

UNIVERSITÉ DE MONTRÉAL

OUTIL INTÉGRÉ DE COLLECTE, D'ANALYSE ET DE
VISUALISATION DE DONNÉES DE MOBILITÉ

PIERRE-LÉO MONGEON-BOURBONNAIS
DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIAE DOCTOR
(GÉNIE CIVIL)
AVRIL 2016

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

OUTIL INTÉGRÉ DE COLLECTE, D'ANALYSE ET DE
VISUALISATION DE DONNÉES DE MOBILITÉ

présentée par : MONGEON-BOURBONNAIS Pierre-Léo
en vue de l'obtention du diplôme de : Philosophiae Doctor
a été dûment acceptée par le jury d'examen constitué de :

M. SAUNIER Nicolas, Ph. D., président

Mme MORENCY Catherine, Ph. D., membre et directrice de recherche

M. TRÉPANIÉ Martin, Ph. D., membre

M. CORNELIS Éric, Doctorat, membre externe

DÉDICACE

À Isa, Louis et Romy

REMERCIEMENTS

Merci à Isabelle, Louis et Romy pour leur patience, leur présence et leur support.

Merci à Catherine pour sa confiance, son expérience et sa motivation.

Merci à Julien pour son écoute, pour les corrections et les lectures, pour les discussions motivantes et les nombreuses heures passées à me sauver la vie quand les délais étaient trop serrés dans nos projets communs.

Merci à Louiselle pour sa bonne humeur et parce que c'est réconfortant de savoir que quelqu'un a réussi à terminer sa thèse avec brio sur un sujet beaucoup plus polémique et politique que le mien.

Merci à Hubert pour son écoute et son expérience fabuleuse dans les enquêtes OD.

Merci à Pierre Tremblay pour sa confiance et son support, même en période d'austérité.

Merci à mes parents pour avoir encouragé ma créativité et ma détermination.

Merci à la musique et au piano pour m'avoir montré la discipline et la rigueur. Un merci particulier à Bach, Mozart, Chopin et Prokofiev.

Merci à 50% au TDAH pour la créativité et l'intuition (parce que l'autre 50% n'est pas toujours drôle...)

RÉSUMÉ

Depuis plusieurs décennies au Québec et ailleurs dans le monde, les planificateurs du transport dressent le portrait de la mobilité des personnes au moyen d'enquêtes Origine-Destination. De telles enquêtes permettent de mettre en commun un grand nombre de données relatives aux comportements de déplacements des ménages et des personnes. Or, les ménages contactés habituellement par téléphone sont de plus en plus difficiles à rejoindre et de nombreux jeunes ménages ne possèdent plus de ligne téléphonique fixe au domicile. De ce fait, de nouvelles méthodes de recrutement, mais surtout d'entrevue, sont nécessaires pour répondre à cette problématique. D'autre part, la quantité toujours plus grande de données collectées en parallèle aux enquêtes fait en sorte que la complexité des systèmes informationnels et des bases de données de mobilité est en forte croissance.

Ce travail de recherche a pour objectif de proposer des solutions à ces deux problématiques. En premier lieu, un nouvel outil de collecte de données de mobilité sur le web a été développé, et d'autre part, un entrepôt de données de mobilité a été construit, le tout au sein d'une plateforme intégrée.

Deux hypothèses sont énoncées dans le cadre de la thèse. La première expose que les enquêtes de mobilité peuvent demeurer des outils de collecte valables pour l'étude des comportements de mobilité, pour autant que l'on mette de l'avant de nouveaux questionnaires web plus flexibles et plus conviviaux pour les répondants et les administrateurs. En parallèle, la mise en valeur des par données collectées pendant les entrevues web peut supporter une analyse plus fine du comportement des répondants, et en particulier rend possible la modélisation de la durée d'entrevue pour des fins de suivi, de prédiction en temps réel et d'optimisation des ressources.

La deuxième hypothèse exprime le fait que la modélisation dimensionnelle favorise l'intégration des données de mobilité dans un même entrepôt de données et assure une structure plus compréhensible et plus robuste à ces données. Plus particulièrement, cette modélisation permet de présenter des faits saillants rapidement, de réaliser des analyses multiples et ciblées, de favoriser la comparabilité d'un ensemble de données à l'autre et d'automatiser la génération d'objets de visualisation.

La modélisation dimensionnelle, utilisée en intelligence d'affaires, distingue les objets d'une base de données en tables de faits et de dimensions. Les tables de faits comptent seulement des clés permettant de les joindre aux tables de dimensions, qui comprennent les attributs descriptifs des objets. Dans un modèle dimensionnel (aussi appelé schéma en étoile), il n'existe habituellement qu'un seul degré de séparation entre les différentes tables, ce qui accélère les requêtes d'analyse et facilite la compréhension du schéma.

La plateforme logicielle développée intègre les phases de collecte, de suivi, de traitement, d'analyse, de visualisation et de diffusion des données de mobilité au moyen d'un questionnaire web et soutenu par un entrepôt de données de mobilité. Un questionnaire web de type Origine-Destination est proposé pour des enquêtes-ménage et des enquêtes-personne. Ensuite, une modélisation des durées d'entrevue est présentée pour optimiser les ressources et effectuer un suivi en temps réel des entrevues. Finalement, un schéma dimensionnel d'entrepôt de données est proposé et défini.

La plateforme supporte la construction d'un questionnaire flexible. En effet, la configuration permet la création et la modification des questions, des sections, des types de réponses et des choix de réponse à tout moment. De plus, elle comprend un module de validation en temps réel que les administrateurs peuvent configurer pour chaque question ou pour un ensemble de questions reliées entre elles.

L'adaptation de la modélisation dimensionnelle au domaine du transport permet de structurer, d'enrichir et de mettre en valeur les données provenant de plusieurs sources et propose des processus automatisés pour le traitement et la validation des données. De plus, la plateforme peut créer automatiquement des cartes animées de déplacements pour des besoins de diffusion et pour faciliter les analyses temporelles et spatiales. Finalement, elle génère automatiquement des faits saillants à partir des données validées sauvegardées dans l'entrepôt de données.

Bien que développée pour le contexte québécois, la plateforme a été conçue pour être générique et utilisée ailleurs et dans de nombreux contextes.

L'intégration de ces modules dans une même plateforme s'inscrit dans l'évolution attendue des outils d'aide à la planification et a comme objectif principal de faciliter la gestion des enquêtes

de mobilité et de simplifier le traitement des données. De cette manière, il sera possible de mieux soutenir les analyses et les simulations qui contribueront à comprendre et à prédire les comportements de mobilité.

ABSTRACT

For several decades in Quebec and elsewhere in the world, transportation planners studied the mobility of people through Origin-Destination surveys. Such surveys enable the sharing of a large amount of data related to the travel behavior of households and people. However, households usually contacted by telephone are increasingly difficult to reach and many young families no longer own a landline at home. Therefore, new methods of recruitment, but mainly new interviewing method, are needed to address this issue. On the other hand, the growing amount of data collected in parallel to surveys increases the complexity of databases and makes analysis even more difficult.

This research project was proposed to find solutions to these two problems. First, a new mobility data collection tool on the web was developed, and secondly, a mobility datawarehouse was built, both within the same integrated platform.

Two hypotheses are set out in the framework of the thesis. The first states that travel surveys may remain valid collection tools for the study of travel behavior, provided that we put forward new, more flexible and user-friendly web questionnaires. In parallel, the enhancement of paradata collected while the interviews are conducted can support a finer analysis of respondents' behavior, and enable the modeling of the interview duration for monitoring purposes and server optimization.

The second hypothesis expresses that building a datawarehouse based on dimensional modeling for travel data can promote a more understandable structure and generate more robust results. In fact, the datawarehouse will provide faster access to surveys results and assist in the publication of highlights and in the task of performing targeted analyzes. It will also promote data comparability between surveys and simplify the generation of visualizations.

Dimensional modeling is mainly used in business intelligence. It distinguishes objects in a database using fact and dimension tables. Fact tables only include keys to join dimension tables, which include descriptive attributes related to objects. In a dimensional model (also called star schema), there is usually only one degree of separation between the various tables, which accelerates queries and facilitates the understanding of the schema.

The software platform developed integrates the collection, tracking, processing, analysis, visualization and dissemination of travel data through a web questionnaire and supported by a datawarehouse. Specifically, an Origin-Destination web questionnaire is proposed for both household-based and person-based surveys. Then a model of interview duration is presented to allow better server management and real-time prediction for monitoring purposes. Finally, a datawarehouse based on dimensional modeling is designed to simplify and accelerate analysis, visualization and dissemination of travel data.

The platform enables the formulation of flexible questionnaires. Indeed, the configuration allows the creation and modification of questions, sections, types of answers and response choices at any time. In addition, it includes a real-time validation module that administrators can configure for each question or for a set of interrelated questions.

The adaptation of dimensional modeling to transportation data encourages a better structure while integrating, enriching and enhancing data. It provides automated data processing and validation, and can automatically create animated maps of trips for dissemination purposes and to facilitate temporal and spatial analyzes. Finally, it automatically generates survey highlights from the validated data saved in the datawarehouse.

Although developed for the Quebec context, the platform is designed to be generic and used elsewhere and in different contexts.

Integrating data into a single software platform follows the expected development of transportation planning tools. It also has the objectives of facilitating analysis and creating more robust simulations for better prediction of travel behavior.

TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS	IV
RÉSUMÉ	V
ABSTRACT.....	VIII
TABLE DES MATIÈRES	X
LISTE DES TABLEAUX.....	XVII
LISTE DES FIGURES.....	XVIII
LISTE DES ANNEXES.....	XXI
LISTE DES SIGLES ET ABRÉVIATIONS.....	XXII
CHAPITRE 1 INTRODUCTION.....	1
1.1 Hypothèses de recherche	3
1.2 Description générale, contexte et portée de la thèse.....	4
1.3 Objectifs principaux.....	4
1.4 Schéma général de la plateforme	4
1.5 Aperçu du document	6
CHAPITRE 2 ÉTAT DES CONNAISSANCES.....	7
2.1 Collecte de données de mobilité.....	7
2.1.1 Enquêtes Origine-Destination	7
2.1.2 Deux types d'enquêtes Origine-Destination	7
2.1.3 Mode de recrutement et mode de collecte	8
2.1.4 Suivi des entrevues par des parodonnées	11
2.1.5 Enquêtes web sur la mobilité	12
2.2 Modèles de données.....	22
2.2.1 Modèles de bases de données • Relationnel / SQL, NoSQL et NewSQL	22
2.2.2 Objets d'études associées à la mobilité des personnes	30

2.3	Entrepôts de données.....	32
2.3.1	Modélisation dimensionnelle	33
2.3.2	Faits	36
2.3.3	Dimensions	36
2.4	Entrepôt de données et visualisation.....	37
2.5	La visualisation en transport.....	38
2.5.1	Automatiser la création d’objets de visualisation cartographiques	39
CHAPITRE 3 QUESTIONNAIRE WEB ORIGINE-DESTINATION.....		40
3.1	Adaptation du questionnaire téléphonique	40
3.1.1	Du déplacement vers le lieu visité	40
3.2	Types de questions et choix de réponses.....	40
3.3	Base de données de collecte.....	44
3.3.1	Choix du modèle de base de données et de la plateforme de développement web.....	44
3.3.2	Schémas	46
3.3.3	Description des tables de la base de données de collecte	49
3.4	Configuration du questionnaire	55
3.4.1	Format YAML	55
3.4.2	Langues	56
3.4.3	Insertion de valeurs déjà déclarées par le répondant	56
3.4.4	Styles.....	56
3.4.5	Fichiers de configuration.....	57
3.5	Sections du questionnaire Origine-Destination conventionnel.....	65
3.5.1	Domicile.....	66
3.5.2	Profil • <i>Entrevue personne</i>	67
3.5.3	Membres du ménage	68

3.5.4	Déplacements.....	71
3.5.5	Section finale	82
3.6	Validations effectuées pendant l’entrevue.....	84
3.7	Module d’administration	85
3.7.1	Suivi des entrevues	85
3.7.2	Observation et validation des entrevues.....	88
3.7.3	Commentaires des répondants	89
3.7.4	Importation et exportation des données.....	89
3.8	Paradonnées enregistrées	91
3.9	Analyse des commentaires émis par les répondants à la fin des questionnaires	92
3.10	Enquêtes web réalisées dans le cadre de la thèse	95
3.10.1	Polytechnique Montréal • <i>Enquête auprès des étudiants et employés</i> Enquête- personne • Automne 2010.....	96
3.10.2	Ministère des transports du Québec • <i>Enquête régionale de Trois-Rivières</i> • <i>Enquête auprès des propriétaires de téléphones cellulaires</i> Prototype d’enquête-personne • Printemps 2011	97
3.10.3	Polytechnique Montréal et Université de Montréal • <i>Enquête auprès des étudiants et employés</i> Enquête-personne • Automne 2011	98
3.10.4	Ministère des transports du Québec • <i>Enquête régionale de Québec</i> Prototype d’enquête-ménage • Automne 2011	98
3.10.5	Ministère des transports du Québec • <i>Enquête régionale de Sherbrooke</i> • <i>Enquête auprès des étudiants post-secondaire de la région de Sherbrooke</i> Enquête-personne • Automnes 2012 et 2013	98
3.10.6	Agence métropolitaine de transport • <i>Enquête régionale de Montréal</i> Enquête- ménage • Automne 2013.....	99
3.10.7	Communauto et Bixi Montréal • <i>Enquête auprès des membres de l’autopartage et du vélopartage</i> Enquête-personne • Automne 2013	99

3.10.8	Forum vies mobiles • <i>Enquête sur la mobilité des 16-35 ans</i> Enquête-personne (non Origine-Destination) • Automne 2014.....	99
3.10.9	Enquête longue-distance • <i>Enquête prototype sur les déplacements longue-distance</i> Enquête-personne • 2014-2016	100
3.10.10	Universités de Toronto • <i>Enquête auprès des étudiants universitaires de la région de Toronto</i> Enquête-personne • Automne 2015.....	100
CHAPITRE 4	MODÉLISATION DE LA DURÉE D'ENTREVUE	101
4.1	Statistiques reliées à la durée d'entrevue.....	101
4.2	Méthodologie	104
4.2.1	Calcul des taux de réponse et d'achèvement	104
4.2.2	Critères d'achèvement des entrevues	104
4.2.3	Modèle de durée d'entrevue	105
4.3	Résultats de la modélisation de la durée d'entrevue.....	111
4.3.1	Modèle de durée d'entrevue pour enquête-ménage.....	111
4.3.2	Modèle de durée d'entrevue pour enquête-personne	113
4.4	Optimisation des ressources.....	116
4.5	Suivi des entrevues en temps réel par les administrateurs.....	116
4.6	Distribution temporelle des entrevues.....	116
CHAPITRE 5	ENTREPÔT DE DONNÉES DE MOBILITÉ	119
5.1	Modélisation dimensionnelle des données de mobilité	119
5.2	Attributs comparables	121
5.2.1	Procédure de correspondance des attributs comparables de la base de données de collecte vers l'entrepôt de données	122
5.3	Enrichissement des données	123
5.4	Validation et vérifications des données • ETL.....	124
5.4.1	Configuration des validations avant importation dans l'entrepôt de données	125

5.4.2	Audits	126
5.5	Cohérence des données dans l'entrepôt de données	127
5.6	Les tables de faits et leurs dimensions	127
5.6.1	Les dimensions enquête et échantillon	127
5.6.2	Les dimensions de dates	128
5.6.3	Les dimensions d'heures et d'heures étendues	128
5.6.4	Les dimensions d'objets géographiques	129
5.6.5	Les dimensions de calculs de chemin	129
5.6.6	Le ménage	130
5.6.7	La personne	131
5.6.8	Le lieu visité	132
5.6.9	Le séjour	133
5.6.10	La chaîne de déplacements	134
5.6.11	Le déplacement	135
5.6.12	Le segment de déplacement	137
5.7	Attributs personnalisés	138
5.8	Pondération des objets	139
5.9	Exportation des fichiers	139
5.9.1	Dictionnaire implicite des attributs	139
5.9.2	Choix des objets et des attributs à exporter	140
5.10	Processus de traitement et de validation des données	140
5.11	Contributions de l'entrepôt de données	141
5.12	Contenu de l'entrepôt de données actuel et importation de données externes	142
CHAPITRE 6	OBJETS DE VISUALISATION	143
6.1	Cartes animées de déplacements	143

6.1.1	Application utilisée pour la production de cartes animées	145
6.1.2	Méthodologie de production de cartes animées de déplacements	145
6.2	Cartes animées de la distribution géographique et temporelle d'objets sur un territoire.....	148
6.2.1	Préparation des données	148
6.3	De la collecte à la présentation des faits saillants	149
CHAPITRE 7 CONCLUSION.....		152
7.1	Contributions principales de la plateforme développée.....	152
7.2	Limitations	153
7.2.1	Représentativité des échantillons	153
7.2.2	Mesure de l'évolution du questionnaire au moyen des commentaires des répondants	154
7.2.3	Modélisation des durées d'entrevue à partir de variables déclarées pendant l'entrevue	154
7.2.4	Relative rigidité du modèle de données proposé pour l'entrepôt de données.....	154
7.3	Questions de recherche futures.....	154
7.3.1	Conception de l'interface d'entrevue et ergonomie cognitive.....	155
7.3.2	Analyse et modélisation des temps de réponse par type de question	155
7.3.3	Publication des méthodes de traitement, de validation et de stockage des données de mobilité existantes	156
7.3.4	Comparabilité des enquêtes	157
7.4	Poursuite du développement.....	157
7.4.1	Préparer la persistance polyglotte	157
7.4.2	Poursuite de la production de faits saillants	157
7.4.3	Adaptation du questionnaire web en une interface de saisie pour intervieweur téléphonique ou face-à-face.....	158
7.4.4	Importation de données de cartes à puce	158

7.5 Conclusion.....	158
BIBLIOGRAPHIE	160
ANNEXES	168

LISTE DES TABLEAUX

Tableau 2.1 : Avantages et inconvénients de l'utilisation du mode web dans les enquêtes de mobilité	14
Tableau 3.1 : Éléments de formulaire, types de données et apparence permis par la plateforme d'enquête	41
Tableau 3.2 : Nombre et proportion de commentaires positifs et négatifs associés aux différentes versions de l'outil.....	94
Tableau 3.3 : Contexte de réalisation pour 8 enquêtes-personnes et 2 enquêtes-ménages conduites entre 2010 et 2015	96
Tableau 4.1 : Statistiques descriptives, taux de réponse et taux d'achèvement pour les 2 enquêtes-ménage réalisées en 2011 et 2013.	102
Tableau 4.2 : Statistiques descriptives, taux de réponse et taux d'achèvement pour les 7 enquêtes-personne réalisées entre 2010 et 2013.....	103
Tableau 4.3 : Hypothèses utilisées pour la modélisation de la durée d'entrevue des enquêtes-ménage et enquêtes-personne.....	106
Tableau 4.4 : Statistiques descriptives des variables indépendantes utilisées dans la modélisation de la durée d'entrevue des enquêtes-ménage QC11 et MTL13	109
Tableau 4.5 : Statistiques descriptives des variables indépendantes utilisées dans la modélisation de la durée d'entrevue des enquêtes-personne PY11, UM11, TR11, SH13, CM13 et BX13	109
Tableau 4.6 : Modèle de durée d'entrevue pour les enquêtes-ménage	113
Tableau 4.7 : Modèle de durée d'entrevue pour les enquêtes-personne.....	115
Tableau 5.1 : Extrait de la table d'audits.....	126

LISTE DES FIGURES

Figure 1-1 : Schéma général de la plateforme intégrée	5
Figure 2-1 : Diagramme entité-association des objets relatifs à l'analyse de la mobilité	30
Figure 2-2 : Modèle relationnel SAQE, Ministère des transports du Québec	31
Figure 2-3 : Exemple de modèle dimensionnel en étoile pour la vente au détail	35
Figure 2-4 : Comparaison entre le modèle normalisé et le modèle dimensionnel	35
Figure 2-5 : De la réalité au cerveau.....	37
Figure 3-1 : Schéma simplifié des objets et du processus d'entrevue	47
Figure 3-2: Modèle de données de la base de données de collecte.....	48
Figure 3-3 : Section du questionnaire associée au domicile du répondant	67
Figure 3-4 : Section du questionnaire associée au profil du répondant • <i>Entrevue personne</i>	68
Figure 3-5 : Section du questionnaire associée au profil de tous les membres du ménage	70
Figure 3-6 : Section du questionnaire pour la sélection d'un membre du ménage.....	71
Figure 3-7 : Section d'introduction demandant au répondant s'il a effectué au moins un déplacement lors du dernier jour ouvrable	72
Figure 3-8 : Section d'introduction à la saisie des lieux visités	73
Figure 3-9 : Section de sélection du lieu de départ de la journée	74
Figure 3-10 : Section indiquant la liste des lieux visités et des activités	75
Figure 3-11 : Section permettant de géolocaliser un lieu visité.....	75
Figure 3-12 : Section montrant la liste des lieux visités accompagnée de boutons permettant d'insérer, de modifier ou de supprimer des lieux	76
Figure 3-13 : Section permettant la saisie des heures de départ et d'arrivée des lieux visités	77
Figure 3-14 : Section montrant la saisie des modes de transport.....	78

Figure 3-15 : Section de liste des modes de transport empruntés	79
Figure 3-16 : Section montrant les questions spécifiques aux déplacements utilisant le transport collectif.....	80
Figure 3-17 : Section montrant les questions spécifiques aux déplacements de type « auto-conducteur »	81
Figure 3-18 : Section permettant de sélectionner les autoroutes et ponts et tunnels empruntés lors d'un déplacement de type « auto-conducteur ».....	82
Figure 3-19 : Section finale et ses questions habituelles.....	84
Figure 3-20: Module d'administration des enquêtes web • <i>Section de suivi des entrevues</i>	86
Figure 3-21: Suivi des entrevues • <i>Distribution des entrevues débutées et complétées par jour</i>	87
Figure 3-22: Suivi des entrevues • <i>Distribution des entrevues débutées et terminées par heure</i> ...	87
Figure 3-23: Suivi des entrevues • <i>Distribution des durées d'entrevue</i>	87
Figure 3-24: Suivi des entrevues • <i>Durée d'entrevue en fonction du nombre de déplacements</i>	88
Figure 3-25: Suivi des entrevues • <i>Taux d'achèvement des sections du questionnaire</i>	88
Figure 3-26: Interface de sélection des attributs à exporter pour les objets « déplacements » ..	90
Figure 3-27: Extrait d'un fichier au format CSV exporté pour les objets « déplacements » visualisé à l'aide de <i>Microsoft Excel</i>	91
Figure 3-28 : Évolution de la proportion de commentaires positifs relatifs à l'interface	94
Figure 4-1 : Histogrammes de la durée d'entrevue et de son logarithme naturel et diagramme quantile-quantile de la durée d'entrevue comparée à une distribution normale pour les enquêtes-ménage	111
Figure 4-2 : Distribution temporelle des entrevues pour les enquêtes-ménage	118
Figure 4-3 : Distribution temporelle des entrevues pour les enquêtes-personne.....	118
Figure 5-1 : Correspondance des attributs comparables dans la base de données de collecte et dans l'entrepôt de données (Exemple pour les activités)	123
Figure 5-2 : Schéma dimensionnel • Ménage	131

Figure 5-3 : Schéma dimensionnel • Personne	132
Figure 5-4 : Schéma dimensionnel • Séjour.....	133
Figure 5-5 : Schéma dimensionnel • Lieu visité	134
Figure 5-6 : Schéma dimensionnel • Chaîne de déplacements	135
Figure 5-7 : Schéma dimensionnel • Déplacement	137
Figure 5-8 : Schéma dimensionnel • Segment.....	138
Figure 6-1 : Capture d'une carte animée de déplacements montrant l'achalandage sur le métro à l'heure de pointe du matin	143
Figure 6-2 : Capture d'une carte animée de déplacements montrant l'achalandage sur les routes vers Polytechnique	144
Figure 6-3 : Capture d'une carte animée de déplacements montrant les déplacements de l'échantillon complet de l'enquête régionale de Québec 2011.....	144
Figure 6-4 : Carte animée de la distribution géographique et temporelle des usagers du transport collectif et des conducteurs pour la région de Montréal.....	149
Figure 6-5 : Module de présentation des faits saillants	151
Figure 7-1 : Quartiles des temps de réponse par type de question	156
Figure B-1 : Schéma des tables de l'entrepôt de données	172

LISTE DES ANNEXES

ANNEXE A	ÉCHANTILLON DE COMMENTAIRES POSITIFS, NÉGATIFS ET NEUTRES CONCERNANT L'INTERFACE ET LE QUESTIONNAIRE	168
ANNEXE B	TABLES DE L'ENTREPÔT DE DONNÉES	172
ANNEXE C	CALCULATEURS DE CHEMIN ET DONNÉES CARTOGRAPHIQUES LIBRES ...	179

LISTE DES SIGLES ET ABRÉVIATIONS

ACID	Propriétés d'une transaction informatique (atomicité, cohérence, isolation, durabilité)
AMT	Agence métropolitaine de transport
BI	Business Intelligence (Intelligence d'affaires)
CAPI	Computer-Aided Personal Interviewing (Entrevues en personne assistées par ordinateur)
CATI	Computer-Aided Telephone Interviewing (Entrevues téléphoniques assistées par ordinateur)
CAWI	Computer-Aided Web Interviewing (Entrevues web assistées par ordinateur)
COST	European Cooperation in Science and Technology (Coopération européenne en sciences et technologies)
CSS	Cascading Style Sheets (Feuilles de style en cascade)
CSV	Comma Separated Value (Format de fichier de données)
ETL	Extract, Transform, Load (Extraction, Transformation, Importation)
GTFS	General Transit Feed Specification (Spécification pour publication d'horaires de transport collectif)
MTQ	Ministère des transports du Québec
OD	Origine-Destination
OGC	Open Geospatial Consortium (Consortium géospatial ouvert)
OLAP	Online Analytical Processing (Traitement analytique en ligne)
OSM	Open Street Map (Données cartographiques libres)
OSRM	Open Source Routing Machine (Logiciel libre de calcul de chemins)
POLY	Polytechnique Montréal
SAQE	Saisie assistée des questionnaires d'enquête (Logiciel utilisé au MTQ)
SGBD	Système de gestion de bases de données
SHANTI	Survey Harmonisation with New Technologies Improvement (Harmonisation des enquêtes dans un contexte d'innovation technologique)
SIG	Système d'information géographique
SQL	Structured Query Language (Langage de requête structurée)
UdeM	Université de Montréal
UdeS	Université de Sherbrooke
UUID	Universally Unique Identifier (Identifiant universel unique)
XML	Extensible Markup Language (Langage de balisage extensible)
YAML	YAML Ain't Markup Language, anciennement Yet Another Markup Language (Format de représentation de données par sérialisation)

CHAPITRE 1 INTRODUCTION

Que ce soit pour en accroître l'efficacité et la durabilité ou pour en planifier l'évolution, il est essentiel pour les différents acteurs œuvrant dans le domaine du transport des personnes de dresser un portrait juste et précis de la mobilité et des réseaux sur lesquels ils interviennent. Sachant cela, l'augmentation incessante de la quantité de données, la multiplication et la complexité croissante des plateformes logicielles utilisées et l'imputabilité grandissante des intervenants en regard des décisions qu'ils prennent soulèvent des questionnements sur l'utilisation efficace des ressources humaines et matérielles dans ce domaine. De plus, la problématique de la durabilité en transport suscite le besoin de mesurer l'empreinte environnementale des déplacements, l'accessibilité, les effets sur la santé et sur l'équité sociale ainsi que l'influence de la mobilité sur les économies locale et globale. Conséquemment, les intervenants du milieu doivent planifier leurs projets simultanément à court et à long termes sous de nombreuses facettes, tout en s'assurant que les changements apportés aux systèmes de transport dont ils ont la responsabilité puissent répondre aux besoins de mobilité actuels et futurs.

En planification des transports, la collecte de données est la première et l'une des plus complexes des étapes d'un long processus qui permet de dresser le portrait de la mobilité d'une population. En effet, l'analyse des comportements de mobilité des personnes requiert d'emblée la collecte de données qui décrivent autant les attributs socio-démographiques que les déplacements des personnes. Parmi les sources de données les plus utilisées pour répondre à cet objectif, on retrouve les recensements, les enquêtes de circulation routière, les enquêtes à bord des véhicules de transport collectif, les relevés GPS et les enquêtes Origine-Destination. Ces dernières, en particulier, visent à évaluer les tendances des comportements de déplacements et à obtenir un diagnostic des problématiques entourant les réseaux de transport empruntés par la population. Or, de nombreux ménages et personnes sont difficiles à rejoindre pour une entrevue téléphonique ou face-à-face pour des raisons de confidentialité ou de disponibilité (Behrens, Freedman, & McGuckin, 2009). D'autre part, de plus en plus de ménages, les jeunes en particulier¹, ne possèdent

¹ La proportion de ménages canadiens de 35 ans et moins utilisant le cellulaire seulement est passé de 26,1% en 2008 à 60,6% en 2013, selon l'enquête sur le service téléphonique résidentiel 2013 de Statistiques Canada.

plus de ligne téléphonique fixe au domicile et utilisent exclusivement le cellulaire (Statistiques Canada, 2014). Par ailleurs, lorsque un ménage est rejoint par téléphone, il n'y a habituellement qu'une seule personne qui répond pour l'ensemble des membres de son ménage, ce qui provoque des biais significatifs (Badoe & Steuart, 2002; Verreault & Morency, 2014). Ces problématiques nouvelles affectent la représentativité des échantillons tirés des annuaires téléphoniques utilisés couramment lors les enquêtes Origine-Destination. En guise de solution, de nouvelles méthodes de recrutement et d'entrevue ont été proposées, comme les recrutements par courriel auprès de grands générateurs de déplacements, les entrevues web et la collecte de données GPS.

Une fois la collecte complétée, le traitement, la validation et le stockage des données de mobilité sont les étapes essentielles à réaliser pour permettre aux analystes et planificateurs d'étudier les comportements de mobilité et de proposer des modèles et des outils d'aide à la gestion et à l'optimisation des réseaux de transport. Dans ce domaine, la modélisation totalement désagrégée, qui modélise de manière individuelle le comportement de mobilité des personnes, a permis de répondre en partie aux défis technologiques des dernières décennies (Trépanier, 1999). De plus, elle a favorisé la construction de modèles plus précis et plus justes en mettant en valeur les données désagrégées. Finalement, elle a permis de mieux structurer les relations entre les objets de transport en proposant un modèle orienté objet pour les données relatives à la mobilité. Toutefois, dû à l'accroissement du nombre de données de mobilité collectées et à la multiplication des sources de collecte (Rousseau, 2016), le fardeau et les défis associés à la fusion des données de mobilité n'a fait que s'amplifier depuis quelques années (Bayart, Bonnel, & Morency, 2009; Mohammadian & Zhang, 2007; Venigalla, 2004). De plus, la nécessité d'intégrer de nouveaux modes de transport (autopartage, vélopartage, covoiturage commercial, multimodalité, etc., (Sioui, Morency, & Trépanier, 2013)) et de nouveaux indicateurs de mobilité durable (Sioui, 2014) aux modèles et aux analyses a rendu encore plus complexes le traitement et le stockage des données. Conséquemment, l'analyse de la mobilité est devenue une tâche plus lourde qui se traduit souvent par de longs délais dans la production de faits saillants² et par un retard de l'analyse des tendances dans les

² Par exemple, un an s'est écoulé entre la fin de la collecte de données et la présentation des faits saillants lors de l'enquête régionale de Montréal de 2013 conduite par l'AMT et ses partenaires.

comportements de mobilité. Pourtant, dans le domaine de l'intelligence d'affaires, la mise en place d'entrepôts de données conçus pour faciliter l'analyse a permis d'accélérer et de simplifier l'intégration de données de provenances multiples (Jukic, 2006; Kimball, 1997). Du même coup, ils ont contribué à simplifier la consultation des bases de données d'analyse. De surcroît, ces entrepôts ont permis l'enrichissement des données et la génération de visualisations innovatrices au moyen de procédures automatisées et de pré-calculs, tout en favorisant les analyses comparatives et tendanciennes grâce à des attributs durables, c'est-à-dire des échelles et des plages de valeurs qui ne changent pas d'un ensemble de données à l'autre et d'une année à l'autre.

1.1 Hypothèses de recherche

La thèse s'inscrit en réponse aux deux problématiques énoncées plus haut : la première concernant la difficulté à rejoindre certains répondants et la deuxième qui s'appuie sur la complexité de gestion et de traitement des données de mobilité. D'une part, la thèse étudie et propose un nouvel outil de collecte de données de mobilité, et d'autre part, elle présente une nouvelle approche pour le stockage et l'enrichissement des données de mobilité.

En premier lieu, la thèse pose comme hypothèse que les enquêtes de mobilité peuvent demeurer des outils de collecte valables pour l'étude des comportements de mobilité, pour autant que l'on mette de l'avant de nouveaux questionnaires web plus flexibles et plus conviviaux pour les répondants et les administrateurs. En parallèle, la mise en valeur des paradonnées collectées pendant les entrevues web peut supporter une analyse plus fine du comportement des répondants, et en particulier rend possible la modélisation de la durée d'entrevue pour des fins de suivi, de prédiction en temps réel et d'optimisation des ressources.

Dans un deuxième temps, la thèse énonce l'hypothèse que la construction d'un schéma de modélisation dimensionnelle (Kimball, 1997; Kimball & Ross, 2013) favorisera l'intégration des données de mobilité dans un même entrepôt de données et assurera une structure plus compréhensible et plus robuste à ces données. Plus particulièrement, cette modélisation permettra de présenter des faits saillants rapidement, de réaliser des analyses multiples et ciblées, de favoriser la comparabilité d'un ensemble de données à l'autre et d'automatiser la génération d'objets de visualisation.

1.2 Description générale, contexte et portée de la thèse

Considérant la complexité élevée rencontrée lors de la collecte et du traitement des données d'enquêtes Origine-Destination en particulier, le projet de recherche associé à la thèse comprend la réalisation d'une plateforme logicielle favorisant l'intégration des phases de collecte, de suivi, de traitement, d'analyse, de visualisation et de diffusion des données de mobilité au moyen d'un questionnaire web et soutenu par un entrepôt de données de mobilité.

Plus précisément, un questionnaire en ligne permettant la tenue d'enquêtes de type Origine-Destination est proposé. Une modélisation des durées d'entrevue est ensuite réalisée au moyen des données relatives au processus d'entrevue. Par la suite, un schéma d'entrepôt de données de mobilité facilitant l'analyse, la visualisation et la diffusion des données validées est présenté.

Le projet a été réalisé au Québec en utilisant des données locales et régionales. Toutefois, la plateforme a été, dès le départ, conçue de manière à pouvoir être générique et utilisée dans plusieurs contextes, et elle a été réalisée avec l'objectif d'accroître la portée des analyses qui la mettront à profit.

1.3 Objectifs principaux

Les objectifs principaux de la thèse sont d'une part de proposer un outil de collecte de données de mobilité sur le web plus convivial et flexible, par ailleurs de modéliser la durée d'entrevue au moyen de données sur le processus d'entrevue, et enfin de normaliser, simplifier, accélérer et automatiser les processus de traitement, d'enrichissement et de stockage des données de mobilité au moyen d'un entrepôt de données. En fait, la plateforme développée dans le cadre de la thèse a pour objectif général de permettre une collecte et une gestion des données de mobilité qui, au final, auront davantage de sens pour les planificateurs qui les utiliseront.

1.4 Schéma général de la plateforme

Le schéma général de la plateforme réalisée dans le cadre de la thèse est illustré à la Figure 1-1.

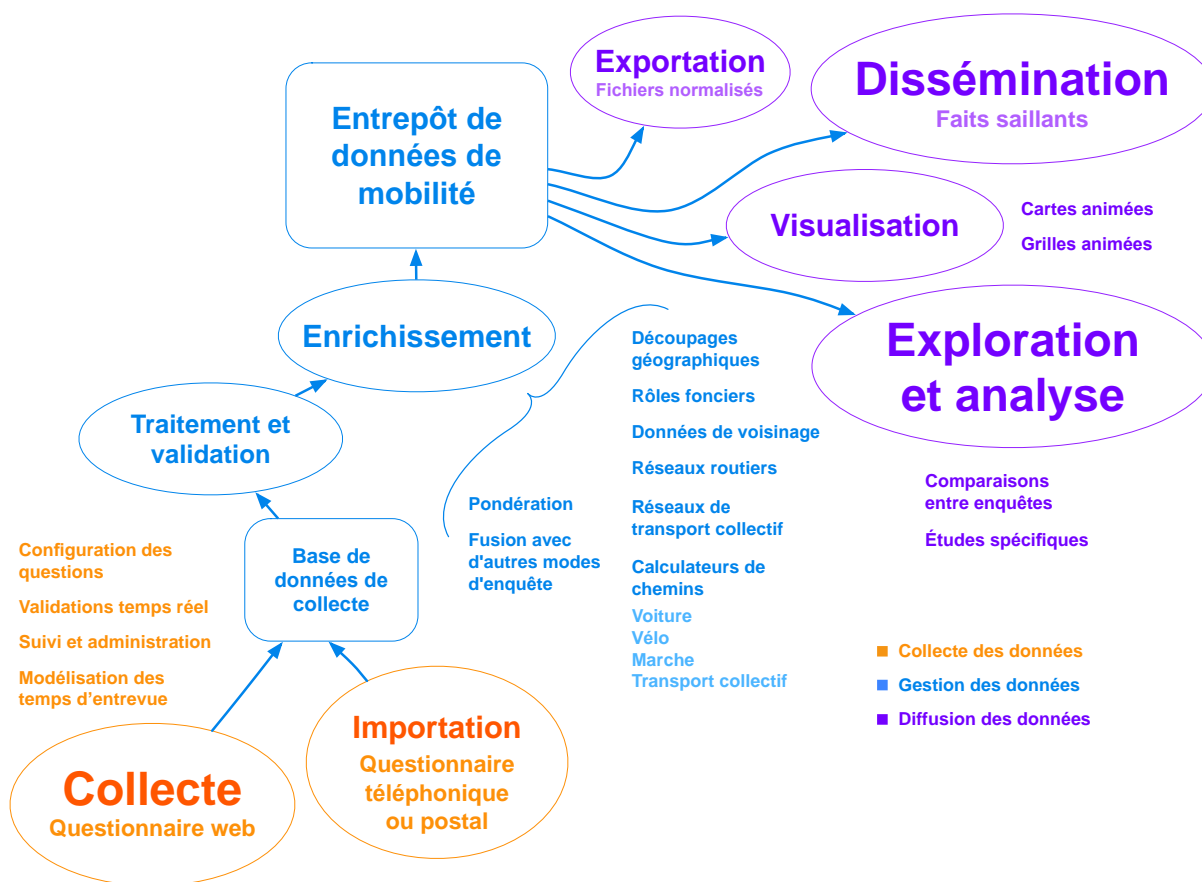


Figure 1-1 : Schéma général de la plateforme intégrée

Lors de la collecte de données ou de l'importation de données provenant d'autres modes d'enquêtes, les données sont enregistrées dans la base de données de collecte et, après traitement, validation et enrichissement, sauvegardées dans l'entrepôt de données. L'étape d'enrichissement des données comprend un ensemble de modules et fait appel à des sources de données externes. Une fois les données entreposées, la dissémination des résultats au moyen de faits saillants et de visualisations peut être mise de l'avant. En parallèle, un module d'exportation des fichiers de résultats permet l'utilisation des données validées dans d'autres outils de modélisation et d'analyse.

1.5 Aperçu du document

Au Chapitre 2, une revue de littérature à deux volets sera présentée. Le premier volet portera sur les enquêtes Origine-Destination et la collecte de données de mobilité — en particulier les questionnaires web — et le deuxième sur les modèles de données et la modélisation dimensionnelle, suivi d'une courte introduction des défis de visualisation et de création de cartes animées en transport. Le questionnaire web Origine-Destination développé sera décrit et illustré au Chapitre 3, suivi au Chapitre 4 de la modélisation de la durée d'entrevue. Le Chapitre 5 présentera la spécification de l'entrepôt de données de mobilité. Des exemples d'objets de visualisation et de faits saillants soutenus par l'entrepôt de données suivront au Chapitre 6. Enfin, les conclusions, les propositions de développements futurs et les perspectives seront énoncées au Chapitre 7.

CHAPITRE 2 ÉTAT DES CONNAISSANCES

La revue de littérature est construite en deux volets principaux. Au premier volet, elle résume les méthodes actuelles de collecte de données de mobilité et les problématiques sous-jacentes, en mettant en relief les avancées récentes dans l'utilisation des entrevues web. Au second volet, les paradigmes entourant la gestion des données de mobilité sont énoncés, suivis d'une introduction à la modélisation dimensionnelle empruntée au domaine de l'intelligence d'affaires pour la mise sur pied d'entrepôts de données performants et durables.

2.1 Collecte de données de mobilité

En guise d'introduction, un bref résumé de l'état de l'art dans le domaine des enquêtes de mobilité depuis l'apparition des premières enquêtes téléphoniques assistées par ordinateur (CATI) sera exposé. Par la suite, une revue des progrès réalisés dans la tenue d'enquêtes web (tous domaines confondus) viendra mettre en contexte les enquêtes web axées sur la mobilité.

2.1.1 Enquêtes Origine-Destination

Les enquêtes Origine-Destination représentent une source de données importante pour l'étude de la mobilité des populations. Lors des entrevues, les répondants sont invités à fournir leur profil socio-démographique et à déclarer les lieux qu'ils ont visités pendant une période de temps déterminée — habituellement une journée complète de 24 heures —, tout en prenant soin de mentionner les activités effectuées sur ces lieux ainsi que les modes de transport empruntés pour s'y rendre. Ces informations, quelquefois accompagnées d'autres questions sur la mobilité et les déplacements, assurent un diagnostic ponctuel de la mobilité de la population étudiée. Afin de rendre possible les comparaisons et le suivi de l'évolution des tendances, l'exercice est, la plupart du temps, répété sur une base décennale, quinquennale ou annuelle, et/ou conduit en continu avec des échantillons plus modestes.

2.1.2 Deux types d'enquêtes Origine-Destination

Lorsque réalisées pour analyser la mobilité des habitants d'une région dans son ensemble, les enquêtes Origine-Destination s'adressent généralement aux ménages de la région — ce sont des

enquêtes-ménage — et cherchent à collecter, lors d’une même entrevue, des données sur les déplacements de la totalité des membres du ménage. Par ailleurs, les enquêtes-personne recueillent des informations sur la mobilité d’une seule personne par entrevue.

2.1.2.1 Enquête-ménage

Dans le cadre d’une enquête-ménage, il est habituel de recueillir les caractéristiques démographiques de chaque membre du ménage ainsi que les déplacements effectués par ceux-ci durant la période spécifiée. Généralement, lorsque ces enquêtes sont téléphoniques et font appel à un seul répondant dans le ménage, ce dernier est appelé à révéler les informations des autres membres de son ménage après avoir répondu lui-même à l’entrevue. On l’appelle alors le répondant proxy. Du reste, pour des raisons de dépendance aux déplacements des adultes, de confidentialité et de sécurité, les questions relatives à la mobilité — les activités et déplacements — sont rarement posées pour les enfants en bas âge (les moins de 5 ans par exemple) (Stopher 1996).

2.1.2.2 Enquête-personne

Lorsqu’une enquête-personne est menée, par exemple auprès de la communauté d’un grand générateur de déplacement — une entreprise ou un établissement institutionnel — les questions portant sur la mobilité ne concernent généralement qu’un seul individu par entrevue. Il est néanmoins d’usage de recueillir des informations d’ordre démographique sur les autres membres du ménage du répondant dans un souci éventuel de comparabilité avec une enquête-ménage régionale ou pour des besoins de pondération au moyen d’un recensement de la population locale.

2.1.3 Mode de recrutement et mode de collecte

Le mode de recrutement est défini comme étant le médium utilisé pour rejoindre, dans un premier temps, les répondants. Le mode de collecte est le médium utilisé pour conduire l’entrevue elle-même.

Les modes actuellement utilisés autant pour le recrutement que pour la collecte sont le téléphone, le questionnaire sur papier, l’entrevue en personne (ou face-à-face), le courriel et le web. Le GPS n’est utilisé que pour la collecte de données de mobilité. Il arrive souvent qu’un mode

distinct soit sélectionné pour le recrutement et pour la collecte. Dans le cas où des modes différents sont mis de l'avant pour une même phase (habituellement la collecte), on parle alors d'enquête à modes mixtes. Afin de mettre en contexte les différents modes de collecte, ils sont définis brièvement dans les sections suivantes.

2.1.3.1 Enquêtes en personne (face-à-face)

Les enquêtes de mobilité étaient originalement conduites en entrevues face-à-face. L'intervieweur se rendait directement chez le répondant, ou l'interceptait à un lieu déterminé (sur la route ou à un lieu de transfert modal) (Bonnell, Lee-Gosselin, Madre, & Zmud, 2009; Stopher, 2009). On retrouve encore ce mode de collecte, notamment en France, bien que son coût élevé ait encouragé les administrateurs d'enquêtes de mobilité à proposer d'autres modes de collecte.

2.1.3.2 Enquêtes sur papier

Le mode de collecte sur papier est habituellement utilisé pour les envois postaux de questionnaires ou pour les enquêtes à bord de véhicules de transport collectif. Il a par contre tendance à être remplacé par le mode téléphonique ou le mode web pour réduire les coûts et rendre le questionnaire plus convivial.

2.1.3.3 Enquêtes téléphoniques assistées par ordinateur (CATI) sur la mobilité

Les enquêtes téléphoniques assistées par ordinateur (Computer-Assisted Telephonic Interview : CATI) ont été proposées dans le but de bonifier et de normaliser les entrevues téléphoniques (Groves, Berry, & Mathiowetz, 1980; Palit & Sharp, 1983). En demandant à l'intervieweur de suivre un script d'entrevue prédéfini et en instrumentant de manière précise le processus de collecte de données, les enquêtes CATI ont permis d'encadrer et de standardiser le processus d'entrevue (Kelly, 2008). De plus, elles ont encouragé la collecte de données de meilleure qualité (Leeuw, Hox, & Snijkers, 1998) lorsque des questionnaires complexes étaient nécessaires comme dans le cas des enquêtes Origine-Destination. Toutefois, en raison du nombre élevé de supervisions supplémentaires prescrites et des ressources matérielles et connaissances techniques requises, un minimum d'environ 1000 entrevues doivent être complétées pour obtenir

une réduction appréciable des coûts lors de l'utilisation d'une plateforme CATI (Leeuw et al., 1998).

2.1.3.4 Enquêtes web/courriel assistées par ordinateur (CAWI)

Dès le début des années 2000, les taux de réponse de plusieurs enquêtes ont commencé à baisser, ce qui a conduit les chercheurs à se pencher sur cette problématique et, par la suite, à évaluer de nouvelles méthodologies ou à proposer de nouveaux modes d'enquête pour la collecte de données (Couper, 2011): les entrevues par courriel et sur le web (Computer-Assisted Web Interview : CAWI). Comme les questionnaires courriel fonctionnent essentiellement de la même façon que les questionnaires web³, ils sont considérés comme équivalents aux enquêtes web de type CAWI.

2.1.3.5 Enquêtes GPS

Les enquêtes GPS permettent d'obtenir les déplacements des individus de manière passive. Soit l'on prête un appareil GPS au répondant pour qu'il enregistre ses traces pendant une période déterminée, soit on lui fournit une application à installer sur son propre appareil muni d'un GPS. Dans certains cas, on tente d'obtenir les activités effectuées à chacun des lieux visités ainsi que les modes empruntés pour effectuer les déplacements détectés à l'aide d'un questionnaire web ou d'une interface de saisie intégrée dans l'application utilisée. Toutefois, de plus en plus de projets de recherche s'articulent autour de la détection automatique des activités (Rasouli, 2014; Schüssler, N., & Axhausen, K. W., 2008; Zhao, Pereira, Ball, & Kim, 2015) et des modes (Stenneth, Wolfson, Yu, & Xu, 2011; Xiao, Juan, & Zhang, 2015; Zong, Bai, Wang, Yuan, & He, 2015) à partir des traces GPS.

2.1.3.6 Enquêtes de type carnet de déplacements (journal)

Les enquêtes avec questionnaire répété sur plusieurs jours, semaines, mois ou années pour un même ménage ou une même personne permettent d'étudier le rythme dans les comportements

³ Les questionnaires courriel utilisent le navigateur intégré du lecteur de courriel du répondant, ce navigateur étant construit sur le même moteur que les navigateurs web.

de mobilité, de même que les différences dans les motifs de déplacements d'une période à l'autre (Axhausen, Zimmermann, Schönfelder, Rindsfuser, & Haupt, 2002). De surcroît, les enquêtes de type carnets de déplacements favorisent l'analyse de l'influence des choix antérieurs d'un individu sur sa mobilité (Schlich & Axhausen, 2003). Ces types d'enquêtes peuvent utiliser un ou plusieurs modes de collecte.

