler (STEM): A Way towards Precision Public Health. *Journal of environmental and public health*, 2017, 7602301. <https://doi.org/10.1155/2017/7602301>

Savage, M. (2013). The social life of methods: a critical introduction. *Theory, Culture & Society*, 30(4), 3–21. <https://doi.org/10.1177/0263276413486160>

Savage, M. et Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41(5), 885–899. <https://doi.org/10.1177/0038038507080443>

Savoie-Zajc, Lorraine. (2000). Chapitre 13. L’entrevue semi-dirigée. Dans Gauthier, B. (2008). *Recherche sociale, 5e édition: De la problématique à la collecte des données*. (5è éd.). Presses de l'Université du Québec. <https://ebookcentral.proquest.com/lib/uqam/detail.action?docID=3263794>

Science Direct (SD). (2023a). *Spatial Data 4.6.1 Spatial versus nonspatial data*. <https://www.sciencedirect.com/topics/engineering/spatial-data>. Consulté le 1^{er} avril 2023

----- (2023b). *Image Analysis, 1.1.2.3 Image Analysis*. <https://www.sciencedirect.com/topics/computer-science/image-analysis>. Consulté le 1^{er} avril 2023

Sénat du Canada. (2023). *La fracture numérique au Canada pénalise les populations autochtones et rurales : sénateur Klyne*. <https://sencanada.ca/fr/sencaplus/opinion/la-fracture-numerique-au-canada-penalise-les-populations-autochtones-et-rurales-senateur-klyne/>. Consulté le 30 octobre 2023

Siddiqui, S., Alhamdi, H. W. S. et Alghamdi, H. A. (2022). Recent chronology of covid-19 pandemic. *Frontiers in Public Health*, 10, 778037–778037. <https://doi.org/10.3389/fpubh.2022.778037>

Smith-Doerr, Laurel. (2017). Organizing and Governing Science. Dans Felt, U., Fouché, R., Miller, C. A. et Smith-Doerr, L. (Eds.). *The handbook of science and technology studies*. MIT Press. <https://ebookcentral.proquest.com/lib/uqam/detail.action?docID=5052910>

Smolinski, M. S., Crawley, A. W., Olsen, J. M., Jayaraman, T. et Libel, M. (2017). Participatory Disease Surveillance: Engaging Communities Directly in Reporting, Monitoring, and Responding to Health Threats. *JMIR public health and surveillance*, 3(4), e62. <https://doi.org/10.2196/publichealth.7540>

Statistique Canada (2023). (27 septembre 2023). *Estimations démographiques du Canada au 1er juillet 2023 : la plus forte croissance de la population depuis 1957*. <https://www150.statcan.gc.ca/n1/daily-quotidien/230927/dq230927a-fra.htm>. Consulté le 30 octobre 2023

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N. et Somani, B. K. (2022). The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish journal of medical science*, 191(4), 1473–1483. <https://doi.org/10.1007/s11845-021-02730-z>

Thoër, Christine et Millerand, Florence. (2016). Internet comme terrain d'investigation et mode de collecte des données qualitatives en santé. Dans J. Kivits, F. Balard, C. Fournier et M. Winance (dirs.) *Les recherches qualitatives en santé*. Paris : Armand Colin, 151-164. <https://www-cairn-info.proxy.bibliotheques.uqam.ca/les-recherches-qualitatives-en-sante--9782200611897.htm>

Tourassi, G., Yoon, H. J., Xu, S. et Han, X. (2016). The utility of web mining for epidemiological research: studying the association between parity and cancer risk. *Journal of the American Medical Informatics Association : JAMIA*, 23(3), 588–595. <https://doi.org/10.1093/jamia/ocv141>

Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., Jackson, J., Judson-Garcia, J., Khan, M., Lin, M., Mamada, R., Moreno, V. M., Nazari, F., Okuneye, K., Ross, M. L., Rodriguez, C., Medlock, J., Ebert, D. et Castillo-Chavez, C. (2015). Mass Media and the Contagion of Fear: The Case of Ebola in America. *PloS one*, 10(6), e0129179. <https://doi.org/10.1371/journal.pone.0129179>

Tremblay, André. (1991). *Sondages : histoire, pratique et analyse*. G. Morin.