2.1.3.7 Enquêtes à modes mixtes

Les enquêtes de mobilité à mode mixte sont de plus en plus mises de l'avant pour obtenir des données de meilleure qualité, pour minimiser différents biais et accroître les taux de réponse globaux (Bayart & Bonnel, 2012; 2015). Par contre, de nouveaux biais de comparabilité et de compatibilité apparaissent puisque les questionnaires sont différents, les questions ne sont pas toujours formulées de la même manière et les répondants qui utilisent un mode ou un autre sont souvent de profils socio-démographiques différents (Bayart & Bonnel, 2008).

2.1.3.8 Harmonisation des enquêtes de mobilité

Pour faire face à la prolifération de questionnaires et de méthodologies d'enquêtes en Europe notamment, le COST (European Cooperation in Science and Technology) a proposé un document de référence pour harmoniser les méthodes d'enquêtes sur la mobilité (Armoogum & INRETS, 2014). Cet effort de normalisation est prometteur pour assurer la mise en valeur des enquêtes de mobilité en Europe et ailleurs dans le monde. Toutefois, bien que l'un des objectifs de ce document soit de proposer des données harmonisées, les méthodologies de traitement, de validation et de stockage des données n'y sont pas définies.

2.1.4 Suivi des entrevues par des paradonnées

Le terme spécialisé paradonnée (paradata), proposé pour la première fois par (Couper, 1998) et précisé dans (Kreuter, Couper, & Lyberg, 2010), est défini comme l'information collectée de manière automatique lors d'entrevues assistées par ordinateur (CATI et CAWI) pour des besoins de surveillance, de gestion et d'évaluation de la qualité. Le comportement des répondants pendant l'entrevue est en ce sens évalué grâce à l'analyse des paradonnées. Alors que les durées d'entre-

vue et les abandons sont presque toujours archivés, des informations plus précises peuvent également être rassemblées durant les entrevues web notamment. En particulier, les touches de clavier enfoncées, le mouvement et les clics de la souris sont quelquefois conservés (l'enregistrement de ce type de données doit toutefois faire l'objet d'un consentement de la part du répondant pour des questions de confidentialité), de même que les caractéristiques et spécifications du navigateur, du système d'exploitation ou de l'appareil utilisés. Toutefois, la tendance à conserver de plus en plus de parodonnées — tendance encouragée par de meilleurs serveurs et de nouveaux outils de collecte — requiert un grand nombre de ressources pour intégrer, valider et analyser les parodonnées. C'est pour cette raison que des auteurs préconisent un meilleur encadrement et une gestion plus structurée des parodonnées (Kreuter, 2013).

2.1.4.1 Parodonnées des enquêtes-ménages régionales Origine-Destination téléphoniques de la région de Montréal

Les enquêtes-ménages régionales Origine-Destination téléphoniques sont tenues à tous les cinq ans depuis les années 1970 dans la grande région de Montréal. La première enquête de type CATI a été conduite en 1993 et depuis, des améliorations techniques et méthodologiques ont permis d'encadrer autant l'administration des enquêtes que les entrevues elles-mêmes (Trépanier, Champleau, & Morency, 2008). Avec les années, une grande quantité de parodonnées ont été rassemblées et analysées, permettant à (Morency, 2008) et à (Champleau, 2003) d'établir une méthodologie structurée d'évaluation de la qualité des entrevues. De surcroît, en sauvegardant les statuts d'appels, le début et la fin des entrevues et en enregistrant la voix et toutes les actions de l'intervieweur à l'intérieur de l'interface de saisie, il était même possible, ultérieurement, de simuler entièrement l'entrevue.

2.1.5 Enquêtes web sur la mobilité

L'intégration du mode web aux enquêtes de mobilité comporte certes des atouts considérables, mais également plusieurs inconvénients. Certains des bienfaits et des désavantages vont s'estomper ou, au contraire, s'accroître au fil des années selon les tendances observées actuellement.

2.1.5.1 Avantages et inconvénients du mode web pour la tenue d'enquêtes

Le Tableau 2.1 présente les avantages et les désavantages du mode web, ainsi que la tendance observée pour chacun (à la hausse, à la baisse ou plutôt stable). La suite reprend en détail les avantages et inconvénients.

Tableau 2.1 : Avantages et inconvénients de l'utilisation du mode web dans les enquêtes de mobilité (Bourbonnais & Morency, 2013b)

Avantages		Inconvénients	
Description	Tendance	Description	Tendance
Légende			
↑ / ↑↑ : avantages qui devraient s'accroître avec le temps → : devraient rester plus ou moins stables		↑ / ↑↑ : désavantages qui devraient s'améliorer ↓ : désavantages qui pourraient s'accroître → : devraient rester plus ou moins stables	
Coût faible (coût marginal presque nul) (Armoogum, Axhausen, & Madre, 2009)	→	Accès Internet et vitesse de connexion inégaux	↑↑
Mode moins contraignant (le moment pour répondre est choisi par le répondant) (Armoogum et al., 2009)	→	Les taux de réponse sont déjà en baisse (sollicitation des répondants en forte hausse) (Armoogum et al., 2009)	↓
Collecte rapide des données	→	Facilité d'utilisation des répondants inégale (Alsnih, 2007)	↑
Validation des données en temps réel (moins d'erreur dans les données recueillies) (Armoogum et al., 2009; Timmermans & Hato, 2009)	↑	La concentration des répondants est mise à l'épreuve (multitâches) (Adler & Benbunan-Fich, 2012; Lynn & Kaminska, 2013)	↓
Adaptabilité et flexibilité (le questionnaire peut s'adapter au répondant et aux réponses précédentes)	↑↑	Le questionnaire doit être adapté au mode web, ce qui rend la comparaison/intégration plus difficile (Braunsberger, Wybenga, & Gates, 2007)	→
Permet d'étudier le comportement des répondants lors de l'entrevue	↑↑	Les listes de courriels sont de mauvaise qualité et/ou difficiles à obtenir	?
Possibilité de modifier le questionnaire pendant la tenue de l'enquête	↑↑	Le recrutement doit souvent être effectué en utilisant un autre mode	?
Rend la gestion de l'enquête moins lourde	↑	Plusieurs biais peuvent être induits (Alsnih, 2007) (Biais de couverture, d'échantillonnage, de mesure et de non-réponse)	↑
Permet la diffusion de résultats en temps réel et la présentation de données statistiques préliminaires aux répondants à la fin de leur entrevue	↑↑	Différences techniques d'un répondant à l'autre (navigateur, système d'exploitation, connexion Internet, appareil, etc.) (Armoogum et al., 2009)	↑
Permet un nombre élevé d'entrevues simultanées (prévoir l'équipement et la connexion adéquate pour répondre aux heures d'achalandage élevé)	↑	La compréhension des questions par les répondants est difficile à évaluer et peut causer des réponses de mauvaise qualité ou des biais selon le niveau de scolarité, l'âge ou le profil social	→
Possibilité de poser des questions aléatoires	→		
Accroît le niveau d'interactivité et de visualisation (Armoogum et al., 2009; Bonnel et al., 2009)	↑↑		
Possibilité de capter des groupes difficiles à rejoindre (Riandey & Quaglia, 2009)	↑↑		

Coût • Le faible coût des enquêtes en ligne est une des raisons principales pour lesquelles elles sont apparues. En fait, les enquêtes web réduisent massivement le coût marginal des entrevues (Armoogum et al., 2009), alors que leur coût fixe de développement et d'administration s'approche de celui des enquêtes conventionnelles.

Qualité des données et validation • La qualité technique des données obtenues d'un questionnaire web dépend autant de la formulation des questions, c'est-à-dire de leur précision, clarté et concision, que des procédés de validation et de filtrage intégrés à l'application d'entrevue. Comme la véracité des informations déclarées est pratiquement impossible à vérifier, il importe à tout le moins de diminuer les erreurs de mesure et de planifier avec rigueur le schéma de base de données ainsi que les processus de sauvegarde du questionnaire web. L'attention apportée à l'interface utilisateur et le degré de raffinement de la conception graphique influencent la confiance et la motivation des répondants (Dillman, 2011), et, conséquemment, la qualité des données recueillies. En comparaison avec les autres modes d'enquête, certaines informations personnelles sont livrées plus facilement en ligne ou par la poste que lors d'une entrevue face-à-face ou téléphonique (De Leeuw & Hox, 2011), mais le fait que l'entrevue soit auto-administrée ne permet pas de diriger le répondant ou de le guider lorsque des problèmes techniques ou de compréhension surviennent.

Flexibilité et adaptation du questionnaire • Dans la plupart des cas, il est recommandé de favoriser la plus grande flexibilité possible lors du déroulement de l'entrevue en adaptant les questions au profil du répondant. Des filtres permettent habituellement de retirer les sections et les questions non applicables ou de modifier la formulation et l'ordre des questions en fonction des réponses précédentes dans le but de conserver l'attention du répondant et de minimiser le temps d'entrevue. Par ailleurs, comme la sauvegarde s'effectue habituellement à la fin de chacune des sections du questionnaire ou en continu, le répondant peut ainsi quitter l'entrevue et y retourner plus tard. Il peut toutefois y avoir une certaine confusion auprès des répondants d'un même ménage qui pourrait provoquer une déclaration de réponses pour la mauvaise personne.

Conception du questionnaire et interactivité • Tout en suivant les standards applicables à la conception de sites web et les pratiques courantes, l'ergonomie du questionnaire en ligne doit être conçue de façon à conserver la motivation du participant tout en s'assurant d'obtenir des

réponses complètes et d'optimiser la fluidité de l'entrevue (Dillman, 2011). Pour capter l'attention du répondant alors qu'il est sollicité de toutes parts lorsqu'il navigue sur le web, il est crucial de porter une attention particulière au design et de simplifier l'interface de manière à rendre l'entrevue plus rapide et plus limpide pour les répondants.

Suivi du comportement des répondants • Un avantage substantiel du mode web est la possibilité d'améliorer le questionnaire en modifiant l'ordre des questions pour optimiser le temps de réponse et favoriser l'ergonomie la plus fluide et la plus claire possible. Pour ce faire, le comportement détaillé des répondants pendant la durée de l'entrevue doit être analysé à l'aide des paradonnées. C'est en instrumentant chacun des boutons, des champs de texte et des sélections à l'aide de minuteries et de détecteurs, et en sauvegardant dans un journal les messages d'erreur et les valeurs entrées avant la confirmation du participant que l'on peut analyser son comportement et effectuer les ajustements nécessaires, au prix d'un développement plus long et plus coûteux à court terme. Une instrumentation équivalente est déjà utilisée au sein de certains systèmes CATI. ("Transports Québec : Le progiciel MADITUC," n.d.)

Accès à Internet, vitesse de connexion et aisance des répondants • Pour l'instant, la couverture incomplète des services Internet à haute vitesse empêche de rejoindre l'ensemble de la population, surtout en dehors des grands centres urbains. Cette limite tend par contre à s'estomper avec le temps, puisque l'accès à Internet sera vraisemblablement disponible à l'ensemble de la population dans un futur proche. En effet, l'accès internet est de plus en plus mentionné comme faisant partie des droits humains universels et de l'accès au savoir (United Nations specialized agency for information and communication technologies, 2003). Toutefois, l'habileté de certains à se servir d'un navigateur web et à remplir des questionnaires complexes en ligne peut être limitée ; c'est le cas par exemple des personnes âgées et des familles à faible revenu qui n'ont pas toujours les moyens de s'équiper en matériel informatique adéquat. Finalement, il est important de noter que les analphabètes fonctionnels et les personnes possédant certaines limitations physiques ou cognitives ne peuvent répondre aux questionnaires web conventionnels, à moins de proposer une entrevue rendue accessible en respectant plusieurs critères (Capra, Ferreira, da Silveira, Ribeiro, & Modesto, 2011).

L'aspect technique • Par rapport à la technologie téléphonique conventionnelle, la technologie web est beaucoup plus complexe et les spécifications techniques des ordinateurs des répondants varient énormément. Il existe plusieurs navigateurs donnant le moyen d'accéder aux questionnaires en ligne et chaque version possède ses points forts et ses faiblesses. Les standards, dans le domaine du web, ont été lents à s'établir et leur respect n'est pas toujours assuré. Le navigateur Internet Explorer, longtemps le plus populaire, a été le dernier à respecter les normes établies par le consortium W3C ("World Wide Web Consortium (W3C)," n.d.). Les anciennes versions sont encore utilisées ("Top 5 Browsers from Nov 2011 to Nov 2012 | StatCounter Global Stats," n.d.), ce qui rend la tâche des développeurs plus difficile puisqu'ils doivent prévoir plusieurs versions de leur questionnaire afin de limiter les problèmes techniques. Cette problématique s'estompera dans un futur relativement proche puisque les différents navigateurs offerts depuis le début des années 2010 respectent en grande partie les standards dictés par le W3C. Par contre, la multiplication des téléphones intelligents, des tablettes et, du même coup, de navigateurs adaptés à la technologie mobile ajoute une nouvelle couche de complexité. En effet, l'interface du questionnaire doit être adaptée aux multiples dimensions des écrans, notamment les petits écrans des téléphones et les écrans de taille moyenne des tablettes. Pour simplifier le processus, il existe maintenant des outils intégrés qui créent des applications en ligne compatibles avec la plupart des téléphones intelligents et des tablettes numériques ("Enyo JavaScript Application Framework," n.d.; "PhoneGap," n.d.).

Par ailleurs, les problèmes techniques que rencontrent possiblement les répondants lors d'une entrevue en ligne peuvent être relativement fréquents si une attention particulière n'est pas portée à la compatibilité de la plateforme d'enquête utilisée avec les systèmes d'exploitation et les navigateurs les plus utilisés.

Concentration des répondants et degré d'achèvement des entrevues • Une des plus grandes nouveautés occasionnées par l'apparition des technologies informatiques et mobiles est l'encouragement à une attention partagée des individus et à une capacité à exécuter plusieurs tâches simultanément. Que ce soit par l'omniprésence de la publicité ou par l'introduction d'applications multitâches dans les appareils informatiques modernes, les utilisateurs sont sans cesse bombardés d'information et d'appels à l'action, ce qui les conduit à effectuer plusieurs tâches en

même temps, mais qui, du même coup limite de manière notable leur performance à compléter chacune des tâches (Adler & Benbunan-Fich, 2012). Dans le cadre d'une entrevue en ligne, une perte de concentration peut influencer grandement la qualité des réponses aux différentes questions relativement complexes que contient un questionnaire d'enquête sur la mobilité. La durée de l'entrevue et l'attention que le répondant y porte jouent également sur la motivation du candidat à terminer le questionnaire.

Malheureusement, les téléphones intelligents et les différentes technologies mobiles accentuent ce phénomène et, de ce fait, pourraient rendre la problématique de la concentration limitée des répondants de plus en plus importante dans le futur (Lynn & Kaminska, 2013).

Objectif ultime : réduire le taux de non-réponse • Dans la grande majorité des cas, l'élaboration d'une version web du questionnaire d'enquête dans le domaine du transport est promue dans le but d'augmenter le taux de réponse et de compléter les échantillons incomplets. Que ce soit en offrant une alternative lors de l'entrevue téléphonique ou de l'envoi postal dans le cadre d'une enquête à modes mixtes ou réalisée exclusivement par le mode web, celui-ci fournit l'opportunité au répondant d'accéder au questionnaire en tout temps, ce qui accroît sa flexibilité, au prix d'une augmentation des entrevues partiellement complétées ou complétées de manière distraite.

2.1.5.2 Les biais du mode web et de la mixité des modes

Les biais du mode web et de la mixité des modes ont été soulevés et résumés par (Couper, 2001). Les sections suivantes représentent une synthèse de ses travaux.

Biais de couverture, d'échantillonnage ou de sélection • Ce biais concerne la représentativité de l'échantillon choisi par rapport à l'univers d'étude. Il est pratiquement impossible à éliminer dans le cadre d'enquêtes régionales du fait des univers comportant un très grand nombre de ménages et d'individus. Lorsque le recrutement est effectué au moyen du mode web, le biais de couverture (à quel point l'échantillon diffère de la population étudiée) est difficile à mesurer puisque les échantillons ne sont habituellement pas contrôlés (les administrateurs de l'enquête ne sélectionnent pas les répondants) à moins de ne permettre l'accès que sur invitation. Le biais

de couverture est influencé également par la sélection de l'échantillon. Est-ce que tous les individus compris dans l'échantillon ont les mêmes chances de participer ? Dans le cas d'une enquête web, un biais de couverture apparaît notamment lorsque l'accès à internet et la vitesse d'accès sont restreints chez certains.

Biais de non-réponse • Il est important de pouvoir connaître, au moins en partie, certaines informations élémentaires sur les non-répondants, dans le but de mesurer la représentativité de ceux-ci au sein de l'univers d'étude ou de l'échantillon. Pour le mode web, certains groupes sociaux et certaines cohortes ont de fortes chances de manifester des taux élevés de non-réponse (personnes âgées, ménages à faible revenu par exemple). Dans ce cas, le mode mixte est de rigueur, bien qu'il faille assurer une compatibilité des données et un suivi des autres biais associés au mode web.

Biais de mesure • Du point de vue du questionnaire lui-même, la motivation du répondant, la qualité graphique de l'interface et le déroulement de l'entrevue peuvent influencer la collecte de données. Par exemple, certains répondants exagèrent ou, au contraire, atténuent certaines de leur caractéristiques pour des raisons sociales, culturelles ou pour éviter d'être jugés (Olson, 2006).

Autres sources d'incertitude • Il existe également plusieurs autres sources d'incertitudes associées à la tenue d'entrevues en ligne. Par exemple, le contrôle du répondant est inexistant, c'est-à-dire qu'il est impossible de connaître la véritable identité du répondant. Un individu peut répondre pour une autre personne ou tenter d'influencer les résultats en modifiant ses réponses. Bien que certaines de ces incertitudes soient présentes également avec d'autres modes d'enquête, il est beaucoup plus laborieux de contrôler le répondant lorsqu'aucun contact par téléphone ou en personne n'est possible.

Effet des incitatifs • On retrouve des enquêtes pour lesquelles des prix (matériels ou monétaires) sont distribués à tous les répondants ou sous forme de tirage au sein de l'échantillon de répondants. Les résultats de la recherche démontrent en majorité que les incitatifs pour les enquêtes web ont moins d'effet sur les taux de réponse que pour les modes conventionnels (en personne,

par la poste, etc.) (Goritz, 2006), surtout si l'incitatif est distribué à la fin de l'enquête plutôt qu'au moment de l'entrevue (Porter & Whitcomb, 2003).

Qualité de conception • La qualité de conception influence les taux de réponse de manière significative (Dillman, 2011). Le degré d'attraction permet en outre de conserver la motivation du répondant en rendant l'entrevue plus agréable.

Adaptation des questionnaires d'un autre mode vers le web • La littérature ne permet pas d'évaluer l'effort fourni pour adapter les questions au mode web. Il est tout de même essentiel de mentionner qu'il faut éviter à tout prix de calquer le questionnaire web sur la version téléphonique ou postale de celui-ci lorsque des modes mixtes sont privilégiés. En effet, le contexte est par exemple différent d'une entrevue téléphonique pendant laquelle le répondant peut obtenir des précisions sur le type de réponses à fournir à l'intervieweur. Chacune des questions doit être formulée de manière courte et précise afin d'éviter toute ambiguïté. Un système de validation en temps réel doit également conduire à éviter les réponses erronées ou invalides.

Utilisation d'outils cartographiques pour la localisation des lieux • Les technologies de cartographie en ligne sont de plus en plus performantes et répandues. Intégrées à un questionnaire à l'intérieur duquel les répondants doivent localiser des lieux, elles simplifient l'interface et peuvent servir à la validation des lieux. Grâce à elles, le géocodage automatique des adresses et des noms de lieux assure des gains de temps, au risque d'obtenir certaines localisations erronées ou imprécises. Une évaluation rigoureuse des forces et faiblesses de chaque fournisseur de cartes en lignes — les plus connus étant Google Maps, Bing Maps, Yahoo Maps et OpenStreetMap — est de mise. La précision du géocodage, l'actualité des données sur les réseaux de transport et les lieux commerciaux et publics varient d'un service à l'autre et d'une région à l'autre.

Mise en commun des efforts de réalisation d'enquêtes de mobilité en ligne • À première vue, outre les outils d'enquêtes en ligne génériques comme SurveyMonkey et SurveyGizmo ("Online Survey Software | SurveyGizmo - Affordable Enterprise Survey Software," n.d.), il n'existe pas d'outil de collecte de données spécialisé dans le domaine du transport et de la mobilité qui soit disponible dans un contexte général. Pourtant, les questionnaires web de ce type sont de plus en plus utilisés à mesure que les populations étendent leur accès rapide et fiable au réseau Internet.

Pour l'instant, aucune initiative ne peut être évaluée de manière indépendante, puisque les questionnaires web sur la mobilité sont habituellement éphémères et aucun démonstrateur n'est actuellement accessible en ligne. Or, la grande majorité des organisations qui mettent de l'avant de tels enquêtes profiteraient d'une plateforme intégrée, flexible et ouverte. Les coûts de développement seraient réduits et la fiabilité, améliorée. C'est entre autres à ce besoin particulier que souhaite répondre la thèse.

2.2 Modèles de données

La qualité et la structure de la base de données est l'élément qui influence probablement le plus l'efficacité et la validation des analyses qui sont effectuées grâce aux données collectées. Pourtant, bien qu'il existe de nombreuses normes pour la création et l'optimisation des bases de données en général, très peu de publications scientifiques explorant la conception des bases de données ont été présentées dans le domaine des enquêtes Origine-Destination en particulier. Pour cette raison, cette section comprendra d'abord un aperçu des modèles de bases de données les plus robustes et les plus populaires. Par la suite, un schéma orienté-objet déjà proposé pour la représentation des données de mobilité ainsi que le schéma de données utilisé au Ministère des transports du Québec (MTQ) pour la gestion des données d'enquêtes Origine-Destination seront présentés.

2.2.1 Modèles de bases de données • Relationnel / SQL, NoSQL et NewSQL⁴

Bien que le modèle relationnel demeure le plus utilisé à ce jour, un nouveau mouvement appelé NoSQL s'est intégré au domaine de la gestion de données avec l'ambition de proposer des alternatives. En effet, le modèle relationnel a montré des signes de faiblesses sur les plans de la flexibilité et de la capacité à s'adapter à un modèle de données évolutif. Les bases de données NoSQL comprennent les modèles orientés clé-valeur, les modèles orientés documents, les modèles orientés colonnes et les modèles orientés graphe. Une courte description du modèle relationnel / SQL et des modèles NoSQL les plus populaires permettra de mettre en contexte le concept de persistance polyglotte, qui consiste à utiliser plusieurs modèles de bases de données à l'intérieur d'un même système, pour finalement introduire les nouveaux modèles NewSQL, qui tentent d'offrir le meilleur compromis entre la robustesse du modèle relationnel et la flexibilité et les performances élevées des modèles NoSQL.

⁴ Deux ouvrages principaux ont permis de dresser le portrait des modèles de bases de données les plus utilisés actuellement dans les systèmes de gestion de bases de données (SGBD) : (Redmond & Wilson, 2012) et (Sadalage & Fowler, 2012).

2.2.1.1 Modèle relationnel • Langage SQL

Le modèle relationnel a été proposé en 1974 par (Codd, 1970). Soutenu par l'algèbre relationnelle, il fait preuve d'une crédibilité qu'aucun autre modèle n'a atteint jusqu'à maintenant. Dans le modèle relationnel, les relations sont représentées habituellement sous forme de tables dans lesquelles chaque attribut fait office de colonne. Les enregistrements, appelés « tuples », constituent les lignes de chacune des tables. Le modèle relationnel encourage le respect des formes normales (normalisation des données). La normalisation des modèles de données permet d'éviter la redondance et assure une meilleure cohérence. Cette normalisation implique la décomposition des relations en tables plus petites comportant des attributs isolés faisant en sorte que la modification d'un attribut particulier ne se fait que dans une seule table. Le langage qui lui est associé, SQL (Structured Query Language) possède un standard établi (ISO, 2011). Il définit les opérations que l'on peut effectuer sur des relations et implémente l'algèbre relationnelle. De plus, il s'articule sous trois facettes : la définition du schéma (création, modification et suppression des tables, colonnes et relations entre les tables), la manipulation des données (insertion, mise à jour et suppression des enregistrements) et le contrôle des données (permissions de lecture et d'écriture définies pour chaque catégorie d'utilisateur).

La grande majorité des bases de données relationnelles respectent également les propriétés ACID (Atomicité Cohérence Isolation Durabilité), qui garantissent la cohérence des données après une transaction. Les propriétés ACID sont historiquement liées au modèle relationnel, même si elles peuvent être, en tout ou en partie, vérifiées par d'autres modèles de données. L'atomicité assure qu'une transaction comprenant une ou plusieurs commandes est entièrement complétée. En cas d'arrêt inattendu du serveur ou d'erreur dans une ou plusieurs des commandes comprises dans la transaction, toutes les commandes sont alors annulées afin de s'assurer que les données reviennent à leur état initial. La cohérence permet de s'assurer que la transaction ne provoquera pas d'invalidité dans les données en vérifiant les contraintes d'intégrité préalablement établies. L'isolation est une mesure de sécurité qui empêche la dépendance entre plusieurs transactions. Le fait que les transactions doivent être totalement indépendantes rend possible les opérations simultanées ou en série avec le même résultat final sur les données (ce qui permet à plusieurs utilisateurs ou applications d'effectuer en même temps des transactions sur la même

base de données). Enfin, la durabilité est la propriété qui s'assure que les données soient enregistrées de manière permanente dans la mémoire physique après confirmation d'une transaction, par opposition à un stockage dans la mémoire volatile (RAM) qui s'efface lors du redémarrage du serveur ou en cas de panne.

Les avantages du modèle relationnel sont :

- Documentation exhaustive, crédibilité.
- Fondements basés sur l'algèbre relationnelle
- Compatibilité élevée avec la plupart des applications de gestion de données
- Structure rigide et robuste
- Normalisation des données pour éviter la redondance
- Propriétés ACID
- Index performants sur différentes colonnes pour faciliter les tris et les filtres sur les données
- Relations simples explicites entre les enregistrements et leurs tables au moyen de références entre les identifiants des enregistrements liés (clés étrangères)
- Requêtes relativement performantes
- Mécanismes de sécurité et de gestion des utilisateurs avec privilèges définis

En contrepartie, les inconvénients sont :

- Rigidité (peu enviable lorsque les attributs et leurs types de données sont variables) ou avec des schémas de données en évolution continue
- Complexe avec des données hiérarchisées (en arbre) ou imbriquées (Silberschatz, Korth, & Sudarshan, 2002). Représenter des hiérarchies avec un modèle relationnel demande de créer une table pour chaque branche et la répétition des attributs communs. Les branches qui comportent plus d'un parent sont également complexes à mettre en relation sans plusieurs tables de jointures.
- L'intégration à un modèle de données orienté objet nécessite une plateforme de conversion, ce qui provoque une perte de performance
- Moins efficace lorsque distribué sur plusieurs serveurs

2.2.1.2 Modèle orienté clé-valeur

Le modèle orienté clé-valeur, pour lequel chaque enregistrement (valeur) est associé à une clé unique, demeure le plus simple et un des modèles NoSQL les plus performants (Cattell, 2010; Tudorica & Bucur, 2011). Les valeurs peuvent être de différents types, par exemple un nombre, un texte, une liste, une image, un vidéo, un objet JSON ou un objet XML. Les attributs ne sont pas définis de manière explicite, ils peuvent être différents d'un enregistrement à l'autre et peuvent changer dynamiquement.

Les avantages du modèle orienté clé-valeur sont :

- Performance élevée qui peut être optimisée pour la mémoire physique de type flash (Hyeontaek Lim, 2011)
- Simplicité
- Flexibilité du type de données (schéma flexible)
- Peut fonctionner facilement sur plusieurs serveurs (clusters)
- Efficace lorsqu'il y a peu de relations entre les enregistrements
- Efficace pour stocker des données de sessions web, des préférences et des profils d'utilisateurs ou des paniers d'achats en ligne

Les inconvénients du modèle orienté clé-valeur sont :

- Manque souvent d'index et de fonctions de tri ou de filtre (seule la clé est indexée)
- Ne permet pas de requêtes complexes en utilisant les attributs ou le contenu des valeurs elles-mêmes (tous les enregistrements doivent être lus à chaque requête par attribut)
- Les opérations simultanées sur plusieurs clés ne sont pas possibles de manière automatique
- Peu ou pas de mécanismes standards prévus pour la gestion de la sécurité et des utilisateurs

2.2.1.3 Modèle orienté documents

Bien qu'il ressemble au modèle orienté clé-valeur, le modèle orienté documents fonctionne généralement avec des types de données divers, mais prédéfinis. La valeur associée à chaque clé est, dans ce cas, appelée « document », mais le principe est le même.

Les avantages du modèle orienté documents sont :

- Peut contenir des structures de données imbriquées comme des arbres
- Schéma flexible, s'adapte à tout type de données, bien que certaines restrictions s'appliquent sur les données entrantes, puisque le type de document accepté est prédéfini.
- Facilité d'ajouter de nouveaux attributs et de les modifier
- S'intègre bien aux langages et applications orientés objet
- Efficace pour les systèmes de gestion de contenu (content management systems), la gestion des produits pour le commerce en ligne, l'analyse de l'audience des sites web (web analytics) et l'analyse en temps-réel dans le domaine de la finance, dans la recherche de fraude et la surveillance, dans l'analyse des comportements des internautes, dans la proposition d'annonces publicitaires ciblées, etc.

Les inconvénients du modèle orienté documents sont :

- Pas de vérification de la cohérence et du schéma
- Les données ne peuvent être normalisées comme dans un modèle relationnel, ce qui provoque de la redondance
- Ne permet pas de relations complexes entre les enregistrements
- Peu ou pas de mécanismes standards prévus pour la gestion de la sécurité et des utilisateurs

2.2.1.4 Modèle orienté colonnes

Créé au départ pour la gestion de grande quantité de données, le modèle orienté colonnes stocke les données sous forme de colonnes plutôt que sous forme de lignes afin d'améliorer, entre autre, les performances globales. En effet, dans un modèle orienté colonnes, les valeurs associées à une même colonne sont sauvegardées les unes à la suite de l'autre, ce qui accélère les comparaisons d'un même attribut enregistré dans une colonne particulière pour les grands ensembles de données. Par exemple, une table comportant trois colonnes (id, nom, age) et trois objets : [1, Paul, 24], [2, Jeanne, 35], [3, Violette, 13] serait stockée physiquement dans une base de données relationnelle conventionnelle suivant cet ordre :

1,Paul,24,2,Jeanne,35,3,Violette,13

alors que dans un modèle orienté colonnes, il serait plutôt stocké par colonne :

1,2,3,Paul,Jeanne,Violette,24,35,13

Les avantages du modèle orienté-colonnes sont :

- La meilleure performance pour les grandes quantités de données (Hadjigeorgiou, 2013; Han, E, Le, & Du, 2011; Tudorica & Bucur, 2011)
- Intègre des fonctions de suivi de versions (versioning), de compression, de récupération de mémoire (garbage collection) et de création de tables en mémoire (in-memory tables⁵)
- Permet la redondance sur plusieurs serveurs appelés « nœuds »
- Très pertinent pour l'indexation de pages web (engins de recherche), la gestion de contenu (content management systems), la gestion des journaux d'événements (event logging)

Les inconvénients du modèle orienté-colonnes sont :

- Administration et configuration complexes, apprentissage difficile
- Peu efficace pour les bases de données relativement petites
- Besoin de connaître les types de requêtes à l'avance pour bien structurer le schéma, difficile de modifier le schéma une fois lancé

⁵ Certaines implémentations du modèle relationnel peuvent également générer des tables en mémoire

- Peu utile pour des applications en développement ou des prototypes
- Peu ou pas de mécanismes standards prévus pour la gestion de la sécurité et des utilisateurs

2.2.1.5 Modèle orienté graphe

Le modèle orienté graphe (Angles & Gutierrez, 2008; Gyssens, Paredaens, Van den Bussche, & Van Gucht, 1994; Vicknair et al., 2010) est le moins commun des modèles NoSQL. Conçu spécialement pour l'analyse des données interconnectées de manière complexe, sa structure est constituée de nœuds et de liens (relations) formant un réseau. Partant d'un schéma flexible, chaque nœud et chaque lien peuvent posséder des attributs différents et constituent, en ce sens, un enregistrement clé-valeur distinct.

Les avantages du modèle orienté graphe sont :

- Idéal pour les données interconnectées, aucune limite sur les relations entre les objets
- Permet de représenter toutes sortes de réseaux de manière précise et efficace
- Très performant pour naviguer dans un réseau et analyser les liens entre les nœuds, notamment les réseaux sociaux, les calculs de chemin, la gestion des parcours ou les services basés sur la localisation spatiale (location-based services)
- S'intègre bien à une conception orientée objet lorsque les attributs sont prédéfinis pour chaque catégorie d'objet (classe)

Les inconvénients du modèle orienté graphe sont :

- Peut difficilement se diviser en sous-graphes
- Difficile à partager sur plusieurs serveurs distincts (clusters) à cause de l'interconnexion des éléments
- Peu efficace lorsque des changements doivent être apportés à un grand nombre de nœuds ou de liens simultanément
- Peu ou pas de mécanismes standards prévus pour la gestion de la sécurité et des utilisateurs

2.2.1.6 Données spatiales

La plupart des systèmes de gestion de bases de données fonctionnant sur les modèles présentés proposent des extensions spécifiques pour la gestion de données géographiques. L'extension la

plus connue et une des plus complètes et des plus documentées est l'extension PostGIS fonctionnant sous le SGBD (Système de gestion de bases de données) relationnel PostgreSQL. Cette extension permet par exemple d'utiliser des attributs géographiques (points, lignes et polygones) et d'effectuer des requêtes spatiales. Le modèle orienté graphe est toutefois considéré comme étant le plus prometteur et le mieux adapté pour représenter des données spatiales ayant de multiples relations entre elles, bien qu'il ne puisse pas encore répondre à tous les besoins d'analyse auxquels répondent les extensions spatiales associées aux SGBD relationnels. En ce sens, il demeure pour l'instant incomplet.

2.2.1.7 Utilisation de multiples modèles dans un même système • Persistance polyglotte

Suivant la nouvelle tendance appelée programmation polyglotte (Ford, 2006; MacDonald, 2016) qui suggère l'utilisation de plusieurs langages de programmation au sein d'un même système pour une meilleure efficacité et une meilleure performance de chaque module, un paradigme semblable a été proposé pour la gestion de données: la persistance polyglotte (Sadalage & Fowler, 2012). En effet, l'utilisation de plusieurs modèles au sein d'une même application ou d'un même projet accroît la flexibilité du système dans son ensemble tout en améliorant les performances, ce qui est rendu possible lorsque le modèle le plus efficace est favorisé pour chaque besoin spécifique de stockage de données. Toutefois, la persistance polyglotte implique généralement une plus grande complexité de mise en œuvre et peut engendrer la redondance de certaines données.

2.2.1.8 Modèles NewSQL

Les modèles de type NewSQL ont été proposés récemment (Aslett, 2011a; 2011b) pour répondre au besoin de collecte de données transactionnelles tout en respectant les propriétés ACID que garantissent les bases de données relationnelles. Les modèles NewSQL supportent le modèle relationnel et le langage SQL, mais facilitent en même temps l'adaptation à l'augmentation de la quantité d'enregistrements et d'attributs appelée extensibilité horizontale à la manière du modèle orienté colonnes. Les modèles de type NewSQL demeurent relativement nouveaux et, en ce sens, leur utilisation est limitée pour l'instant à des prototypes ou à des utilisations spécialisées. Cela dit, le potentiel des bases de données NewSQL est probablement sous-estimé, d'une

part à cause du manque de documentation exhaustive et d'autre part du fait de la complexité de leur configuration et de leur gestion.

2.2.2 Objets d'études associées à la mobilité des personnes

Suivant l'approche totalement désagrégée mise de l'avant pour la validation, l'analyse et la modélisation des données provenant d'enquêtes de mobilité, (Trépanier, 1999) a proposé un diagramme entité-association⁶ représentant les objets associés à la tenue d'une enquête ménage Origine-Destination (Figure 2-1).

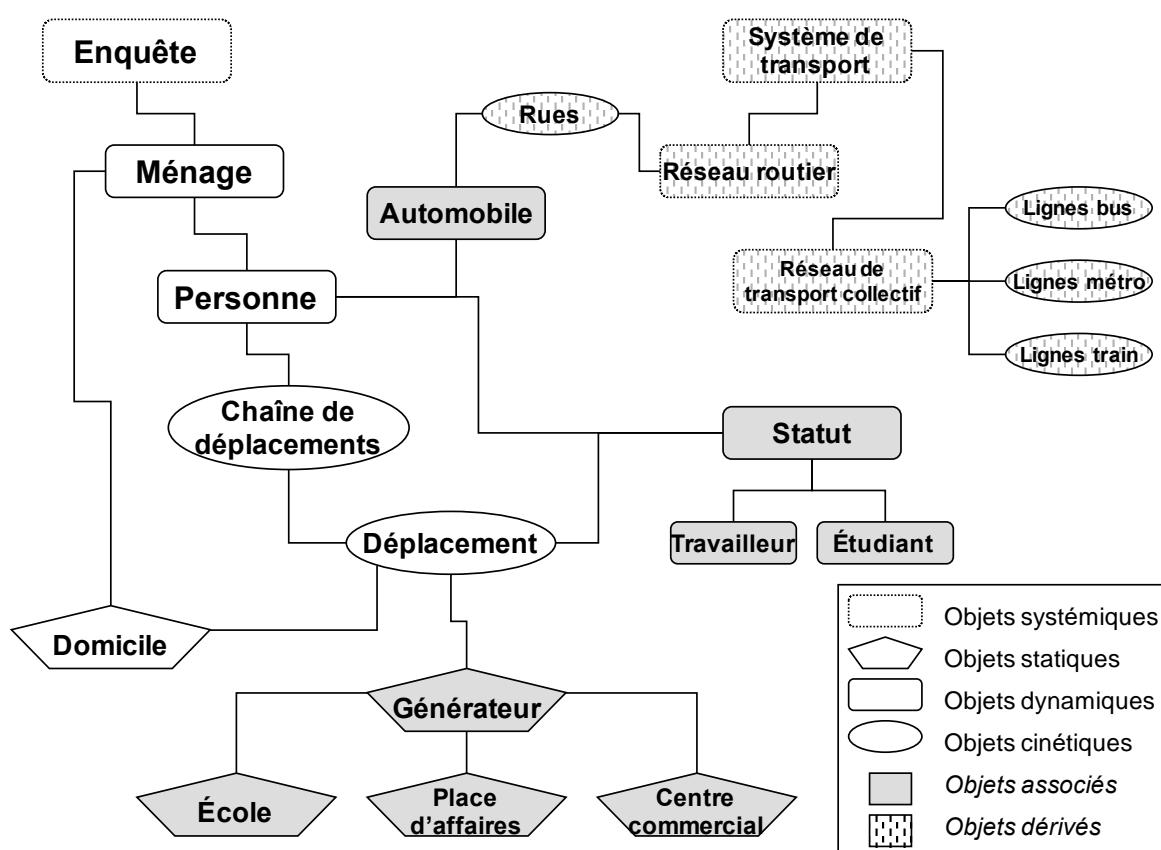


Figure 2-1 : Diagramme entité-association des objets relatifs à l'analyse de la mobilité (Trépanier, 1999) (reproduit avec la permission de l'auteur)

⁶ Un diagramme entité-association permet de décrire le type d'information à être enregistré dans une base de données. Il comprend les objets et les associations qui relient ces objets entre eux.

Saisie Assistée des Questionnaires d'Enquête (SAQE)

Version Québec 2011

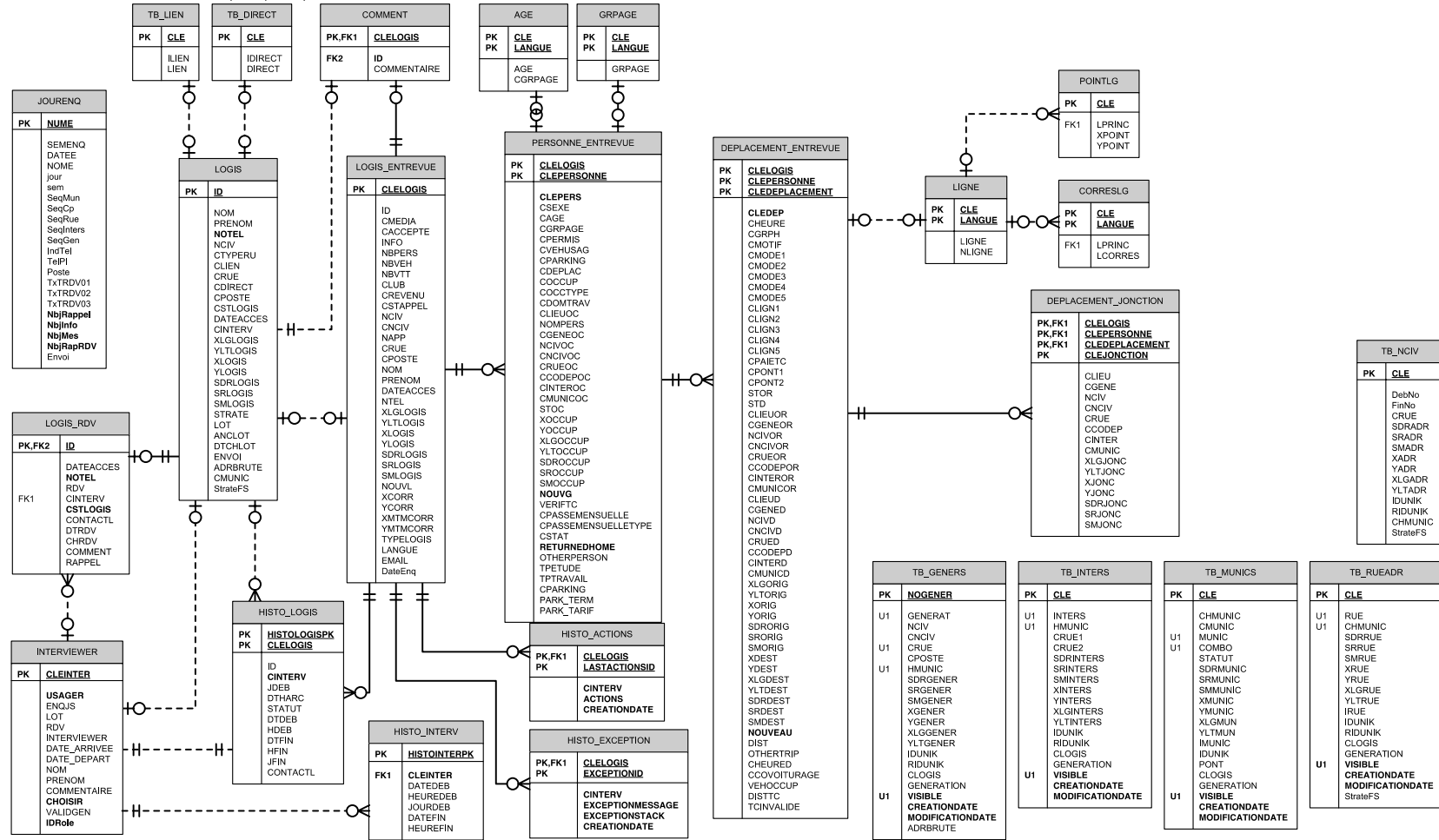


Figure 2-2 : Modèle relationnel SAQE, Ministère des transports du Québec

En parallèle, le schéma relationnel utilisé au MTQ (Ministère des transports du Québec) dans l'application SAQE (Saisie assistée des questionnaires d'enquête) pour la tenue des enquêtes régionales (Figure 2-2) reprend les mêmes éléments en ajoutant certaines tables associées plus spécifiquement au processus d'entrevue. De plus, ce schéma dresse la liste des attributs collectés pour chaque objet (table). Dans les enquêtes régionales Origine-Destination téléphoniques, les attributs des ménages, des personnes et des déplacements, incluant les modes de transport et l'activité à destination, sont demandés au répondant par l'intervieweur. Les chaînes de déplacements⁷ ne font pas partie du modèle SAQE. Habituellement, elles sont générées à partir de l'ordre et des caractéristiques des déplacements déclarés et ajoutées aux fichiers d'analyse après le traitement et la validation des données.

2.3 Entrepôts de données

Un entrepôt de données est un outil d'intelligence d'affaires⁸ de stockage de données conçu pour faciliter et accélérer l'analyse et la mise en valeur des données validées provenant de plusieurs sources. Il existe deux principales approches pour la gestion des données au sein d'un entrepôt de données : le modèle normalisé et le modèle dimensionnel, proposé par (Kimball & Ross, 2013), les deux étant basés sur le modèle de données relationnel. Le modèle normalisé respecte les formes normales suggérées par le modèle relationnel afin d'éviter la redondance de l'information. Toutefois, il augmente le nombre de tables, entraîne un grand nombre de recombinaisons des enregistrements reliés entre eux lors des requêtes, et ne permet pas de gains d'efficacité lors de l'analyse puisque l'interprétation des résultats n'est pas facilitée par le schéma lui-même. De son côté, le modèle dimensionnel a été conçu pour faciliter l'analyse et l'interprétation des données, au prix d'une certaine redondance des données, d'un plus grand espace de stockage et d'un

⁷ Pour tenir compte de l'influence des déplacements entre eux et de leur interdépendance, le concept de chaîne de déplacements a été étudié et détaillé par (Valiquette, 2010) afin de mieux représenter les contraintes sous-jacentes à l'enchaînement des déplacements.

⁸ L'intelligence d'affaires vise à analyser et à clarifier les données brutes de transactions et d'opérations dans le but de mieux comprendre les processus en œuvre dans un domaine en particulier.

temps de traitement plus élevé lors de la conversion à partir des données opérationnelles d'origine. Par contre, les requêtes effectuées auprès d'un entrepôt de données basé sur le modèle dimensionnel sont beaucoup plus rapides parce que les tables ont déjà été dénormalisées en partie et, dans certains cas, agrégées pour certaines analyses. La base de données n'a donc pas à relier toutes les tables mises en relation dans un schéma relationnel normalisé. En fait, le modèle dimensionnel recommande de ne permettre qu'un seul degré de séparation entre les tables. La Figure 2-4 illustre cette différence fondamentale entre les deux modèles. Par exemple, le modèle normalisé, à gauche, possède trois degrés de séparation entre la table de transaction de vente et la table de représentants alors que le modèle dimensionnel ne comporte qu'un seul degré de séparation pour la même relation.

2.3.1 Modélisation dimensionnelle

Selon (Kimball & Ross, 2013), la modélisation dimensionnelle répond à deux besoins principaux : fournir des données qui sont compréhensibles et effectuer rapidement les requêtes demandées. Les modèles dimensionnels sont également appelés schémas en étoile. Au centre de l'étoile, on retrouve une table de faits qui représente les objets élémentaires au centre du domaine d'étude. Par exemple, pour une entreprise de vente au détail, un enregistrement dans la table de faits s'inscrit pour chaque vente d'un produit spécifique à un moment donné (une transaction de vente). Aux pointes de l'étoile, on retrouve les dimensions, c'est-à-dire l'ensemble des descriptions et des éléments de contextes relatifs au fait. Dans le cas d'une vente au détail, on relie une table de dimension pour la date, une seconde pour le produit et sa description, une troisième pour le vendeur, une quatrième pour le client, une autre pour le magasin dans lequel a eu lieu la vente, et ainsi de suite, selon les besoins de l'utilisateur de la base de données. Dans l'étude de la mobilité en transport, le fait peut par exemple être associé à chaque segment de déplacement (un segment étant équivalent à un mode particulier utilisé lors du déplacement), ce qui représente l'élément le plus précis étudié. Comme dimensions, on pourrait retrouver le profil de la personne ayant effectué le déplacement, les informations sur le mode emprunté, la description

du déplacement et de la chaîne de déplacement, le domicile de la personne, son ménage, etc⁹. La Figure 2-3 illustre le schéma en étoile basé sur le modèle dimensionnel pour la vente au détail. La Figure 2-4 montre un extrait du schéma complet selon le modèle normalisé, à gauche, et le modèle dimensionnel, à droite. Pour la dimension Produit, les attributs des dimensions liées sont intégrés dans la table, alors que du côté normalisé, ces attributs sont séparés en plusieurs tables distinctes.

⁹ La modélisation dimensionnelle complète appliquée aux données de mobilité est présentée au Chapitre 5.

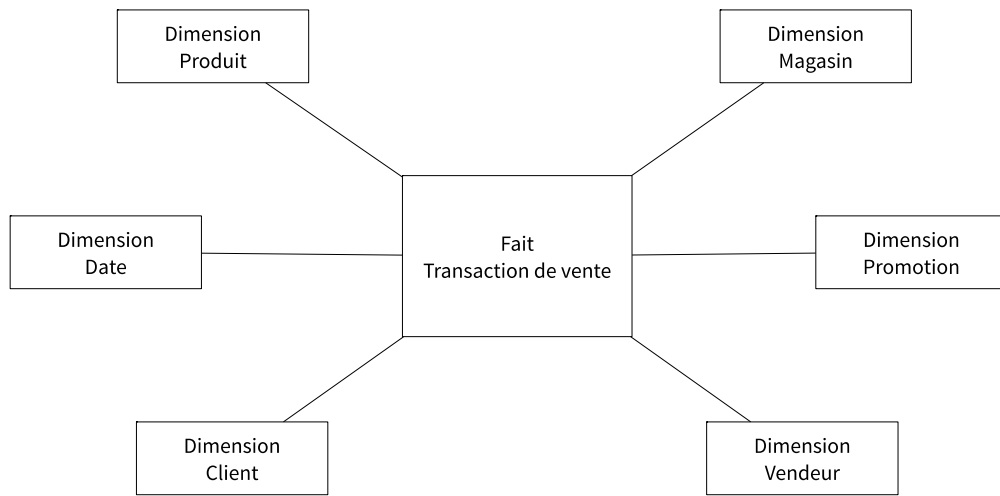
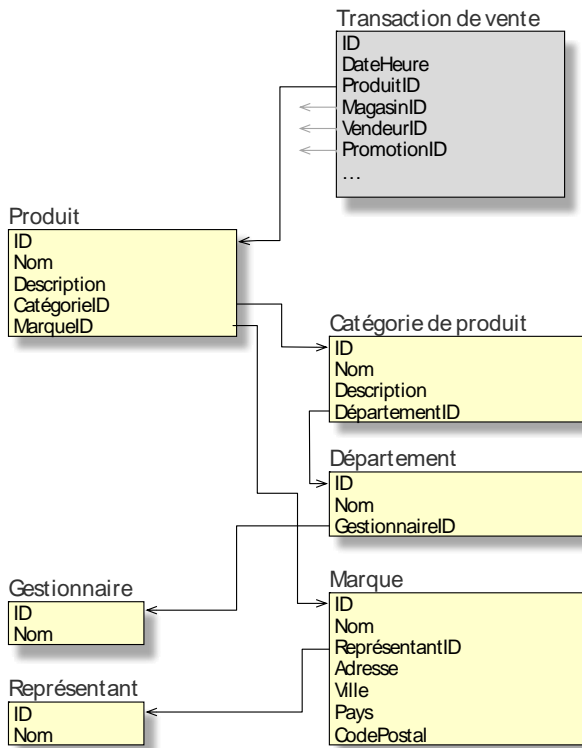


Figure 2-3 : Exemple de modèle dimensionnel en étoile pour la vente au détail

Modèle normalisé



Modèle dimensionnel

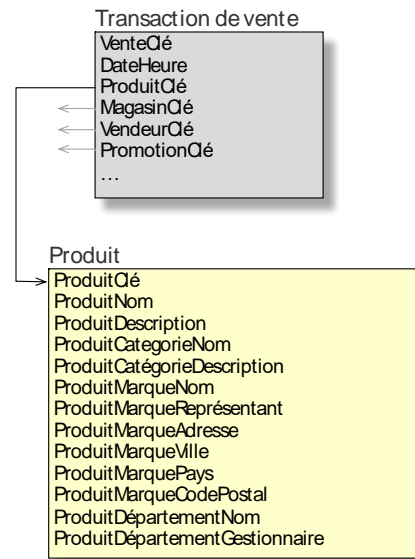


Figure 2-4 : Comparaison entre le modèle normalisé et le modèle dimensionnel

2.3.2 Faits

La table de faits ne contient, la plupart du temps, que des valeurs mesurables (habituellement des valeurs numériques) et des clés étrangères permettant de relier les faits aux dimensions associées. Comme c'est la table qui contiendra le plus grand nombre d'enregistrements, la règle de ne pas inclure d'informations textuelles ou superflues permet de réduire sa taille de façon importante. Seules les tables de faits sont normalisées et peuvent être reliées à d'autres tables (les dimensions).

2.3.3 Dimensions

Les tables de dimensions, quant à elles, contiennent des descriptions textuelles, des attributs numériques et, habituellement, des champs pré-calculés qui enrichissent les données pour en faciliter la compréhension et l'analyse. Comme elles sont liées à l'aide de clés dans la table de faits et qu'une même dimension est normalement associée à plus d'un fait, le nombre d'enregistrements dans la table de dimensions est relativement faible par rapport au nombre d'enregistrements dans la table de faits. Il est important de noter que les tables de dimensions sont dé-normalisées. Par exemple, dans le cas d'une dimension représentant un produit en particulier, les informations sur la catégorie de produit sont intégrées directement dans la dimension (Figure 2-4). De ce fait, les informations sont répétées pour chaque produit faisant partie de la même catégorie. Ce procédé augmente la taille des tables de dimensions, mais permet une meilleure compréhension des attributs dimensionnels tout en favorisant de meilleures performances lors des requêtes.

2.4 Entrepôt de données et visualisation

L'arrivée du phénomène « Big Data¹⁰ » a contribué à exacerber une des problématiques les plus tenaces associées à l'utilisation d'un grand nombre de données pour clarifier des enjeux particuliers : la surcharge informationnelle (information overload) (Wurman, 2000):

People still have anxiety about how to assimilate a body of knowledge that is expanding by the nanosecond. [...] Information anxiety is produced by the ever-widening gap between what we understand and what we think we should understand.¹¹

La Figure 2-5, tirée de (Cairo, 2012), montre la structure du processus d'acquisition de connaissances en relation avec le défi du Big Data et celui de la visualisation.

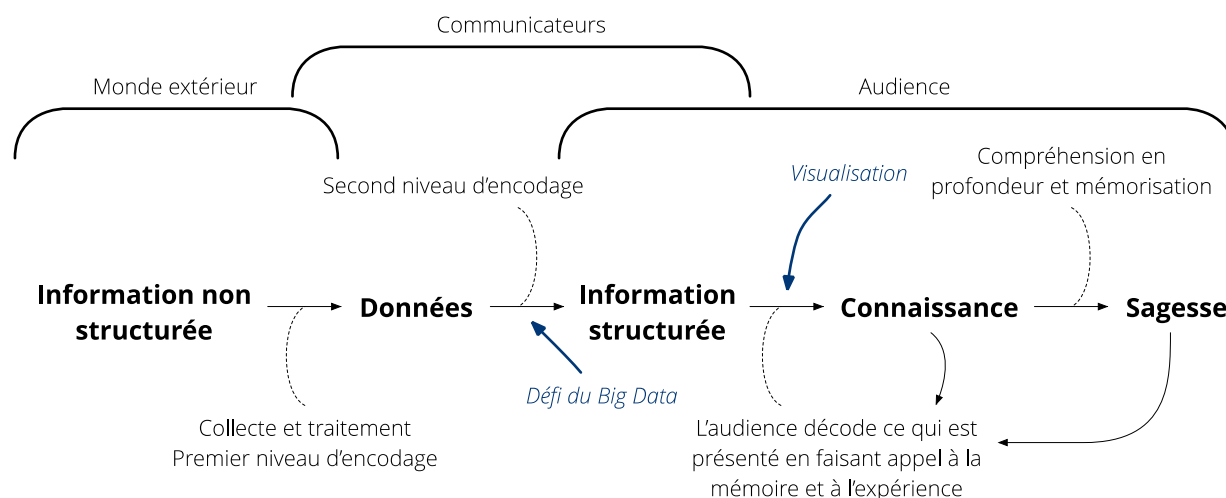


Figure 2-5 : De la réalité au cerveau (traduit de l'anglais depuis From reality to people's brain) (Cairo, 2012), figure 1.8, p. 16.¹²

L'entrepôt de données permet de structurer l'information de manière à faciliter l'acquisition de connaissances et à simplifier le traitement des données d'un domaine à l'autre (Adamson, 2010).

¹⁰ Le phénomène Big Data est un terme utilisé pour représenter des ensembles de données de taille si grande que les outils traditionnels sont inefficaces pour traiter les données.

¹¹ (Wurman, 2000) p. 14.

¹² Figure 1.8 présentée avec la permission de l'éditeur de The Functional Art: An introduction to information graphics and visualization par Alberto Cairo. Copyright © 2013. All rights reserved by Pearson Education, Inc. and New Riders.

Plus la quantité de données augmentera au fur et à mesure que les technologies de collecte le permettront, plus le besoin de structure et de comparabilité entre les différentes sources de données deviendra crucial pour réaliser des analyses rigoureuses sous un échéancier serré. En fait, la structure d'un entrepôt de données basé sur la modélisation dimensionnelle permet aux scientifiques et au grand public de pouvoir sélectionner des ensembles de données personnalisés dans le but de conduire leurs propres explorations (Manoochehri, 2013). La redondance des données qui accompagne le modèle dimensionnel constitue tout-de-même une limite à la quantité de données que l'on peut sauvegarder. Le stockage des dimensions comportant des attributs répétés pourrait être optimisé par l'utilisation de pointeurs¹³ plutôt que des champs textes enregistrés en format binaire, mais copiés plusieurs fois sur le disque.

2.5 La visualisation en transport

(Buliung & Morency, 2010) expliquent que la visualisation est un exercice exploratoire qui peut encourager la proposition de nouvelles hypothèses de recherche et, lorsque bien encadrée, prévenir les mauvaises interprétations par des non spécialistes du domaine étudié. La clarté, la précision et l'efficacité sont les critères les plus importants pour valoriser un objet de visualisation et faciliter la compréhension de phénomènes humains et physiques complexes. Plus particulièrement, la géovisualisation (MacEachren & Kraak, 2001) intègre la dimension spatiale aux visualisations et constitue un outil essentiel pour l'exploration et l'analyse de données spatiales. Des groupes de recherche entiers et des associations de spécialistes de la visualisation (ARUP, n.d.; Citylab, n.d.; van der Wijk, n.d.) proposent un grand nombre d'objets de visualisation pour éclaircir les enjeux liés au transport. Selon (Pack, 2010), le simple fait de rendre les données plus accessibles au moyen d'objets de visualisation efficaces convainc les agences en charge de la gestion du transport de financer des programmes spéciaux de création d'outils de visualisation.

¹³ Le pointeur, en programmation, permet de relier une variable (champ de texte par exemple) à un espace mémoire précis. De cette façon, le même champ de texte répété dans plusieurs tables peut simplement être remplacé par un pointeur unique, ce qui réduit de manière importante la quantité de mémoire nécessaire pour le stockage des données.

Toutefois, la création de tels outils demande beaucoup de temps et un grand nombre de ressources. De plus, les nouvelles fonctionnalités offertes s'accompagnent d'attentes de plus en plus grandes de la part des utilisateurs, ce qui a pour effet d'accroître encore davantage les besoins en ressources et en temps. Or, pour limiter cette escalade, un certain degré d'automatisation dans la création des objets de visualisation est nécessaire.

2.5.1 Automatiser la création d'objets de visualisation cartographiques

Une fois les données enregistrées dans un entrepôt de données, un grand nombre d'objets de visualisation cartographiques peuvent être construits. Toutefois, la création de ces visualisations — des cartes animées par exemple — demande un effort considérable et souvent, un long travail manuel (Harrower, 2004). L'automatisation de la création de cartes animées réduirait ce fardeau et favoriserait leur diffusion. Connaissant l'objectif premier des visualisations qui est de révéler les tendances et les relations qui ne peuvent être déduites sans l'aide de représentations visuelles (Meirelles, 2013), le fait de simplifier le processus de création de ces représentations accélère l'acquisition de connaissances et permet de mieux comprendre les phénomènes que la communication verbale ou écrite et les illustrations conventionnelles statiques ne permettent pas de discerner. Les phénomènes spatio-temporels comme les déplacements des personnes sur un territoire constituent un défi de taille pour les créateurs de visualisations. En effet, les concepteurs doivent représenter des ensembles de données dynamiques en conservant la fluidité du temps (Meirelles, 2013). Grâce aux cartes animées, il est possible d'illustrer des trajectoires et des interactions, le tout au moyen d'un seul et même médium. En parallèle, (G. Andrienko, Andrienko, Bak, Keim, & Wrobel, 2013; N. Andrienko & Andrienko, 2006) confirment que la combinaison de fonctions interactives permet la comparaison et la détection de tendances. En particulier, l'échelle de temps sélectionnée et la vitesse de l'animation choisie favorisent la détection d'événements particuliers ou de phénomènes non détectables autrement. Les cartes animées devraient conséquemment inclure une interface de contrôle des différents paramètres temporels et spatiaux, ce qui est rarement le cas actuellement.

CHAPITRE 3 QUESTIONNAIRE WEB ORIGINE-DESTINATION

3.1 Adaptation du questionnaire téléphonique

Dans la plupart des expériences réalisées dans le cadre de la thèse, le questionnaire d'enquête web Origine-Destination devait être en grande partie compatible avec la version téléphonique afin de permettre une fusion éventuelle des données provenant des deux sources. Toutefois, étant donné les différences importantes entre la tenue d'une entrevue téléphonique faisant appel à un intervieweur et une entrevue web réalisée sans intervention humaine autre que le répondant lui-même, des ajustements ont été apportés dans le but d'optimiser le processus d'entrevue pour le répondant web.

3.1.1 Du déplacement vers le lieu visité

À la suite de la tenue d'une enquête pilote réalisée en 2010 à Polytechnique, il était d'emblée évident que le concept de déplacement était ambigu pour certains répondants. Le concept de lieu visité accompagné d'une activité effectuée à ce lieu a donc été proposé pour remplacer le terme « déplacement » dans le questionnaire web. Après avoir localisé géographiquement les lieux qu'il a visités pendant la journée spécifiée et puisque les déplacements sont alors générés automatiquement par la plateforme, le répondant est en mesure de déclarer les activités effectuées et les heures d'arrivée et de départ des lieux ainsi que les modes empruntés lors de chaque déplacement entre ces lieux. Une étude approfondie de l'influence du choix du concept présenté au répondant (déplacement, lieu visité et/ou activité) sur les taux de mobilité et les activités déclarées devrait permettre de mesurer l'impact réel des changements de formulation dans le questionnaire, mais cette analyse n'a pas encore été réalisée.

3.2 Types de questions et choix de réponses

Le questionnaire généré par la plateforme est générique, en ce sens où les questions ainsi que leur type et l'ordre dans lequel elles sont posées peuvent être entièrement personnalisés. La plateforme actuelle permet l'utilisation des éléments de formulaire décrits au Tableau 3.1.

Tableau 3.1 : Éléments de formulaire, types de données et apparence permis par la plateforme d'enquête

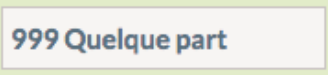
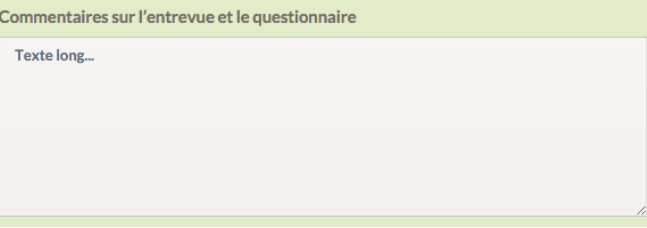


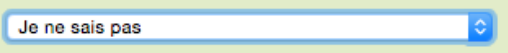
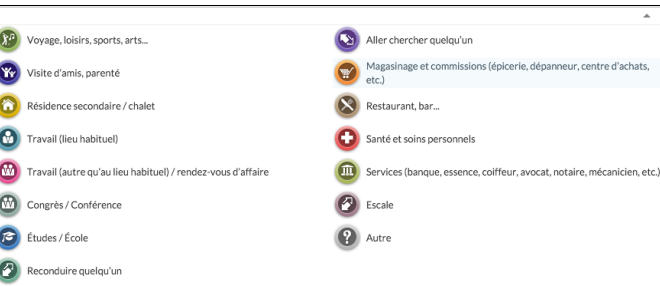
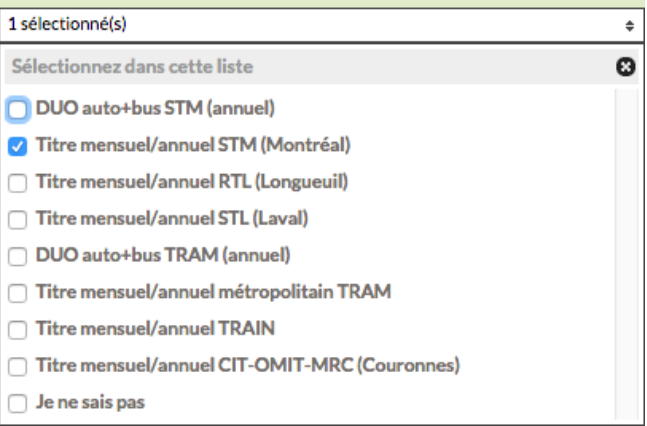

Élément de formulaire	Type de données récupéré	Apparence
Champ de texte sur une ligne	Chaîne de caractères	
Champ de texte long	Texte	
Champ numérique	Numérique	
Boutons radio	Booléen (Vrai/Faux, Oui/Non) ou choix multiples avec réponse unique	
Menu déroulant simple	Choix de réponse unique	
Menu déroulant simple avec icônes	Choix de réponse unique	
Menu déroulant multiple avec cases à cocher	Choix de réponse multiple	
Boutons colorés avec titre	Choix de réponse unique ou multiple	

Tableau 3.1 : Éléments de formulaire, types de données et apparence permis par la plateforme d'enquête (suite)

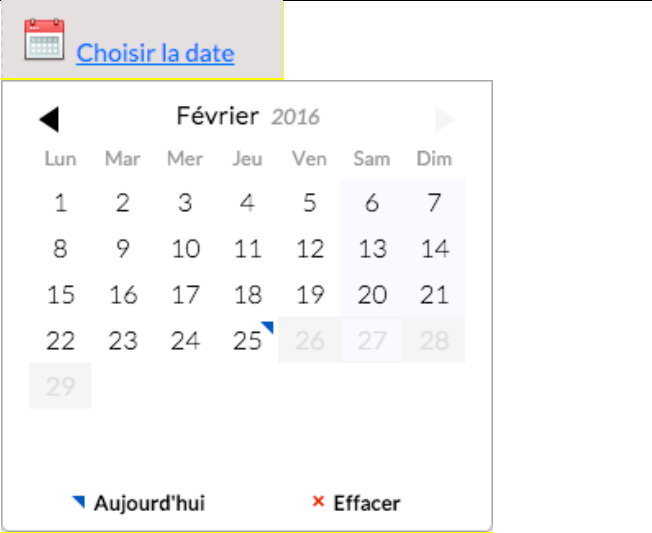

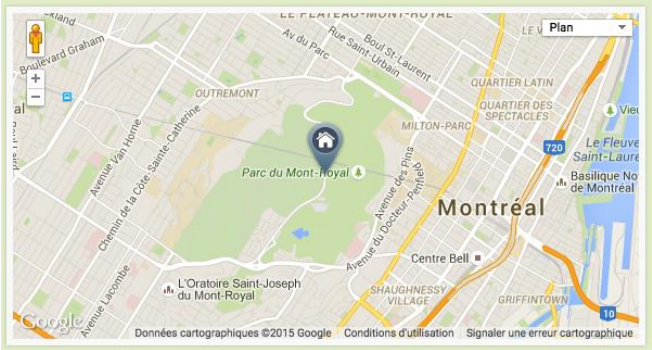
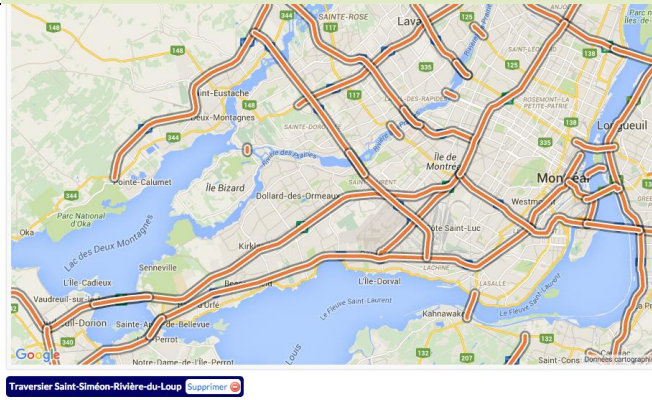




Saisie de date (ca- lendrier)	Date	
Saisie d'heure	Heure (Heure, minutes)	
Carte cliquable	Lieu géographique unique	
Carte avec éléments géographiques cliquables	Lieux, polygones ou polygones géographiques uniques ou multiples	

Tableau 3.1 : Éléments de formulaire, types de données et apparence permis par la plateforme d'enquête (suite)

<p>Carte de réseau ou de bâtiments (carte d'un réseau de métro, de train d'un campus universitaire)</p> <p>Le répondant clique sur la carte pour sélectionner un lieu en particulier</p>	<p>Lieux géographiques (stations de métro, gares ou bâtiments) uniques</p>	
<p>Curseur coulissant</p>	<p>Numérique à partir d'une plage définie</p>	
<p>Boutons ordonnés (ordre peut être modifié par le répondant)</p>	<p>Choix de réponses multiples ordonnées</p>	
<p>Champ de texte autocomplété</p>	<p>Choix de réponses uniques ou multiples</p>	

3.3 Base de données de collecte

La base de données de collecte suit le modèle relationnel et est normalisée. Elle est à distinguer de l'entrepôt de données, qui est construit selon le modèle dimensionnel. Les deux bases de données sont séparées dans le but de les optimiser selon leurs fonctions principales. La base de données de collecte doit éviter les redondances et être performante en écriture des données (sauvegarde des entrevues en temps réel) alors que l'entrepôt de données est conçu pour être performant en lecture des données (requêtes pour analyse et visualisation). L'entrepôt de données sera décrit au Chapitre 5.

3.3.1 Choix du modèle de base de données et de la plateforme de développement web

Connaissant les divers modèles décrits à la section 2.2.1, une étude des avantages et inconvénients reliés à chacun a permis de déterminer que le modèle le plus pertinent à utiliser pour l'instant est le modèle relationnel. Les facteurs prédominants ayant influencé ce choix sont, d'une part, la robustesse des mécanismes de maintien de la cohérence entre les données (propriétés ACID) et d'autre part, la compatibilité des bases de données relationnelles et du langage SQL avec la plupart des applications, ce qui permettrait à un administrateur de se connecter à la base de données de collecte en utilisant l'outil de son choix. De plus, sa documentation complète et l'historique de son utilisation permettent de simplifier le développement et la construction du schéma de données. Toutefois, il existe maintenant des modèles plus performants qui pourront être intégrés lorsqu'ils auront atteint un certain degré de maturité, soit dans des modules connexes, soit en remplacement de la base de données principale destinée à rassembler les données collectées lors des entrevues. En ce sens, une étude plus approfondie pourrait se pencher sur la sélection des tables qui pourraient être séparées de la base de données principale et reconstruites au moyen d'un modèle de données plus optimisé.

La plateforme web utilisée pour le développement de l'outil est Ruby on Rails¹⁴, (Hansson, n.d.), qui est l'une des plus flexibles pour la création d'applications web (Meenakshi, 2015). En particulier, elle permettra un changement de modèle rapide qui ne compromet pas la stabilité de l'application. À ce jour, Ruby on Rails peut utiliser les bases de données relationnelles PostgreSQL¹⁵, MySQL¹⁶, Sqlite¹⁷ ainsi que la base de données orientée-document MongoDB¹⁸, bien que de nombreux modules d'extension permettent d'utiliser d'autres modèles de données.

3.3.1.1 Base de données PostgreSQL

Autant la base de données de collecte que l'entrepôt de données utilisent la base de données relationnelle PostgreSQL ("PostgreSQL: The world's most advanced open source database," n.d.) pour plusieurs raisons. D'abord, elle est disponible gratuitement grâce à une licence libre (open source). Ensuite, elle est considérée comme étant la plus avancée et la plus facile à convertir en cas de migration ultérieure vers d'autres modèles SQL. Enfin, elle comprend une extension pour la gestion de données géographiques (PostGIS) qui est probablement la plus complète (plus de 300 fonctions spatiales) que l'on retrouve dans les systèmes de gestion de données relationnels.

3.3.1.2 Extension spatiale PostGIS

L'extension spatiale PostGIS¹⁹ (PostGIS Development Group, n.d.) intégrée à PostgreSQL a été utilisée pour la gestion de tous les objets à caractère spatial se retrouvant dans le schéma de données de l'outil. Elle respecte plusieurs standards de l'organisme Open Geospatial Consortium²⁰ (OGC) (Geospatial, n.d.) et peut être utilisée pour la gestion ou la consultation de données spatiales au sein d'un grand nombre d'applications SIG.

¹⁴ <http://rubyonrails.org/>

¹⁵ <http://www.postgresql.org/>

¹⁶ <https://www.mysql.com/>

¹⁷ <https://www.sqlite.org/>

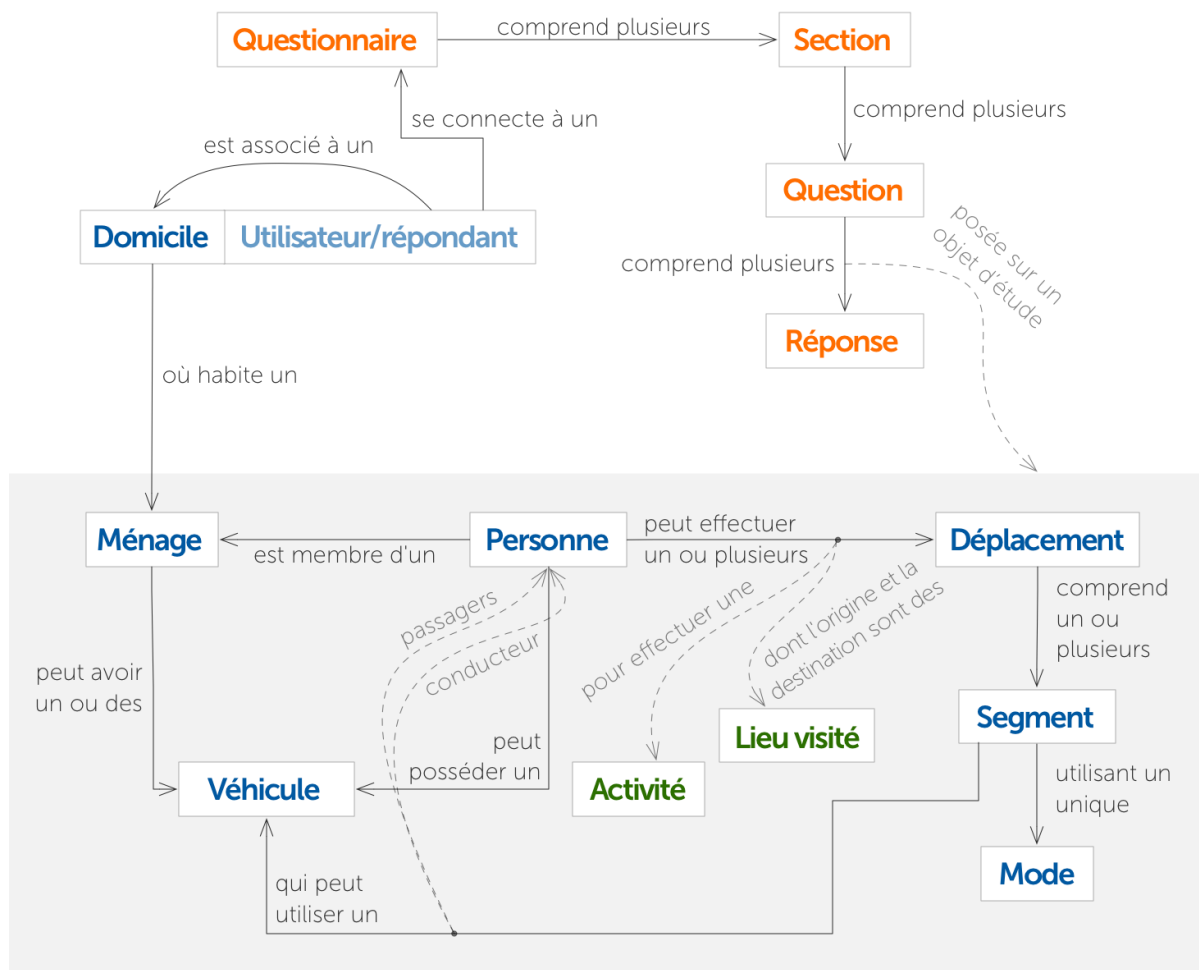
¹⁸ <https://www.mongodb.org/>

¹⁹ <http://postgis.net/>

²⁰ <http://www.opengeospatial.org/>

3.3.2 Schémas

La Figure 3-1 représente le schéma simplifié des objets faisant partie du processus d'enquête. On y retrouve les objets reliés directement à l'entrevue (questionnaire, section, question et réponse), les objets d'étude de la mobilité des personnes (ménage et domicile, personne, déplacement, segment, lieu visité) et les objets descriptifs (l'activité associée à chaque lieu visité et le mode associé à chaque segment). Les déplacements sont divisés en segments, un segment étant associé à l'utilisation d'un mode en particulier. De cette façon, un déplacement multimodal comportera plusieurs segments alors qu'un déplacement à un seul mode ne contiendra qu'un seul segment. La Figure 3-2 contient le modèle de données complet de la base de données de collecte. L'utilisation de l'anglais dans la nomenclature des tables et des attributs a été privilégiée par souci de convention.



Légende des couleurs

Objets relatifs au questionnaire et à l'entrevue

Objets d'étude de la mobilité des personnes

Objets descriptifs

Figure 3-1 : Schéma simplifié des objets et du processus d'entrevue

3.3.3 Description des tables de la base de données de collecte

3.3.3.1 od_survey

Cette table ne contient que les informations permettant de décrire l'enquête. Comme une base de données de collecte distincte est créée pour chaque enquête, la table n'est pas reliée aux autres, et la relation avec les autres tables est implicite.

3.3.3.2 od_samples

Cette table comprend les différents échantillons associés à l'enquête. Son utilisation est facultative. On y retrouve notamment des attributs caractérisant la taille de la population, la taille de l'échantillon, la méthode de recrutement, la méthode de collecte ainsi que des informations sur les sources de données qui ont permis de constituer l'échantillon.

3.3.3.3 od_sample_batches

Comme il est d'usage de diviser les échantillons en lots dans certaines enquêtes (par période d'entrevue, par découpage géographique, par langue d'entrevue, etc.), on retrouve dans cette table les différents lots associés à chaque échantillon. Cette table est également facultative.

3.3.3.4 od_users

La table des utilisateurs comprend toutes les informations relatives au répondant et aux procédures de connexion au questionnaire. Elle permet entre autres de distinguer les utilisateurs valides et non-valides, les répondants effectuant simplement des tests ou les administrateurs de l'enquête, qui ont accès aux modules d'administration et de suivi.

3.3.3.5 od_geographic_features et od_geographic_feature_categories

La table d'objets géographiques est utilisée pour stocker des découpages géographiques, des lieux utilisés comme raccourcis ou toute autre couche géographique de type point, polygone ou polygone. Elle est accompagnée d'une table représentant les catégories d'objets géographiques, qui servent à créer différents regroupements. Il est possible, pour chaque catégorie, de spécifier si celle-ci doit faire l'objet d'une intersection spatiale avec les objets de mobilité géographiques (domiciles, lieux visités, lieux habituels, etc.) lors de l'importation dans l'entrepôt de données de

mobilité au moyen de l'attribut booléen nommé `intersect_in_datawarehouse` (seulement pour les polygones²¹).

3.3.3.6 `od_sections`

La table des sections sert à enregistrer les différentes pages du questionnaire, leur numéro de séquence (ordre d'apparition durant l'entrevue) et les classes d'objets requises (domicile, ménage et/ou personne) pour qu'elles puissent être activées.

3.3.3.7 `od_interview_section_completed`

Cette table enregistre les états d'achèvement des différentes sections, pour chaque répondant. Au démarrage d'une entrevue, un enregistrement est créé pour le répondant et pour chaque section et l'attribut booléen `completed` est initialisé à faux. Lorsque le répondant termine une section, l'attribut booléen `completed` devient vrai. Cette table permet de suivre l'évolution des entrevues pour chaque répondant, de même que le taux d'achèvement des différentes sections du questionnaire.

3.3.3.8 `od_questions`

Les questions et leurs attributs sont enregistrés dans la table `od_questions`. Les questions peuvent comporter des choix de réponse définis (`has_choices` = vrai) et l'attribut `has_multiple_choices` permet de spécifier s'il est possible de choisir plus d'un choix lorsque le répondant répond (réponses à choix multiples). La classe d'objet et l'attribut associés à chaque question sont également définis dans cette table. Par exemple, une question permettant de déclarer la taille du ménage aura comme `model_class` « `od_household_interview` » (ménage) et comme `model_attribute` « `size` » (taille du ménage).

3.3.3.9 `od_question_choices`

Les choix de réponse de chaque question qui en possède sont spécifiés dans cette table. Les choix de réponses d'une même question peuvent être regroupés en utilisant un `group_id` facultatif.

²¹ L'intersection spatiale détermine quel polygone de chaque catégorie contient le lieu représenté par un point (domicile, lieu visité ou lieu habituel de travail ou d'études).

Cela permet aux choix de réponse d'apparaître en groupes dans un menu déroulant par exemple. Lorsque des échelles de valeurs sont utilisées, il est possible de choisir les valeurs de départ et de fin de chaque intervalle de valeur en format nombres entiers, nombres réels, dates ou heures. Il est également possible d'associer un objet géographique à un choix de réponse en particulier au moyen de la clé étrangère `geographic_feature_id`. Finalement, une clé étrangère auto-référencée appelée `durable_comparable_id` peut être utilisée pour spécifier un choix de réponse faisant référence à un attribut utilisé dans l'entrepôt de données pour des fins de comparaisons entre enquêtes.

3.3.3.10 od_responses

Les réponses déclarées pour chaque question sont sauvegardées dans cette table. Chaque réponse est associée à une question et à un objet particulier au moyen des attributs `model_class`, qui représente le type d'objet (ménage, personne, déplacements, etc.) et `object_id`, qui fait référence à son identifiant unique. La valeur associée à la réponse peut être un nombre entier, un nombre réel, une chaîne de caractères, un texte long, un booléen vrai/faux, une heure, une date, une date et heure ou simplement un choix de réponse prédéfini dans la table `od_question_choices`.

3.3.3.11 od_versions

À chaque fois qu'une réponse est ajoutée, supprimée ou modifiée par un répondant, l'historique est conservé dans la table de versions (attribut nommé `value`). Le type d'événement (création, modification, suppression) et l'objet concerné sont alors spécifiés (`model_class` et `object_id`).

3.3.3.12 od_audits

Les audits sont utilisés lors des validations effectuées sur les entrevues. Lorsque la plateforme vérifie les critères de validation configurés par les administrateurs — habituellement après la phase de collecte —, les erreurs et les incohérences sont associées à chaque objet fautif et conservées dans cette table au moyen d'une description textuelle.

3.3.3.13 od_interview_logs

Cette table comprend l'intégralité des événements qui ont été enregistrés pendant les entrevues de tous les répondants (on appelle ces données d'événement les parodonnées). Selon les paramètres de configuration, le comportement des répondants pendant l'entrevue peut être reproduit en tout ou en partie. Par défaut, chaque apparition d'un message d'erreur ou d'une fenêtre explicative, chaque clic sur un bouton, chaque début et fin de section, chaque mouvement d'un lieu géographique sur une carte et chaque sélection dans un menu est enregistré pendant l'entrevue et accompagné d'informations sur le navigateur et l'appareil utilisés par répondant (adresse IP, nom et version du navigateur, type d'appareil, système d'exploitation, etc.). La langue de l'interface d'entrevue sélectionnée au moment de l'événement est également conservée, de même que la section active.

3.3.3.14 od_places

Chaque lieu géographique localisé pendant l'entrevue (lieu de domicile, lieu visité, lieu habituel, etc.) est sauvegardé dans cette table. Lorsqu'une adresse a été déclarée, elle est incluse au moyen des attributs relatifs à l'adresse.

3.3.3.15 od_usual_places

La table des lieux habituels comprend les lieux habituels d'études et/ou de travail déclarés par les répondants pendant l'entrevue. Un attribut permet de conserver le nom du lieu lorsque le répondant le spécifie, de même que la chaîne de caractère utilisée par le répondant pour rechercher le lieu au moyen de l'outil de géocodage (`place_search_name`).

3.3.3.16 od_households

Un nouvel enregistrement est créé dans cette table lors de la création de chaque utilisateur dans la table `od_users`. Elle ne comprend que des informations sur le domicile et son lieu géographique (`home_place_id`).

3.3.3.17 od_household_interviews

La base de données de collecte permet la tenue de multiples entrevues pour un même ménage habitant un domicile spécifique. C'est pour cette raison qu'une table distincte a été ajoutée pour

sauvegarder les entrevues ménage. Toutes les entrevues relatives à un même ménage ont la même valeur de clé étrangère vers la table `od_households`, qui comprend la localisation du domicile. Des attributs de la table `od_household_interviews` permettent de sauvegarder la date et l'heure de début et de fin de l'entrevue, la date de déclaration des déplacements lors de l'utilisation d'un questionnaire de type Origine-Destination, ainsi que plusieurs caractéristiques du ménage (présence et type de connexion internet au domicile, présence et type de ligne téléphonique résidentielle, informations de contact, etc.)

3.3.3.18 od_user_interviews

Cette table permet de sauvegarder les caractéristiques des membres du ménage ou du répondant unique pour une enquête-personne. Elle comprend notamment des attributs socio-démographiques (âge, groupe d'âge, sexe, occupation principale, etc.) et des informations générales sur la mobilité de la personne (possession d'un permis de conduire, d'un abonnement de transport collectif, d'un abonnement d'autopartage ou de vélopartage, etc.). Des attributs permettent de spécifier si les déclarations ont été produites par un répondant proxy (`is_responded_by_proxy`, `responded_by_proxy_id`), dans le cas où une question spécifique est ajoutée à ce sujet à la fin des sections relatives à une personne.

3.3.3.19 od_vehicles

Dans le cas où un questionnaire comprendrait des questions sur les véhicules possédés ou utilisés par les répondants, cette table peut enregistrer les différentes caractéristiques des véhicules déclarés et spécifiés. Les véhicules peuvent être associés soit à une personne, soit à un ménage. Un véhicule peut être une voiture, une camionnette, un véhicule lourd, un vélo ou un deux-roues motorisé.

3.3.3.20 od_visited_places

Les lieux visités déclarés par les répondants sont inclus dans cette table et ordonnés chronologiquement au moyen de l'attribut `sequence`. L'activité effectuée à chaque lieu est conservée au moyen d'une clé externe vers la table d'activités. Lorsque déclarées, la date et l'heure d'arrivée et de départ du lieu sont également sauvegardées.

3.3.3.21 od_activities

Les activités pouvant être sélectionnées et associées à chaque lieu visité déclaré sont configurées à l'avance et conservées dans cette table. Des clés étrangères vers les activités comparables présentes dans la table od_question_choices peuvent être spécifiées pour des besoins de comparabilité avec d'autres enquêtes au moyen de l'entrepôt de données.

3.3.3.22 od_trips et od_trip_chains

Pendant l'entrevue, les déplacements sont automatiquement générés à partir des lieux visités déclarés par les répondants, puis archivés dans la table od_trips. Lorsque tous les déplacements ont été générés, les chaînes de déplacements sont déterminées et sauvegardées dans la table od_trip_chains. Les attributs de ces chaînes (nombre de boucles, contrainte, type de chaîne ouverte, simple ou complexe) sont alors automatiquement définis.

3.3.3.23 od_segments

Pour chaque mode emprunté lors d'un déplacement, le questionnaire peut inclure des questions relatives aux segments de déplacements. Pour les déplacements effectués en voiture, le type de stationnement au début et à la fin du segment peut être spécifié, de même que le nombre d'occupants dans le véhicule. Pour les segments effectués en transport collectif, la ligne utilisée peut également être déclarée par le répondant, de même que les arrêts de départ et d'arrivée. Ces informations sont alors converties en identifiants faisant référence aux fichiers GTFS (General Transit Feed Specification) des agences locales. Les fichiers GTFS sont utilisés par les agences de transport collectif pour diffuser les horaires planifiés des différentes lignes de leurs réseaux. Lorsque demandée au répondant, le lieu d'origine et de destination du segment peut être localisé sur une carte et sauvegardé, ce qui permet de déterminer, par exemple, le lieu de stationnement des véhicules empruntés. Cette table permet également de spécifier l'identifiant du membre du ménage qui conduisait le véhicule lorsque le mode déclaré est auto passager et que le type de conducteur « membre du ménage » est sélectionné.

3.3.3.24 od_modes

De la même manière que les activités, les modes de transport sont définis lors de l'élaboration du questionnaire et sauvegardés dans cette table. Des attributs permettent de sélectionner le type de mode (mode actif, mode de transport collectif, mode associé à la conduite d'un véhicule, etc.), et les distances minimale et maximale spécifiées permettent de filtrer la liste des modes présentés au répondant en fonction des distances de déplacement estimées lors de la génération des déplacements.

3.3.3.25 od_journeys

Cette table est utilisée dans le cadre d'enquêtes pour lesquelles il est demandé de déclarer des informations sur un séjour s'échelonnant sur une longue période ou sur un voyage comportant plusieurs déplacements. Le répondant peut donner un nom à chaque séjour, choisir la date et heure de début et de fin, puis sélectionner les lieux visités associés à ce séjour. Les chaînes de déplacement concernées seront alors générées et associées au séjour dans la table od_journeys.

3.4 Configuration du questionnaire

Le questionnaire est administré et préparé au moyen de fichiers de configuration au format YAML²² (YAML Ain't Markup Language, anciennement Yet Another Markup Language). Une fois ceux-ci complétés, le questionnaire est prêt à être testé, puis envoyé aux répondants potentiels. Il est possible de modifier les questions, les choix de réponse, les validations et conditions d'affichage ainsi que l'ordre d'apparition des questions à tout moment grâce à ces fichiers.

3.4.1 Format YAML

Le format YAML est utilisé couramment pour des fichiers de configuration et remplace ou cohabite habituellement avec le format XML (Extensible Markup Language)²³. Il permet de faciliter la

²² <http://yaml.org/>

²³ <https://www.w3.org/XML/>

déclaration des types de données les plus communs utilisés dans la plupart des langages de programmation. La structure du fichier fonctionne par indentation et est plus épurée que XML, tout en conservant la grande majorité de ses fonctions.

3.4.2 Langues

Les champs de texte inclus dans les fichiers de configuration peuvent être déclarés dans plusieurs langues. Par exemple, pour chaque langue permise dans le questionnaire, une ligne distincte sera incluse, précédée d'un acronyme de deux lettres respectant la norme ISO 639-1²⁴:

```
name:
  fr: "Possédez-vous un permis de conduire?"
  en: "Do you have a driver's license?"
```

3.4.3 Insertion de valeurs déjà déclarées par le répondant

Lorsqu'une partie d'une zone de texte apparaissant dans la configuration doit être remplacée par une valeur précise déclarée par le répondant précédemment, il suffit de mettre le nom de l'attribut à insérer dans le texte accompagné d'un signe de pourcentage et d'accolades ouvrante et fermante (ici, « nickname » représente l'alias donné au début de l'entrevue à chacun des membres du ménage pour les identifier plus tard, cet alias étant enregistré dans la table `od_user_interviews` au moyen de l'attribut `nickname`) :

```
driving_license:
  label:
    fr: "Est-ce que %{nickname} détient un permis de conduire?"
    en: "Does %{nickname} have a driver's license?"
```

3.4.4 Styles

Des styles peuvent être insérés également dans les zones de texte en utilisant le signe de pourcentage suivi de parenthèses à l'intérieur desquelles sont inscrits les noms des styles à utiliser.

²⁴ https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes

Les différents styles sont des classes CSS (Cascading Style Sheets)²⁵, le format standard de présentation de contenu web. Un exemple de texte explicatif apparaissant sous une question est présenté ici (une première partie en rouge gras suivie d'une seconde en vert gras) :

```
transit_pass_owner:
  help:
    fr: "%(red bold)Ne pas tenir compte des billets%
      ↳ %(green bold)Inclure les titres mensuels et annuels%"
```

3.4.5 Fichiers de configuration

Les différents fichiers de configuration seront présentés dans cette section, accompagnés d'exemples. Une fois les différents fichiers complétés ou après que des modifications aient été apportées, une commande permet de mettre à jour la base de données. Pour ce faire, la plateforme lit les différents fichiers et insère les données dans les tables concernées, en prenant soin de supprimer les enregistrements qui ne se retrouvent plus dans le fichier de configuration (éléments supprimés).

3.4.5.1 Configuration générale de l'enquête

La configuration générale de l'enquête se fait au moyen d'un fichier appelé `od_survey.yml`. On y retrouve le nom court (`shortname`) de l'enquête, les dates de début et de fin, le nom, le chemin d'accès web, le courriel de contact, l'organisation en charge ainsi que le fuseau horaire utilisé pour l'affichage et le calcul des heures.

```
survey:
  shortname: od_sherbrooke_2013
  start_on: 2012-11-06
  end_on: 2013-04-22
  name:
    fr: "Enquête Origine-Destination Sherbrooke 2012-2013"
    en: "Origin-Destination Sherbrooke 2012-2013 Survey"
  url:
    fr: "http://www.etude-mobilite.net/sherbrooke/fr"
    en: "http://www.etude-mobilite.net/sherbrooke/en"
  survey_email:
    fr: demo@demo.demo
    en: demo@demo.demo
```

²⁵ <https://www.w3.org/Style/CSS/Overview.en.html>

```

organization:
  fr: "MTQ | Transport Québec"
  en: "MTQ | Transport Québec"
timezone: "Eastern Time (US & Canada)"

```

3.4.5.2 Configuration générale des sections

La configuration générale des sections du questionnaire (les pages) se fait au moyen d'un fichier appelé `od_sections.yml`. On y retrouve le nom court (shortname) de chacune des sections, accompagné de ses différents attributs (séquence dans le questionnaire, nom apparaissant dans le menu, objets requis pour l'affichage de cette section, etc.). Les objets requis (`required_classes`) peuvent être, dans l'ordre : `od_user` (présence de l'objet utilisateur dans la base de données, c'est-à-dire que le répondant est bel et bien enregistré), `od_household` (présence de l'objet ménage dans la base de données), `od_household_interview` (présente de l'objet entrevue du ménage, disponible dès que le répondant a débuté la section ménage), `od_user_interview` (présence de l'objet personne) :

```

start:
  sequence: 100
  menu_name:
    fr: "Bienvenue!"
    en: "Welcome!"
  name:
    fr: "Introduction"
    en: "Introduction"
  required_classes: "od_user"

home:
  sequence: 200
  menu_name:
    fr: "Votre domicile"
    en: "Your home"
  name:
    fr: "Votre domicile"
    en: "Your home"
  required_classes: "od_user od_household od_household_interview"

household_members:
  sequence: 300
  menu_name:
    fr: "Votre ménage"
    en: "Your household"
  name:
    fr: "Les membres de votre ménage"
    en: "Your household members"
  required_classes: "od_user od_household od_household_interview"

```



```

profile:
  sequence: 400
  menu_name:
    fr: "Profil"
    en: "Profile"
  name:
    fr: "Profil"
    en: "Profile"
  required_classes:
    "od_user od_household od_household_interview od_user_interview"
...

```

3.4.5.3 Configuration des questions

La configuration des questions se fait au moyen d'un fichier appelé `od_questions.yml`. On y retrouve le nom court (shortname) de chacune des questions, accompagné de ses différents attributs :

```

gender:
  model_class: od_user_interview
  model_attribute: gender
  field_type: radio
  datatype: string
  label:
    fr: "Sexe"
    en: "Gender"

age:
  model_class: od_user_interview
  model_attribute: age
  field_type: string
  datatype: integer
  label:
    fr: "Âge"
    en: "Age"

age_group:
  model_class: od_user_interview
  model_attribute: age_group
  field_type: select
  hidden: true
  datatype: string
  label:
    fr: "Groupe d'âge"
    en: "Age group"

household_size:
  model_class: od_household_interview
  model_attribute: size
  field_type: string
  datatype: integer
  label:

```

```

fr: "Combien de personnes habitent dans votre domicile de façon
  ↳ permanente, y compris vous-même, pendant la semaine?"
en: "Including yourself, how many people live permanently in
  ↳ your household during the week?"
help_popup:
  title:
    fr: "Qui inclure dans ce nombre?"
    en: "Who to include?"
  content:
    fr: "%(green bold)Vous devez vous inclure dans ce nombre%
      ↳ %(green bold)Inclure les chambreurs et colocataires%
      ↳ %(red bold)Ne pas inclure quelqu'un qui habite dans un
      ↳ appartement séparé à l'intérieur de votre demeure%"
    en: "%(green bold)You need to include yourself%
      ↳ %(green bold)Include roomers and roommates%
      ↳ %(red bold)Do not include those who live in a separate
      ↳ suite within your home%"

home_geography:
  model_class: od_household.od_place
  model_attribute: geography
  field_type: map
  marker_icon: home
  datatype: point
  default_lat: 45.54
  default_lon: -73.68
  default_zoom: 9
  geocoding: true
  geocoding_fields: [home_street_address, home_postal_code,
    ↳ home_city, home_region, home_country]
  minimum_geocoding_precision: postal_code
  minimum_manual_placement_zoom: 12
  label:
    fr: "Localisation de votre domicile"
    en: "Home location"

...