Universities Canada. (2023). *Search Programs*. <https://www.universitystudy.ca/search-programs/>. Consulté le 4 novembre, 2023

Vardi, M. (2012). “What is an algorithm?”, *Communications of the ACM*, 55(3): 5. [doi:10.1145/2093548.2093549](https://doi.org/10.1145/2093548.2093549)

Vayena, E., Dzenowagis, J., Brownstein, J. S. et Sheikh, A. (2018). Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, 96(1), 66–68. <https://doi.org/10.2471/BLT.17.197426>

Velasco E. (2018). Disease detection, epidemiology and outbreak response: the digital future of public health practice. *Life sciences, society and policy*, 14(1), 7. <https://doi.org/10.1186/s40504-018-0071-4>

Vigfusson, Y., Karlsson, T. A., Onken, D., Song, C., Einarsson, A. F., Kishore, N., Mitchell, R. M., Brooks-Pollock, E., Sigmundsdottir, G. et Danon (2021). Cell-phone traces reveal infection-associated behavioral change. *Proceedings of the National Academy of Sciences of the United States of America*, 118(6), e2005241118. <https://doi.org/10.1073/pnas.2005241118>

Vinck, D. (2016). Les humanités numériques, c'est la quantification des sciences humaines. Dans D. Vinck, *Humanités Numériques: La culture face aux nouvelles technologies* (p. 39-46). Paris: Le Cavalier Bleu. <https://www-cairn-info.proxy.bibliotheques.uqam.ca/humanites-numeriques--9782846708883.htm>

----- (2007). *Sciences et société : sociologie du travail scientifique* (Ser. Collection u. sociologie). Armand Colin. <https://www.dunod.com/sciences-humaines-et-sociales/sciences-et-societe-sociologie-du-travail-scientifique>

Western University (WU). (2023). *Clinical Epidemiology*. https://www.schulich.uwo.ca/epibio/research/research_clusters/methodological_approaches_and_disciplines/clinical_epi.html#:~:text=Clinical%20epidemiology%20is%20the%20application%20of%20epidemiology%20principles,decision-making%2C%20for%20the%20purpose%20of%20improving%20patient-level%20outcomes. Consulté le 8 avril 2023

Whitehead, J., Smith, M., Anderson, Y., Zhang, Y., Wu, S., Maharaj, S. et Donnellan, N. (2021). Improving spatial data in health geographics: a practical approach for testing data to measure children's physical activity and food environments using Google Street View. *International journal of health geographics*, 20(1), 37. <https://doi.org/10.1186/s12942-021-00288-8>

Younesi, E. et Hofmann-Apitius, M. (2013). From integrative disease modeling to predictive, preventive, personalized and participatory (P4) medicine. *European Association for Predictive, Preventive and Personalised Medicine - The EPMA journal*, 4(1), 23. <https://doi.org/10.1186/1878-5085-4-23>

Young, S. D., Rivers, C. et Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive medicine*, 63, 112–115. <https://doi.org/10.1016/j.ypmed.2014.01.024>

Zhang, D., Yin, C., Zeng, J., Yuan, X. et Zhang, P. (2020). Combining structured and unstructured data for predictive models: a deep learning approach. *BMC medical informatics and decision making*, 20(1), 280. <https://doi.org/10.1186/s12911-020-01297-6>

Zhao, N., Cao, G., Vanos, J. K. et Vecellio, D. J. (2018). The effects of synoptic weather on influenza infection incidences: a retrospective study utilizing digital disease surveillance.

International journal of biometeorology, 62(1), 69–84. <https://doi.org/10.1007/s00484-017-1306-4>

Zuboff, S. (2019). The Definition. Dans *The age of surveillance capitalism : the fight for a human future at the new frontier of power* (First). PublicAffairs

Zwitter, A. (2014). Big Data ethics. *Big Data & Society*, 1(2).
<https://doi.org/10.1177/2053951714559253>