```

Plusieurs attributs peuvent être inclus dans la configuration des questions, par exemple :

- **Fenêtre surgissante (help_popup)** • La question s'accompagne d'un bouton d'aide. Lorsque le répondant clique sur ce bouton, une fenêtre comportant le texte explicatif (content) apparaît.
- **Aide (help)** • Un texte explicatif est inséré sous la question et demeure visible.
- **Question cachée (hidden: true)** • La question est cachée par défaut à l'ouverture de la section. Cela permet, par exemple, de montrer cette question seulement si certaines conditions sont remplies pendant que le répondant est à l'intérieur de la section. Les conditions sont définies dans les configurations spécifiques des sections (3.4.5.5).

- **Attributs spécifiques aux cartes** • Les questions comportant une carte à cliquer possèdent plusieurs attributs permettant de configurer la carte (centre de la carte par défaut, type d'icône pour le marqueur à cliquer/déplacer, activation du géocodage automatique, champs utilisés pour le géocodage, précision minimale et zoom minimal de saisie)

3.4.5.4 Configuration des choix de réponse

Pour chaque question possédant des choix de réponse, ceux-ci doivent être définis dans le fichier de configuration `od_question_choices.yml`, précédés du nom court de la question associée :

```
age_group:
  "0-4":
    name:
      fr: "0 à 4 ans"
      en: "0 - 4 years old"
    range_integer_start_value: 0
    range_integer_end_value: 4
    internal_id: 1
    sequence: 100
    durable_comparable_sn: "0-4"
  "5-9":
    name:
      fr: "5 à 9 ans"
      en: "5 - 9 years old"
    range_integer_start_value: 5
    range_integer_end_value: 9
    internal_id: 2
    sequence: 200
    durable_comparable_sn: "5-9"
  "10-14":
    name:
      fr: "10 à 14 ans"
      en: "10 - 14 years old"
    range_integer_start_value: 10
    range_integer_end_value: 14
    internal_id: 3
    sequence: 300
    durable_comparable_sn: "10-14"
  ...

driving_license:
  driving_license_yes:
    internal_id: 1
    sequence: 100
    name:
      fr: "Oui"
      en: "Yes"
  driving_license_no:
    internal_id: 2
    sequence: 200
```

```

name:
  fr: "Non"
  en: "No"
i_dont_know:
  internal_id: 3
  sequence: 300
name:
  fr: "Je ne sais pas"
  en: "I don't know"
...

```

Chaque choix de réponse peut inclure un identifiant interne (`internal_id` : en format numérique ou chaîne de caractères) utilisé par l'organisation en charge de l'enquête, ce qui permet de sauvegarder ce code dans la base de données et de l'inclure, au besoin, dans les fichiers d'exportation pour analyse. De plus, les choix de réponse peuvent inclure un lien vers un attribut comparable (`durable_comparable_sn`) qui est inclus dans toutes les enquêtes pour lesquelles cet attribut est compatible. Le numéro de séquence permet de déterminer l'ordre d'apparition des choix de réponse dans les menus déroulants par exemple. Pour modifier cet ordre, il suffit de remplacer cette valeur sans déplacer le contenu dans le fichier de configuration.

3.4.5.5 Configuration spécifique des sections

À chaque section est associé un fichier de configuration distinct permettant de déterminer les conditions d'affichage des questions et les validations à effectuer en temps réel pendant l'entrevue. Le fichier porte le nom court de la section (exemple : `home.yml` pour la section « Domicile »). Lorsque aucune condition ou validation n'est associée à une question, le caractère `~` (équivalent à une valeur nulle ou une négation en format YAML) est inséré. Pour chaque validation, une condition doit être incluse, de même qu'un message d'erreur qui apparaîtra au-dessus de la question si la validation échoue pendant l'entrevue. En utilisant les accolades, les conditions peuvent faire intervenir des variables représentant les valeurs associées à une question en particulier. Les con-

ditions sont traitées par le langage javascript pendant l'entrevue. Toute condition javascript valide peut être utilisée. Il est également possible d'utiliser des expressions régulières²⁶ valides en javascript en utilisant le format /EXPRESSION_REGULIERE/.test(VARIABLE) :

```

home_apartment_number: ~

home_street_address:
  validators:
    -
      condition: "!{home_street_address}"
      message:
        fr: "L'adresse est requise."
        en: "Address is required."

home_city:
  validators:
    -
      condition: "!{home_city}"
      message:
        fr: "La ville est requise."
        en: "City is required."

home_region: ~

home_country: ~

home_postal_code:
  conditionals_popups:
    -
      shortname: pere_noel
      condition: "{home_postal_code} == 'H0H 0H0'"
      button_name:
        fr: "Oui!"
        en: "Yes!"
      message:
        fr: "Vous êtes vraiment le père Noël???"
        en: "Are you really Santa Claus???"
  validators:
    -
      condition: "{home_postal_code} && !(/^[A-Za-z][0-9][A-Za-z]
        ↳ *[0-9][A-Za-z][0-9]$/.test({home_postal_code}))"
      message:
        fr: "Le code postal est invalide."
        en: "Postal code is invalid."

home_geography:
  validators:
    -
      condition: "!{home_geography}"

```

²⁶ https://fr.wikipedia.org/wiki/Expression_rationnelle

```

message:
  fr: "La localisation du domicile a échoué ou n'est pas assez
      ↳ précise. Assurez-vous d'être à un zoom suffisant."
  en: "Home localization failed or is not precise enough.
      ↳ Please make sure you are at the required zoom level."

household_size:
  validators:
  -
    condition: "{household_size}"
    message:
      fr: "La taille du ménage est requise."
      en: "Household size is required."
  -
    condition: "{household_size} &&
                ↳ !(/^\\d+$/).test({household_size})"
    message:
      fr: "La taille du ménage est invalide."
      en: "Household size is not valid."
  -
    condition: "{household_size} && ({household_size} > 20 ||
                ↳ {household_size} == 0)"
    message:
      fr: "La taille du ménage doit être au moins de 1 personne
          ↳ (vous devez vous inclure dans ce nombre) et au maximum
          ↳ de 20 personnes."
      en: "Household size must be greater than 1 (you must include
          ↳ yourself) and less or equal to 20."

household_car_number:
  validators:
  -
    condition: "{household_car_number}"
    message:
      fr: "Le nombre de véhicules est requis."
      en: "Number of vehicles is required."
  -
    condition: "{household_car_number} &&
                ↳ !(/^\\d+$/).test({household_car_number})"
    message:
      fr: "Le nombre de véhicules est invalide."
      en: "Number of vehicles is not valid."
  -
    condition: "{household_car_number} && ({household_car_number}
                ↳ > 20 || {household_car_number} < 0)"
    message:
      fr: "Le nombre de véhicules doit être plus grand ou égal à 0
          ↳ et plus petit ou égal à 20."
      en: "Number of vehicles must be greater or equal to 0 and
          ↳ less than or equal to 20."

```

3.4.5.6 Configuration du questionnaire et de l'ordre des questions

Enfin, un dernier fichier de configuration (od_questionnaire.yml) permet de déterminer le format d'affichage des différentes sections, d'insérer des textes d'introduction, des boutons et de déterminer l'ordre d'affichage des questions :

```
home:
  divs:
    intro:
      -
        width: 100%
        align: center
        content:
          fr: "Prière de répondre aux questions suivantes"
          en: "Please respond to the following questions"
    questions:
      -
        width: 70%
        content:
          question.household_size: ~
          question.household_car_number: ~
          question.home_postal_code: ~
          question.home_apartment_number: ~
          question.home_street_address: ~
          question.home_city: ~
          question.home_region: ~
          question.home_country: ~
          question.home_geography: ~
    buttons:
      -
        width: 100%
        align: center
        content:
          button.button_save_and_continue: ~
  ...
```

3.5 Sections du questionnaire Origine-Destination conventionnel

Même si les différentes sections et questions présentes dans le questionnaire peuvent être modifiées et déplacées, un canevas de questionnaire Origine-Destination a été proposé et utilisé pour la plupart des enquêtes web réalisées dans le cadre du projet de recherche soutenant la thèse. Afin de mieux comprendre le déroulement d'une entrevue Origine-Destination en ligne, le contenu de ces sections sera décrit et illustré.

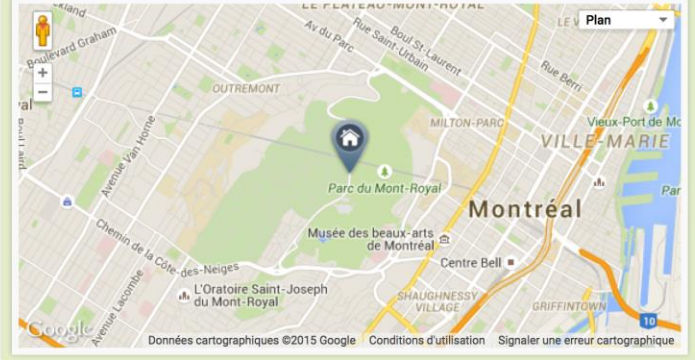
3.5.1 Domicile

La première section concerne le domicile du répondant et de son ménage, dans le cas d'une entrevue ménage (Figure 3-3). On demande d'abord le nombre de personnes habitant le domicile de même que le nombre de véhicules accessibles aux personnes y habitant (motorisation du ménage). Ensuite, le répondant indique le type de bâtiment et entre son adresse qu'il peut visualiser au moyen d'une carte grâce à la géolocalisation de son domicile en fonction de l'adresse déclarée. Il peut également déplacer l'icône de son domicile sur la carte ou cliquer à l'endroit précis où se situe son domicile.

[🏠 Votre domicile](#)
[👤 Votre profil](#)
[👨‍👩‍👧‍👦 Votre ménage](#)
[🚗 Vos déplacements](#)
[Fin](#)

Prière d'entrer les informations suivantes:

Combien de personnes habitent dans votre domicile de façon permanente, y compris vous-même, pendant la semaine? Qui inclure dans ce nombre?	<input type="text" value="3"/>	✓
Combien de véhicules sont à la disposition des personnes qui habitent chez vous (ne pas inclure les véhicules Communauto)? Quel(s) véhicule(s) inclure?	<input type="text" value="0"/>	✓
Dans quel type d'habitation ou de propriété vivez-vous?		✓
	Propriétaire Maison jumelée (semi-détaché)	
Code postal	<input type="text" value="H2C2C4"/>	✓
Numéro d'appartement Laisser vide si aucun	<input type="text"/>	
Adresse	<input type="text" value="999 Quelque part"/>	✓
Ville	<input type="text" value="Montréal"/>	✓
Province	<input type="text" value="Québec"/>	✓
Pays	<input type="text" value="Canada"/>	✓
Localisation de votre domicile Vous pouvez cliquer sur la carte ou déplacer l'icône pour préciser l'emplacement de votre domicile. Utilisez le zoom pour plus de précision (+/-). <i>La localisation de votre domicile doit être la plus précise possible.</i>		✓



[Sauvegarder et continuer](#)

Figure 3-3 : Section du questionnaire associée au domicile du répondant

3.5.2 Profil • *Entrevue personne*

Lors d'une entrevue personne, la seconde section comprend une liste de questions relatives au profil du répondant (Figure 3-4). Elle contient des questions d'ordre démographique (âge ou

groupe d'âge, sexe, occupation, etc.) de même que des questions relatives à la mobilité quotidienne (possession d'un permis de conduire, possession d'un abonnement de transport collectif, etc.) et, dans certains cas, des questions sur l'accès à la téléphonie cellulaire (possession d'un téléphone cellulaire ou d'un téléphone intelligent). Pour les travailleurs et les étudiants, le questionnaire demande également de géolocaliser le lieu habituel de travail ou d'études et de spécifier si le travail du répondant implique des déplacements fréquents ou si le répondant travaille au domicile sur une base régulière.

Votre profil

Sexe	<input checked="" type="radio"/> Homme <input type="radio"/> Femme	✓
Âge	33	✓
Occupation principale Travailleur à temps plein: 30 heures et plus par semaine Travailleur à temps partiel: moins de 30 heures par semaine	Travailleur à temps complet	✓
Possédez-vous un permis de conduire?	<input checked="" type="radio"/> Oui <input type="radio"/> Non	✓
Possédez-vous un titre mensuel ou annuel de transport en commun pour ce mois-ci?	<input checked="" type="radio"/> Oui <input type="radio"/> Non	✓

Sauvegarder

Figure 3-4 : Section du questionnaire associée au profil du répondant • *Entrevue personne*

3.5.3 Membres du ménage

Les entrevues ménages possèdent une section concernant le profil de tous les membres du ménage (Figure 3-5). Le répondant ou les répondants, lorsque plusieurs membres du ménage participent à l'entrevue, doivent au minimum y déclarer l'âge, le sexe, l'occupation ainsi que la possession d'un permis de conduire et d'un abonnement de transport collectif pour chacune des personnes habitant au domicile. Pour une entrevue personne, cette section est habituellement présente dans le but de mieux connaître les membres du ménage du répondant pour des besoins de

comparaison avec les enquêtes-ménage. Lorsque cette section a été complétée pour une entrevue ménage, le répondant choisit le membre du ménage pour lequel il désire poursuivre l'entrevue. À tout moment pendant la suite de l'entrevue, il peut sélectionner un autre membre de son ménage pour modifier ou compléter le questionnaire pour cette personne. La Figure 3-6 illustre l'interface de sélection des membres du ménage.

[🏠 Votre domicile](#)
[👤 Votre ménage](#)
[👤 Profil](#)
[👤 Déplacements](#)
[Fin](#)

Prière d'entrer les informations suivantes sur tous les membres de votre ménage:

Personne 1 – Simon

Identifiant Sert d'identifiant pour les autres membres du ménage lors de leur entrevue. Exemple: prénom, surnom, etc. <i>Cette information confidentielle ne sera pas conservée pour l'analyse.</i>	<input type="text" value="Simon"/>	✓
Sexe	<input type="radio"/> Femme <input checked="" type="radio"/> Homme	✓
Âge <i>Je préfère donner seulement le groupe d'âge</i>	<input type="text" value="31"/>	✓
Occupation principale Travailleur à temps plein: 21 heures et plus par semaine Travailleur à temps partiel: moins de 30 heures par semaine	<input type="text" value="Travailleur à temps complet"/>	✓
Est-ce que cette personne possède un permis de conduire?	<input checked="" type="radio"/> Oui <input type="radio"/> Non <input type="radio"/> Je ne sais pas	✓
Est-ce que cette personne possède une passe/abonnement de transport en commun valide? Ne pas tenir compte des billets ou des passes de semaine Inclure les titres mensuels et/ou annuels	<input checked="" type="radio"/> Oui <input type="radio"/> Non <input type="radio"/> Je ne sais pas	✓
Quelle(s) passe(s) ou abonnement(s) de transport en commun cette personne possède-t-elle?	<input type="text" value="1 sélectionné(s)"/>	✓

Personne 2 – Violette

Identifiant Sert d'identifiant pour les autres membres du ménage lors de leur entrevue. Exemple: prénom, surnom, etc. <i>Cette information confidentielle ne sera pas conservée pour l'analyse.</i>	<input type="text" value="Violette"/>	✓
Sexe	<input checked="" type="radio"/> Femme <input type="radio"/> Homme	✓
Groupe d'âge <i>Je voudrais plutôt donner l'âge exact</i>	<input type="text" value="30 à 34 ans"/>	✓
Occupation principale Travailleur à temps plein: 21 heures et plus par semaine Travailleur à temps partiel: moins de 30 heures par semaine	<input type="text" value="Travailleur à temps complet"/>	✓
Est-ce que cette personne possède un permis de conduire?	<input checked="" type="radio"/> Oui <input type="radio"/> Non <input type="radio"/> Je ne sais pas	✓
Est-ce que cette personne possède une passe/abonnement de transport en commun valide? Ne pas tenir compte des billets ou des passes de semaine Inclure les titres mensuels et/ou annuels	<input type="radio"/> Oui <input checked="" type="radio"/> Non <input type="radio"/> Je ne sais pas	✓

Figure 3-5 : Section du questionnaire associée au profil de tous les membres du ménage

Personne 3 – **Bob**

Identifiant Sert d'identifiant pour les autres membres du ménage lors de leur entrevue. Exemple: prénom, surnom, etc. <i>Cette information confidentielle ne sera pas conservée pour l'analyse.</i>	<input type="text" value="Bob"/>	✓
Sexe	<input type="radio"/> Femme <input checked="" type="radio"/> Homme	✓
Âge <i>Je préfère donner seulement le groupe d'âge</i>	<input type="text" value="2"/>	✓

Sauvegarder et continuer

Figure 3-5 : Section du questionnaire associée au profil de tous les membres du ménage (suite)

Pour quelle personne désirez-vous continuer l'entrevue?




	Simon	31 ans
	Violette	30 à 34 ans
	Bob <i>Vous n'avez pas à répondre à l'entrevue pour cette personne</i>	2 ans

Figure 3-6 : Section du questionnaire pour la sélection d'un membre du ménage

3.5.4 Déplacements

De manière usuelle, dans le cadre d'une entrevue ménage de type Origine-Destination, on demande au ménage en entier de déclarer les déplacements qu'il a effectués lors du dernier jour ouvrable. Cette section comprend d'abord une page d'introduction présentée dans le but de clarifier la méthode et les caractéristiques de saisie des lieux visités et des déplacements associés.

Elle est suivie de la page de sélection du lieu de départ de la journée, puis de l'interface de déclaration et de géolocalisation des lieux visités. Lorsque tous les lieux visités ont été ajoutés, on demande au répondant de sélectionner les heures de départ et d'arrivée à ces différents lieux. Finalement, la page relative aux modes de transport permet de déclarer les moyens de transport utilisés lors de chacun des déplacements effectués pendant la journée.

3.5.4.1 Introduction

L'introduction demande d'abord au répondant de déclarer s'il a effectué au moins un déplacement lors du dernier jour ouvrable (Figure 3-7). Si tel est le cas, elle explique ensuite en quelques phrases au répondant le déroulement prévu pour la section relative aux déplacements et aux lieux visités (Figure 3-8). C'est à cet endroit que sont précisés les types de lieux à déclarer et ceux à ignorer afin de simplifier la saisie. La date du dernier jour ouvrable, ou toute autre date choisie par le système en fonction des paramètres fournis par l'administrateur de l'enquête, y est également indiquée. Dans le cas où le répondant a déclaré travailler régulièrement sur la route, des précisions concernant les déplacements sont ajoutées afin de simplifier et de rendre plus rapide la déclaration des lieux visités pour ces personnes au sujet desquelles on demande seulement les lieux de départ et d'arrivée de chaque boucle de type « sur la route » effectuée pendant la journée. Le texte d'introduction présenté ici constitue un exemple pour une enquête de type Origine-Destination. Il peut être modifié ou remplacé au besoin.



🏠 Votre domicile → 👤 Votre profil → 👨‍👩‍👧‍👦 Votre ménage → 🚗 Vos déplacements → Fin

Avez-vous effectué au moins un déplacement **lundi le 28 octobre 2013**?

- **tous moyens de transport confondus** (y compris la marche et le vélo)
- **en incluant même les petits arrêts** (exemple: station-service, garderie/école, dépanneur ou tout lieu où vous êtes allés chercher ou reconduire quelqu'un)

Oui Non

Figure 3-7 : Section d'introduction demandant au répondant s'il a effectué au moins un déplacement lors du dernier jour ouvrable

Votre domicile

Votre profil

Votre ménage

Vos déplacements

Fin

Vos déplacements lors du dernier jour ouvrable

Lundi le 28 octobre 2013

Nous allons vous demander de localiser les lieux où vous êtes allés, **lundi le 28 octobre 2013, entre 0:00 (minuit) et 4:00 du matin le lendemain** (28h plus tard):

- **inclure les petits arrêts** (Exemple: station-service, garderie/école, dépanneur ou tout lieu où vous êtes allés chercher ou reconduire quelqu'un)
- **ne pas inclure les lieux de transfert** (arrêts d'autobus, stations de métro, gares, terminus, stationnements incitatifs, etc.) **sauf si vous êtes allés reconduire ou chercher quelqu'un à cet endroit**

Ne pas utiliser une autre journée, même si le dernier jour ouvrable n'était pas représentatif de vos déplacements quotidiens.

[Pourquoi cette journée, même si elle n'est pas représentative?](#)

Continuer

Figure 3-8 : Section d'introduction à la saisie des lieux visités

3.5.4.2 Lieu de départ de la journée

Le lieu de départ de la journée, c'est-à-dire à quel endroit était le répondant avant le premier déplacement, permet de démarrer la déclaration des lieux visités et de situer le point de départ de la journée sur la carte des déplacements. Un exemple de section permettant au répondant de sélectionner le lieu de départ est présenté à la Figure 3-9. Lorsque le répondant ne choisit pas le domicile comme point de départ, il doit situer le lieu sur la carte avant de débiter la saisie des autres lieux qu'il a visités pendant la journée.






Figure 3-9 : Section de sélection du lieu de départ de la journée

3.5.4.3 Lieux visités et activités

Cette section sert à localiser tous les lieux que le répondant a visités pendant la journée sélectionnée par le système, dans l'ordre chronologique (Figure 3-10). Pour ce faire, une carte lui permet de situer le lieu et des boutons raccourcis lui permettent de sélectionner soit son lieu habituel de travail ou d'étude, soit tout autre lieu déjà localisé par lui ou un autre membre de son ménage (Figure 3-11). Il est également possible de saisir une intersection précise ou de rechercher un endroit particulier au moyen d'un service de géolocalisation. Une fois le lieu sélectionné, le répondant peut, de manière facultative, entrer un nom pour ce lieu afin de simplifier la visualisation subséquente des déplacements. Il doit également choisir l'activité effectuée à ce lieu à l'aide d'une liste munie d'icônes. À chaque lieu localisé, le système demande au répondant d'indiquer s'il est retourné au domicile après avoir visité ce lieu. Cette procédure a été ajoutée pour éviter que le répondant oublie de déclarer explicitement à chaque fois qu'il est retourné à son domicile. La liste des lieux visités se met automatiquement à jour à chaque nouvelle localisation, et, comme le montre la Figure 3-12, le répondant peut, au besoin, y insérer, y modifier ou y retirer des lieux.

[Votre domicile](#)
[Votre ménage](#)
[Votre profil](#)
[Déplacements](#)
[Opinions](#)

Lieux où vous êtes allés :
mardi le 20 octobre 2015
L'ordre chronologique doit être respecté
Ne pas oublier le point d'arrivée de la journée (domicile ou autre)

-  Point de départ: Domicile
-  Aller chercher quelqu'un
-  Domicile

[Continuer \(ajouter le prochain lieu\)](#)

[Modifier, insérer ou supprimer un ou des lieux](#)

Lorsque tous les lieux visités dans la journée ont été ajoutés (incluant la destination finale de la journée):

[J'ai terminé d'ajouter les lieux](#)

[Je me suis trompé, je n'ai effectué aucun déplacement ce jour-là](#)

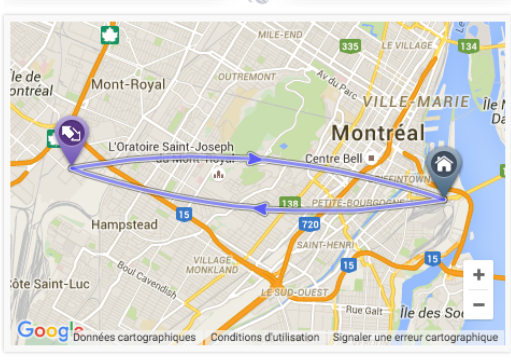


Figure 3-10 : Section indiquant la liste des lieux visités et des activités

[Votre domicile](#)
[Votre ménage](#)
[Votre profil](#)
[Déplacements](#)
[Opinions](#)

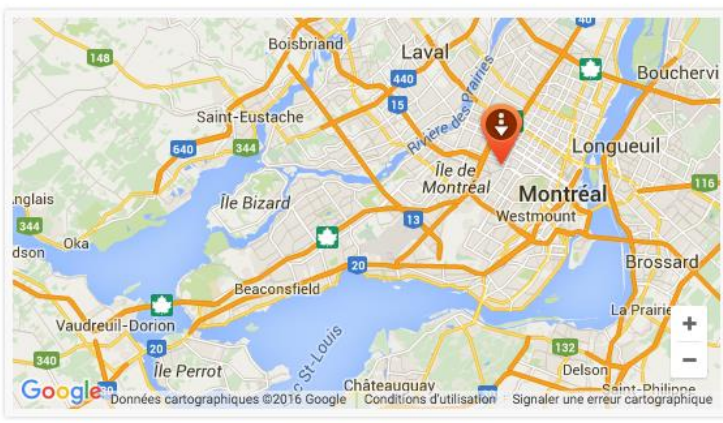
Naviguez et localisez le lieu en cliquant à l'endroit approprié sur la carte ou utilisez les raccourcis de droite. Vous pouvez cliquer plusieurs fois sur la carte, seul le dernier clic sera utilisé pour confirmer le lieu.

Utilisez ces boutons pour chercher un lieu ou accélérer la sélection des lieux déjà visités

[Confirmer le lieu](#)

- [Trouver une intersection](#)
- [Trouver une adresse](#)
- [Naviguer vers une autre région/ville](#)
- [Domicile](#)

Lieux déjà visités:
[Aller chercher quelqu'un](#)



[Confirmer le lieu](#)

[Annuler](#)

Figure 3-11 : Section permettant de géolocaliser un lieu visité



Figure 3-12 : Section montrant la liste des lieux visités accompagnée de boutons permettant d’insérer, de modifier ou de supprimer des lieux

3.5.4.4 Horaire de la journée

Après avoir indiqué les lieux visités, le répondant peut sélectionner les heures de départ et d’arrivée à chacun de ces lieux au moyen d’un tableau (Figure 3-13). Le système valide la chronologie et vérifie, en temps réel, qu’aucune période d’activité ou de déplacement ne se superpose à une autre. Une ligne du temps située au-dessus du tableau illustre l’horaire de la journée afin de simplifier et de préciser la sélection des heures de départ et d’arrivée.

Votre horaire de la journée:
mardi le 20 octobre 2015

Si vous ne vous souvenez pas de l'heure exacte, choisissez celle qui vous semble la plus logique selon votre mémoire

		À quelle heure êtes-vous arrivé(e) à ce lieu?		À quelle heure avez-vous quitté ce lieu?	
	Point de départ: Domicile			quitté à	7 h : 10
	Aller chercher quelqu'un	arrivé à	8 h : 05	quitté à	9 h : 05
	Domicile	arrivé à	10 h : 15	quitté à	11 h : 05
	Restaurant	arrivé à	11 h : 25	quitté à	13 h : 00
	Point d'arrivée: Domicile	arrivé à	13 h : 25		

[Modifier les lieux](#)

[Sauvegarder et continuer](#)

Figure 3-13 : Section permettant la saisie des heures de départ et d'arrivée des lieux visités

3.5.4.5 Modes de transport

Grâce à la liste complète des lieux visités et à l'horaire associé, le système génère automatiquement les déplacements et en présente la liste au répondant. Ce dernier doit alors déclarer tous les modes de transport utilisés lors de ces déplacements, dans l'ordre (Figure 3-14 et Figure 3-15). Des questions spécifiques apparaissent en fonction des modes choisis. Par exemple, lorsqu'un segment de déplacement utilise le transport collectif, les stations de métro empruntées ou les lignes de bus utilisées peuvent être précisées, comme illustré à la Figure 3-16. Lors d'un déplacement de type « auto conducteur », le questionnaire demande entre autre les types de stationnement à l'origine et à destination, de même que le nombre d'occupants à l'intérieur du véhicule (Figure 3-17). Pour les déplacements effectués en tant que passager d'un véhicule, on demande au répondant de sélectionner le type de conducteur (membre du ménage, voisin, collègue, etc.). Dans certaines entrevues, les autoroutes et les ponts ou tunnels empruntés étaient demandés aux conducteurs, comme illustré à la Figure 3-18.

The screenshot displays a travel planning application interface. On the left, a vertical sidebar contains icons for home, a speech bubble, and a question mark. The main content area is divided into sections for different trips:

- Domicile**: DÉPLACEMENT 1 (12:05 →) from Domicile to Sur la route (15:20). Moyens de transport: 2: Train de banlieue, 3: Autobus.
- Sur la route**: DÉPLACEMENT 2 (20:15 →) from Sur la route to Domicile (21:15).
- Domicile**: DÉPLACEMENT 3 (21:30 →) from Domicile to Autre (22:05).
- Autre**: DÉPLACEMENT 4 (23:35 →) from Autre to Domicile (02:10).
- Domicile**: (Empty section)

A dropdown menu is open over the 'Moyen de transport' field of the first trip, listing the following options:

- Auto - Conducteur
- Auto - Passager
- Méto
- Autobus
- Train de banlieue
- Marche (déplacement entier)
- Vélo
- Moto ou scooter
- Taxi
- Autobus scolaire (autobus jaune)
- Transport adapté
- Autocar interurbain ou nolisé
- Train interurbain (de type Via Rail)
- Avion
- Autre
- Je ne sais pas

On the right, a map titled 'Carte des déplacements' shows the Montreal area with a route highlighted in green and blue. The map includes labels for 'Montreal', 'Longueuil', 'Brossard', and 'La Prairie'. At the bottom, there are two links: [Modifier les lieux](#) and [Modifier les heures de déplacements](#).

Figure 3-14 : Section montrant la saisie des modes de transport

The screenshot displays a user interface for planning a multi-stop journey. On the left, a vertical list of trips is shown, each with a location icon, a title, a 'Modify transport' link, and departure/arrival times. The trips are: 1. Domicile to Sur la route (12:05 to 15:20), 2. Sur la route to Domicile (20:15 to 21:15), 3. Domicile to Autre (21:30 to 22:05), 4. Autre to Domicile (23:35 to 02:10). Each trip is accompanied by icons representing the transport modes used. Below the list are links to 'Modify locations' and 'Modify departure times', and a large green button to 'Confirm trips and continue'. On the right, a Google Map of the Montreal area shows a blue route connecting the locations: Domicile (home icon), Sur la route (bus icon), Domicile (home icon), Autre (question mark icon), and Domicile (home icon). The map includes labels for various areas like Longueuil, Brossard, and La Prairie.

Domicile

DÉPLACEMENT 1 [Modifier les moyens de transport](#)
De: Domicile (12:05 →)
vers: Sur la route (→ 15:20)

Sur la route

DÉPLACEMENT 2 [Modifier les moyens de transport](#)
De: Sur la route (20:15 →)
vers: Domicile (→ 21:15)

Domicile

DÉPLACEMENT 3 [Modifier les moyens de transport](#)
De: Domicile (21:30 →)
vers: Autre (→ 22:05)

Autre

DÉPLACEMENT 4 [Modifier les moyens de transport](#)
De: Autre (23:35 →)
vers: Domicile (→ 02:10)

Domicile

[Modifier les lieux](#) [Modifier les heures de déplacements](#)

Confirmer les déplacements et continuer

Carte des déplacements

Figure 3-15 : Section de liste des modes de transport empruntés

Domicile

DÉPLACEMENT 1
De: **Domicile** (12:05 -->)
vers: **Sur la route** (--> 15:20)
Moyen(s) de transport:

2: **Train de banlieue**
3: **Autobus**

Moyen de transport:

Ligne(s) de bus: [La ligne n'apparaît pas dans la liste](#)
[Je ne connais pas le numéro ou le nom](#)

- STM 131 Assomption
- STM 31 Saint-Denis
- STL 31 Direction Terminus Henri-Bourassa
- STL 31 Direction Auteuil
- CITSO CIT Sud-ouest 31 Châteauguay - Montréal
- CITROUS CIT Roussillon 31 Sainte-Catherine est
- RTL 31 Secteurs R-S-T / boul. St-Laurent
- CRTL CRT Lanaudière 31 Saint-Gabriel - Joliette
- CITPI CIT La Presqu'île 31 Pincourt - Régent
- CITSV CIT Sorel-Varenes 731 Saint-Amable - Longue...
- CITLR CIT Le Richelain 31 Candiac - La Prairie - M...

Sur la route

DÉPLACEMENT 2
De: **Sur la route** (20:00 -->)
vers: **Domicile** (--> 20:00)

Domicile

Carte des déplacements

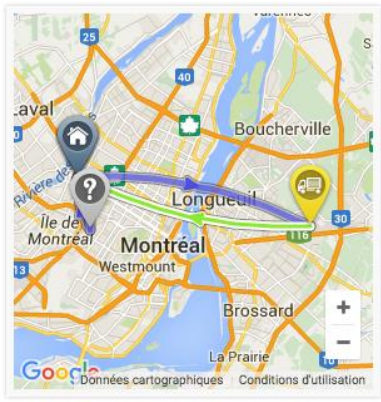
Figure 3-16 : Section montrant les questions spécifiques aux déplacements utilisant le transport collectif

Sur la route

DÉPLACEMENT 2
 De: **Sur la route** (20:15 →)
 vers: **Domicile** (→ 21:15)
 Moyen(s) de transport:

1: **Train de banlieue**
 2: **Auto - Conducteur**

Moyen de transport	Auto - Conducteur	✓
Type de stationnement avant le départ	Sur rue	✓
Est-ce que Simon avait payé pour garer le véhicule avant le départ?	Vignette de résident	✓
Type de stationnement à destination	Stationnement intérieur	✓
Est-ce que Simon avait payé pour garer le véhicule à destination?	Fourni ou payé par l'employeur	✓
Nombre total de personnes dans le véhicule	2	✓
Autoroute(s), pont(s) et tunnel(s) empruntés(s) Ne rien sélectionner si aucun n'a été emprunté ou si vous ne le savez pas	Sélectionner sur la carte A-35 Aut. de la Vallée des Forts Supprimer A-10 Autoroute des Cantons de l'Est Supprimer	✓



Carte des déplacements

Figure 3-17 : Section montrant les questions spécifiques aux déplacements de type « auto-conducteur »

de vérifier si le ménage peut être rejoint par les méthodes conventionnelles de recrutement utilisées lors des enquêtes téléphoniques régionales. Lorsque l'enquête s'adresse à une population spécifique, par exemple la communauté d'un établissement d'enseignement, les instigateurs y insèrent habituellement des questions spécifiques (précisions sur les services offerts par l'établissement ou commentaires au sujet des déplacements effectués vers ou à partir de l'établissement, etc.). Finalement, on demande au répondant d'apprécier le déroulement de l'entrevue à l'aide d'une boîte de texte lui permettant d'entrer des commentaires et d'inscrire son courriel, dans le cas où il aimerait participer à d'autres études de mobilité. Lorsque l'entrevue est complétée, le répondant peut, à tout moment, modifier les réponses déclarées à n'importe quelle section du questionnaire. Il peut en outre quitter l'entrevue et s'y reconnecter plus tard pour effectuer des ajouts ou des précisions.

Dernières questions relatives à votre ménage

Possédez-vous un accès internet à la maison?	<input checked="" type="radio"/> Oui <input type="radio"/> Non	✓
Est-ce que votre accès internet résidentiel est à haute vitesse?	<input checked="" type="radio"/> Oui <input type="radio"/> Non	✓
Revenu global du ménage avant impôts en 2012 (facultatif)	60 000 à 89 999 \$	✓
Seriez-vous prêt à participer à d'autres études universitaires sur la mobilité? <small>Si vous répondez oui, votre courriel sera ajouté à la liste de la Chaire Mobilité de l'École Polytechnique de Montréal et ne sera jamais partagé avec d'autres instances.</small>	<input checked="" type="radio"/> Oui <input type="radio"/> Non	✓
Votre courriel <small>Ne sera jamais partagé avec d'autres instances.</small>	test@test.com	✓
Commentaires généraux sur l'entrevue et le questionnaire	Mes commentaires sur l'entrevue...	✓

Terminer l'entrevue

Figure 3-19 : Section finale et ses questions habituelles

3.6 Validations effectuées pendant l'entrevue

Pendant toute l'entrevue, les validations spécifiées dans les fichiers de configuration sont effectuées en temps réel afin d'améliorer la qualité des données collectées. L'affichage ou non de certaines questions respecte les conditions définies dans les fichiers de configuration. De cette façon, un répondant non-travailleur ne verra pas les questions associées au type de travail effectué ou au lieu habituel de travail. De la même façon, les questions relatives au permis de conduire ou aux abonnements de transport collectifs seront posées seulement aux membres du ménage assez âgés pour être concernés. Comme expliqué à la section 3.4, l'administrateur peut également

créer ses propres validations et ses propres conditions, et ce, pour n'importe quelle question posée à toute section du questionnaire. Tout au long de l'entrevue, les réponses sélectionnées ou indiquées sont sauvegardées en temps réel (à chaque sélection ou à chaque seconde lorsqu'un champ de texte est activé), ce qui réduit le nombre de réponses perdues en cas de problème technique et qui permet un suivi des entrevues pour faciliter et préciser l'analyse subséquente des comportements de réponse.

3.7 Module d'administration

Le module d'administration est conçu pour les organisateurs de l'enquête et les analystes. Il comprend un ensemble de graphiques et de tableaux permettant d'effectuer le suivi des entrevues. On peut également y trouver la liste des commentaires des répondants ainsi qu'une interface destinée à l'importation de fichiers géographiques de type shapefile et à l'exportation des fichiers de résultats.

3.7.1 Suivi des entrevues

La section de suivi des entrevues est illustrée à la Figure 3-20 et comprend des graphiques montrant le nombre d'entrevues débutées et complétées à chaque jour (Figure 3-21) et en fonction de l'heure de la journée (Figure 3-22). Les statistiques sur les durées moyennes d'entrevue (Figure 3-23) et les durées en fonction du nombre de déplacements déclarés sont révélées à cet endroit (Figure 3-24). Le taux d'achèvement des différentes sections (Figure 3-25) permet d'estimer de manière générale les sections provoquant le plus grand nombre d'abandons. Les données enregistrées dans la table `od_interview_section_completed` sont utilisées pour générer ce graphique.

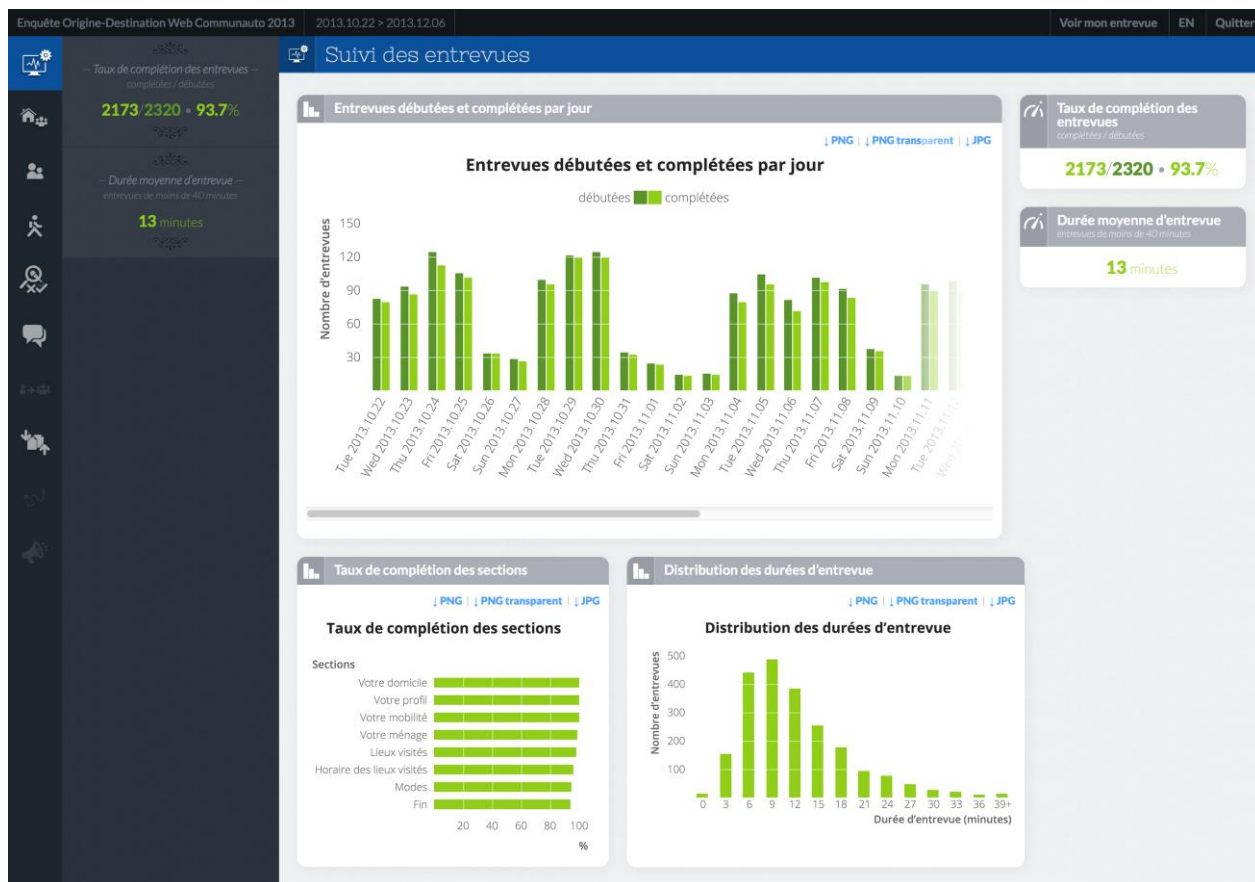


Figure 3-20: Module d'administration des enquêtes web • Section de suivi des entrevues

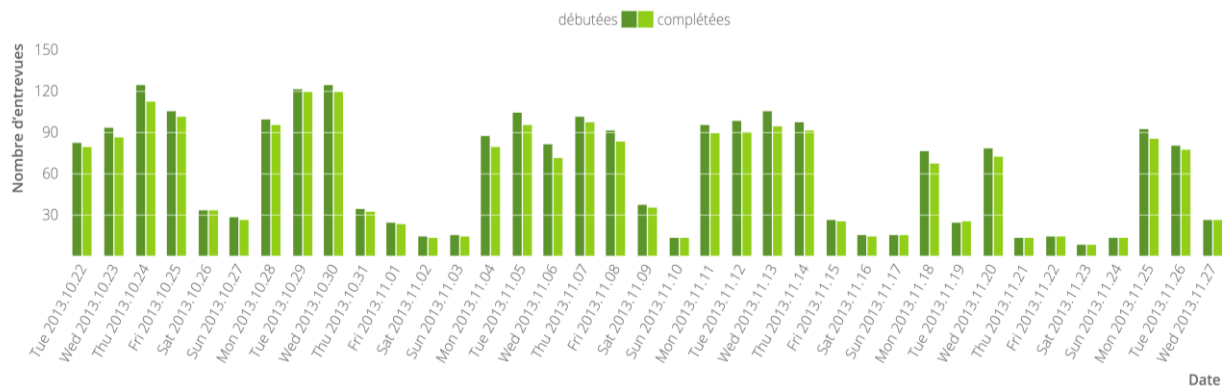


Figure 3-21: Suivi des entretiens • *Distribution des entretiens débutés et complétés par jour*

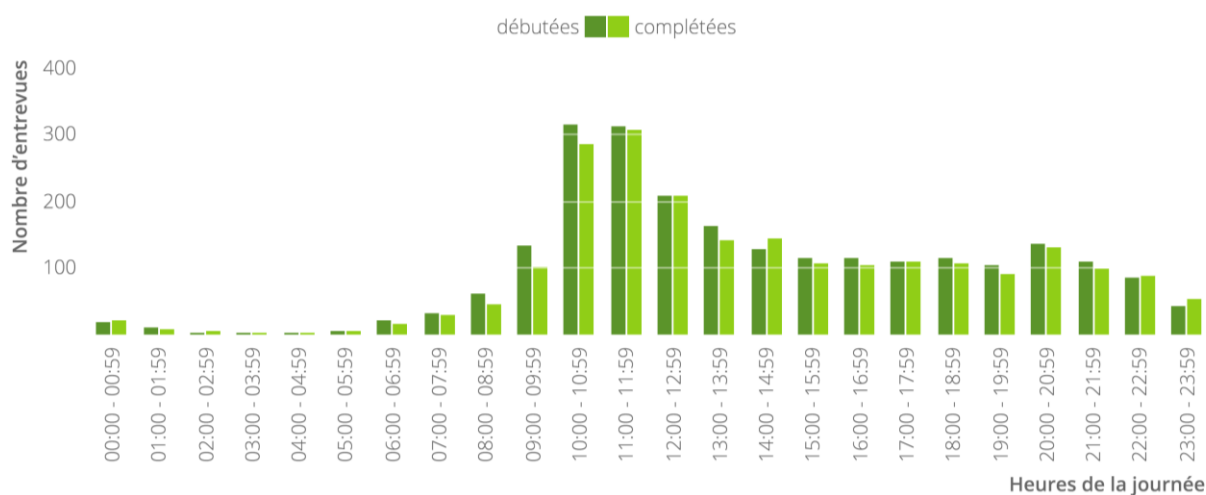


Figure 3-22: Suivi des entretiens • *Distribution des entretiens débutés et terminés par heure*

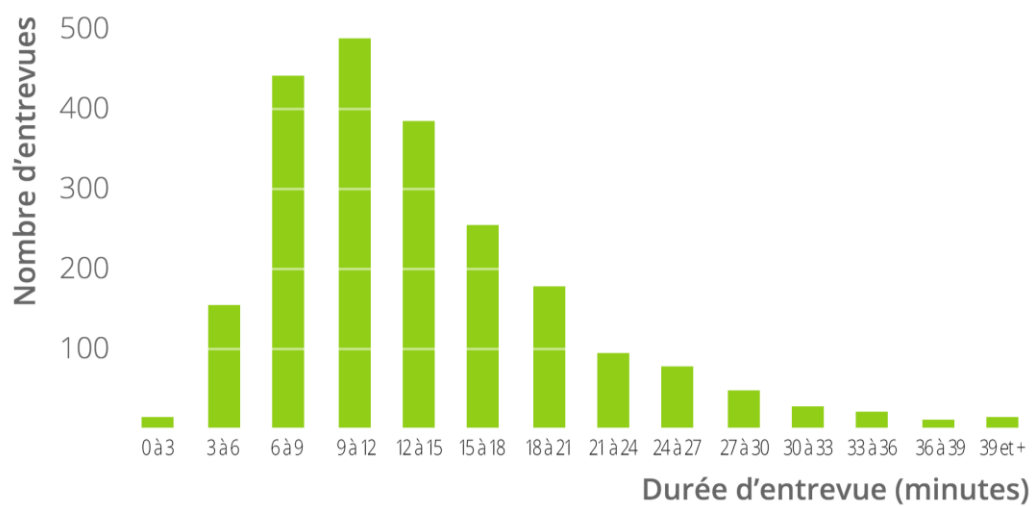


Figure 3-23: Suivi des entretiens • *Distribution des durées d'entrevue*

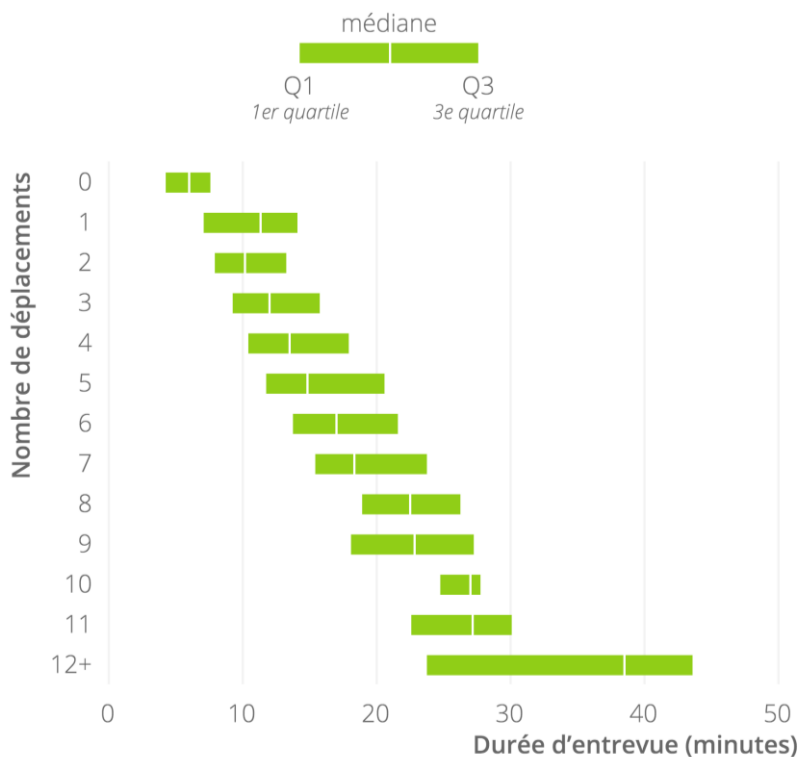


Figure 3-24: Suivi des entretiens • *Durée d'entrevue en fonction du nombre de déplacements*

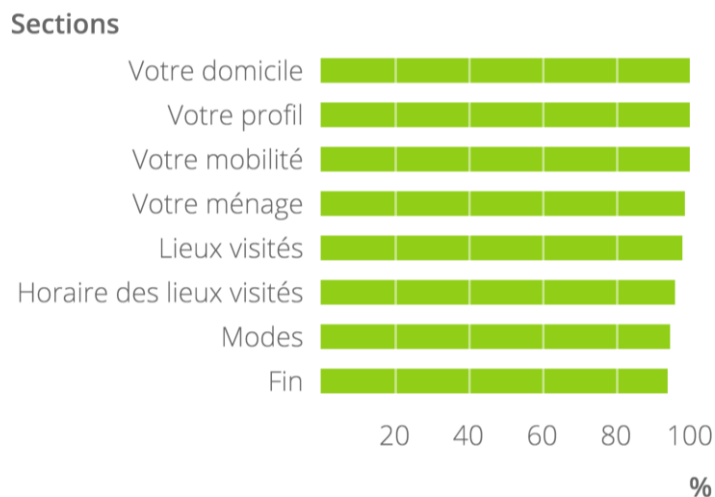


Figure 3-25: Suivi des entretiens • *Taux d'achèvement des sections du questionnaire*

3.7.2 Observation et validation des entretiens

La section d'observation et de validation comprend une liste des entretiens comportant des erreurs ou possédant des informations déclarées qui sortent de la normale après avoir été traitées

par le processus de validation automatisé. Après avoir consulté cette liste, l'analyste ou l'administrateur peut, en sélectionnant une entrevue en particulier, consulter celle-ci comme s'il était le répondant lui-même, c'est-à-dire en utilisant la même interface. Le processus de validation est présenté plus en détail à la section 5.10.

3.7.3 Commentaires des répondants

Les commentaires écrits par les répondants sont rassemblés et rendus disponibles dans une section distincte. Cela permet de comparer rapidement et de mettre en contexte ces commentaires afin de proposer des améliorations à l'expérience vécue lors des entrevues et de déceler les problèmes techniques qui auraient pu avoir lieu.

3.7.4 Importation et exportation des données

La dernière section du module d'administration concerne les importations et exportations de fichiers et de données relatives à l'enquête. C'est à cet endroit qu'il est possible d'importer des fichiers d'objets géographiques et des données de recensement qui seront utiles pour la pondération des entrevues et l'enrichissement des données collectées. Lorsque les entrevues ont été validées et archivées dans l'entrepôt de données, une interface d'exportation des fichiers finaux permet de sélectionner le type d'objet (ménage, personne, déplacement, segment, etc.) ainsi que les attributs à exporter. Les fichiers peuvent être exportés en format CSV (Comma Separated Value) avec encodage Unicode (permettant les caractères latins et spéciaux) ou avec l'encodage prévu pour l'utilisation facilitée dans Microsoft Excel. La Figure 3-26 présente un exemple de l'interface de sélection des attributs pour le type d'objet « déplacement », suivie de la Figure 3-27, qui montre un fichier exporté ouvert dans Microsoft Excel. Les attributs disponibles proviennent de l'entrepôt de données défini et présenté au Chapitre 5.



Déplacements

▸ Déplacement (clés)

▼ Déplacement

- Déplacement | IPERE
- Déplacement | Nombre de segments
- Déplacement | Nombre de modes uniques
- Déplacement | Nombre de segments en voiture
- Déplacement | Nombre de segments à vélo
- Déplacement | Premier ou dernier déplacement?
- Déplacement | Mode unique ou multimode?
- Déplacement | Catégorie de mode comparable
- Déplacement | Modes
- Déplacement | Date de début | Jour de la semaine
- Déplacement | Date de début étendue (24h+) | Jour de la semaine
- Déplacement | Date de fin | Jour de la semaine
- Déplacement | Date de fin étendue (24h+) | Jour de la semaine
- Déplacement | Motifdet à l'origine
- Déplacement | MOTIFDET
- Déplacement | Durée déclarée de l'activité à l'origine (secondes)
- Déplacement | Durée déclarée de l'activité à l'origine (minutes)
- Déplacement | Durée déclarée de l'activité à destination (secondes)
- Déplacement | Durée déclarée de l'activité à destination (minutes)
- Déplacement | Premier stationnement | Type
- Déplacement | Premier stationnement | Type | Identifiant interne
- Déplacement | Premier stationnement | Type comparable
- Déplacement | Premier stationnement | Type de paiement
- Déplacement | Premier stationnement | Type de paiement | Identifiant interne
- Déplacement | Premier stationnement | Type de paiement comparable
- Déplacement | Premier stationnement | Période de paiement
- Déplacement | Premier stationnement | Période de paiement | Identifiant interne
- Déplacement | Premier stationnement | Période de paiement comparable
- Déplacement | Premier stationnement | Montant payé
- Déplacement | Premier stationnement | Devise du montant payé
- Déplacement | Qui conduisait pour le premier segment passager?
- Déplacement | Qui conduisait pour le premier segment passager (comparable)?
- Déplacement | Nombre d'occupants dans le véhicule lors du premier segment voiture
- Déplacement | Distance à vol d'oiseau (mètres)
- Déplacement | DIST
- Déplacement | Distance à vol d'oiseau (miles)
- Déplacement | Temps de parcours déclaré (secondes)
- Déplacement | Temps de parcours déclaré (minutes)
- Déplacement | Trajectoire à vol d'oiseau | WKT des coordonnées WGS 84
- Déplacement | Trajectoire à vol d'oiseau | Géographie Postgis
- Déplacement | Trajectoire à vol d'oiseau | SRID des coordonnées projetées
- Déplacement | Trajectoire à vol d'oiseau | WKT des coordonnées projetées
- Déplacement | Trajectoire à vol d'oiseau | Géométrie Postgis projetée

Figure 3-26: Interface de sélection des attributs à exporter pour les objets « déplacements »

A	B	C	D	E	F	G
Déplacement IPERE	Déplacement Nombre de segments	Déplacement Premier ou dernier déplacement?	Déplacement Mode unique ou multimode?	Catégorie de mode comparable	Déplacement Qui conduisait pour le premier segment passager?	Déplacement Temps de parcours déclaré (minutes)
52	1	Premier déplacement de la personne	Un seul mode	Transport collectif		45
53	1	Dernier déplacement de la personne	Un seul mode	Transport collectif		60
54	1	Premier déplacement de la personne	Un seul mode	Transport collectif		20
55	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Transport collectif		20
56	1	Dernier déplacement de la personne	Un seul mode	Transport collectif		25
57	1	Premier déplacement de la personne	Un seul mode	Marche		35
58	1	Dernier déplacement de la personne	Un seul mode	Marche		45
59	1	Premier déplacement de la personne	Un seul mode	Transport collectif		30
60	1	Dernier déplacement de la personne	Un seul mode	Transport collectif		45
61	1	Premier déplacement de la personne	Un seul mode	Transport collectif		45
62	1	Dernier déplacement de la personne	Un seul mode	Transport collectif		45
63	1	Premier déplacement de la personne	Un seul mode	Auto conducteur		30
64	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto conducteur		60
65	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto conducteur		30
66	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto conducteur		30
67	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto conducteur		0
68	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto conducteur		0
69	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto passager	Un membre de votre famille n'habitant pas chez vous	0
70	1	Dernier déplacement de la personne	Un seul mode	Auto passager	Un membre de votre famille n'habitant pas chez vous	0
71	1	Premier déplacement de la personne	Un seul mode	Marche		5
72	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Transport collectif		15
73	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Transport collectif		25
74	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Transport collectif		15
75	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto passager	Un collègue de travail	10
76	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Transport collectif		15
77	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Transport collectif		15
78	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Auto passager	Un ami	20
79	1	Dernier déplacement de la personne	Un seul mode	Auto passager	Un ami	30
80	1	Premier déplacement de la personne	Un seul mode	Transport collectif		45
81	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Transport collectif		50
82	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Marche		10
83	1	Déplacement intermédiaire (ni le premier ni le dernier)	Un seul mode	Marche		10

Figure 3-27: Extrait d'un fichier au format CSV exporté pour les objets « déplacements » visualisé à l'aide de *Microsoft Excel*

3.8 Paradonnées enregistrées

En plus des réponses déclarées par les répondants, la plateforme enregistre un grand nombre de paradonnées sur le processus d'entrevue lui-même. Bien que la quantité et le type de paradonnées sauvegardées au terme de chaque entrevue aient évolué au cours du développement de l'outil, la liste suivante indique les paradonnées conservées dans la version actuelle de l'outil :

- Horodatage (date et heure) de début et de fin de l'entrevue ;
- Horodatage de début et de fin de chacune des sections du questionnaire ; Lorsque le répondant revient à une section précédente, un enregistrement est également sauvegardé.
- Horodatage et valeur enregistrée à chaque modification d'une réponse à une question ; Pour les questions demandant au répondant d'entrer un nombre ou du texte, la valeur entrée est sauvegardée à chaque seconde. Un historique du texte déclaré par le répondant est donc rendu disponible par l'enregistrement de ces paradonnées. Pour les autres types

de question, la valeur est enregistrée à chaque modification de la réponse (élément choisi dans un menu déroulant, bouton sélectionné, case cochée, etc.).

- Horodatage et description de chaque message d'erreur ou de validation montrée au répondant pendant l'entrevue ;
- Horodatage et positionnement de chaque géolocalisation, de chaque clic et de chaque déplacement d'un marqueur effectués sur une carte.

L'ensemble de ces parodonnées servent à analyser en profondeur les comportements de réponse des répondants et peuvent en outre servir à reproduire presque en totalité une entrevue. Le mouvement de la souris et les différentes touches du clavier pourront également être enregistrés dans une version ultérieure, en autant que le répondant accorde son consentement au moyen de formulaire inséré au début du questionnaire (confidentialité et respect de la vie privée). Pour l'instant, un tel processus aurait un effet trop important sur les performances du navigateur chez la plupart des répondants.

3.9 Analyse des commentaires émis par les répondants à la fin des questionnaires

À la lumière des commentaires émis au sujet de l'interface par les répondants, la convivialité et la simplicité du système de collecte a été mentionnée par plusieurs utilisateurs. Sachant qu'il est difficile d'évaluer de manière empirique l'expérience vécue par les répondants durant l'entrevue, les commentaires qu'ils ont émis ont été tout de même classés en quatre catégories principales : les commentaires positifs et négatifs relatifs au questionnaire et à l'interface utilisateur, les commentaires neutres, c'est-à-dire comportant une facette positive et une facette négative ou pour lesquels il est impossible de tirer une conclusion claire, et finalement, les commentaires non reliés au questionnaire lui-même, mais qui concernent plutôt l'organisation en charge de l'enquête ou le réseau de transport local. Afin de mesurer l'évolution de la plateforme au cours des années, les commentaires positifs, négatifs et neutres concernant spécifiquement l'interface ont été comptabilisés et comparés en fonction de la version de la plateforme utilisée (Tableau 3.2 et Figure 3-28). Bien que les conclusions que l'on puisse en tirer soient limitées, la tendance suivie par la proportion de commentaires positifs est à la hausse avec l'évolution de la plateforme d'en-

quête. Dans la littérature, il existe très peu de publications qui traitent de l'évaluation de la qualité des questionnaires web à partir de la satisfaction des répondants. (Schonlau, 2015) est le seul exemple permettant de mettre en contexte la méthode d'évaluation des commentaires en les catégorisant (positifs, négatifs, neutres). Dans les deux enquêtes web qu'ils ont analysées, la proportion de commentaires positifs sur le nombre de commentaires positifs, négatifs et neutres concernant l'interface est de moins de 4% (1,8% si l'on inclut les commentaires neutres), en comparaison des proportions entre 70% et 88% obtenues pour les enquêtes de mobilité réalisées dans le cadre de cette thèse. La différence est probablement due à la nature des échantillons et des sujets traités. Une analyse plus pointue serait nécessaire pour tenter d'évaluer de manière plus précise et objective l'évolution du questionnaire et la qualité de l'interface proposée, car aucune mesure absolue de la qualité ne peut être effectuée pour l'instant.

Tableau 3.2 : Nombre et proportion de commentaires positifs et négatifs associés aux différentes versions de l'outil

Version	Enquêtes	Type de questionnaire	Nombre de positifs	Nombre de négatifs	Proportion des commentaires positifs
2010	Polytechnique 2010	Personne	128	45	74.0%
2011	Polytechnique et Université de Montréal 2011	Personne	345	76	81.9%
2013A	Communauto et BIXI 2013	Personne	403	92	81.4%
2013B	Sherbrooke 2013	Personne	75	12	86.2%
2015	StudentMoveTO 2015	Personne	1270	174	88.0%
2011	Québec 2011	Ménage	12	5	70.6%
2013A	AMT 2013	Ménage	44	14	75.9%

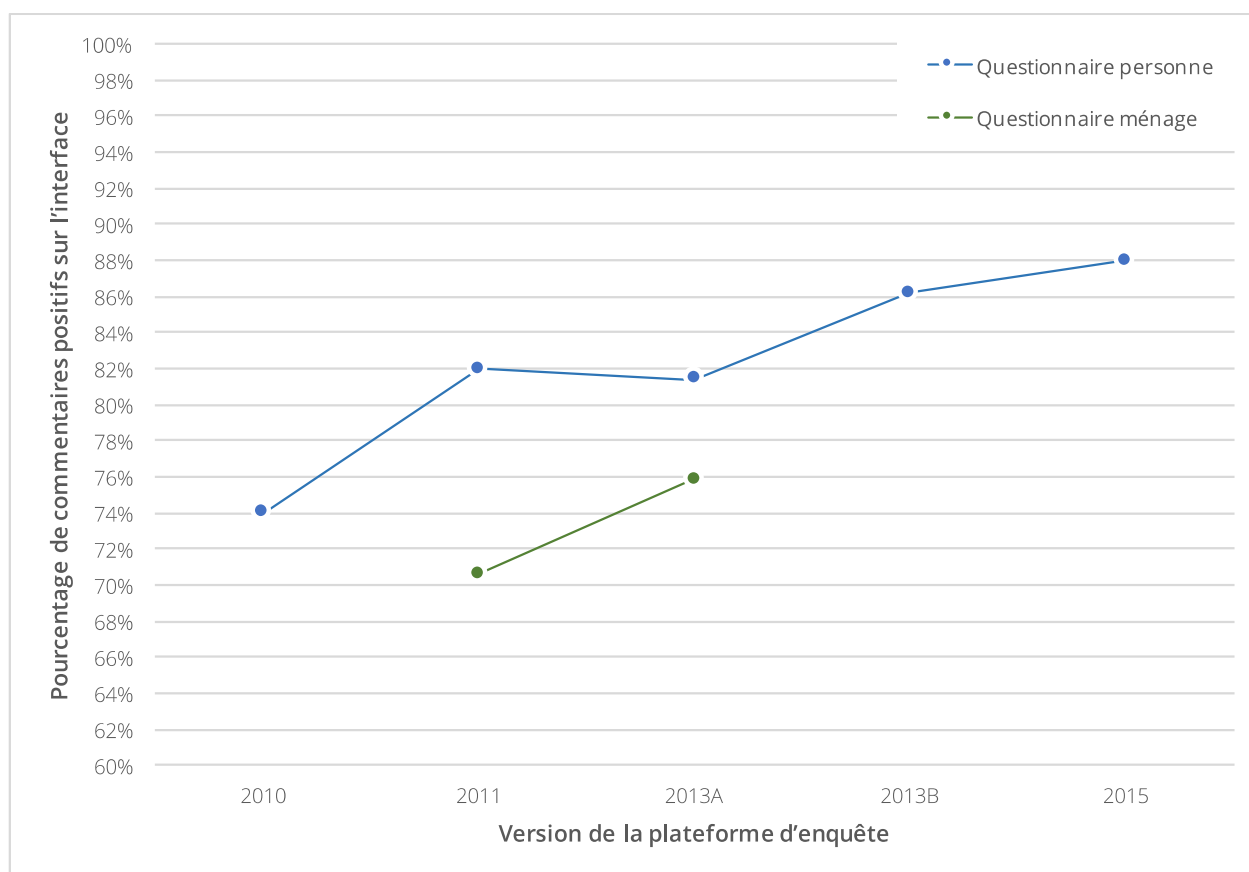


Figure 3-28 : Évolution de la proportion de commentaires positifs relatifs à l'interface

Un échantillon de commentaires²⁷ positifs, négatifs et neutres relatifs au questionnaire et à l'expérience du répondant sont présentés à l'Annexe A et permettent d'évaluer la pertinence du classement proposé.

3.10 Enquêtes web réalisées dans le cadre de la thèse

Plusieurs enquêtes Origine-Destination ont été conduites en utilisant la plateforme développée. D'abord un prototype, l'outil de collecte a évolué au cours des années pour devenir plus robuste, plus complet et plus convivial, autant pour les répondants que pour les administrateurs. Le Tableau 3.3 rassemble les périodes d'enquête, les méthodes de recrutement, les échantillons et les objectifs principaux de ces enquêtes. Une courte description de chacune des enquêtes ayant été réalisées permet ensuite de mettre en contexte le développement de l'outil. Des acronymes sont ajoutés pour simplifier la description des différentes enquêtes. Les taux de réponse et d'achèvement des entrevues pour chacune des enquêtes sont présentés au Tableau 4.1 et au Tableau 4.2, dans le cadre de la modélisation des durées d'entrevue (Chapitre 4).

²⁷ Les commentaires énoncés sont présentés dans leur langue d'origine (en français ou en anglais, selon la langue du questionnaire utilisée). Les éléments ne concernant pas l'entrevue elle-même ont été retirés.

Tableau 3.3 : Contexte de réalisation pour 8 enquêtes-personnes et 2 enquêtes-ménages conduites entre 2010 et 2015

Enquête	Période	Recrutement	Échantillon	Objectif principal
Polytechnique Montréal PY10 <i>Questionnaire personne</i>	Automne 2010	Courriel (liste officielle de l'établissement)	Étudiants et employés	Plan de mobilité durable
Polytechnique Montréal PY11 <i>Questionnaire personne</i>	Automne 2011	Courriel (liste officielle de l'établissement)	Étudiants et employés	Plan de mobilité durable
Université de Montréal UM11 <i>Questionnaire personne</i>	Automne 2011	Courriel (liste officielle de l'établissement)	Étudiants et employés	Plan de mobilité durable
Trois-Rivières – MTQ TR11 <i>Questionnaire personne</i>	Printemps 2011	Téléphones cellulaires (liste locale) Par la poste	Utilisateurs de téléphones cellulaires Un répondant par ménage	Test de l'outil dans un contexte régional
Étudiants de la région de Sherbrooke – MTQ SH13 <i>Questionnaire personne</i>	Printemps 2013	Courriel (listes officielles des établissements)	Étudiants (université principale et collèges)	Fusion des données avec l'échantillon conventionnel CATI
Communauto CM13 <i>Questionnaire personne</i>	Automne 2013	Courriel (liste officielle des membres)	Membres du service d'autopartage	Étude des comportements de mobilité des membres
Bixi BX13 <i>Questionnaire personne</i>	Automne 2013	Courriel (liste officielle des membres)	Membres du service de vélopartage	Étude des comportements de mobilité des membres
Étudiants universitaires de la région de Toronto TT15 <i>Questionnaire personne</i>	Automne 2015	Courriel (listes officielles des établissements)	Population étudiante universitaire	Étude des comportements de mobilité des étudiants universitaires
Québec – MTQ QC11 <i>Questionnaire ménage</i>	Automne 2011	Par la poste Courriel (résidences étudiantes) Facebook et Twitter (invitations effectuées par les agences de transport locales)	Membres du ménage de 5 ans et plus Étudiants habitant dans une résidence Utilisateurs des réseaux sociaux	Premier test du questionnaire ménage
Montréal – AMT MTL13 <i>Questionnaire ménage</i>	Automne 2013	Refus téléphoniques Site web des partenaires	Membres du ménages de 5 ans et plus	Deuxième test du questionnaire ménage

3.10.1 Polytechnique Montréal • *Enquête auprès des étudiants et employés* Enquête-personne • Automne 2010

L'outil d'enquête en ligne a été développé pour la première fois en 2010 et bénéficiait notamment du potentiel des cartes en ligne pour le géocodage et le positionnement des lieux d'origine et de destination des déplacements enquêtés. Le questionnaire a été élaboré en se basant en grande partie sur celui utilisé par l'AMT (Agence métropolitaine de transport) dans le cadre des enquêtes téléphoniques Origine-Destination de la grande région de Montréal. Ce choix assurait une meilleure compatibilité des données pour des fins de recherche et de comparaison et permettait du même coup aux différents intervenants du projet de se familiariser avec les enquêtes conventionnelles sur la mobilité. Par contre, à l'opposé de l'enquête régionale de l'AMT lors de laquelle le répondant doit répondre pour tous les membres de son ménage (entrevue ménage), l'entrevue web préparée pour Polytechnique était individuelle (entrevue personne).

3.10.2 Ministère des transports du Québec • *Enquête régionale de Trois-Rivières* • *Enquête auprès des propriétaires de téléphones cellulaires* Prototype d'enquête-personne • Printemps 2011

En raison de la disparition accélérée des lignes téléphoniques résidentielles fixes, notamment au sein des jeunes ménages au profit de la téléphonie cellulaire depuis quelques années²⁸ (Statistiques Canada, 2014), le ministère des Transports a proposé d'évaluer un recrutement à partir d'un annuaire commercial de numéros cellulaires et d'offrir aux ménages rejoints de cette façon de répondre à l'enquête régionale exclusivement par le mode web, le tout dans une optique expérimentale. Pour ce faire, les personnes jointes par cellulaire ayant accepté de participer faisaient connaître leur adresse courriel. On leur envoyait par la suite un court message décrivant les objectifs de l'enquête et les procédures de connexion au site web sur lequel était hébergé le prototype. Un certain nombre de répondants ayant refusé de participer par téléphone ont aussi été invités à participer sur le web. Un code d'accès et un mot de passe étaient fournis à tous les participants, ce qui donnait l'occasion de déterminer avec davantage de précision les taux de réponse et les périodes privilégiées pour répondre à l'enquête, le tout en fonction du moment où le ménage avait été contacté par téléphone. D'autre part, un lot de 1000 personnes tiré de l'échantillon de l'enquête régionale conventionnelle a été attribué au mode web. Les personnes étaient alors contactées par la poste. Encore une fois, un code d'accès et un mot de passe étaient associés à chaque résidence et favorisait un suivi du taux de réponse en fonction des dates d'envoi postal. Une grève de Postes Canada a par contre rendu laborieux le suivi des envois puisque des délais considérables ont été rencontrés, certaines lettres ayant même pu être perdues ou, dans le meilleur des cas, réacheminées à une date ultérieure après la fin de la période d'enquête prévue.

²⁸ 21 % des ménages ont déclaré n'utiliser qu'un téléphone cellulaire en 2013, comparé à 13 % en 2010.

3.10.3 Polytechnique Montréal et Université de Montréal • *Enquête auprès des étudiants et employés*

Enquête-personne • Automne 2011

Après le succès de 2010 à Polytechnique, l'Université de Montréal a émis le souhait de faire participer l'ensemble de sa communauté à une enquête de mobilité en 2011 afin d'élaborer un plan de transport. Dans le but d'amasser des données sur le plus grand nombre possible d'étudiants et d'employés de l'ensemble du campus universitaire, Polytechnique a reconduit son enquête pour une deuxième année consécutive. Lors de cette enquête conjointe en mode personne, une grande quantité de données ont été amassées et analysées. Des commentaires sur l'interface et le questionnaire ont été collectés afin d'encourager l'amélioration de l'outil dans ses versions subséquentes.

3.10.4 Ministère des transports du Québec • *Enquête régionale de Québec* Prototype d'enquête-ménage • Automne 2011

Afin de tester pour une première fois l'outil de collecte en mode ménage, le Ministère des transports du Québec a sélectionné un sous-échantillon de l'enquête Origine-Destination de la région de Québec. Pour l'occasion, environ 750 ménages ont été contactés par la poste pour participer à l'étude.

3.10.5 Ministère des transports du Québec • *Enquête régionale de Sherbrooke* ***• Enquête auprès des étudiants post-secondaire de la région de Sherbrooke***

Enquête-personne • Automnes 2012 et 2013

Pour faire face à la sous-représentation du milieu étudiant dans l'enquête régionale de Sherbrooke, le ministère des Transports a mandaté l'équipe de Polytechnique pour réaliser une version en ligne en mode personne pour rejoindre les étudiants de la région par l'intermédiaire des institutions d'enseignement. Les données de cette version web de l'enquête ont été les premières à être fusionnées à celles recueillies par téléphone auprès des ménages de la région de Sherbrooke. La fusion n'a pas été effectuée au moyen de la plateforme développée dans le cadre de la thèse.

3.10.6 Agence métropolitaine de transport • *Enquête régionale de Montréal* Enquête-ménage • Automne 2013

Dans le cadre de l'enquête régionale téléphonique sur la mobilité des ménages tenue par l'AMT à l'automne 2013, les organisateurs ont proposé aux personnes ayant refusé de répondre au téléphone d'accéder à une version en ligne du questionnaire. D'autre part, cette même version du questionnaire en ligne était proposée sur les réseaux sociaux des partenaires de la région dans le but de mesurer l'attrait de l'enquête auprès des utilisateurs de ces réseaux.

3.10.7 Communauto et Bixi Montréal • *Enquête auprès des membres de l'autopartage et du vélopartage* Enquête-personne • Automne 2013

En parallèle à l'enquête régionale de l'AMT, le questionnaire web développé a été proposé aux membres du service d'autopartage Communauto. Les résultats permettront de mieux comparer les habitudes de mobilité des membres de l'autopartage avec les habitudes de déplacements de l'ensemble de la population. En utilisant le modèle proposé aux membres de Communauto, une version a également été conçue pour les membres des services de vélopartage de BIXI et répondait aux mêmes besoins de comparaisons.

3.10.8 Forum vies mobiles • *Enquête sur la mobilité des 16-35 ans* Enquête-personne (non Origine-Destination) • Automne 2014

Dans le cadre du forum vies mobiles, les chercheurs associés au projet ont élaboré un questionnaire destiné aux jeunes de 16 à 35 ans résidant dans les régions de Montréal et de Lyon, en France. Le questionnaire ne comportait pas de section de type Origine-Destination, mais plutôt un ensemble de questions qualitatives et quantitatives portant sur les habitudes et l'historique de mobilité des répondants. Le questionnaire a été présenté exclusivement en mode web et a permis aux chercheurs québécois et français de mieux mesurer les changements de mobilité des nouvelles générations d'adultes, notamment sur le plan de leur rapport à la voiture et aux technologies de l'information.

3.10.9 Enquête longue-distance • *Enquête prototype sur les déplacements longue-distance*

Enquête-personne • 2014-2016

Afin de répondre à un besoin de mieux comprendre les dynamiques de transport interurbain et régional, un prototype de questionnaire axé sur les déplacements longue distance des individus sur une période de plusieurs semaines est en cours de développement pour être testé auprès de différents échantillons d'utilisateurs des réseaux de transport régionaux et interurbains. La collecte de données devrait débuter au cours de l'année 2016.

3.10.10 Universités de Toronto • *Enquête auprès des étudiants universitaires de la région de Toronto*

Enquête-personne • Automne 2015

En prévision de l'enquête régionale torontoise Transportation Tomorrow Survey (TTS) de 2016, une version du questionnaire a été proposée aux 160 000 étudiants des quatre principales institutions universitaires de la région de Toronto (University of Toronto, Ryerson University, York University et OCAD University).

CHAPITRE 4 MODÉLISATION DE LA DURÉE D'ENTREVUE

Sachant que la durée de chaque entrevue a été mesurée pour toutes les enquêtes réalisées depuis 2011, un modèle explicatif et de prédiction de la durée d'entrevue en fonction des réponses données pendant l'entrevue a été évalué. L'objectif principal de cette modélisation n'est pas de prédire la durée d'entrevue en ne connaissant que les attributs socio-démographiques du répondant, mais bien de mieux comprendre la relation entre les durées d'entrevue et les informations déclarées par le répondant. Elle permettra un meilleur partage des ressources matérielles des serveurs utilisés pour la conduite des enquêtes. De plus, elle favorisera une évaluation, en temps réel, de la plage de temps prévue avant de compléter chaque entrevue. Cet intervalle de temps pourra être intégré au module d'administration pour des besoins de suivi, ou au questionnaire pour permettre au répondant d'estimer l'intervalle de durée d'entrevue restante en fonction des réponses précédentes.

4.1 Statistiques reliées à la durée d'entrevue

Un sommaire des statistiques descriptives des différentes enquêtes est présenté au Tableau 4.1 (enquête-ménage) et au Tableau 4.2 (enquête-personne). Le Tableau 3.3 présente le contexte de chacune des enquêtes ainsi que les acronymes utilisés dans cette section. La durée d'entrevue médiane est de 18,1 minutes pour les enquêtes-ménage (18,3 si on exclut les entrevues des étudiants habitant une résidence universitaire, qui sont des ménages à une seule personne exclusivement), et de 12,2 minutes pour les enquêtes-personne.

Tableau 4.1 : Statistiques descriptives, taux de réponse et taux d'achèvement pour les 2 enquêtes-ménage réalisées en 2011 et 2013.

Enquête	Univers d'étude	# Invitations (# rappels)	# Entrevues débutées	# Entrevues complétées	Durée d'entrevue							
					Moyenne	Écart-Type	Interquartile IQR	Min	Q1 25%	Médiane	Q3 75%	Max
QC11	<i>Par la poste</i>	1000 lettres envoyées 749 reçues (aucun rappel)	138	98	23.7	13.4	14.2	5.6	14.7	20.1	29.0	63.0
	<i>Étudiants en résidence</i>	Inconnu (aucun rappel)	60	41	11.4	7.2	11.3	2.2	5.6	9.7	16.9	32.2
	<i>Réseaux sociaux</i>	Inconnu (aucun rappel)	36	26	12.7	6.4	8.8	4.0	7.6	12.2	16.4	28.1
MTL13	<i>Refus téléphoniques</i>	553 ont accepté de participer 456 courriels valides ont reçu une invitation (1 rappel)	70	46	23.0	13.1	19.3	3.9	12.6	22.0	31.9	56.8
	<i>Sites web des partenaires</i>	Inconnu (aucun rappel)	599	443	21.4	12.5	13.8	1.4	12.8	19.0	26.6	92.0
Total enquête-ménage		Inconnu	903	654	20.9	12.6	14.2	1.4	12.1	18.1	26.3	92.0

Enquête	Échantillon	# Contacts	# Abandons	# Entrevues complétées	Taux de réponse (entrevue débutée)	Taux d'achèvement	Taux d'abandon
QC11	<i>Par la poste</i> 1000 lettres envoyées	749	40	98	18.4%	9.8%	28.9%
MTL13	<i>Refus téléphoniques</i> 553 ont accepté de participer	456 courriels valides ont reçu l'invitation	24	46	15.4%	8.6%	34.3%
	<i>Sites web des partenaires</i>	Inconnu	156	443	N/A	N/A	26.0%

Tableau 4.2 : Statistiques descriptives, taux de réponse et taux d'achèvement pour les 7 enquêtes-personne réalisées entre 2010 et 2013

Enquête	Univers d'étude	# Invitations (# rappels)	# Entrevues débutées	# Entrevues complétées	Durée d'entrevue								
					Moyenne	Écart-Type	Interquartile IQR	Min	Q1 25%	Médiane	Q3 75%	Max	
PY10		8 618	8 618 (2 rappels)	1 972	1 530	13.2	6.5	7.4	2.5	8.7	11.6	16.1	39.8
PY11		8 576	8 576 (2 rappels)	1 929	1 673	12.6	5.9	7.0	1.7	8.5	11.4	15.4	75.9
UM11		58 483	58 483 (2 rappels)	7 948	6 501	14.1	6.9	8.0	1.8	9.4	12.8	17.4	74.3
TR11	Téléphones cellulaires: N/A	1 644 (aucun rappel)	333 ont accepté de participer 109 entrevues débutées	81		11.9	7.5	7.5	2.5	7.3	10.0	14.8	47.6
	Par la poste: N/A	< 1 000 (aucun rappel)	54	35		13.0	7.2	9.9	2.7	7.9	11.8	17.8	33.0
SH13	Université: 16 194	16 194 (1 rappel)	2 399	1 838		13.0	5.5	6.7	2.5	9.1	12.1	15.7	48.1
	Collèges: 6 820	6 820 (aucun rappel)	683	467		13.7	6.5	7.3	3.2	9.4	12.3	16.7	53.7
CM13	~30 000	Inconnu (aucun rappel)	3 143	2 527		13.8	7.1	8.2	1.9	8.9	12.3	17.1	63.2
BX13	78 068	74 905 (1 rappel)	6,191	4 423		13.6	7.3	8.2	2.1	8.7	12.1	16.9	69.6
Total enquête-personne		> 180 000	24 428	19 075		13.6	6.8	7.8	1.7	9.0	12.2	16.7	75.9

Enquête	Échantillon	# Contacts	# Abandons	# Entrevues complétées	Taux de réponse (entrevue débutée)	Taux d'achèvement	Taux d'abandon
PY10	Étudiants et employés	8 618	442	1 530	● 22.9%	● 17.8%	● 22.4%
PY11	Étudiants et employés	8 576	256	1 673	● 22.5%	● 19.5%	● 13.3%
UM11	Étudiants et employés	58 483	1 447	6 501	● 13.6%	● 11.1%	● 18.2%
TR11	Téléphones cellulaires	1 644 333 ont accepté de participer	28	81	● 6.6%	● 4.9%	● 25.7%
	Par la poste	< 1 000	19	35	N/A	N/A	● 35.2%
SH13	Université	16 194	561	1 838	● 14.8%	● 11.3%	● 23.4%
	Collèges	6 820	216	467	● 10.0%	● 6.8%	● 31.6%
CM13	Membres: ~30 000	Inconnu	616	2 527	N/A	N/A	● 19.6%
BX13	Membres: 78 068	74 905	1 768	4 423	● 8.3%	● 5.7%	● 28.6%

4.2 Méthodologie

4.2.1 Calcul des taux de réponse et d'achèvement

Pour BX13, la liste des courriels ne contenait pas d'information sur le statut de l'abonnement. Conséquemment, il était impossible d'obtenir le nombre exact de membres actifs, ce qui explique probablement le faible taux de réponse. D'autre part, le taux de réponse pour CM13 n'a pu être calculé, puisque le nombre de membres ayant reçu une invitation demeure inconnu. En fait, une fenêtre apparaissait sur le site de Communauto lorsqu'un membre effectuait une réservation. Ainsi, seuls les membres actifs durant la période d'enquête ont pu accéder à l'entrevue. De son côté, l'enquête TR11 comprenait deux échantillons différents. Le premier était un échantillon provenant d'une liste de téléphones cellulaires achetée par le MTQ pour évaluer la faisabilité d'inclure les usagers du cellulaire lors des prochaines enquêtes régionales sur la mobilité. Les intervieweurs appelaient chaque numéro apparaissant sur la liste et, lorsque l'interlocuteur acceptait de participer, lui indiquaient l'adresse du site web à utiliser pour débiter l'entrevue. Le second échantillon comptait 1000 adresses postales sélectionnées de manière aléatoire sur le territoire d'étude. Une lettre d'introduction était envoyée et incluait un lien vers le site web d'entrevue, tout en prenant soin de mentionner qu'un seul membre du ménage pouvait répondre. Malheureusement, une grève des employés du service postal a provoqué un délai dans la distribution du courrier. Certaines lettres ont même pu être égarées. De ce fait, le nombre exact de lettres reçues par les ménages demeure inconnu et le taux de réponse n'a donc pas été calculé dans ce cas. Du côté de QC11, le taux de réponse n'a été calculé que pour l'échantillon postal puisque le nombre d'invitations envoyées par le biais des autres échantillons était également inconnu. Les taux de réponse pour MTL13 sont donnés pour les ménages ayant accepté de répondre en ligne après avoir refusé de répondre à l'entrevue téléphonique. Ils ne sont pas fournis pour les personnes contactées sur les réseaux sociaux des partenaires de l'enquête régionale étant donné l'absence d'information sur le nombre d'individus ayant reçu une invitation.

4.2.2 Critères d'achèvement des entrevues

Dans l'analyse des taux d'achèvement des enquêtes étudiées, une entrevue est considérée complète lorsque le répondant a terminé toutes les sections, qu'il a répondu à toutes les questions

requis et que toutes les validations du questionnaire ont été effectuées et ne retournent pas d'erreur. Dans le cas des enquêtes-ménage, ces trois conditions doivent de surcroît être vérifiées pour les entrevues de tous les membres du ménage.

4.2.3 Modèle de durée d'entrevue

Les parodonnées utilisées pour mesurer la durée d'entrevue incorporent les horodatages (*timestamps*) inscrits dans le journal des entrevues (logs). La durée totale est calculée entre le clic sur le bouton « Débuter l'entrevue » et le clic sur le bouton « Terminer l'entrevue ». Pour exclure de cette durée les moments pendant lesquels le répondant ne répond pas activement à l'entrevue, les durées de plus de 5 minutes²⁹ entre deux actions consécutives enregistrées dans le journal sont remplacées par la durée moyenne des actions comparables du même répondant. Bien que cette méthode apporte un biais important dans l'analyse statistique des temps de réponse, elle demeure une manière simple pour ne conserver que les durées pendant lesquelles le répondant est attentif. Il n'existe pas de mécanisme fiable dans les navigateurs web pour vérifier quel onglet ou quelle page est active, et il est impossible de savoir si le répondant est bel et bien devant son écran lorsqu'aucune action n'est détectée. Une analyse plus précise des durées d'entrevue pourra être réalisée seulement après obtention de journaux d'entrevues complétées par un panel de répondants supervisés dans un lieu contrôlé. Les durées d'entrevues de l'enquête PY10 n'ont pas été incluses dans la création du modèle, car la méthode d'enregistrement des parodonnées de durées n'était pas assez robuste pour obtenir des durées précises et comparables.

4.2.3.1 Hypothèses

Plusieurs hypothèses ont été proposées en se basant sur les résultats de (Chapleau, 2003) et (Morncy, 2008) (entrevues Origine-Destination téléphoniques CATI), et de (Gummer & Rossmann, 2014) (entrevues web non reliées au transport). Les entrevues provenant des enquêtes-ménage ont fait l'objet d'une modélisation distincte des entrevues des enquêtes-personne, puisque leurs durées d'entrevue moyennes sont significativement différentes et parce que l'entrevue est com-

²⁹ 5 minutes constituent le seuil à partir duquel la distribution des durées entre les actions se stabilise.

plétée par un ou plusieurs membres du ménage, ce qui rend difficile l'inclusion des attributs démographiques du répondant. En ce sens, pour les enquêtes-ménage, seul l'âge moyen des adultes de plus de 18 ans dans le ménage est inclus en tant que variable démographique dans le modèle, puisque la personne qui répond au questionnaire et son âge ne sont pas connus. Le Tableau 4.3 présente les hypothèses reliées à la durée d'entrevue des enquêtes-ménage et des enquêtes-personne.

Tableau 4.3 : Hypothèses utilisées pour la modélisation de la durée d'entrevue des enquêtes-ménage et enquêtes-personne

<i>EM: enquête-ménage</i>		
<i>EP: enquête-personne</i>	Source	Hypothèse
Hypothèse EM1 / EP1	CATI	Le nombre de déplacements est proportionnel à la durée d'entrevue (effet prédit: grand)
Hypothèse EM2 / EP2	CATI	La taille du ménage augmente proportionnellement la durée d'entrevue (effet prédit: grand)
Hypothèse EM3a	(G. & R. 2014)	Les ménages constitués de personnes plus âgées ont des durées d'entrevue plus longues (effet prédit: moyen)
Hypothèse EM3b	CATI	Les ménages constitués de personnes plus âgées ont des durées d'entrevue moins longues (effet prédit: moyen)
Hypothèse EP3a	(G. & R. 2014)	Les personnes plus âgées ont des durées d'entrevue plus longues (effet prédit: moyen)
Hypothèse EP3b	CATI	Les personnes plus âgées ont des durées d'entrevue moins longues (effet prédit: moyen)
Hypothèse EM4 / EP4	CATI	Les entrevues débutées après 17h prennent davantage de temps à compléter (effet prédit: léger)
Hypothèse EM5 / EP5	Nouvelle proposition	La présence d'une connexion internet haute vitesse au domicile réduit le temps d'entrevue (chargement plus rapide des pages et plus grande expérience de l'utilisation d'internet) (effet prédit: léger)
Hypothèse EM6 / EP6	(G. & R. 2014)	L'utilisation d'un appareil à petit écran augmente le temps d'entrevue (effet prédit: léger)
Hypothèse EM7 / EP7	Nouvelle proposition	Le fait que le répondant ne déclare aucun déplacement réduira la durée d'entrevue encore davantage qu'avec EM1 / EP1 puisqu'il n'aura pas à lire les introductions à la déclaration des déplacements. En fait, l'hypothèse prédit qu'il existe un temps fixe requis pour lire ces introductions et textes explicatifs. (effet prédit: léger)
Hypothèse EP8 <i>PY11, SH13, CM13 et BX13 seulement</i>	(G. & R. 2014)	Les répondants intéressés davantage par le sujet auront des durées d'entrevue plus longues à cause de leur motivation à être plus précis dans leurs réponses. La réponse à la question: <i>Voudriez-vous participer à d'autres études sur la mobilité?</i> est utilisée pour connaître cet intérêt (effet prédit: léger)
Hypothèse EP9	CATI	Le fait que le répondant soit une femme augmente la durée d'entrevue (effet prédit: léger)
Hypothèse EP10	Nouvelle proposition	Le nombre de questions supplémentaires spécifique à l'enquête augmente de manière proportionnelle la durée d'entrevue (effet prédit: moyen)
Hypothèse EP11	Nouvelle proposition	Un grand nombre de modes de transport uniques déclarés augmente la durée d'entrevue, à cause de la complexité accrue. (des questions différentes sont posées en fonction du mode sélectionné). (effet prédit: léger)

Comme la déclaration des déplacements est considérée comme étant l'étape la plus complexe du questionnaire, l'analyse des entrevues CATI a révélé que le nombre de déplacements déclarés

influençait de manière significative et avec un grand effet la durée d'entrevue globale. Conséquemment, la première hypothèse considère que le nombre de déplacements déclarés est proportionnel avec la durée d'entrevue (EM1 et EP1). Comme le nombre de questions associées aux déplacements change en fonction du mode déclaré, ce nombre est séparé en trois attributs distincts : le nombre de segments effectués en voiture, en transport collectif et selon un autre mode de transport. D'autre part, pour tenir compte de la complexité de localiser les lieux visités sur la carte, le nombre de lieux visités uniques est également inclus. En fait, les lieux visités de mêmes coordonnées sont comptés une seule fois puisque des raccourcis permettent de sélectionner rapidement un lieu déjà géolocalisé.

Dans les hypothèses EM2 et EP2, la taille du ménage est incluse, suivant les résultats obtenus des entrevues CATI. En effet, dans les questionnaires CATI et dans le questionnaire web ménage et personne, un groupe de questions permet de spécifier le profil de chacun des membres du ménage.

Selon (Gummer & Rossmann, 2014), les personnes plus âgées nécessitent davantage de temps pour compléter une entrevue en ligne, alors que du côté des entrevues CATI, les ménages plus âgés ont des temps d'entrevue moyens plus faibles. Sachant cela, les hypothèses EM3 et EP3 sont séparées en deux versions (A et B), ce qui permet de distinguer le sens de l'effet sur la durée d'entrevue. Comme les modèles CATI indiquent une durée moins longue pour les entrevues effectuées après 17h, les hypothèses EM4 et EP4 ont été testées pour vérifier ce résultat dans le cadre d'entrevues web.

Les hypothèses EM5 et EP5 tentent de vérifier si la possession d'une connexion internet à haute vitesse au domicile tend à réduire la durée d'entrevue. En effet, les répondants possédant un accès haute vitesse devraient être plus à l'aise avec les technologies web que les utilisateurs de connexions à basse vitesse, d'autant plus que la connexion haute vitesse permet d'accélérer le chargement des pages lors de l'entrevue. Toutefois, il est difficile, voire impossible dans certains cas, de vérifier si l'entrevue a été complétée à la maison, au travail ou dans un autre lieu. En utilisant les résultats de (Gummer & Rossmann, 2014), les hypothèses EM6 et EP6 tentent de vérifier si l'utilisation d'un appareil à petit écran tend à augmenter le temps de réponse. Les petits

écrans nécessitent davantage de manipulations lors de l'entrevue pour être en mesure de localiser les lieux sur la carte ou de lire les textes d'introduction.

Lorsque le répondant déclare n'avoir effectué aucun déplacement durant la journée enquêtée, plusieurs textes d'introduction n'apparaissent pas et, en ce sens, cela devrait davantage réduire la durée d'entrevue. Les hypothèses EM7 et EP7 permettent de vérifier ce raisonnement.

Les quatre dernières hypothèses sont spécifiques aux enquêtes-personne. La première, EP8, est proposée pour vérifier si, comme l'ont démontré (Gummer & Rossmann, 2014), un répondant avec un intérêt plus grand pour le sujet sera plus précis dans ses réponses, et de ce fait, prendra davantage de temps pour compléter l'entrevue. Pour mesurer cet intérêt, la réponse donnée à la question « Seriez-vous intéressé à participer à d'autres études sur la mobilité ? » est utilisée. Comme cette question n'a été posée que pour les enquêtes PY11, SH13, CM13 et BX13, une variable nominale binaire (dummy) a été ajoutée pour tenir compte des réponses inconnues. L'hypothèse EP9 utilise le sexe du répondant pour tenter de confirmer les résultats obtenus des entrevues CATI (Chapleau, 2003; Morency, 2008) pour lesquelles il a été vérifié que les femmes répondaient en moyenne plus rapidement à l'entrevue que les hommes. L'hypothèse EP10 tente de mesurer l'effet du nombre de questions ajoutées à la fin du questionnaire sur la durée totale d'entrevue. Enfin, l'hypothèse EP11 est proposée afin d'examiner si le nombre de modes distincts déclarés pendant l'entrevue influence le temps de réponse. À chaque nouveau mode choisi, les questions demandées sont différentes et de complexité variables, ce qui devrait, en ce sens, augmenter le temps de réponse global.

Le Tableau 4.4 et le Tableau 4.5 comprennent les statistiques descriptives pour chaque variable indépendante utilisée dans les modèles des enquêtes-ménage et personne.

Tableau 4.4 : Statistiques descriptives des variables indépendantes utilisées dans la modélisation de la durée d'entrevue des enquêtes-ménage QC11 et MTL13

Variable	Moyenne \bar{X}	Écart-type	Min	Max
Nombre de lieux visités uniques (entier)	5.041	2.905	0	17
Nombre de segments voiture	3.199	3.818	0	25
Nombre de segments transport collectif	2.760	3.479	0	28
Nombre de segments autre mode	1.687	2.587	0	15
Taille du ménage	2.291	1.123	1	7
Âge moyen 18-29 (référence)	0.226	0.419	0	1
Âge moyen 30-39	0.378	0.485	0	1
Âge moyen 40-49	0.190	0.392	0	1
Âge moyen 50-59	0.113	0.317	0	1
Âge moyen 60+	0.093	0.291	0	1
Entrevue débutée après 17h	0.327	0.470	0	1
Accès internet haute vitesse au domicile	0.861	0.346	0	1
Appareil à petit écran utilisé	0.017	0.129	0	1

Tableau 4.5 : Statistiques descriptives des variables indépendantes utilisées dans la modélisation de la durée d'entrevue des enquêtes-personne PY11, UM11, TR11, SH13, CM13 et BX13

Variable	Moyenne \bar{X}	Écart-type	Min	Max
Nombre de lieux visités uniques	3.611	1.956	0	19
Nombre de segments voiture	1.145	1.876	0	16
Nombre de segments transport collectif	1.679	1.950	0	15
Nombre de segments autre mode	1.036	1.614	0	14
Nombre de modes uniques	1.635	1.018	0	6
Taille du ménage	2.533	1.289	1	20
Nombre de questions ajoutées par l'organisation	19.226	18.532	0	42
Âge 18-29 (référence)	0.507	0.497	0	1
Âge 30-39	0.234	0.424	0	1
Âge 40-49	0.121	0.326	0	1
Âge 50-59	0.099	0.299	0	1
Âge 60+	0.039	0.195	0	1
Aucun déplacement déclaré (non-mobile)	0.092	0.289	0	1
Entrevue débutée après 17h	0.328	0.469	0	1
Accès internet haute vitesse au domicile	0.862	0.345	0	1
Appareil à petit écran utilisé	0.042	0.202	0	1
Sexe: femme	0.592	0.492	0	1
Intéressé à participer à d'autres études = oui (référence)	0.317	0.465	0	1
Intéressé à participer à d'autres études = non	0.222	0.416	0	1
Intéressé à participer à d'autres études = inconnu	0.461	0.498	0	1

Afin de construire le modèle, une régression linéaire multivariée a été appliquée et les variables indépendantes ont été ajoutées en ordre décroissant de la taille de l'effet prédit selon la littérature ou estimé (méthode hiérarchique³⁰ d'ajouts de variables dans le modèle). Comme dans (Gummer & Rossmann, 2014), le logarithme naturel de la durée d'entrevue a été privilégié dans le but d'atténuer l'asymétrie. En effet, le coefficient d'asymétrie de la variable non transformée pour les enquêtes-ménage est de 1.477, alors que son logarithme naturel présente un coefficient de -0.414 (Figure 4-1). Du côté des enquêtes-personne, le coefficient d'asymétrie passe de 1.683 à -0.092.

³⁰ Avec la méthode hiérarchique, l'ordre d'entrée des variables dans le modèle est déterminé à l'avance par le chercheur.

Diagramme Q-Q de la durée d'entrevue comparée à la distribution normale
 avec effet de la transformation logarithmique naturelle sur la distribution des durées d'entrevue
 Modélisation de la durée d'entrevue web • Enquête ménage

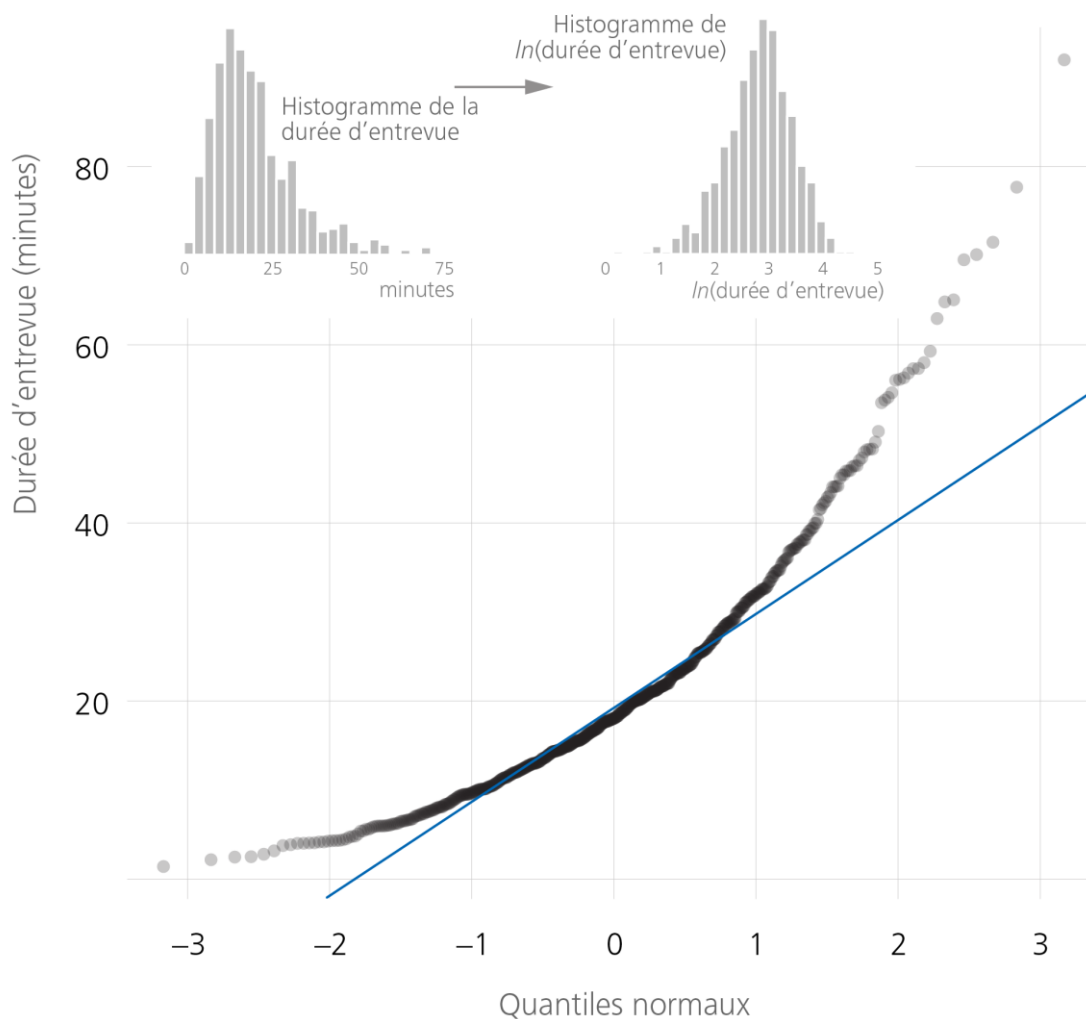


Figure 4-1 : Histogrammes de la durée d'entrevue et de son logarithme naturel et diagramme quantile-quantile de la durée d'entrevue comparée à une distribution normale pour les enquêtes-ménage

4.3 Résultats de la modélisation de la durée d'entrevue

4.3.1 Modèle de durée d'entrevue pour enquête-ménage

Le Tableau 4.6 résume les résultats et les statistiques du modèle de durée d'entrevue pour les enquêtes-ménage. À l'exception de la variable représentée dans l'hypothèse EM4, toutes les va-

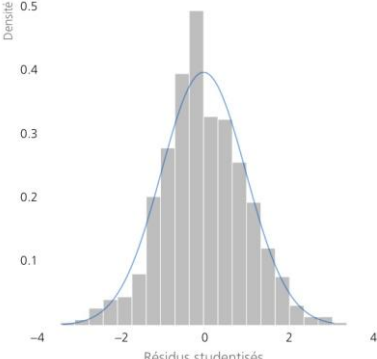
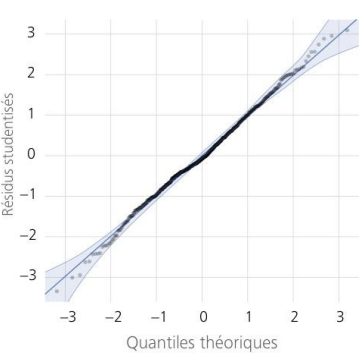
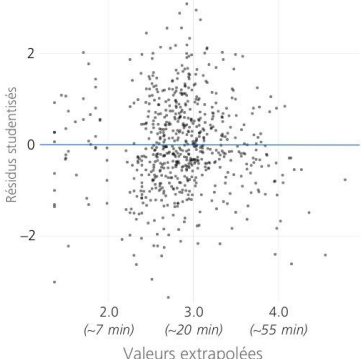
riables indépendantes sont significatives. De ce fait, seule l'hypothèse EM4 indiquant si l'entrevue a été débutée après 17h a été retirée du modèle. L'hypothèse EM3a est confirmée par opposition à EM3b. Conséquemment, les répondants plus âgés semblent avoir des durées d'entrevues plus longues de manière significative.

L'hypothèse d'absence de multicollinéarité est remplie puisque aucune variable indépendante ne présente un facteur VIF (variance inflation factor) plus grand que 5 (équivalente à une tolérance plus petite que 0.2). Les résidus sont distribués de manière aléatoire à l'exception des entrevues de plus de 55 minutes (logarithme naturel > 4.0) (Tableau 4.6, troisième figure). Conséquemment, l'utilisation du modèle pour les entrevues de très longue durée est à éviter.

En analysant les cas influents, il est possible de vérifier qu'il n'existe aucune entrevue pour laquelle la distance de *Cook* approche 1, ce qui confirme qu'il n'y a aucune durée d'entrevue qui influence de manière trop importante l'estimation des variables indépendantes. L'histogramme et le diagramme quantile-quantile des résidus indiquent que la normalité et la linéarité des résidus est acceptable.

Tableau 4.6 : Modèle de durée d'entrevue pour les enquêtes-ménage

Taille de l'échantillon <i>n</i>	Régression SS	Résidus SS	R^2 ajusté	Erreur standard des résidus	Test de Durbin-Watson
654	173.321 (10 df)	77.128 (643 df)	0.687	0.346	● 1.957

Dépendante: <i>ln</i> (durée d'entrevue)	Coef. B	Intervalles de confiance B		Erreur std. B	Coef. std. β	t	q-value*	Statistiques de collinéarité	
		2.5%	97.5%					Tolérance	VIF
<i>Constante</i>	2.124	2.015	2.232	0.055		38.597	< .0001		
EM1 Nombre de segments voiture	0.060	0.047	0.073	0.006	● 0.370	9.359	< .0001	● 0.306	● 3.267
EM1 Nombre de segments transport collectif	0.046	0.029	0.065	0.009	● 0.261	9.296	< .0001	● 0.608	● 1.644
EM1 Nombre de lieux visités uniques	0.047	0.037	0.056	0.005	● 0.219	5.085	< .0001	● 0.258	● 3.874
EM1 Nombre de segments ni voiture ni transport collectif	0.046	0.031	0.061	0.008	● 0.191	6.033	< .0001	● 0.478	● 2.091
EM2 Taille du ménage	0.082	0.053	0.111	0.015	● 0.148	5.546	.0013	● 0.669	● 1.495
<i>Variables dummy:</i>									
EM7 Nombre de déplacements = 0	-0.701	-0.826	-0.576	0.064	● -0.275	-11.001	< .0001	● 0.768	● 1.302
EM3a Âge: 60+	0.401	0.300	0.503	0.052	● 0.189	7.750	< .0001	● 0.809	● 1.236
EM3a Âge: 50-59	0.258	0.169	0.346	0.045	● 0.132	5.723	< .0001	● 0.901	● 1.110
EM6 Appareil à petit écran utilisé	0.504	0.296	0.712	0.106	● 0.105	4.766	< .0001	● 0.991	● 1.009
EM5 Internet haute vitesse au domicile	-0.133	-0.214	-0.053	0.041	● -0.075	-3.243	.0032	● 0.904	● 1.106

* le q-value est le p-value normalisé sur un échantillon de taille 100: $q = p \sqrt{\frac{n}{100}}$ (Woolley, 2003)

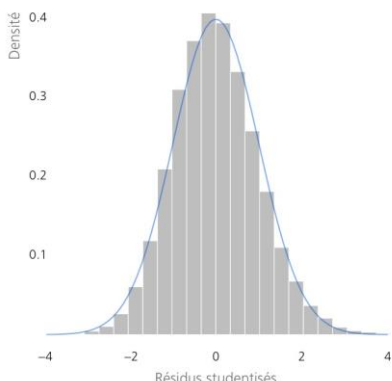
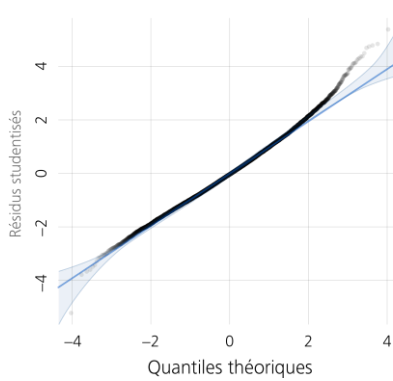
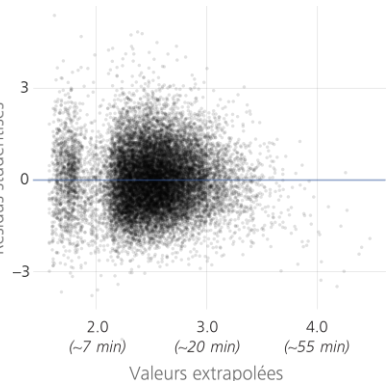
4.3.2 Modèle de durée d'entrevue pour enquête-personne

Le Tableau 4.7 montre les mêmes informations, mais cette fois pour les enquêtes-personne. Dans ce cas, toutes les variables indépendantes testées sont significatives et l'hypothèse EP3a est également vérifiée. Contrairement aux résultats obtenus lors des enquêtes de type CATI, la durée d'entrevue des femmes n'est pas plus courte, mais bien un peu plus longue en moyenne,

mais avec un petit effet. Les hypothèses d'absence de multicollinéarité sont également vérifiées, et aucune entrevue n'influence de manière trop importante les résultats. Les résidus présentent la même tendance que pour les enquêtes-ménage, mais cette fois, les entrevues de plus de 40 minutes (logarithme naturel > 3.7) présentent une certaine hétéroscédasticité. Les résidus ont une certaine déviation par rapport à la normalité et à la linéarité. En fait, lorsque seules les durées d'entrevue inférieures au 95^e centile sont incluses dans le modèle, le diagramme quantile-quantile des résidus ne présente aucune déviation significative par rapport à la normale.

Tableau 4.7 : Modèle de durée d'entrevue pour les enquêtes-personne

Taille de l'échantillon <i>n</i>	Régression SS	Résidus SS	<i>R</i> ² ajusté	Erreur standard des résidus	Test de Durbin-Watson
17545	2178.68 (16 df)	1783.67 (17528 df)	0.549	0.319	● 1.990

Dépendante: <i>ln</i> (durée d'entrevue)	Coef. B	Intervalles de confiance B		Erreur std. B	Coef. std. β	t	q-value*	Statistiques de collinéarité		
		2.5%	97.5%					Tolérance	VIF	
<i>Constant</i>	1.799	1.776	1.821	0.012		156.369	<.0001			
EP1 Nombre de segments voiture	0.071	0.067	0.076	0.002	● 0.282	32.067	<.0001	● 0.332	● 3.012	
EP 1 Nombre de lieux visités uniques	0.065	0.061	0.070	0.003	● 0.269	26.054	<.0001	● 0.240	● 4.165	
EP 1 Nombre de segments transport collectif	0.061	0.057	0.066	0.002	● 0.252	29.838	<.0001	● 0.359	● 2.784	
EP 1 Nombre de segments ni voiture ni transport collectif	0.054	0.049	0.059	0.002	● 0.184	21.867	<.0001	● 0.363	● 2.752	
EP 10 Nombre de questions spécifiques	0.003	0.002	0.003	< 0.001	● 0.101	15.890	<.0001	● 0.636	● 1.573	
EP 2 Taille du ménage	0.026	0.023	0.030	0.002	● 0.072	13.464	<.0001	● 0.908	● 1.101	
EP 11 Nombre de modes uniques	0.027	0.020	0.034	0.004	● 0.058	7.167	<.0001	● 0.397	● 2.521	
<i>Variables dummy:</i>										
EP6 Appareil à petit écran utilisé	0.401	0.378	0.425	0.012	● 0.170	33.188	<.0001	● 0.977	● 1.023	
EP3a Âge: 60+	0.349	0.324	0.374	0.013	● 0.143	27.472	<.0001	● 0.949	● 1.054	
EP3a Âge: 50-59	0.215	0.198	0.231	0.008	● 0.135	26.030	<.0001	● 0.953	● 1.049	
EP7 Nombre de déplacements = 0	-0.215	-0.236	-0.194	0.011	● -0.131	-19.776	<.0001	● 0.587	● 1.704	
EP3a Âge: 40-49	0.117	0.102	0.132	0.008	● 0.080	15.513	<.0001	● 0.958	● 1.044	
EP8 Voudriez-vous participer à d'autres études?	0.046	0.034	0.058	0.006	● 0.045	7.430	<.0001	● 0.707	● 1.414	
EP4 Début de l'entrevue >= 17h	0.039	0.029	0.049	0.005	● 0.038	7.502	<.0001	● 0.982	● 1.018	
EP9 Femme	0.034	0.024	0.044	0.005	● 0.035	6.834	<.0001	● 0.952	● 1.051	
EP5 Internet haute vitesse au domicile	-0.031	-0.045	-0.017	0.007	● -0.023	-4.422	.0001	● 0.975	● 1.025	

* le q-value est le p-value normalisé sur un échantillon de taille 100: $q = p \sqrt{\frac{n}{100}}$ (Woolley, 2003)

4.4 Optimisation des ressources

La modélisation présentée permet d'estimer les durées d'entrevue en fonction des caractéristiques des répondants, mais également à partir des réponses déjà déclarées pendant l'entrevue. Avec les nouveaux services d'optimisation de serveurs à la demande de type nuages, il est possible de modifier en temps réel la puissance de calcul du serveur et la mémoire vive utilisée en fonction des besoins. En utilisant de tels services, la connaissance de la durée d'entrevue moyenne prévue selon les informations déclarées jusqu'à maintenant permettra de n'utiliser que les ressources nécessaires, ce qui aura pour effet de réduire les coûts d'hébergement des serveurs, qui sont souvent facturés à l'heure ou à la minute. De plus, lorsque les répondants potentiels sont recrutés par courriel, les envois pourront être mieux répartis dans la journée afin d'éviter les engorgements et de limiter le nombre d'entrevues simultanées.

4.5 Suivi des entrevues en temps réel par les administrateurs

Grâce au modèle de durée d'entrevue, il sera possible de montrer, dans le module d'administration, les entrevues qui s'éloignent de façon significative de la durée prédite par le modèle. L'administrateur pourra alors vérifier les journaux du serveur (logs) associés aux entrevues dont la durée est anormale pour détecter les problèmes techniques.

4.6 Distribution temporelle des entrevues

Dans un souci d'optimisation des ressources matérielles et logicielles utilisées lors de la tenue d'enquêtes en ligne sur la mobilité, la distribution temporelle des entrevues permet de mesurer l'achalandage sur le serveur pendant une journée moyenne. La Figure 4-2 et la Figure 4-3 illustrent ces distributions temporelles pour les enquêtes-ménage et les enquêtes-personne. En premier lieu, il est possible d'observer que la plupart des entrevues débutent durant les heures conventionnelles de travail et d'étude, ce qui suggère qu'une majorité de répondants remplissent le questionnaire en dehors du domicile. Un grand nombre d'entrevues sont tout de même réalisées tard le soir, en dehors des plages normales d'appels des enquêtes téléphoniques conventionnelles. En ce sens, la plateforme web permet davantage de flexibilité pour les répondants en

termes de disponibilité. Une analyse plus approfondie serait à envisager afin de préciser la disponibilité des répondants en fonction de leur sexe, de leur occupation, de leur âge ou de leurs comportements de mobilité.

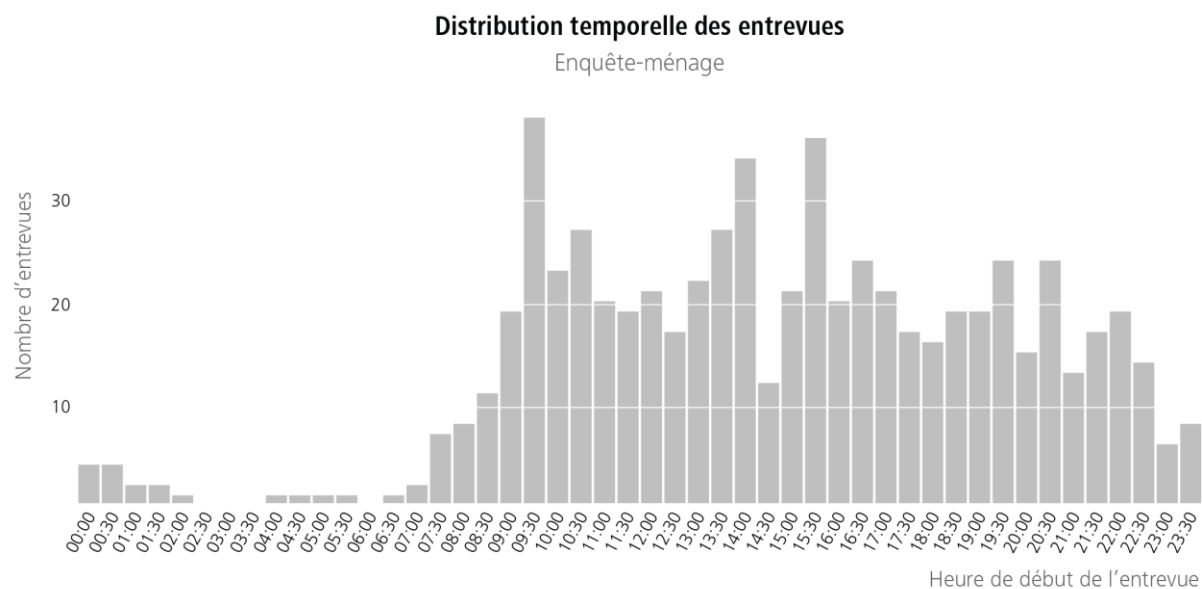


Figure 4-2 : Distribution temporelle des entrevues pour les enquêtes-ménage

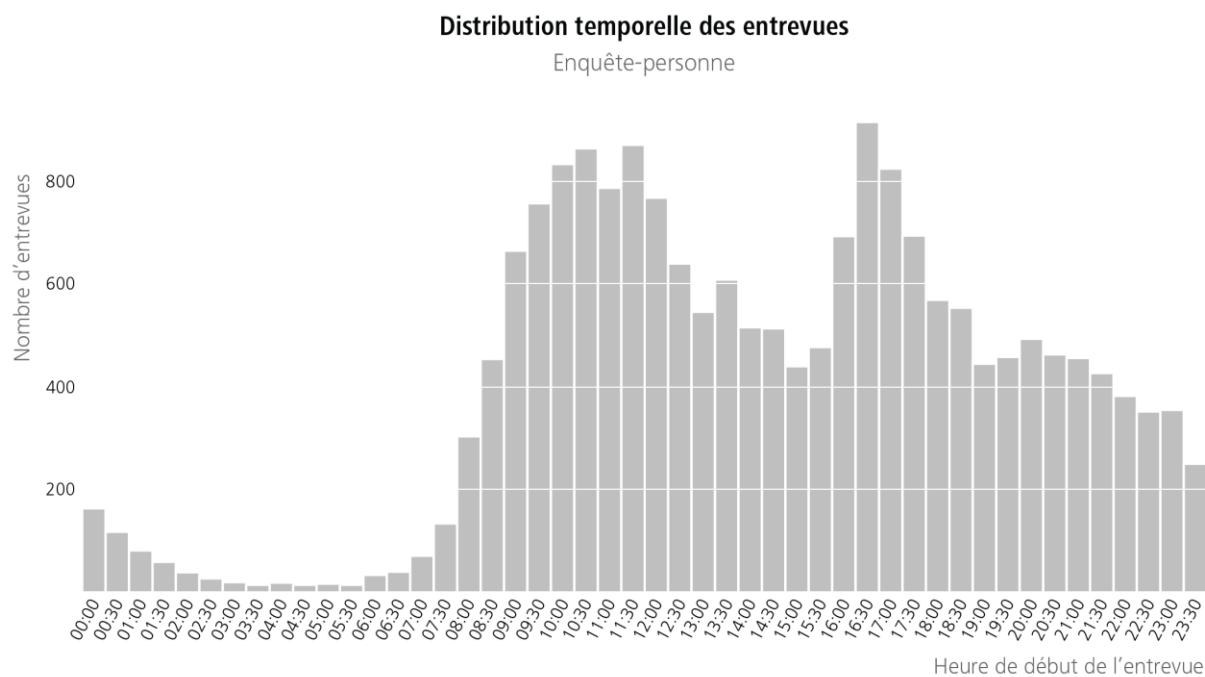


Figure 4-3 : Distribution temporelle des entrevues pour les enquêtes-personne

CHAPITRE 5 ENTREPÔT DE DONNÉES DE MOBILITÉ

Afin de mettre en contexte et de documenter la création d'un entrepôt de données de mobilité, le processus de modélisation proposé sera d'abord énoncé. Les attributs assurant la comparabilité des données d'une enquête à l'autre seront définis et le processus d'enrichissement des données sera décrit. La méthode de traitement et de validation utilisée pour effectuer la conversion des données de collecte vers l'entrepôt de données sera précisée. Par la suite, les caractéristiques des différentes tables présentes dans l'entrepôt seront détaillées et l'interface d'exportation de fichiers pour fins d'analyse sera décrite.

5.1 Modélisation dimensionnelle des données de mobilité

À l'origine de la création d'un entrepôt de données, la modélisation dimensionnelle permet de répondre adéquatement et efficacement à deux besoins fondamentaux : produire des données qui sont compréhensibles pour l'utilisateur et répondre aux requêtes de manière performante (Kimball & Ross, 2013). La première étape afin de construire un modèle dimensionnel — aussi appelé schéma en étoile — consiste à déterminer les objets élémentaires, c'est-à-dire les niveaux de granularité du système. Ce sont ces éléments qui se retrouveront dans les tables de faits. Dans l'étude de la mobilité des personnes, on retrouve les objets suivants, du plus grossier au plus fin : le ménage, la personne, le séjour, la chaîne de déplacements, le déplacement, le lieu visité (au même niveau que le déplacement puisqu'un déplacement compte deux lieux visités : l'origine et la destination) et finalement le segment (un segment de déplacement représente la partie d'un déplacement utilisant un mode de transport unique).

Lors de l'étude de la mobilité des personnes, l'objet élémentaire au centre de l'analyse n'est pas toujours le segment de déplacement. En fait, certaines analyses ne nécessitent pas d'information aussi précise, et en ce sens, peuvent se contenter de données de granularité plus élevée (le ménage, la personne, le séjour, la chaîne de déplacements, le déplacement ou le lieu visité). Comme mentionné dans (Kimball & Ross, 2013), la classification des différents objets d'un système en faits ou en dimensions dépend du contexte. Il est même possible de retrouver des objets qui sont à la fois faits et dimensions. Dans le cas de l'étude de la mobilité, la classification dépend des besoins de l'utilisateur. Par exemple, un analyste désirant étudier la démographie des personnes

sans égard aux déplacements précis effectués par chacun, pourra bénéficier d'un modèle intégrant les personnes comme table de faits. Un autre, intéressé par les horaires, les durées et les distances de déplacements, mais pas par les modes précis empruntés, favorisera un modèle dans lequel les déplacements sont les faits. Conséquemment, dans un souci de flexibilité auprès des utilisateurs de l'entrepôt de données, les objets principaux du système seront tous représentés dans des tables de faits. De cette façon, l'analyste choisira simplement le niveau de granularité désiré et effectuera ses requêtes de la manière la plus performante et la plus efficace en considération de ses besoins. Le désavantage associé à ce choix est la plus grande quantité d'espace disque requise pour stocker les données. Les tables constituant l'entrepôt de données sont présentées à la Figure B-1 de l'Annexe B. Chaque table de faits comprend l'ensemble des clés faisant référence à ses dimensions associées, de même que des champs booléens (suffixe Flag) et une clé naturelle (suffixe NaturalKey) permettant de retrouver l'objet original dans la base de données de collecte (l'association des tables de l'entrepôt de données avec les tables de la base de données de collecte est définie dans le module ETL). La modélisation dimensionnelle ne prévoit qu'un seul degré de séparation entre les tables de faits et de dimensions. Toutefois, il existe une exception à cette règle puisque les dimensions de dates et d'heures peuvent être référencées au moyen de clés dans les tables de dimensions, ce qui entraîne un second degré de séparation. Ce relâchement des règles pour ce cas particulier permet de simplifier le modèle et est spécifié dans la documentation de Kimball. Enfin, un identifiant universel de type Universally Unique Identifier respectant les standards RFC 4122³¹ et ISO/IEC 9834-8:2005³² a été prévu dans chacune des tables afin de pouvoir intégrer, au besoin, les données de multiples enquêtes dans une même

³¹ <https://www.ietf.org/rfc/rfc4122.txt>

³² http://www.iso.org/iso/catalogue_detail.htm?csnumber=36775

base de données et assurer l'unicité de chaque enregistrement³³. Cet attribut est composé du nom du fait ou de la dimension suivi de l'acronyme UUID (Figure B-1).

5.2 Attributs comparables

Un des objectifs principaux de la création d'un entrepôt de données de mobilité est de faciliter la comparaison des données d'une enquête à l'autre. Pour ce faire, un ensemble d'attributs comparables sont proposés. La première caractéristique de ces attributs est leur durabilité, c'est-à-dire qu'ils sont conçus pour ne pas être modifiés ou, à tout le moins, très rarement. En effet, si une modification à l'un des attributs comparables est prévue, une mise à jour de toutes les enquêtes sauvegardées dans l'entrepôt de données doit être effectuée pour conserver la comparabilité des données dans le temps. La deuxième caractéristique est l'inclusion de toutes les possibilités pour permettre des comparaisons rigoureuses. Par exemple, un attribut de mode comparable devra inclure dans ses choix tous les modes possibles, non seulement sur le plan local, mais également sur le plan international. Pour la même raison, un soin particulier doit être apporté à la définition et aux choix proposés pour chaque attribut comparable afin d'éviter les ambiguïtés. Dans le cadre de la thèse, des attributs et dimensions comparables ont été proposés. Par contre, le choix définitif des attributs et dimensions comparables à intégrer dans l'entrepôt de données devra être le résultat d'un consensus éventuel entre les intervenants du milieu des enquêtes de mobilité. Dans la Figure B-1, les tables de dimensions comparables proposées ont été rassemblées pour simplifier le schéma et parce qu'elles contiennent les mêmes attributs (seul le préfixe des attributs changent en fonction du nom de la dimension).

³³ Un identifiant unique universel (UUID) permet à des systèmes distincts d'identifier l'information de manière unique sans contrôle centralisé. En pratique, les probabilités de conflit d'identifiants uniques universels sont très faibles. Par exemple, si on générait 1 milliard d'UUID standardisés par seconde pour les 100 prochaines années, la probabilité d'obtenir une seule clé dupliquée serait d'environ 50%. La probabilité réelle est toutefois plus élevée du fait que la plupart des générateurs d'UUID ne sont pas totalement aléatoires.

(https://en.wikipedia.org/wiki/Universally_unique_identifier)

5.2.1 Procédure de correspondance des attributs comparables de la base de données de collecte vers l'entrepôt de données

Lorsqu'une enquête utilise des attributs comparables dans son entrepôt de données, ces attributs, ainsi que les choix de réponse associés, doivent être définis au préalable dans la base de données de collecte. Par exemple, les activités comparables définies dans l'entrepôt de données doivent être associées aux activités spécifiées dans la table `od_question_choices` de la base de données de collecte au moyen de la clé nommée `ComparableActivityNaturalKey`, qui correspond à l'identifiant de l'activité comparable dans la table `od_question_choices` de la base de données de collecte. La clé `durable_comparable_activity_id` de la table `od_activities` permet alors de spécifier quelle activité comparable est associée à chaque activité définie dans le questionnaire. Dans la Figure 5-1, l'activité « Services » est prévue dans le questionnaire, mais ne se retrouve pas dans la liste des activités comparables. Les activités « Magasinage » et « Services » ont donc la même correspondance d'attribut vers l'activité comparable « Magasinage » de la table `od_question_choices` et de la table `dw_od_comparable_activity_dimensions` de l'entrepôt de données.

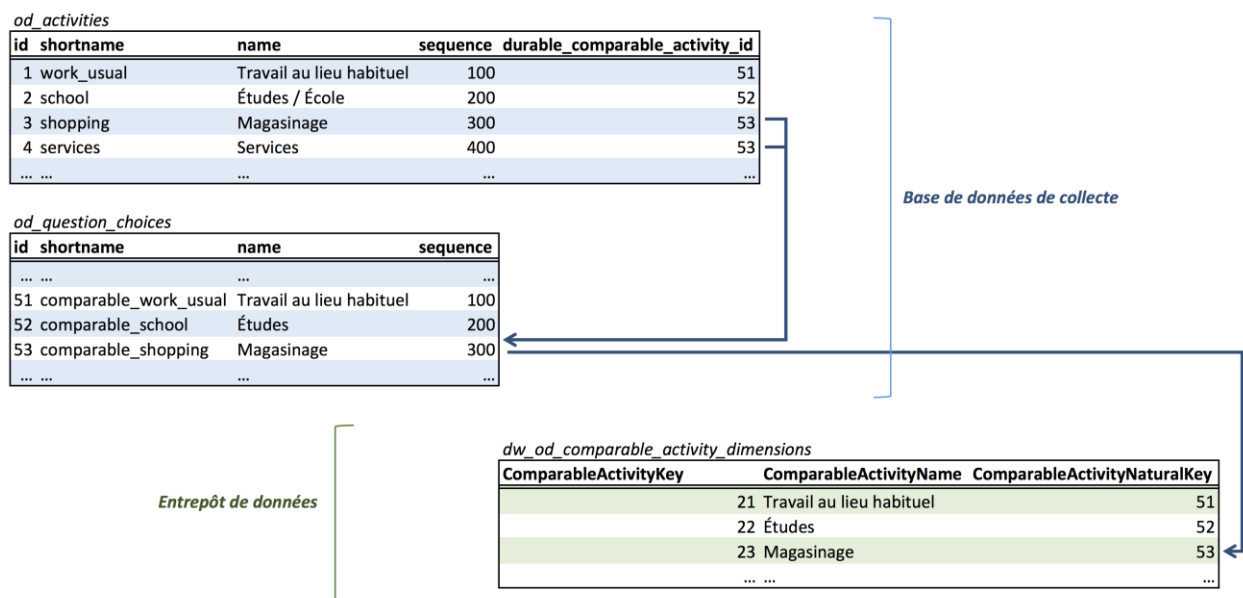


Figure 5-1 : Correspondance des attributs comparables dans la base de données de collecte et dans l'entrepôt de données (Exemple pour les activités)

5.3 Enrichissement des données

La création d'un entrepôt de données s'accompagne d'un processus d'enrichissement des données. Dans cette proposition de schéma dimensionnel pour des données de mobilité des personnes, plusieurs champs sont pré-calculés pendant l'importation dans l'entrepôt de données et un certain nombre de nouvelles informations sont ajoutées afin de faciliter et d'enrichir les analyses. À titre d'exemple, dans la dimension ménage, plusieurs statistiques descriptives sont ajoutées : l'âge moyen du ménage, le nombre de femmes, le nombre d'hommes, le nombre d'étudiants, de travailleurs et de retraités, le nombre de membres de chaque cohorte d'âge, le nombre de déplacements moyens et totaux, les distances moyennes et totales parcourues par l'ensemble des membres du ménage, le nombre minimal de véhicules nécessaires dans le ménage pour effectuer les déplacements en mode auto-conducteur. Aussi, à chaque fois qu'une distance est calculée ou mesurée, trois colonnes sont créées : une première avec la distance en mètres, une deuxième en kilomètres et une troisième en miles. Cela permet à l'entrepôt de données d'être compréhensible dans n'importe quel pays. Pour les attributs de durées, les résultats sont disponibles en secondes et en minutes. Finalement, les données géographiques sont toujours disponibles

autant en format Latitude/Longitude (WGS84) que dans le système de coordonnées cartésiennes³⁴ déterminé à la configuration de l'enquête.

5.4 Validation et vérifications des données • ETL

Lors de la création d'un entrepôt de données, le processus de traitement, de nettoyage et d'importation des données est appelé ETL (Extract, Transform, Load) (Adamson, 2010; Kimball & Ross, 2013). Le module ETL doit être mis à jour lorsque de nouvelles sources de données apparaissent ou lorsque les formats de fichiers changent. Dans le cadre du projet de recherche associé à la thèse, deux sources de données distinctes ont été utilisées pour construire la première version du module ETL : les données provenant du questionnaire web proposé au Chapitre 3 ainsi que les données contenues dans les fichiers de résultat d'enquêtes téléphoniques tenues tous les cinq ans dans la région de Québec depuis 1996. Les données sont validées et traitées pour être enrichies et importées dans l'entrepôt de données après qu'un validateur ait corrigé les erreurs demandant un traitement manuel selon la table d'audits. Un fichier de configuration (5.4.1) permet d'indiquer au module de validation les champs requis et les conditions de validité des différentes informations déclarées par les répondants. Lorsqu'une des règles n'est pas respectée, un enregistrement est ajouté dans la table d'audits de la base de données de collecte. Un ensemble de règles immuables est toutefois proposé dans une optique d'uniformisation. Par exemple, tous les lieux visités par les répondants ainsi que leur domicile doivent être géolocalisés, que ce soit par déclaration par le répondant ou par imputation manuelle lorsque nécessaire. D'autre part, l'âge ou le groupe d'âge et le sexe de chacun des répondants doivent être déclarés ou imputés. Enfin, pour les enquêtes ménage, le nombre de personnes déclarées doit concorder avec la taille du ménage déclarée. Toutes les validations supplémentaires sont flexibles et peuvent changer d'une enquête à l'autre selon les besoins des administrateurs et analystes.

³⁴ La projection vers un système de coordonnées cartésiennes permet de convertir les coordonnées géodésiques sur la surface de la terre en coordonnées planaires afin de faciliter les analyses spatiales et les calculs faisant appel à des données spatiales. Il existe plusieurs systèmes de projection, la plupart étant définis pour être plus précis dans certaines régions du monde.

5.4.1 Configuration des validations avant importation dans l'entrepôt de données

Deux types de validations sont effectués, la première concerne les validations absolument requises pour que les données puissent être transférées vers l'entrepôt de données (required). La deuxième comprend les validations facultatives (auditable). Ces dernières s'accompagnent d'un audit lorsqu'elles ne sont pas respectées, mais elles ne bloquent pas l'importation dans l'entrepôt. Dans le fichier de configuration, les validations sont séparées par type d'objet (ménage : `od_household_interview`, personne : `od_user_interview`, déplacement : `od_trip`, etc.). Lorsque l'attribut est simplement listé, cela veut dire que l'attribut est requis, mais qu'aucune vérification de la valeur déclarée n'est requise. Si une condition est spécifiée, le système vérifiera cette condition et inclura dans l'audit le message indiqué dans la description associée si elle échoue. Le langage Ruby est utilisé pour les conditions puisque celles-ci sont évaluées par le serveur qui fonctionne sous la plateforme Ruby On Rails. Un extrait de fichier de configuration de validation pour l'entreposage est présenté ici :

```
required:
  od_household_interview:
    - home
    - size
    - car_number
    - home:
      - geography
    - condition: "{object}.size.nil? || {object}.od_user_interviews.size
      ↳ == 0 || {object}.size != {object}.od_user_interviews.size"
      description: "Le nombre de personnes déclarées n'est pas égal à la
      ↳ taille du ménage"
    - condition: "{object}.od_user_interviews.count{ |ui| ui.min_age >=
      ↳ 15 } == 0"
      description: "Aucun membre du ménage de 15 ans ou plus"
  od_user_interview:
    - age_group
    - gender
    - occupation
    - driving_license_owner
    - transit_pass_owner
    - condition: "{object}.driving_license_owner && {object}.min_age <
      ↳ 16"
      description: "La personne a moins de 16 ans, mais un permis de
      ↳ conduire est déclaré"
    - condition: "{object}.transit_pass_owner && {object}.min_age < 6"
      description: "La personne a moins de 6 ans, mais un abonnement de
      ↳ transport collectif est déclaré"
    - condition: "{object}.no_trip.nil?"
```

```

description: "La mobilité de la personne est inconnue"
- condition: "{object}.no_trip == true && {object}.trips_number >
  ↳ 0"
description: "La personne est non mobile, mais des déplacements
  ↳ ont été déclarés"
- condition: "{object}.no_trip == false && {object}.trips_number ==
  ↳ 0"
description: "La personne est mobile, mais aucun déplacement n'a
  ↳ été déclaré"
...

auditable:
  od_household_interview:
    - condition: "{object}.car_number &&
      ↳ {object}.household_members_drivers.any? && {object}.car_number /
      ↳ {object}.household_members_drivers.size.to_f > 1"
    description: "Il y a plus d'un véhicule déclaré par membre du ménage
      ↳ possédant un permis de conduire"
    - condition: "{object}.building_type_id.nil?"
    description: "Le type de bâtiment est manquant"
...

```

5.4.2 Audits

Grâce aux audits, toutes les erreurs rencontrées sont archivées et associées à l'objet concerné. Un extrait de la table d'audits est présenté au Tableau 5.1. Le processus de correction des erreurs est présenté à la section 5.10.

Tableau 5.1 : Extrait de la table d'audits

Audit ID	Objet ID	Type d'objet	Description de l'audit
7655	129	OdSegment	Le mode est Auto conducteur, mais le type de stationnement à la destination est manquant
7678	21	OdHouseholdInterview	L'heure de fin de l'entrevue est manquante
7679	93	OdPlace	La géographie du lieu est manquante
7684	151	OdSegment	Le mode est Auto conducteur, mais le nombre d'occupants dans le véhicule est manquant
7734	62	OdUserInterview	La possession d'un permis de conduire est manquante pour une personne en âge de conduire
7791	37	OdHouseholdInterview	Le ménage possède plus d'un véhicule par membre possédant un permis de conduire

5.5 Cohérence des données dans l'entrepôt de données

Le fait de dénormaliser les données de la base de données de collecte et de constituer une base de données (l'entrepôt de données) non normalisée s'accompagne d'une problématique importante sur le plan de la cohérence des données. Par exemple, si un utilisateur modifie le nom d'une activité, celle-ci doit être modifiée dans la dimension activité et dans les tables qui comprennent des noms d'activités (lieux visités et chaînes de déplacements, entre autres). Or, dans un modèle dimensionnel, c'est la responsabilité du module ETL de vérifier et de respecter la cohérence des données en s'assurant de modifier les valeurs dans toutes les tables affectées. En ce sens, l'entrepôt de données doit être utilisé en lecture seule par l'ensemble des utilisateurs. En fait, le module ETL doit être le seul à posséder des droits d'écriture sur l'ensemble des tables. En ce sens, il est essentiel de limiter l'accès au module ETL aux seuls administrateurs qualifiés afin de limiter les risques d'erreurs et de conserver la comparabilité des attributs d'une enquête à l'autre.

5.6 Les tables de faits et leurs dimensions

Dans cette section, les dimensions d'enquête, d'échantillon, de dates et d'heures ainsi que les dimensions de calcul de chemin seront définies. Ensuite, chacune des tables de faits (ménage, personne, lieu visité, séjour, chaîne de déplacements, déplacement et segment) sera présentée à l'aide d'un schéma et décrite par ses attributs principaux.

5.6.1 Les dimensions enquête et échantillon

Deux dimensions sont associées à toutes les tables de faits. En premier lieu, on retrouve la dimension enquête, qui énumère les caractéristiques de l'enquête pour laquelle ont été traités les différents objets de mobilité. Ensuite, on retrouve la dimension décrivant l'échantillon associé à chacun des ménages et individus répondants à l'enquête. Cette dimension comprend des attributs sur le mode de recrutement et de collecte, sur sa taille, ainsi que sur les périodes de recrutement.

5.6.2 Les dimensions de dates

Dans les dimensions de dates, on ajoute plusieurs différents attributs permettant de faire des tris particuliers associés aux dates sans avoir à effectuer de filtres sur plusieurs champs distincts. Entre autres, on y retrouve le nom du jour, le numéro de la date dans la semaine (1 à 7), dans le mois (valeurs possibles de 1 à 31) et dans l'année (valeurs possibles de 1 à 366) ainsi que le numéro de semaine dans le mois (1 à 4 ou 5) et l'année (1 à 52 ou 53). Un autre champ indique si la date fait partie de la semaine (lundi au vendredi) ou de la fin de semaine (samedi ou dimanche). Enfin, l'année, le mois et le jour sont également fournis séparément dans leurs champs respectifs, au long ou en chiffre. En utilisant cette dimension, il est par exemple possible d'effectuer une requête qui groupe les déplacements en fonction du type de jour de semaine (semaine ou fin de semaine) sans avoir à grouper d'abord par jour de la semaine (lundi au dimanche), pour ensuite scinder entre le vendredi et le samedi. Cela permet de simplifier les requêtes et les groupements, surtout lorsque plusieurs autres tris ou filtres sont ajoutés à la requête. La dimension de date sert également à standardiser toutes les dates utilisées dans l'entrepôt de données, ce qui assure l'intégrité et la comparabilité des données. Dans la table de dates, il n'existe qu'un seul enregistrement par date unique. En ce sens, les dimensions et les faits qui référencent la même date possèdent la même clé de date.

5.6.3 Les dimensions d'heures et d'heures étendues

Les dimensions d'heures permettent de faciliter les groupements ou les tris relatifs aux heures ou aux périodes de la journée. Une table de dimension est créée pour les 24 heures de la journée (0 à 23), alors qu'une deuxième comprend l'ensemble des 1440 minutes (0 à 1439) d'une journée. Dans ces tables, on retrouve des attributs permettant de représenter l'heure, avec ou sans les minutes, selon plusieurs formats. Par exemple, des champs distincts sont prévus pour l'heure en format international 24 heures et en format AM/PM sur 12 heures. Une autre colonne comprend le nombre de minutes depuis minuit, facilitant les analyses séquentielles qui ne nécessitent plus de conversion ou de fonctions spécifiques pour mesurer les intervalles de temps. Les dimensions d'heures étendues sont semblables, à la différence qu'elles comprennent les heures de 0 à 47 heures (0 à 2879 minutes). L'ajout de ces dimensions d'heures étendues assure une meilleure

cohérence lors de l'analyse de déplacements de nuit puisque le format préconisé par la spécification GTFS (General Transit Feed Specification) pour les horaires de transport collectif peut dépasser 24 heures (un voyage de nuit peut par exemple débuter à 24 :20 et se terminer à 26 :50, ce qui équivaut à 00 :20 et 02 :50 du matin le lendemain).

5.6.4 Les dimensions d'objets géographiques

Grâce à l'interface d'importation, les administrateurs d'une enquête peuvent téléverser des fichiers de type shapefile (fichiers couramment utilisés pour le stockage de données géographiques), et, de cette façon, inclure des données géographiques comme des points, des lignes ou des polygones correspondant à une catégorie en particulier. Par exemple, il est possible d'importer les découpages géographiques associés à un recensement de population qui inclut des attributs socio-démographiques utiles aux regroupements des résultats de l'enquête. Une fois le fichier importé, les objets qui en font partie et leur catégorie seront ajoutés à la base de données de collecte (table `od_geographic_features` et `od_geographic_feature_categories`) et insérés dans la table de dimension d'objets géographiques de l'entrepôt de données. Lors du traitement vers l'entrepôt de données, chacun des enregistrements possédant un attribut de type géographique (domicile, lieu habituel de travail ou d'études, lieu visité, etc.) sera relié à tous les objets géographiques de type polygones qui lui sont associés (au moyen d'une intersection spatiale vérifiant l'inclusion dans les polygones de chaque catégorie). Une colonne distincte est créée pour chaque catégorie d'objets géographiques importée. Les attributs des objets géographiques seront ainsi disponibles pour toutes les requêtes faites sur les tables de faits associées. Ainsi, si l'on importe les secteurs de recensement du territoire couvert par l'enquête, incluant la population et la taille de chaque cohorte d'âge par exemple, chaque domicile et chaque lieu déclaré par les répondants possèdera une dimension « secteur de recensement » qui comprendra la description du secteur, ses attributs et sa géographie.

5.6.5 Les dimensions de calculs de chemin

Les dimensions de calculs de chemins sont proposées dans le but de recueillir l'information provenant d'un calcul de chemin pour un déplacement en particulier. Le processus d'enregistrement fait appel à un outil de calcul spécifique associé à chacun des modes de transport principaux

(marche, vélo, voiture, transport collectif, Park & Ride utilisant la voiture et le transport collectif, Bike & Ride utilisant le vélo et le transport collectif). Conséquemment, un groupe de colonnes est prévu pour chaque mode et permet de récupérer les résultats au sein d'un même enregistrement dans le but de faciliter les comparaisons entre différents modes pour un même déplacement, sans se limiter au mode déclaré par le répondant, et alimenter l'estimation de modèles de choix modal. Pour chaque mode calculé, la position géographique de l'origine et de la destination, le trajet à vol d'oiseau, la géographie du parcours réel sur le réseau, l'heure de départ, l'heure d'arrivée, le temps de parcours, la distance à vol d'oiseau ainsi que la distance parcourue sur le réseau sont sauvegardés. De plus, pour les calculs de transport collectif, la position, la description et l'heure d'embarquement et de débarquement de tous les lieux d'accès et de transfert ainsi que la description de chacune des lignes utilisées et des modes empruntés (bus, tramway, métro, train, etc.) sont également enregistrés. De surcroît, lorsque des valeurs de vitesse moyenne de marche et de vélo sont fournies pour différentes cohortes d'âge et pour chaque sexe, un attribut comprend les valeurs ajustées des temps de parcours à pied et à vélo pour tenir compte des caractéristiques démographiques de la personne pour laquelle le calcul de chemin est effectué (ces valeurs ajustées de vitesses sont configurées par l'administrateur en fonction des particularités locales ou selon des moyennes proposées dans la littérature). Les calculateurs de chemin utilisés par la plateforme sont présentés à l'Annexe C. Il est important de mentionner que les résultats obtenus par ces calculateurs sont les chemins les plus courts en durée et ne représentent pas nécessairement le chemin exact emprunté par le répondant, qui n'est pas connu. Cet enrichissement sert à estimer les temps de parcours et la géographie des trajets les plus rapides sur le réseau local.

5.6.6 Le ménage

Pour chaque ménage, une dimension représentant le domicile (localisation géographique, adresse déclarée, adresse corrigée, etc.) est associée à l'enregistrement dans la table de faits `dw_od_household_facts` (Figure 5-2). La date de déplacements enquêtée dans le questionnaire Origine-Destination est également précisée dans sa propre dimension de date. Lorsque des objets géographiques sont ajoutées au moyen de l'interface d'importation de fichiers shapefile, une

dimension d'objet géographique (non incluse par défaut) s'ajoute pour chaque catégorie importée (par exemple : une dimension pour le secteur de recensement, une seconde pour l'arrondissement, etc.). Chaque dimension géographique utilise la localisation du domicile pour effectuer une intersection spatiale et associer l'objet retourné au ménage. La dimension ménage, quant à elle, comprend ses caractéristiques (taille, motorisation, type de bâtiment, nombre de membres de chaque cohorte, de chaque sexe et de chaque occupation principale, type de ligne téléphonique résidentielle et d'accès internet, etc.).

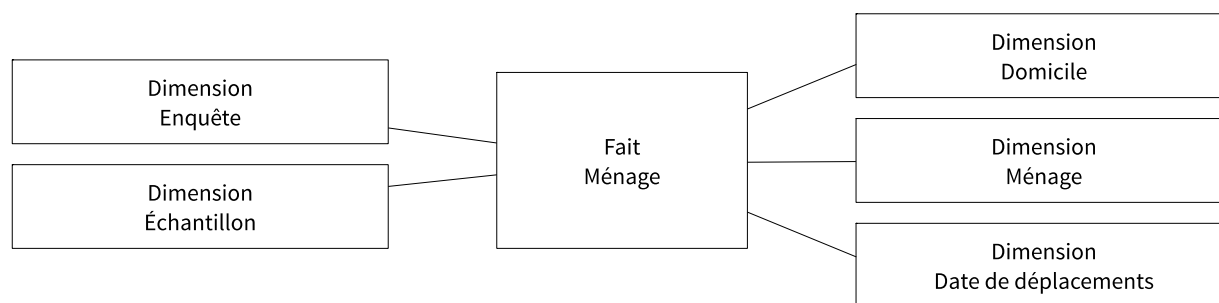


Figure 5-2 : Schéma dimensionnel • Ménage

5.6.7 La personne

La table de fait personne (`dw_od_person_facts`) comprend l'ensemble des répondants et des membres des ménages interrogés valides (Figure 5-3). À chaque personne sont associées une dimension représentant le domicile (identique à celle associée aux ménages), une dimension ménage, deux dimensions sur les lieux habituels (une pour le lieu habituel de travail et l'autre pour le lieu habituel d'étude) ainsi qu'une dimension personne qui inclut les caractéristiques de la personne (âge, groupe d'âge, sexe, occupation principale, possession de permis de conduire, possession d'abonnement de transport collectif, langue d'entrevue, etc.). Cette dimension personne compte également des attributs comparables (groupe d'âge comparable, sexe comparable, occupation comparable, etc.) ainsi que des statistiques descriptives pré-calculées sur les déplacements effectués (distance totale et moyenne, durée totale et moyenne par déplacement, nombre de déplacements effectués pour chaque mode de transport, etc.). Des dimensions de calcul de chemin sont également incluses pour le trajet entre le domicile et le lieu habituel de

travail et/ou d'étude ainsi que le trajet entre le lieu habituel de travail/d'étude et le domicile (déplacement inverse, qui peut être différent en fonction des particularités du réseau de transport local et du mode calculé). Enfin, la dimension personne du répondant dans l'enquête (aussi appelé proxy lorsque la personne ne répond pas elle-même à sa propre entrevue) est également associée pour des besoins d'analyse des comportements de réponse.

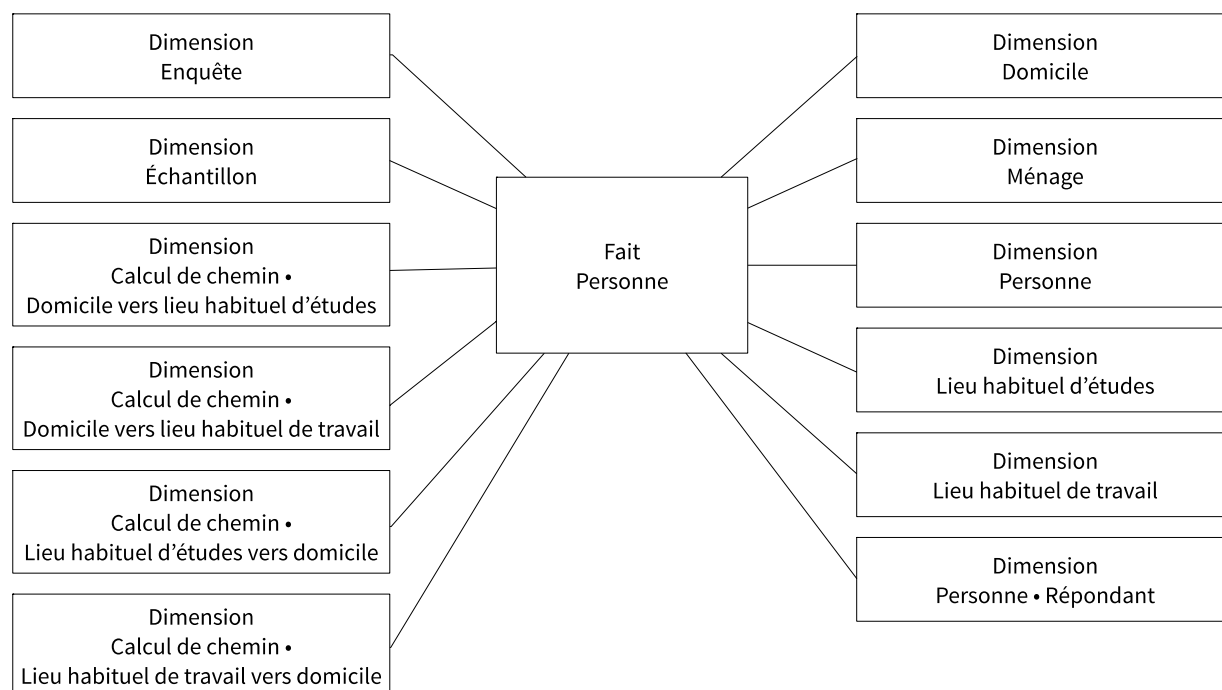


Figure 5-3 : Schéma dimensionnel • Personne

5.6.8 Le lieu visité

Suivant la même logique que la table de fait personne, la table de fait lieu visité (`dw_od_visited_place_facts`) comprend une dimension domicile, ménage, personne ainsi que les dimensions des lieux habituels d'étude et de travail de la personne ayant visité le lieu³⁵ (Figure 5-5). De plus, on retrouve une dimension décrivant l'activité effectuée au lieu visité (version spécifique à l'enquête, accompagnée d'une version comparable permettant la comparabilité de l'enquête avec

³⁵ Si plusieurs répondants visitent le même lieu et y effectuent la même activité, le lieu sera multiplié par le nombre de personnes et relié à chaque dimension personne concernée.

d'autres du même type), de même que les dimensions de dates et heures et heures étendues d'arrivée et de départ. La dimension lieu visité, quant à elle, présente des attributs relatifs tels que la localisation géographique, l'adresse déclarée et l'adresse corrigée lorsqu'applicable.

5.6.9 Le séjour

La table de fait séjour (dw_od_journey_facts) permet de représenter les déplacements de longue durée ou les voyages générant plusieurs chaînes de déplacements (Figure 5-4). Cette table n'est pas utilisée dans le cadre d'une enquête Origine-Destination conventionnelle basée sur un jour unique de déplacements. Elle compte en premier lieu des références vers les dimensions séjour, domicile, ménage, personne et lieux habituels de la personne ayant effectué le séjour. Ensuite, on y retrouve des liens vers les dimensions du premier et du dernier lieu visité du séjour (points de départ et d'arrivée). Finalement, les dimensions de dates et heures (format 24h et étendu 48h) du départ et de l'arrivée sont incluses.

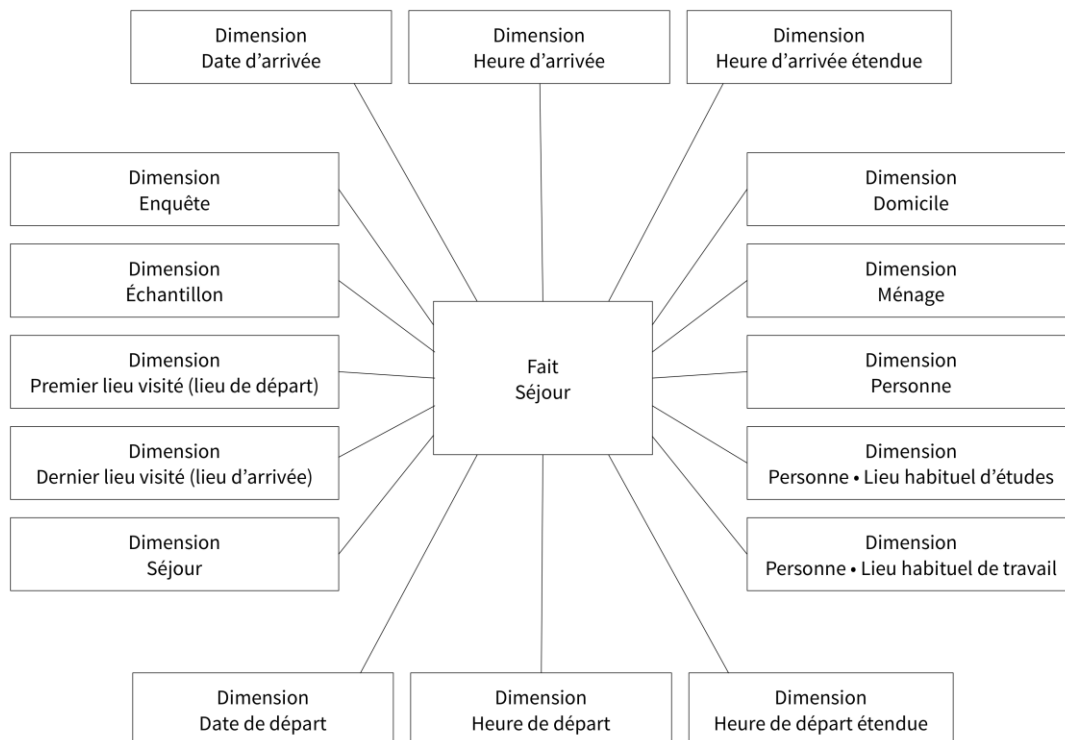


Figure 5-4 : Schéma dimensionnel • Séjour

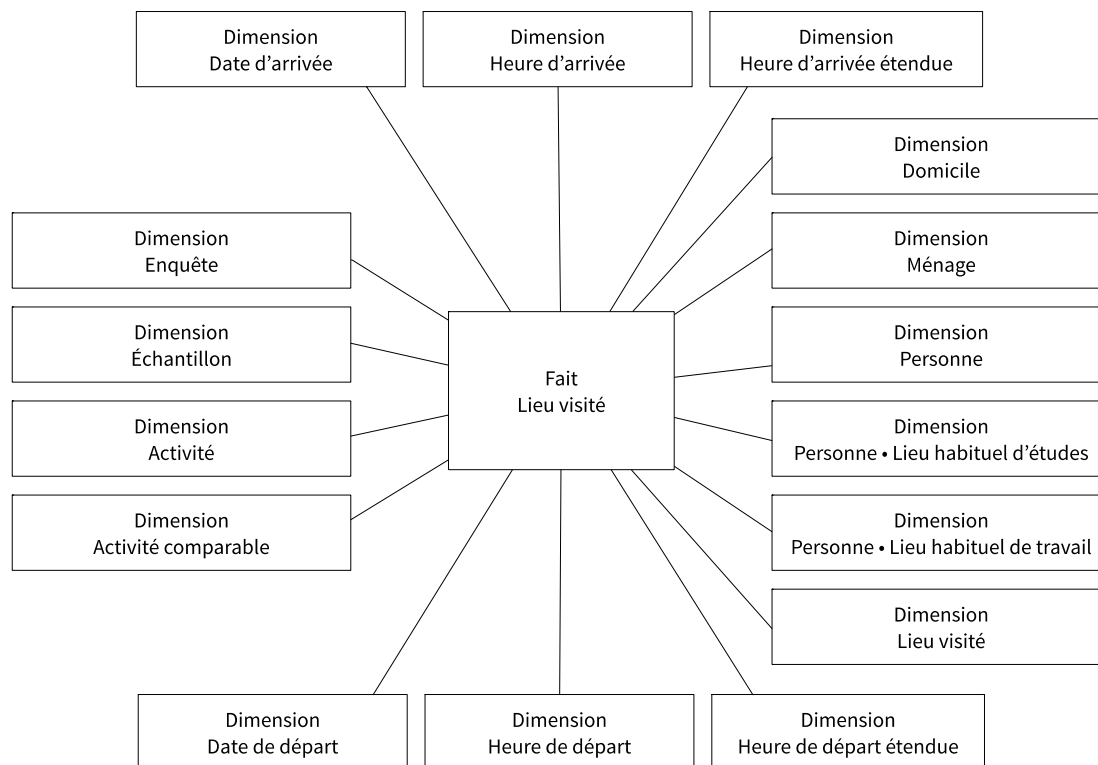


Figure 5-5 : Schéma dimensionnel • Lieu visité

5.6.10 La chaîne de déplacements

En plus des dimensions décrites précédemment (domicile, ménage, personne, lieux habituels), la table de fait chaîne de déplacements (Figure 5-6) est associée aux dates et heures/heures étendues de début et de fin de la chaîne, au point d'ancrage déterminé de manière automatique³⁶ et à l'activité principale effectuée au point d'ancrage de la chaîne de déplacements (accompagnée de l'activité principale comparable). La dimension représentant la chaîne de déplacements comprend les caractéristiques de la chaîne. Elle inclut les attributs relatifs à la catégorie de chaîne selon les définitions de (Valiquette, 2010) (contrainte/non contrainte, simple/complexe/ouverte), le nombre de boucles, le nombre de déplacements, de modes distincts, de segments, la durée

³⁶ Le point d'ancrage est le lieu de l'activité principale effectuée à l'intérieur de la chaîne de déplacement. Les domiciles et les lieux de travail ou d'études sont les points d'ancrage les plus fréquents (Valiquette, 2010).

déclarée de la chaîne, la durée de l'activité principale ainsi que la distance totale parcourue. Lorsque la table de séjour est impliquée dans une enquête, la table de lieux visités s'accompagne également d'une référence vers la dimension séjour concernée.

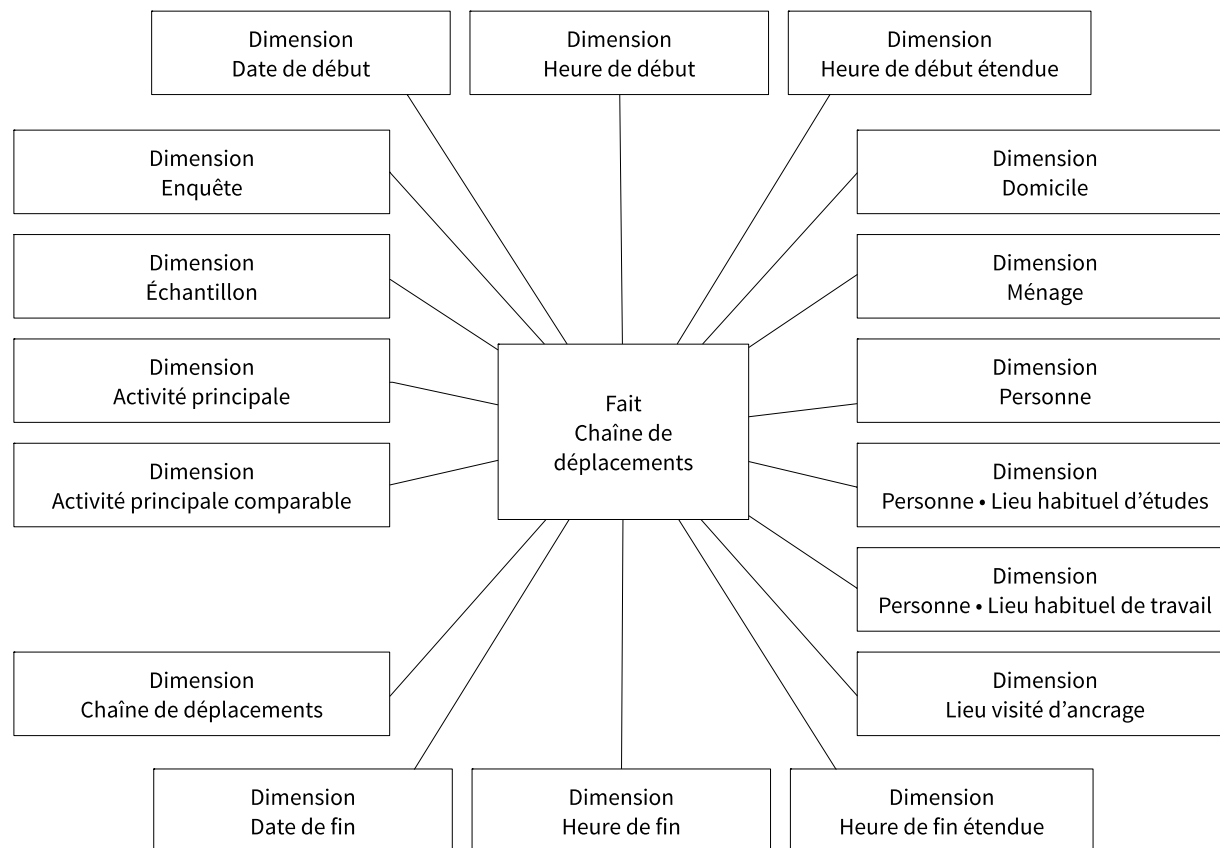


Figure 5-6 : Schéma dimensionnel • Chaîne de déplacements

5.6.11 Le déplacement

En plus des dimensions de domicile, ménage, personne, lieux habituels, la table de fait déplacement (Figure 5-7) comprend la dimension chaîne de déplacement associée (de même que la dimension séjour au besoin) ainsi que des dimensions pour les lieux visités à l'origine et à destination et les dates et heures de début et de fin du déplacement. Les activités effectuées à l'origine et à la destination, accompagnées de leurs dimensions comparables respectives, sont également associées au fait déplacement. Dans le cas où le déplacement compte au moins un transfert modal entre la voiture ou le vélo et le transport collectif, une dimension représente le premier lieu de jonction géolocalisé. De son côté, la dimension déplacement inclut différents attributs comme le nombre de segments, le nombre de modes uniques utilisés, le type de déplacement

(un seul mode ou multimodal), les types de stationnement à l'origine et à destination pour les déplacements effectués en voiture conducteur (type de stationnement, type de paiement, coût, etc.), le type de conducteur pour les déplacements effectués en tant que passager, la distance parcourue et la durée déclarée. La dimension catégorie de mode comparable décrit le mode agrégé imputé en fonction du ou des modes déclarés (mode unique ou mode de type bimodal). Finalement, la table de fait déplacement est liée à une dimension de calcul de chemin, générée pendant la phase d'enrichissement des données. Pour chaque déplacement déclaré, le calculateur de chemins retourne le trajet sur le réseau de transport, le temps de parcours ainsi que plusieurs caractéristiques du chemin théorique le plus rapide en utilisant la voiture, le vélo, la marche et le transport collectif³⁷, peu importe le mode déclaré par le répondant. Cette dimension permet d'éventuelles comparaisons entre les modes de transport qui ont été utilisés ou qui auraient pu l'être, et ce pour tous les déplacements déclarés par les répondants.

³⁷ Des attributs sont prévus pour sauvegarder les résultats de calculs de chemin bimodaux de type Park and Ride (voiture et transport collectif) et Bike and Ride (vélo et transport collectif), mais aucun calculateur de chemins multimodal n'a été intégré à la plateforme pour l'instant.

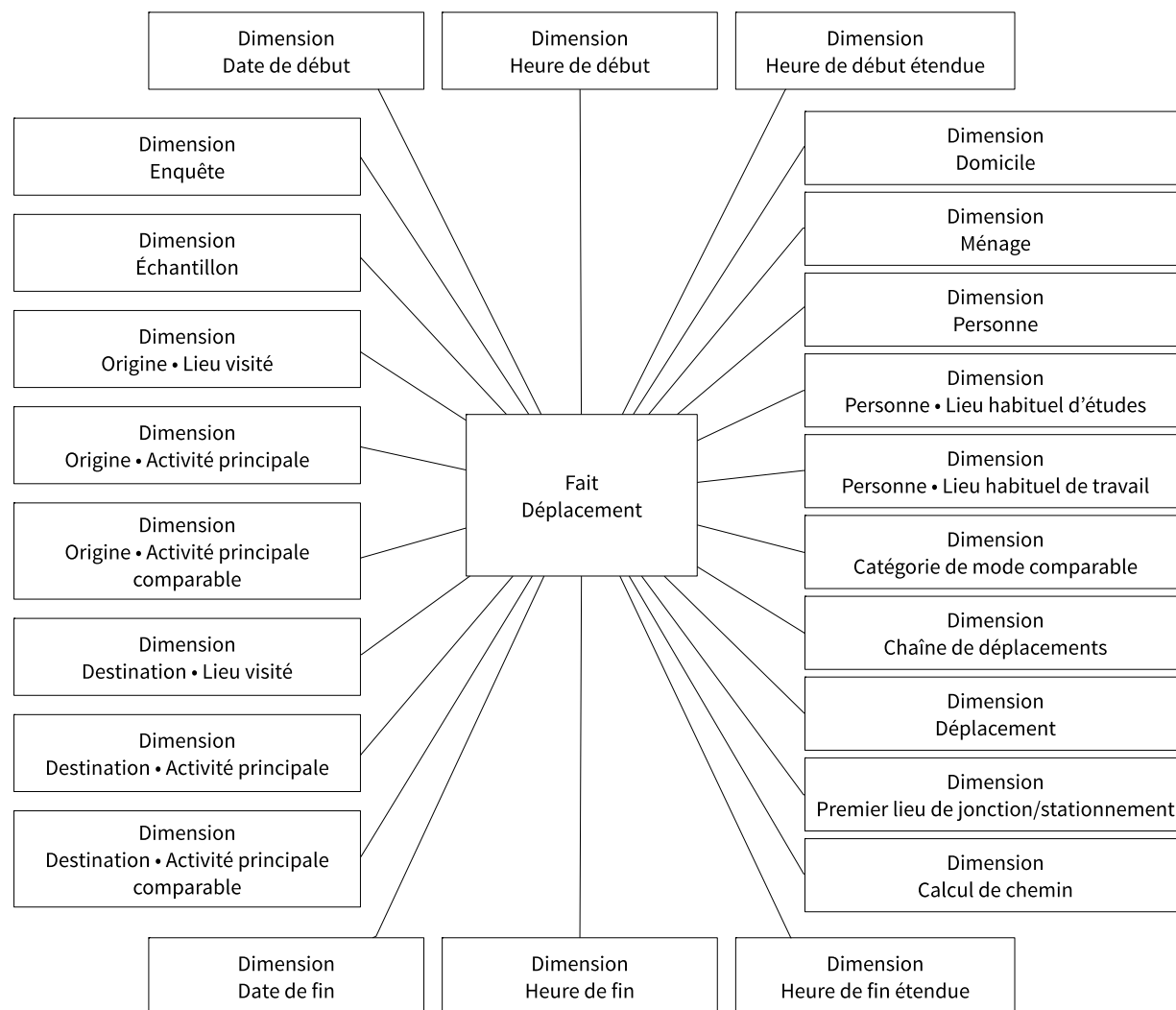


Figure 5-7 : Schéma dimensionnel • Déplacement

5.6.12 Le segment de déplacement

Le segment de déplacement suit le même schéma que celui proposé pour le déplacement, mais comprend également des dimensions relatives aux lignes de transport collectif empruntées et aux arrêts d'embarquement et de débarquement, lorsque déclarés par le répondant (Figure 5-8). La dimension véhicule est ajoutée lorsque les caractéristiques du véhicule utilisé lors du segment sont déclarés (type de véhicule, capacité, année, marque, modèle, odomètre, etc.). De la même façon, si les informations sur le conducteur du véhicule sont disponibles, une dimension représentant le conducteur est incluse. La dimension mode comprend les attributs relatifs au mode

de transport emprunté (nom et description, catégorie de mode comparable, etc.). Enfin, les attributs présents dans la dimension segment contiennent, entre autres, le nombre d'individus ayant effectué le segment avec la personne, le nombre d'occupants dans le véhicule et le type de conducteur (membre du ménage, voisin, collègue, ami, etc.).

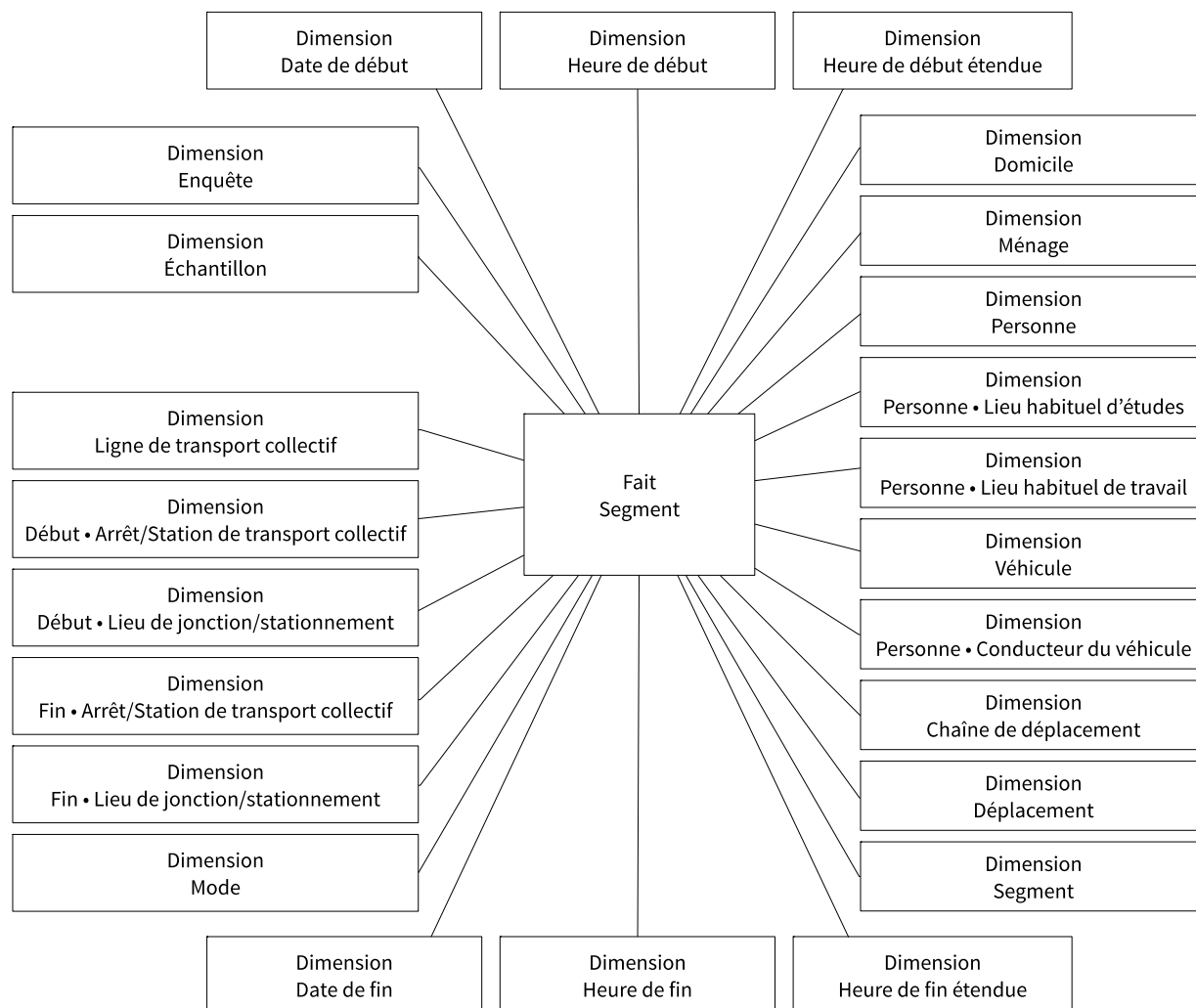


Figure 5-8 : Schéma dimensionnel • Segment

5.7 Attributs personnalisés

Lorsqu'une enquête comprend des questions qui ne sont pas présentes dans un questionnaire de type Origine-Destination conventionnel, le schéma relationnel de l'entrepôt de données sera modifié afin d'inclure les attributs associés aux questions personnalisées configurées dans la base

de données de collecte. Pour ce faire, le module ETL compare la liste des questions de la base de données de collecte avec la liste des attributs de chacune des dimensions. Lorsqu'un attribut est manquant dans l'entrepôt de données, il ajoute une nouvelle colonne à la table de dimension correspondant au type d'objet et au type de données spécifiés dans la configuration des questions. De cette façon, lorsque les données sont par la suite traitées et importées dans l'entrepôt, toutes les réponses à des questions personnalisées sont conservées et sauvegardées à un endroit approprié.

5.8 Pondération des objets

La pondération est le processus d'expansion des données permettant de modéliser la population réelle (l'univers d'étude). Des facteurs d'expansion sont alors attribués à chaque objet que l'on désire pondérer en fonction du taux d'échantillonnage obtenu pour chaque découpage géographique et/ou chaque groupe socio-démographique pour lesquels on possède un recensement exhaustif. Pour représenter le facteur d'expansion lorsque fourni, chaque table de faits comprend un attribut dont le suffixe est `SingleWeight` (prévu pour pondération unique des objets). Par ailleurs, dans le cas où l'analyste désire attribuer de multiples facteurs d'expansion à un même objet, la plateforme peut ajouter des colonnes supplémentaires pour inclure ces facteurs, celles-ci étant automatiquement accompagnées du préfixe `MultipleWeight`.

5.9 Exportation des fichiers

Une fois le traitement, la validation et la pondération effectués, l'utilisateur peut exporter les résultats au moyen de fichiers comportant les enregistrements associés aux différents objets d'étude. Un fichier distinct peut être téléchargé pour les ménages, les personnes, les lieux visités, les chaînes de déplacements, les déplacements et les segments de déplacements.

5.9.1 Dictionnaire implicite des attributs

L'avantage principal de la modélisation dimensionnelle pour la production de fichiers de résultats est la façon dont sont structurés les attributs de chaque dimension. Pour construire une requête d'analyse ou exporter un fichier de résultats, la table de faits demandée effectue automatiquement une jointure relationnelle avec les différentes dimensions qui lui sont associées. Une fois

cette jointure exécutée, les champs des différentes tables de dimensions sont inclus dans le résultat de la requête. De plus, les noms de colonnes sont choisis de telle sorte que le contenu soit évident, ce qui constitue une sorte de dictionnaire implicite des attributs.

5.9.2 Choix des objets et des attributs à exporter

Avant de lancer le téléchargement du fichier de résultats qu'il désire, l'utilisateur peut choisir les différents attributs de chacune des dimensions au moyen d'une interface de saisie comprenant la liste de tous les champs disponibles précédés de cases à cocher. Des groupements pré-configurés sont également disponibles afin de produire rapidement des fichiers dont la structure est utilisée fréquemment. La Figure 3-26 illustre l'interface permettant de choisir les attributs à inclure dans le fichier de résultats à télécharger.

5.10 Processus de traitement et de validation des données

Le traitement et la validation des données en préparation de la génération des faits saillants et des fichiers de résultats est souvent longue et difficile à mettre en œuvre. En effet, pour l'enquête régionale Origine-Destination de Montréal de 2013 par exemple, plus d'un an a été nécessaire entre la fin de la collecte de données et la production des faits saillants pour 231 014 déplacements validés (Agence métropolitaine de transport, 2015), soit approximativement 630 déplacements traités par jour en moyenne. En comparaison, à partir de la fin des entrevues de l'enquête auprès de la communauté étudiante universitaire de Toronto en 2015, moins de 10 jours ont été requis pour traiter, valider et enrichir les données, puis générer les faits saillants pour 36 710 déplacements validés, soit approximativement 3 700 déplacements traités par jour en moyenne, pour un questionnaire de type Origine-Destination relativement comparable³⁸. Bien que plusieurs validations doivent être effectuées manuellement, le module d'observation et de validation des entrevues permet à un validateur d'entrer directement dans l'entrevue du répondant pour corriger ou vérifier les informations déclarées. Une vidéo de démonstration du processus

³⁸ Des données plus précises sur la méthodologie de collecte par téléphone seraient nécessaires pour faire un constat plus précis, mais de telles informations ne sont pas disponibles pour l'instant.

de validation est disponible³⁹. En premier lieu, il est possible de visualiser sur une carte les déplacements des répondants pour chaque entrevue. Les éléments qui nécessitent une vérification apparaissent à droite de la carte, en jaune ou en rouge, selon la gravité du problème détecté (ce sont les audits qui servent à sélectionner les cas problématiques). Par exemple, les déplacements dont le mode déclaré est la marche (vitesse moyenne d'environ 5 km/h), mais qui ont une vitesse de déplacement à vol d'oiseau entre l'origine et la destination de plus de 6 km/h apparaîtront en jaune, alors que ceux qui ont une vitesse de plus de 10 km/h seront affichés en rouge. Le validateur peut utiliser soit le module de correction des heures de départ et d'arrivée des déplacements montré dans la démonstration, soit entrer dans l'entrevue et corriger la déclaration considérée comme erronée. À chaque fois qu'une modification est effectuée dans la base de données de collecte pendant la validation, l'historique est conservé et une mention « modifié par un administrateur » s'ajoute dans les enregistrements concernés.

5.11 Contributions de l'entrepôt de données

La proposition d'un schéma complet de modélisation dimensionnelle pour la construction d'un entrepôt de données constitue en soi une contribution au domaine de l'analyse de la mobilité des personnes. En effet, l'adaptation d'un paradigme ayant fait ses preuves en intelligence d'affaires pour structurer des données permet de simplifier les requêtes soutenant les analyses de mobilité puisqu'un seul degré de séparation est permis dans le modèle de données (exception faite des dimensions de dates et heures) et que les attributs ont des noms compréhensibles. En parallèle, cela permet de structurer et de mettre en valeur les données en facilitant leur traitement au moyen de procédures automatisées et d'une interface de validation dirigée. En effet, les procédures de traitement proposées assurent une cohérence dans la méthodologie utilisée pour la validation des entrevues provenant d'enquêtes distinctes, sans négliger la flexibilité de la configuration de validation prévue pour tenir compte des besoins particuliers propres à chaque exercice. Lors de la création de l'entrepôt de données, les calculs de chemin et les intersections automati-

³⁹ <https://vimeo.com/151051892>

sées avec les zones géographiques importées enrichissent les données de mobilité. Enfin, l'interface d'exportation des fichiers de résultats permet de tenir compte des besoins particuliers de chaque utilisateur puisqu'il est possible de sélectionner seulement les types d'objets désirés et les attributs les plus pertinents selon l'analyse.

5.12 Contenu de l'entrepôt de données actuel et importation de données externes

L'entrepôt de données actuel inclut les données traitées et validées de toutes les enquêtes réalisées dans le cadre de la thèse. De plus, les données des enquêtes Origine-Destination de la région de Québec de 1996, 2001, 2006 et 2011 ont également été importées avec succès⁴⁰. Pour chaque enquête dont les données n'ont pas été collectées au moyen de l'outil de collecte web proposé dans la plateforme, un fichier de conversion spécifique doit être ajouté dans le but de recréer l'ensemble des entrevues. Pour l'instant, un script en langage Ruby créé manuellement effectue la conversion des attributs d'un format à l'autre. Un module de conversion et d'importation de données simplifié sera toutefois proposé dans la prochaine version de la plateforme. Il est également possible d'importer des données constituées de traces GPS de déplacements en autant que les activités et les modes de transport associés aient été imputés ou déclarés. Les traces et les heures de départ et d'arrivée sont alors copiées dans la dimension calcul de chemin reliée à chaque déplacement en remplacement des données habituellement générées par un calculateur de chemin.

⁴⁰ Les données accompagnées de la méthodologie de validation provenant des enquêtes de la région de Montréal n'étaient pas disponibles pour importation dans l'entrepôt de données.

CHAPITRE 6 OBJETS DE VISUALISATION

6.1 Cartes animées de déplacements

Dans le cadre de projets connexes à la mise en place de l'entrepôt de données de mobilité, des cartes animées de déplacements ont été produites pour fins de diffusion et d'analyse. Ces cartes permettent de visualiser dans le temps et dans l'espace le déplacement des personnes lors d'une journée moyenne de semaine, en accéléré. En fait, on peut y voir les traces des personnes ou des véhicules, dans l'objectif d'illustrer l'utilisation des réseaux de transport de la région étudiée. La Figure 6-1, la Figure 6-2 et la Figure 6-3 montrent des captures de cartes animées de déplacements réalisées.



Figure 6-1 : Capture d'une carte animée de déplacements montrant l'achalandage sur le métro à l'heure de pointe du matin (bleu : transport collectif, rouge : voiture conducteur, orange : voiture passager, jaune : marche, vert : vélo). Vidéo complet : <https://vimeo.com/129174639>

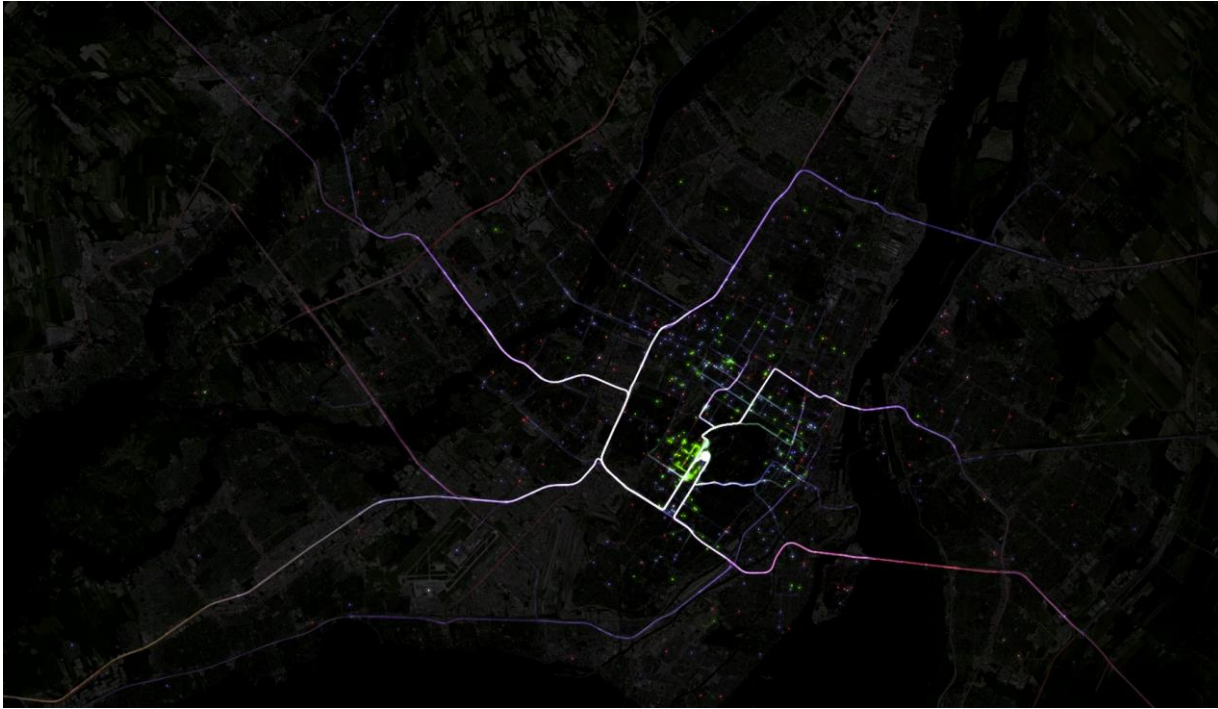


Figure 6-2 : Capture d'une carte animée de déplacements montrant l'achalandage sur les routes vers Polytechnique. Vidéo complet : <https://vimeo.com/119243419>

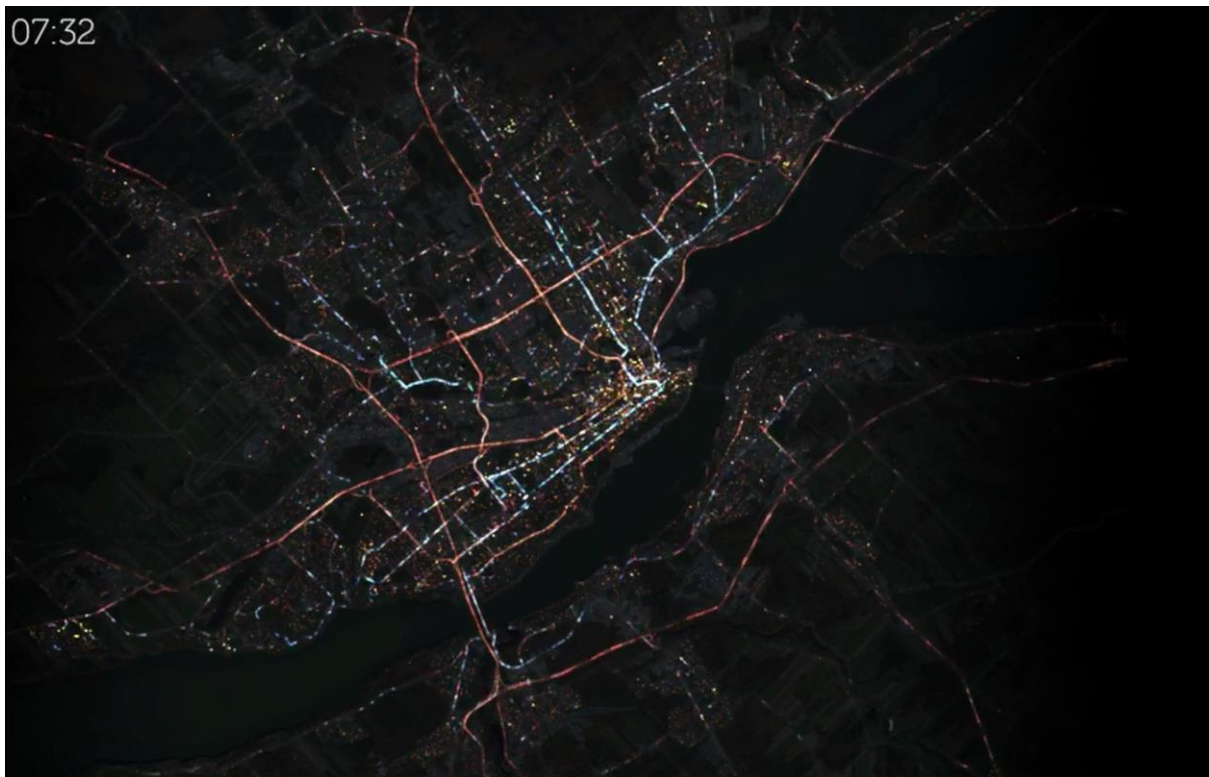


Figure 6-3 : Capture d'une carte animée de déplacements montrant les déplacements de l'échantillon complet de l'enquête régionale de Québec 2011. Vidéo complet : <https://vimeo.com/162916534>

6.1.1 Application utilisée pour la production de cartes animées

Le logiciel libre Processing⁴¹ (Fry, Reas, & Shiffman, n.d.) est utilisé pour produire les cartes animées de déplacements à partir d'un fichier exporté au moyen de l'interface d'exportation de l'entrepôt de données. Processing utilise le langage JAVA et y ajoute des bibliothèques et des fonctions pour faciliter la création d'animations. La performance du logiciel Processing permet actuellement la visualisation à un taux de rafraîchissement appréciable d'environ 30 images par seconde lorsque le nombre de personnes affichées ne dépasse pas 10 000⁴². Pour visualiser davantage de personnes, il est possible de produire une animation statique pouvant être jouée dans le lecteur vidéo de son choix.

6.1.2 Méthodologie de production de cartes animées de déplacements

Dans le but d'automatiser la production de cartes animées, une méthodologie de génération des cartes à partir de l'entrepôt de données est proposée. Les traces des différents déplacements proviennent des résultats fournis par les calculateurs de chemin. En ce sens, les animations générées ne représentent pas nécessairement le mouvement exact des personnes faisant partie de l'échantillon (le chemin emprunté n'étant pas connu), mais bien une estimation selon les chemins les plus courts en écoulement libre (des données de congestion peuvent toutefois être importées dans le calculateur routier pour tenir compte des vitesses réelles sur les différents segments de rue).

6.1.2.1 Préparation des données

Afin de configurer les traces dans Processing, un fichier au format CSV comprenant les déplacements à animer est d'abord importé depuis l'entrepôt de données. La table de fait déplacement est utilisée pour générer le fichier. Par défaut, les attributs suivants sont inclus:

- TripKey (identifiant unique du déplacement)
- PersonKey (identifiant unique de la personne)

⁴¹ <https://processing.org/>

⁴² Vérifié au moyen d'un processeur Intel Xeon Quad-Core 2.8 GHz avec 16 GB de mémoire RAM

- Sequence (séquence du déplacement dans la journée pour la personne)
- ActivityAtDestinationKey (identifiant de l'activité à destination)
- StartAt (Heure de départ)
- EndAt (Heure d'arrivée, qui peut être soit l'heure déclarée soit l'heure d'arrivée calculée par le calculateur de chemin pour le mode déclaré)
- VisitedPlaceOriginX (coordonnée X de l'origine du déplacement)
- VisitedPlaceOriginY (coordonnée Y de l'origine du déplacement)
- VisitedPlaceDestinationX (coordonnée X de la destination du déplacement)
- VisitedPlaceDestinationY (coordonnée Y de la destination du déplacement)
- GeometryWKT (Géométrie projetée du tracé obtenue au moyen du calculateur de chemins en format WKT⁴³, le tracé en voiture étant utilisé par défaut pour les modes autres que marche, vélo ou transport collectif)
- MultimodeCategoryKey (Mode agrégé)

Au besoin, d'autres attributs servant à distinguer les déplacements sur la carte (couleur ou taille du tracé) peuvent également être inclus (groupe d'âge, sexe, occupation, etc.). Par défaut, le mode est utilisé pour sélectionner la couleur d'affichage du tracé (les couleurs peuvent être modifiées dans le fichier de configuration). La requête SQL utilisée pour générer le fichier est la suivante :

```
COPY (
SELECT
    t."TripKey"                AS "TripKey",
    t."TripPersonKey"         AS "PersonKey",
    t."TripSequenceInInterviewDay" AS "Sequence",
    t."TripActivityDestinationKey" AS "ActivityAtDestinationKey",
    t."TripStartTimeExtendedKey"   AS "StartAt",
    t."TripEndTimeExtendedKey"     AS "EndAt",
    o."VisitedPlaceX"           AS "VisitedPlaceOriginX",
    o."VisitedPlaceY"           AS "VisitedPlaceOriginY",
    d."VisitedPlaceX"           AS "VisitedPlaceDestinationX",
```

⁴³ https://en.wikipedia.org/wiki/Well-known_text


```

d."VisitedPlaceY"           AS "VisitedPlaceDestinationY",
CASE t."TripComparableMultimodeCategoryKey"
  WHEN 3 THEN ST_AsText(r."RoutingTransitPostgisGeometry")
  WHEN 4 THEN ST_AsText(r."RoutingWalkingPostgisGeometry")
  WHEN 5 THEN ST_AsText(r."RoutingCyclingPostgisGeometry")
  ELSE ST_AsText(r."RoutingDrivingPostgisGeometry")
END AS "GeometryWKT",
t."TripComparableMultimodeCategoryKey" AS "MultimodeCategoryKey"
FROM dw_od_trip_facts t
LEFT JOIN dw_od_routing_dimensions r
  ON r."RoutingKey" = t."TripRoutingKey"
LEFT JOIN dw_od_visited_place_dimensions o
  ON o."VisitedPlaceKey" = t."TripVisitedPlaceOriginKey"
LEFT JOIN dw_od_visited_place_dimensions d
  ON d."VisitedPlaceKey" = t."TripVisitedPlaceDestinationKey"
ORDER BY
  t."TripPersonKey",
  t."TripSequenceInInterviewDay"
)
TO 'FICHIER.csv' DELIMITER ',' CSV HEADER;

```

Après avoir traité ce fichier, le code Processing développé effectue une interpolation des positions de chaque personne pour chaque intervalle de temps (en fonction du taux de rafraîchissement et de la vitesse d'animation sélectionnée) à partir des points de la trace des déplacements et détermine la couleur d'affichage en fonction des paramètres de configuration.

6.1.2.2 Paramètres

Les paramètres de création de la carte animée de déplacements comprennent d'abord l'heure de début et de fin de l'animation, ainsi que la vitesse d'affichage. Ensuite, la couleur des tracés doit être associée à un attribut du déplacement (une couleur par mode de transport par défaut). Il est également possible de configurer la longueur de la trace après le passage de la personne ou du véhicule, c'est-à-dire pendant combien de temps la trace reste visible sur la carte. La taille et l'opacité du point représentant la personne ou le véhicule peuvent être modifiées en fonction d'un autre attribut. Enfin, la carte de fond doit être importée. Des touches du clavier permettent d'ajuster finement le positionnement de la carte de fond sur le canevas. Pendant l'animation, l'utilisateur peut modifier le zoom, effectuer une rotation ou se déplacer dans l'espace en utilisant les raccourcis clavier ou la souris.

6.2 Cartes animées de la distribution géographique et temporelle d'objets sur un territoire

Un autre type de visualisation utilisant le logiciel Processing est proposé pour illustrer la position des personnes ou des véhicules pendant une journée ou une période déterminée. Les cartes animées de la distribution géographique et temporelle d'objets sur le territoire sont constitués de prismes à base carrée construits au-dessus d'une grille uniforme répartie sur le sol. La hauteur de chaque prisme est associée au nombre d'objets présents dans la cellule de la grille et change en fonction du temps pendant la journée. Ce type d'animation a été utilisé auparavant pour montrer la migration des personnes pendant une journée moyenne de semaine sur un territoire donné (Buliung & Morency, 2010), mais devait être créée manuellement au moyen d'un SIG et de captures d'écran représentant chaque période de la journée. L'utilisation de Processing permet d'accélérer la production de telles cartes et de choisir les paramètres plus facilement. Par exemple, la couleur des prismes peut changer en fonction du nombre d'objets afin d'accentuer les contrastes. De plus, plusieurs cartes peuvent être juxtaposées pour des besoins de comparaison.

6.2.1 Préparation des données

Une structure de fichier particulière doit être préparée pour générer les cartes de distributions géographique et temporelle. Chaque ligne représente une cellule de la grille et chaque colonne la période de la journée. Des colonnes sont ajoutées pour définir les coordonnées géographiques de chaque cellule. La Figure 6-4 montre un exemple de juxtaposition de deux cartes de distribution (usagers du transport collectif en bleu et conducteurs en rouge).

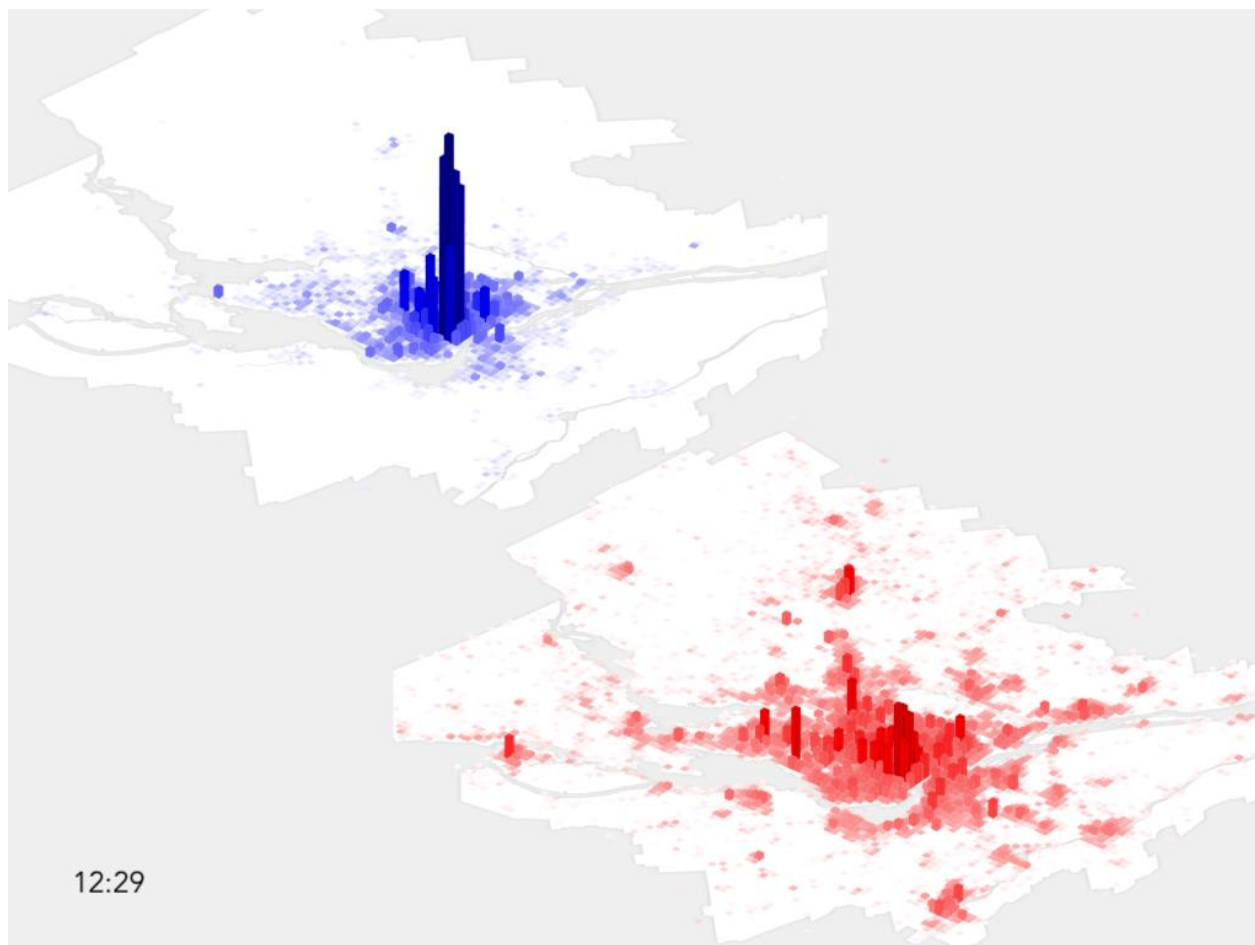


Figure 6-4 : Carte animée de la distribution géographique et temporelle des usagers du transport collectif (bleu) et des conducteurs (rouge) pour la région de Montréal (enquête Origine-Destination 2008) Vidéo complet : <https://vimeo.com/137969467>

6.3 De la collecte à la présentation des faits saillants

La présentation de faits saillants est le premier exercice de diffusion des résultats auprès du public et des intervenants d'une enquête. Sans avoir l'objectif d'analyser en profondeur la question de recherche sous-jacente, les faits saillants doivent servir d'élan et de point de départ au processus d'analyse plus pointu et plus exhaustif qui suivra. Ils ont du même coup le fardeau de valoriser l'effort entrepris dans la bonne conduite de l'enquête et dans le traitement rigoureux des données. Par ailleurs, les faits saillants doivent s'adapter à leur auditoire principal, c'est-à-dire à l'ensemble des répondants, mais également à la population représentée et les décideurs et intervenants qui devront, dans la majorité des cas, prendre des décisions éclairées à la lumière des résultats produits. Ces derniers ne possédant pas toujours des connaissances pointues dans le

domaine d'étude, et surtout ayant rarement le loisir de mettre beaucoup de temps à la compréhension de la méthodologie de collecte, de traitement et de validation, il est primordial que les faits saillants mettent en lumière des résultats rigoureux, honnêtes et présentés de manière claire et conviviale.

La plateforme développée comprend un module de génération et de présentation de faits saillants. Développé conjointement par Julien Faucher et l'auteur de cette thèse en partenariat avec la Ville de Québec, ce module comprend un ensemble de graphiques et de tableaux comparatifs qui facilitent la visualisation des données collectées lors des enquêtes de mobilité (Figure 6-5). Une vidéo de démonstration du module de faits saillants est également disponible⁴⁴. Comme on le voit dans la vidéo, il est possible de comparer les résultats entre différentes enquêtes et entre différents découpages géographiques. Toutes les enquêtes réalisées dans le cadre de la thèse sont compatibles avec le module de génération des faits saillants.

⁴⁴ <https://vimeo.com/164832933>

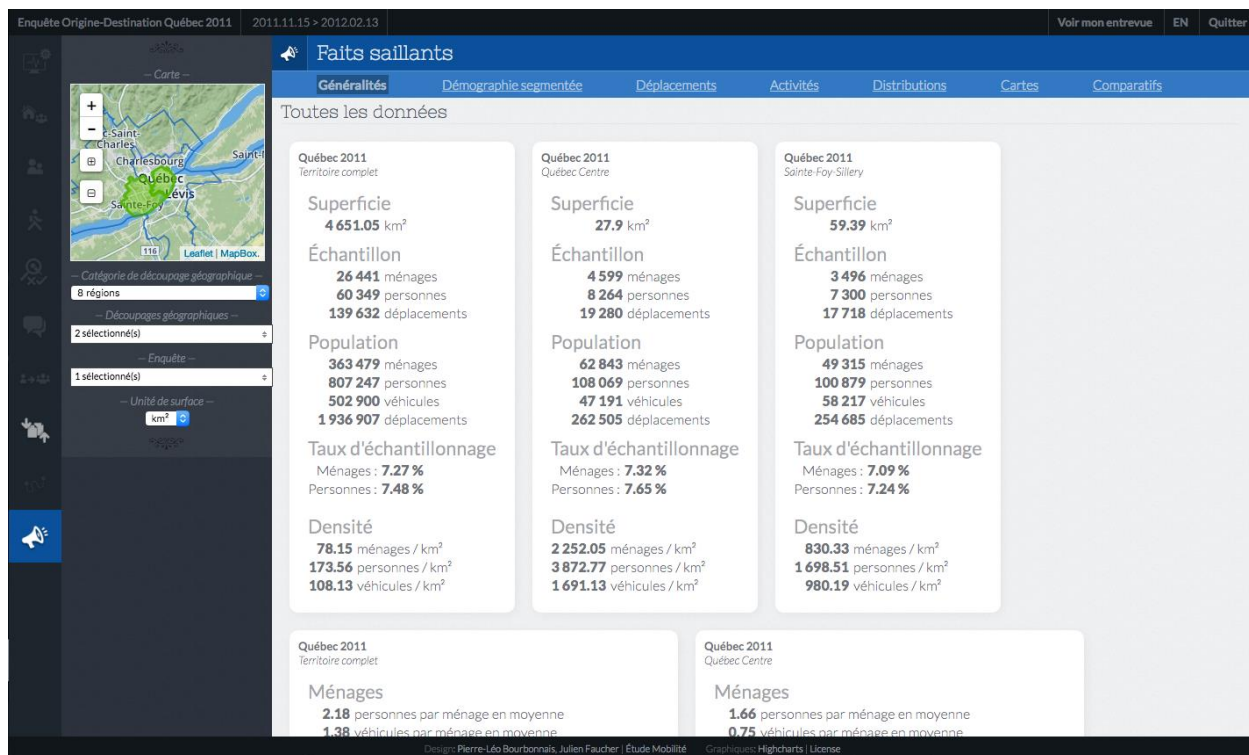


Figure 6-5 : Module de présentation des faits saillants

CHAPITRE 7 CONCLUSION

Étant donnée la difficulté de plus en plus grande à rejoindre certaines cohortes et certains groupes de population lors de la tenue d'enquêtes sur la mobilité des ménages et des personnes, des pistes de solutions ont été proposées au cours des dernières années. Notamment, la mise en œuvre de questionnaires web a été proposée. Toutefois, les questionnaires de type Origine-Destination utilisés pour des entrevues téléphoniques nécessitaient une adaptation qui présentait des défis complexes avant de permettre leur utilisation en ligne auprès de répondants qui y répondent de manière autonome. C'est à ce besoin d'adaptation que répondait la première facette de l'outil développé.

Par ailleurs, une complexité croissante dans le traitement et la gestion des données collectées au moyen d'enquêtes sur la mobilité a stimulé le besoin de simplifier les étapes de préparation, de stockage et de valorisation de ces données. Même si la modélisation totalement désagrégée a permis de représenter fidèlement les objets associés à la mobilité des personnes et de préciser leur structure, la quantité toujours plus grande de données provenant de sources multiples n'a fait qu'exacerber la problématique reliée à la gestion des données qui servent à mieux comprendre les comportements de mobilité et à planifier les réseaux de transport de manière efficace. Pour faire suite à cette difficulté, la seconde facette de la plateforme avait pour fonction principale de rendre plus simple, plus flexible et plus rapide le traitement des données de mobilité, tout en favorisant l'enrichissement des données et la production d'objets de visualisation. Grâce à l'adaptation de la modélisation dimensionnelle utilisée couramment dans le domaine de l'intelligence d'affaire, la mise en place d'un entrepôt de données de mobilité a été proposée dans le but de faciliter l'intégration, le traitement et la validation des données.

7.1 Contributions principales de la plateforme développée

Le questionnaire web développé permet de créer des questionnaires personnalisés et permettant des validations en temps réel. À la lumière des nombreux commentaires mentionnant la qualité et la simplicité du design, c'est probablement la convivialité du questionnaire qui a favorisé son succès auprès des répondants, bien que cet atout soit difficile à mesurer de manière objective.

La modélisation de la durée d'entrevue a été développée dans le but de mieux partager les ressources des serveurs utilisés lors de la conduite d'enquêtes en ligne. De plus, elle contribue à évaluer, en temps réel, la plage de temps prévue avant l'achèvement de chaque entrevue. Le suivi et l'administration des entrevues se trouvent facilités par cette information.

Le questionnaire web a fait l'objet de trois publications. La première, (Bourbonnais & Morency, 2013b), décrivait l'outil de collecte, la deuxième (Bourbonnais & Morency, 2013a) résumait les résultats obtenus à la suite d'une analyse du comportement des répondants et la troisième (Bourbonnais & Morency, 2015) présentait le modèle de durée d'entrevue.

La mise en place d'un entrepôt de données de mobilité permet, dans un tout cohérent, de simplifier la gestion des données de mobilité, d'enrichir les données et de favoriser la comparaison des enquêtes entre elles. De plus, l'utilisation de la plateforme développée pour valider les entrevues complétées permet de traiter plus de 3 700 déplacements par jour, incluant la vérification et la correction manuelle des informations erronées ou invalides, alors que seulement environ 630 déplacements par jour ont été traités entre 2013 et 2015 dans le cadre de l'enquête régionale de Montréal. Finalement, la génération automatisée d'objets de visualisation tels que les cartes animées permet de limiter la quantité de ressources occupées à créer manuellement les objets de visualisation.

7.2 Limitations

7.2.1 Représentativité des échantillons

Les enquêtes réalisées dans le cadre de la thèse comportent une limitation importante en ce sens où la grande majorité des entrevues complétées l'ont été par des répondants jeunes et instruits (étudiants universitaires et membres de services d'autopartage et de vélopartage). Conséquemment, les modèles de durée d'entrevue et l'analyse des commentaires des répondants ne peuvent être généralisés sans inclure des échantillons plus grands provenant d'enquêtes de mobilité dont la collecte a été effectuée au moyen du mode web et pour lesquelles les répondants sont davantage représentatifs de la population.

7.2.2 Mesure de l'évolution du questionnaire au moyen des commentaires des répondants

Bien que l'utilisation des commentaires des répondants ait été le seul moyen, dans l'état actuel, de mesurer l'évolution de la qualité de l'interface de collecte, cette méthode est peu rigoureuse sur le plan scientifique. Elle nécessiterait une analyse plus soutenue par des experts des sciences sociales, cognitives et de l'ergonomie des interfaces homme-machine pour qu'on puisse tirer des conclusions plus précises.

7.2.3 Modélisation des durées d'entrevue à partir de variables déclarées pendant l'entrevue

Le fait de modéliser la durée d'entrevue en utilisant des informations déclarées pendant l'entrevue par le répondant est pertinente pour les utilisations proposées (optimisation des ressources du serveur et prédiction en temps réel pendant l'entrevue), mais en limite les perspectives. Avec le modèle actuel, il est impossible de prédire à l'avance la plage de durée d'entrevue ne connaissant que les attributs démographiques du répondant (groupe d'âge et sexe).

7.2.4 Relative rigidité du modèle de données proposé pour l'entrepôt de données

Bien que la proposition du modèle dimensionnel pour la structure des données de mobilité soit un pas en avant, le modèle demeure rigide sur certains aspects. Notamment, il est difficile d'intégrer de nouvelles tables de dimensions, de nouvelles unités ou de nouveaux modes de calcul de chemin qui n'ont pas été prévues dans le schéma original sans modifier le module ETL.

7.3 Questions de recherche futures

De nombreuses questions de recherches ont été soulevées lors du développement de la plateforme présentée. Les études particulières qu'il conviendrait de mener en priorité afin de poursuivre le développement de la plateforme dans un contexte de recherche riche et rigoureux sont présentées dans cette section.

7.3.1 Conception de l'interface d'entrevue et ergonomie cognitive

L'étude de l'influence des choix de conception de l'interface d'entrevue sur les taux de réponse, la qualité des données et la satisfaction des répondants est sans aucun doute la plus complexe des questions de recherche à poser. Comment mettre en relation la conception du questionnaire avec la qualité des données et la facilité avec laquelle les répondants complètent leur entrevue, le tout de manière objective et rigoureuse dans un contexte où la technologie web est en perpétuelle évolution ? Une étude axée sur l'ergonomie cognitive serait probablement pertinente en ce sens. De plus, des questionnaires modifiés pour vérifier l'effet de différents choix de conception sur le comportement des répondants devront être comparés au moyen d'échantillons de répondants plus petits ou d'enquêtes pilote.

7.3.2 Analyse et modélisation des temps de réponse par type de question

Pour faire suite à la modélisation de la durée d'entrevue et dans le but de mieux comprendre les comportements de réponse, une analyse plus exhaustive des parodonnées du questionnaire web sera entreprise et orientée par type de question. Par exemple, comme la version la plus récente de plateforme — celle utilisée dans le cadre de l'enquête auprès de la communauté étudiante universitaire de Toronto en 2015 — enregistre les parodonnées à chaque action posée par le répondant pendant l'entrevue, il a été possible d'évaluer les temps de réponse, par exemple avec les quartiles présents dans la Figure 7-1. Les boutons radio booléens, les menus déroulants à moins de 5 choix, les échelles de groupes d'âge et les boutons radio pour sélectionner le sexe (femme, homme, autre) sont ceux qui demandent le moins de temps et qui, du même coup, ont la moins grande étendue de temps de réponse. Les champs de texte libre sur une seule ligne font également partie de ce groupe, mais comme ils sont habituellement facultatifs, ils sont peu représentatifs du temps de réponse réellement requis pour entrer une réponse complète, puisque les parodonnées enregistrées comprennent les réponses vides. Par ailleurs, les menus déroulants à choix nombreux, les champs de texte sur plusieurs lignes et les positionnements d'un lieu sur la carte ont des temps de réponse plus long. Les boutons-glissoires ont un temps médian très court, mais un troisième quartile plus élevé, probablement causé par la compatibilité limitée de ces

boutons avec les téléphones intelligents. En effet, plusieurs commentaires ont mentionné la difficulté à déplacer le bouton avec un téléphone intelligent ou un écran de petite taille. Le positionnement d'un lieu sur la carte par un clic produit des quartiles très éloignés les uns des autres. Sachant cela, une analyse particulière des temps de réponse selon le type de question posée sera possible pour les prochaines enquêtes utilisant la plateforme développée. La construction d'un modèle de prédiction des temps de réponse en fonction du type de question et des attributs du répondant favorisera une meilleure compréhension des comportements de réponse et, en ce sens, permettra de mesurer l'influence des temps de réponse de chaque type de question sur les taux d'achèvement des entrevues.

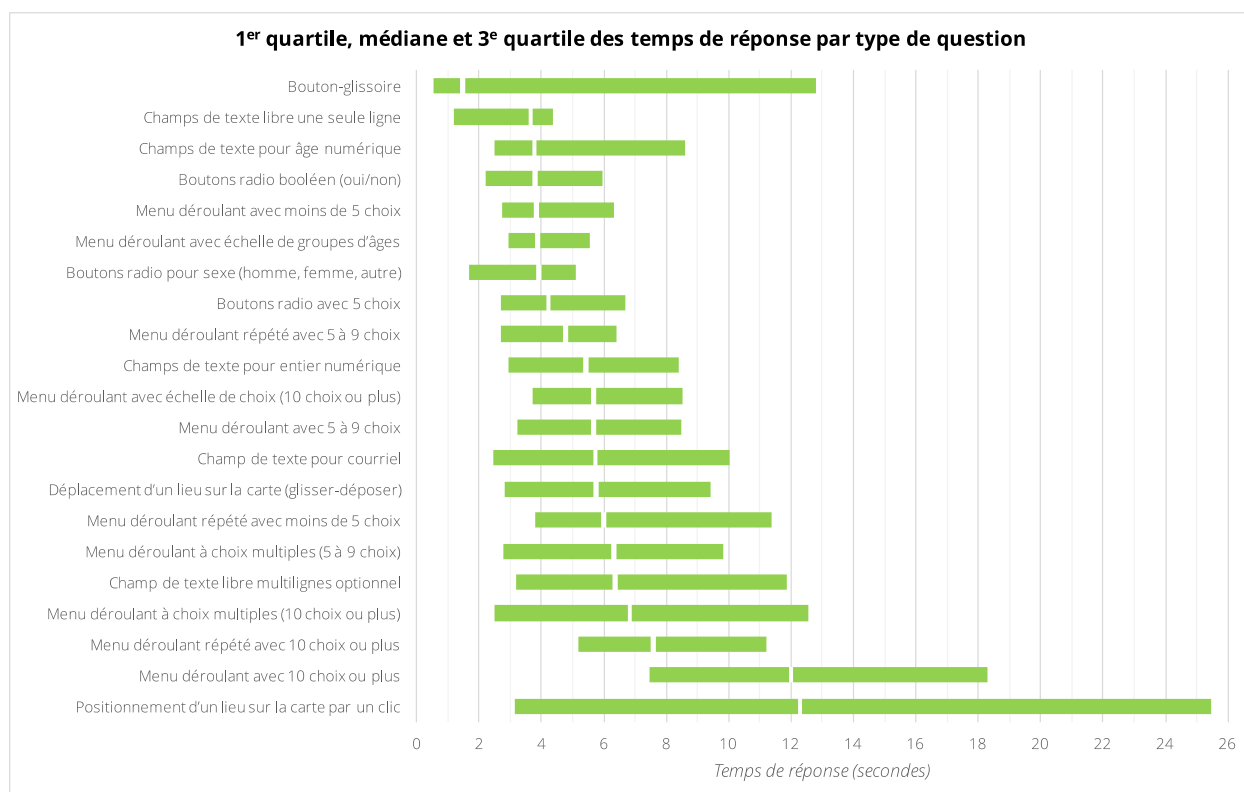


Figure 7-1 : Quartiles des temps de réponse par type de question

7.3.3 Publication des méthodes de traitement, de validation et de stockage des données de mobilité existantes

Une des plus grandes difficultés d'évaluation des contributions de la thèse s'articulait autour de la comparaison des méthodes de traitement, de validation et de stockage proposées avec celles

qui sont utilisées ailleurs dans le domaine. Ces méthodes sont rarement documentées, et quand elles le sont, elles demeurent accessibles seulement aux partenaires qui administrent les enquêtes en question. Sachant cela, la publication des méthodologies de traitement, de validation et de stockage des données en amont de la diffusion des résultats et accessible publiquement serait fortement encouragée.

7.3.4 Comparabilité des enquêtes

La question de la comparabilité des enquêtes exige une étude plus approfondie. De quelle manière doit-on traiter les divergences dans les choix de réponse, comment éviter les ambiguïtés, quelles seraient les pertes de précision dans les réponses si on convertissait la plupart des attributs en attributs comparables ? Même si la plateforme développée permet de gérer plus facilement les attributs comparables, les problématiques associées à l'exercice de comparaison des enquêtes entre elles doivent être évaluées et mesurées.

7.4 Poursuite du développement

Autant l'outil de collecte de données que l'entrepôt de données de mobilité pourront bénéficier d'un développement continu dans les prochaines années. Certains aspects de ce développement sont présentés ici comme étant les plus pertinents et prioritaires à court terme.

7.4.1 Préparer la persistance polyglotte

Afin d'accroître les performances de l'outil de collecte et de permettre son implantation sur des plateformes plus nombreuses et plus récentes, la mise en place de nouveaux modèles de base de données faisant appel à la persistance polyglotte sera une étape nécessaire à franchir dans les prochaines années. L'objectif ultime est d'utiliser le modèle le plus performant et le plus flexible pour chaque module de l'outil, autant pour la collecte de données que pour la modélisation dimensionnelle et le stockage dans l'entrepôt de données.

7.4.2 Poursuite de la production de faits saillants

De nouveaux graphiques et tableaux seront introduits dans le module de faits saillants afin d'enrichir la diffusion des résultats auprès des intervenants du milieu. Une consultation auprès des

chercheurs familiers avec l'analyse des données de mobilité, mais également auprès des experts de la visualisation permettra de déterminer les meilleurs objets de visualisation à intégrer. Une section est également prévue pour créer des graphiques dynamiques en sélectionnant les variables à comparer d'un seul clic. De cette manière, il sera plus rapide d'analyser les résultats et de déterminer les graphiques les plus efficaces pour visualiser les tendances ou éclaircir certains phénomènes relatifs à la mobilité des personnes.

7.4.3 Adaptation du questionnaire web en une interface de saisie pour intervieweur téléphonique ou face-à-face

Une adaptation de l'outil de collecte web pour utilisation auprès d'intervieweurs téléphoniques sera également développée dans le but de simplifier, d'accélérer et d'améliorer la qualité des données recueillies au moyen d'entrevues téléphoniques. Des raccourcis clavier rendront plus rapide la saisie des réponses déclarées et les changements de sections. Des cases de commentaires disponibles en tout temps pendant l'entrevue permettront à l'intervieweur d'y inscrire des notes relatives à l'entrevue et aux réponses et aux localisations ambiguës ou indéterminées. De plus, la mise en place d'une même interface de saisie pour une enquête mixte web/téléphonique facilitera l'administration de l'enquête, le suivi des entrevues, la fusion des données et la comparabilité des résultats.

7.4.4 Importation de données de cartes à puce

L'importation de données de cartes à puce provenant des bases de données des agences de transport collectif est envisagée dans le but de vérifier la précision et la qualité des données déclarées par les répondants utilisant le transport collectif. De plus, ces données pourraient être utilisées pour des fins de modélisation ou pour calibrer les calculateurs de chemins de transport collectif en fonction des parcours réels favorisés par les usagers.

7.5 Conclusion

Les perspectives futures de l'outil de collecte et de la proposition d'une structure d'entrepôt de données de mobilité sont nombreuses. Néanmoins, la mise en commun de données de mobilité enrichies et validées au moyen de la plateforme développée dans le cadre de cette thèse avait

comme objectif général de faciliter l'analyse des comportements de mobilité, de nourrir les modèles et les simulations, mais surtout de mettre en valeur l'ensemble des données collectées par le passé et à venir dans le cadre des enquêtes de mobilité et d'en accroître la portée.

BIBLIOGRAPHIE

- Adamson. (2010). *Star Schema: The Complete Ref.* Tata McGraw-Hill Education.
- Adler, R. F., & Benbunan-Fich, R. (2012). Juggling on a high wire Multitasking effects on performance. *Journal of Human Computer Studies*, 70(2), 156–168.
- Agence métropolitaine de transport. (2015). *Enquête Origine-Destination de Montréal 2013.*
- Alsnih, R. (2007). Characteristics of web based surveys and applications in travel research. In *Travel Survey Methods* (pp. 569–592). Emerald.
- Andrienko, G., Andrienko, N., Bak, P., Keim, D., & Wrobel, S. (2013). *Visual Analytics of Movement.* Springer Science & Business Media.
- Andrienko, N., & Andrienko, G. (2006). *Exploratory Analysis of Spatial and Temporal Data.* Springer Science & Business Media.
- Angles, R., & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys*, 40(1), 1–39.
- Armoogum, J., Axhausen, K. W., & Madre, J.-L. (2009). Lessons from an Overview of National Transport Surveys, from Working Group 3 of COST 355: “Changing Behavior Toward a More Sustainable Transport System.” In P. Bonnel, M. Lee-Gosselin, J. P. Zmud, & J.-L. Madre (Eds.), *The Travel Survey Toolkit: Where to From Here?* (pp. 621–634). 9th International Conference on Transport Survey Methods 2011.
- Armoogum, J., INRETS. (2014). *Survey Harmonisation with New Technologies Improvement, SHANTI.*
- ARUP. (n.d.). Bureau of Transport Statistics Data Visualisation. Retrieved March 10, 2016, from http://www.arup.com/Projects/Bureau_Transport_Statistics_Data_Visualisation.aspx
- Aslett, M. (2011a, April 4). How will the database incumbents respond to NoSQL and NewSQL? Retrieved September 25, 2015, from <https://451research.com/report-short?entityId=66963>
- Aslett, M. (2011b, April 6). What we talk about when we talk about NewSQL. Retrieved September 25, 2015, from https://blogs.the451group.com/information_management/2011/04/06/what-we-talk-about-when-we-talk-about-newsqli/
- Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfuser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29(2), 95–124.
- Badoe, D. A., & Steuart, G. N. (2002). Impact of interviewing by proxy in travel survey conducted by telephone. *Journal of Advanced Transportation*, 36(1), 43–62.

- Bayart, C., & Bonnel, P. (2008). Comparison of web and face-to-face household travel survey-application to Lyon case. *European Transport Conference 2008*.
- Bayart, C., & Bonnel, P. (2012). Combining web and face-to-face in travel surveys: comparability challenges? *Transportation*, 39(6), 1147–1171.
- Bayart, C., & Bonnel, P. (2015). How to Combine Survey Media (Web, Telephone, Face-to-Face): Lyon and Rhône-alps Case Study. *Transportation Research Procedia*, 11, 118–135.
- Bayart, C., Bonnel, P., & Morency, C. (2009). Survey Mode Integration and Data Fusion: Methods and Challenges. In P. Bonnel, M. Lee-Gosselin, J. P. Zmud, & J.-L. Madre (Eds.), (pp. 587–611). Presented at the Transport Survey Methods: Keeping Up with a Changing World, Bingley.
- Behrens, R., Freedman, M., & McGuckin, N. (2009). The Challenges of Surveying “Hard to Reach” Groups: Synthesis of a Workshop. *Keeping up with a Changing World* (pp. 145–152). Emerald Group Publishing Limited.
- Bonnel, P., Lee-Gosselin, M., Madre, J.-L., & Zmud, J. P. (2009). Keeping Up with a Changing World: Challenges in the Design of Transport Survey Methods. In P. Bonnel, M. Lee-Gosselin, J.-L. Madre, & J. P. Zmud (Eds.), (pp. 3–13).
- Bourbonnais, P.-L., & Morency, C. (2013a). Web-based Origin-Destination surveys: an analysis of respondent behaviour (pp. 1–13). Presented at the 13th World Conference on Transport Research.
- Bourbonnais, P.-L., & Morency, C. (2013b). Web-Based Travel Survey: A Demo. In M. Munizaga, J. A. Carrasco, J. Zmud, & M. Lee-Gosselin (Eds.), *The Travel Survey Toolkit: Where to From Here?* (Emerald). Bingley: 9th International Conference on Transport Survey Methods 2011.
- Bourbonnais, P.-L., & Morency, C. (2015). Patterning Respondent Behaviors from 9 Web Travel Surveys (pp. 1–22). Presented at the 10th International Conference on Transport Survey Methods.
- Braunsberger, K., Wybenga, H., & Gates, R. (2007). A comparison of reliability between telephone and web-based surveys. *Journal of Business Research*, 60(7), 758–764.
- Buliung, R. N., & Morency, C. (2010). “Seeing Is Believing”: Exploring Opportunities for the Visualization of Activity–Travel and Land Use Processes in Space–Time
- Cairo, A. (2012). *The Functional Art*. New Riders.
- Capra, E. P., Ferreira, S. B. L., da Silveira, D. S., Ribeiro, B. B., & Modesto, D. M. (2011). Evaluation of web accessibility from the perspective of functional illiteracy. *Ihc+Clihc*, 280–288.
- Cattell, R. (2010). Scalable SQL and NoSQL data stores. *SIGMOD Record*, 39(4), 12–27.
- Chapleau, R. (2003). Measuring the Internal Quality of a CATI Travel Household Survey. Presented at the *International Conference on Transport Survey Quality and Innovation*.

- Citylab. (n.d.). Citylab. Retrieved March 10, 2016, from <http://www.citylab.com/>
- Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*, 13(6), 377–387.
- Couper, M. P. (2001). The promises and perils of Web surveys. In A. Westlake (Ed.), *The Challenge of the Internet* (pp. 35–56). London: Association for Survey Computing.
- Couper, M. P. (2011). The Future of Modes of Data Collection. *Public Opinion Quarterly*, 75(5), 889–908. <http://doi.org/10.1093/poq/nfr046>
- De Leeuw, E. D., & Hox, J. J. (2011). Internet Surveys as Part of a Mixed-Mode Design. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet* (Routledge, pp. 45–76). New York: Routledge. Retrieved from <http://joophox.net/publist/DeLeeuwHox2011.pdf>
- Dibbelt, J., Pajor, T., Strasser, B., & Wagner, D. (2013). Intriguingly simple and fast transit routing. *Experimental Algorithms*.
- Dillman, D. A. (2011). *Mail and Internet Surveys: The Tailored Design Method -- 2007 Update with New Internet, Visual, and Mixed-Mode Guide*. John Wiley & Sons.
- Enyo JavaScript Application Framework. (n.d.). Enyo JavaScript Application Framework. Retrieved December 24, 2012, from <http://enyojs.com>
- Ford, N. (2006, December 5). Polyglot Programming. Retrieved September 25, 2015, from <http://memeagora.blogspot.ca/2006/12/polyglot-programming.html>
- Fry, B., Reas, C., & Shiffman, D. (n.d.). Processing. Retrieved September 1, 2015, from <http://processing.org>
- Geospatial, O. (n.d.). OGC Open Geospatial. Retrieved September 9, 2015, from <http://www.opengeospatial.org/>
- Goritz, A. S. (2006). *Incentives in Web Studies: Methodological Issues and a Review* (pp. 1–13).
- Groves, R. M., Berry, M., & Mathiowetz, N. (1980). Some impacts of computer assisted telephone interviewing on survey methods. *Proceedings of the Joint Statistical Meeting of the American Statistical Association, Section on Survey Research Methods, 1979*.
- Gummer, T., & Rossmann, J. (2014). Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review*.
- Gyssens, M., Paredaens, J., Van den Bussche, J., & Van Gucht, D. (1994). A Graph-Oriented Object Database Model. *IEEE Trans. Knowl. Data Eng.* 6(4), 572–586.
- Hadjigeorgiou, C. (2013). RDBMS vs NoSQL: Performance and Scaling Comparison. MSc in High Performance Computing. *The University of Edinburgh*

Han, J., E, H., Le, G., & Du, J. (2011). Survey on NoSQL database (pp. 363–366). Presented at the 2011 6th International Conference on Pervasive Computing and Applications, IEEE.

Hansson, D. H. (n.d.). Ruby On Rails. Retrieved April 20, 2016, from <http://rubyonrails.org/>

Harrower, M. (2004). A Look at the History and Future of Animated Maps. *Cartographica*, 39(3), 33–42.

ISO. (2011). ISO/IEC 9075-1:2011. *iso.org*.

Jukic, N. (2006). Modeling Strategies And Alternatives For Data Warehousing Projects. *Communications of the ACM*, 49, 1–8.

Karlsruher Institut für Technologie. (n.d.). Fast and Exact Route Planning. Retrieved September 8, 2015, from <http://algo2.iti.kit.edu/routeplanning.php>

Kelly, J. (2008). Computer-Assisted Telephone Interviewing (CATI). *Encyclopedia of Survey Research Methods*. SAGE. Retrieved from <http://srmo.sagepub.com/view/encyclopedia-of-survey-research-methods/n83.xml>

Kimball, R. (1997). A dimensional modeling manifesto. *Dbms*, 10, 58–70.

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit*. John Wiley & Sons.

Kreuter, F. (2013). *Improving Surveys with Paradata*. John Wiley & Sons.

Kreuter, F., Couper, M. P., & Lyberg, L. E. (2010). The use of paradata to monitor and manage survey data collection. *Section on Survey Research Methods – JSM 2010*

Leeuw, E. D., Hox, J. J., & Snijders, G. (1998). The effect of computer-assisted interviewing on data quality. A review. *Journal of the Market Research Society*, 37, 4, pp. 325-344

Lim, H., Fan, B., G. Andersen, D., & Kaminsky, M. (2011). SILT: a memory-efficient, high-performance key-value store. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles (SOSP '11)*. ACM, New York, USA, 1-13.

Luxen, D., & Vetter, C. (2011). Real-time routing with OpenStreetMap data (p. 513). Presented at the GIS '11: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, USA: ACM.

Lynn, P., & Kaminska, O. (2013). The Impact of Mobile Phones on Survey Measurement Error. *Public Opinion Quarterly*, 77(2), 586–605.

MacDonald, A. (2016, November 21). Polyglot Programming: What Is It and Why Should You Be Using It? Retrieved September 25, 2015, from <http://radar.oreilly.com/2013/11/polyglot-programming-what-is-it-and-why-should-you-be-using-it.html>

MacEachren, A. M., & Kraak, M.-J. (2001). Research Challenges in Geovisualization. *Cartography and Geographic Information Science*, 28(1), 3–12.

- Manoochehri, M. (2013). *Data Just Right*. Addison-Wesley.
- Meenakshi, S. (2015). Ruby on Rails – An Agile Developer’s Framework. *International Journal of Computer Applications*, 112(0975-8887), 1–5.
- Meirelles, I. (2013). *Design for Information*. Rockport Pub.
- Mohammadian, A., & Zhang, Y. (2007). Investigating transferability of national household travel survey data. *Transportation Research Record: Journal of the Transportation Research Board*, 1993, 67–79.
- Morency, C. (2008). Enhancing the Travel Survey Process and Data Using the CATI System. *Transportation Planning and Technology*, 31(2), 229–248.
- Olson, K. (2006). Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias. *Public Opinion Quarterly*, 70(5), 737–758.
- Online Survey Software | SurveyGizmo - Affordable Enterprise Survey Software. (n.d.). Online Survey Software | SurveyGizmo - Affordable Enterprise Survey Software. Retrieved December 24, 2012, from <http://surveygizmo.com>
- OpenStreetMap. (n.d.). OpenStreetMap. Retrieved December 24, 2012, from <http://openstreetmap.org>
- Pack, M. L. (2010). Visualization in transportation: challenges and opportunities for everyone., 30(4), 90–96.
- Páez A., Gallo, J., Buliung, R. N., & Dall'erba, S., *Progress in Spatial Analysis* (pp. 119–147). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Palit, C., & Sharp, H. (1983). Microcomputer-Assisted Telephone Interviewing. *Sociological Methods & Research*, 12(2), 169–189.
- PhoneGap. (n.d.). PhoneGap. Retrieved December 24, 2012, from <http://phonegap.com>
- Porter, S. R., & Whitcomb, M. E. (2003). The Impact of Lottery Incentives on Student Survey Response Rates. *Research in Higher Education*, 44(4), 389–407.
- PostGIS Development Group. (n.d.). PostGIS. Retrieved December 24, 2012, from <http://postgis.org>
- PostgreSQL: The world's most advanced open source database. (n.d.). PostgreSQL: The world's most advanced open source database. Retrieved December 24, 2012, from <http://postgresql.org>
- Rasouli, S. (2014). *Mobile Technologies for Activity-Travel Data Collection and Analysis*. IGI Global.
- Redmond, E., & Wilson, J. R. (2012). *Seven Databases in Seven Weeks*. The Pragmatic Bookshelf.

Riandey, B., & Quaglia, M. (2009). Surveying Hard-to-Reach Groups. In *The Travel Survey Toolkit: Where to From Here?* (pp. 127–144). Emerald.

Rousseau, Guy. (2016). Session 380 New Directions in Travel Surveys: Big Data, Smartphones, and Stated Preference. Standing Committee on Travel Survey Methods (ABJ40). Presented at the Transportation Research Board 95th Annual Meeting, Washington D.C.

Sadalage, P. J., & Fowler, M. (2012). *NoSQL Distilled*. Addison-Wesley.

Schlich, R., & Axhausen, K. W. (2003). Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, 30(1), 13–36.

Schonlau, M. (2015). What do web survey panel respondents answer when asked “Do you have any other comment?” *Survey Methods: Insights From the Field (SMIF)*.

Schüssler, N., Axhausen, K. W. (2008). Identifying Trips and Activities and Their Characteristics from GPS Raw Data Without Further Information.

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2002). *Database System Concepts*.

Sioui, L. (2014). Pour une approche pragmatique et opérationnelle de la mobilité durable: Concept, méthodes et outils.

Sioui, L., Morency, C., & Trépanier, M. (2013). How Carsharing Affects the Travel Behavior of Households: A Case Study of Montréal, Canada. *International Journal of Sustainable Transportation*, 7(1), 52–69.

Statistiques Canada. (2014). *Enquête sur le service téléphonique résidentiel, 2013* (pp. 1–2).

Stenneth, L., Wolfson, O., Yu, P. S., & Xu, B. (2011). Transportation mode detection using mobile phones and GIS information. *Gis*, 54–63.

Stopher, P. R. (2009). *The Travel Survey Toolkit: Where to From Here? Keeping up with a Changing World* (pp. 15–46). Emerald Group Publishing Limited.

Stopher, P. R., Metcalf, H. M. A., National Cooperative Highway Research Program, National Research Council (U.S.). Transportation Research Board. (1996). *Methods for household travel surveys*. Washington, D.C. : National Academy Press.

SurveyMonkey: Free online survey software & questionnaire tool. (n.d.). SurveyMonkey: Free online survey software & questionnaire tool. Retrieved December 24, 2012, from <http://surveymonkey.com>

Timmermans, H. J. P., & Hato, E. (2009). Electronic Instrument Design and User Interfaces for Activity-Based Modeling. In P. Bonnel, M. Lee-Gosselin, J. P. Zmud, & J.-L. Madre (Eds.), *Transport Survey Methods* (pp. 437–461). Emerald.

Top 5 Browsers from Nov 2011 to Nov 2012 | StatCounter Global Stats. (n.d.). Top 5 Browsers from Nov 2011 to Nov 2012 | StatCounter Global Stats. Retrieved December 24, 2012, from

<http://gs.statcounter.com/>

Transports Québec : Le progiciel MADITUC. (n.d.). Transports Québec : Le progiciel MADITUC. Retrieved December 24, 2012, from http://www.mtq.gouv.qc.ca/portal/page/portal/ministere/ministere/recherche_innovation/modelisation_systemes_transport/progiciel_madituc

Trépanier, M. (1999). *Modélisation totalement désagrégée et orientée-objet appliquée aux transports urbains*. Publications de Polytechnique Montréal.

Trépanier, M., Chapleau, R., & Morency, C. (2008). Tools and methods for a transportation household survey. *URISA Journal*, 20.

Tudorica, B. G., & Bucur, C. (2011). A comparison between several NoSQL databases with comments and notes, *RoEduNet International Conference 10th Edition: Networking in Education and Research*. IEEE. 1–5.

United Nations specialized agency for information and communication technologies. (2003). Declaration of Principles. Retrieved January 29, 2016, from <http://www.itu.int/net/wsis/docs/geneva/official/dop.html>

Valiquette, F. (2010, September 29). *Typologie des chaînes de déplacements et modélisation descriptive des systèmes d'activités des personnes*.

van der Wijk, J. (n.d.). Flowing City. Retrieved March 10, 2016, from <http://flowingcity.com/>

Venigalla, M. (2004). Household Travel Survey Data Fusion Issues. Presented at the *Transportation Research Board*.

Verreault, H., & Morency, C. (2014). The Chronic Issue of Proxy Respondent Bias. Presented at the 10th International Conference on Transport Survey Methods.

Vicknair, C., Macias, M., Zhao, Z., Nan, X., Chen, Y., & Wilkins, D. (2010). A comparison of a graph database and a relational database: a data provenance perspective. *ACM Southeast Regional Conference*, 42.

Woolley, T. W. (2003). The p-value, the Bayes/Neyman-Pearson Compromise and the Teaching of Statistical Inference in Introductory Business Statistics (pp. 1–7). Presented at the *Journal of the Academy of Business Education*.

World Wide Web Consortium (W3C). (n.d.). World Wide Web Consortium (W3C). Retrieved December 24, 2012, from <http://w3.org>

Wurman, R. S. (2000). *Information Anxiety 2* (2nd ed.). New York: Que.

Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on GPS track data and Bayesian networks. *Ceus*, 54(C), 14–22.

Zhao, F., Pereira, F. C., Ball, R., & Kim, Y. (2015). Exploratory analysis of a smartphone-based travel survey in Singapore. *Transportation Research Record: Journal of the Transportation Research Board*, 2, 45–56.

Zong, F., Bai, Y., Wang, X., Yuan, Y., & He, Y. (2015). Identifying Travel Mode with GPS Data Using Support Vector Machines and Genetic Algorithm. *Information*, 6(2), 212–227.

ANNEXE A ÉCHANTILLON DE COMMENTAIRES POSITIFS, NÉGATIFS ET NEUTRES CONCERNANT L'INTERFACE ET LE QUESTIONNAIRE

Note: les commentaires présentés ont été tirés aléatoirement dans la banque de 2720 commentaires qui concernaient l'interface elle-même.

Échantillon de commentaires positifs

- *Beau visuel, interface agréable!*
- *Très bien développé comme questionnaire!*
- *Visuellement beau et facile de compréhension.*
- *Good job*
- *Questionnaire très clair et bien construit. Facile d'y répondre. Merci!*
- *Pleased with all the factors and options it presented.*
- *Good, simple design*
- *Very neat. Was not frustrating. I enjoyed the questions that were based off my previous answers.*
- *Votre outil de sondage est extrêmement impressionnant. Enfin un outil très, très convivial, simple, intuitif pour la personne qui l'utilise! Wow!*
- *The highlighting green when you've answered a question is really helpful!*
- *Très bon questionnaire!*
- *La collecte des données pour votre étude est extrêmement bien faite. Bravo!*
- *Well organized.*
- *Good survey.*
- *Decently-designed and easy to follow*
- *Bonne idée! Bien montée!*
- *Bravo, excellent format de questionnaire!*
- *Questionnaire très intéressant visuellement, toutes mes félicitations!*
- *Very clear and easy to use.*
- *Questionnaire interactif très bien fait.*
- *It was a well-thought out survey questionnaire*
- *La localisation cartographique pour les points de départ et d'arrivée est très facile d'utilisation. Il en est de même pour toutes les questions du sondage.*
- *Nice initiative. Overall, the survey design is quite clear*
- *It is user friendly!*

- *Enquête adéquate, agréable visuellement.*

Échantillon de commentaires négatifs

- *Tedious*
- *The sliders are not mobile friendly*
- *Prévoir le déplacement du retour inverse au déplacement de l'aller.*
- *The part where we had to recall our day and choose from the map was frustrating because once the location was found. The confirm location button would not pop up and instead I had to go through the options on the side of inputting an address/intersection.*
- *Very cumbersome*
- *Choosing map locations was tricky... You couldn't just click on the map to select, had to enter an address*
- *It was difficult to use the map to pinpoint where I had visited. I felt frustrated and was ready to cancel the survey.*
- *La partie ou il fallait écrire nos déplacements: c'est difficile.*
- *Long et quelque peu répétitif, bien structuré mais lourd à traiter pour un ordinateur, donc beaucoup de bogues*
- *The diary portion of this survey seems a bit tedious to fill out. Exploring ways of simplifying this might benefit the survey participants.*
- *Je ne peux pas modifier un des lieux où je suis allée.*
- *The instructions for the diary part were not super clear.*
- *À la section des véhicules à notre disposition, il ne les liste pas automatiquement. Problématique!*
- *I had difficulty in tracing my home location on the map.*
- *It should be mobile friendly.*
- *Maps didn't work on Chrome, I had to use Safari.*
- *Boring*
- *Quand je clique sur sauvegarder et terminer l'entrevue, ça plante.*
- *Kind of complex specially for the trip journey*
- *Long and annoying*
- *Confusing*
- *L'utilisation de la carte globale de localisation sans zoom n'est pas optimale. J'ai pu me démerder en choisissant des noms de lieu - ce n'était pas 100% intuitif. Pourquoi ne puis-je spécifier que j'étais accompagné à bixi durant mon déplacement ?*
- *Il serait utile de préciser, lors du choix des lieux visités, un choix déplacement pour études et pas seulement pour travail. Aussi, préciser qu'il faut indiquer les destinations finales et non pas les transferts d'un moyen de transport à l'autre*

- *Il devrait y avoir l'option même(s) moyen(s) de transport aller-retour (ou quelque chose du genre), lorsque les déplacements sont les mêmes, pour aller à l'école et revenir à la maison.*
- *It's very noisy*

Échantillon de commentaires neutres

- *Not bad.*
- *It was okay.*
- *Pas toujours évident de répondre mais très bien fait.*
- *À part quelques points difficiles à décortiquer, sondage bien fait.*
- *Très bien fait. Bravo. Cependant, on ne pouvait pas intégrer de déplacement multimodal (ex : métro+communauto, bus+métro, etc)*
- *Très efficace pour la sélection du domicile et la sélection du lieu de déplacement (Polytechnique) déjà pré-mémorisé. Peut être intéressant d'implémenter un trajet aller retour symétrique après avoir saisi le trajet d'aller.*
- *Bonjour. Un peu fastidieux mais bien monté et agréable à remplir. Je suis un grand fan du BIXI et c'est pour cette raison que j'ai répondu. Merci.*
- *L'entrée de l'information est plutôt longue. C'est relativement intuitif, mais très fastidieux.*
- *Average.*
- *Thorough.*
- *Interesting. A bit complex, especially for computer-phobics and people who are not comfortable with technology.*
- *Questionnaire was good, only improvement I would suggest is making it a bit easier to explain the travel diary. Instead of having to pick car for every single trip for example, there should be an obvious way to just select car for every trip.*
- *Quite complicated, but interesting.*
- *Diary was a bit repetitive, wish there was a way to select one mode quickly for all trips.*
- *Questionnaire OK.*
- *Pretty good survey. Maps stuff was a tad glitchy on my phone. I had to use computer instead*
- *I really like your use of the maps. However, it took me a long time to indicate for all the steps of my day that I was always walking. Perhaps, it would be a good idea to have a main mode of transportation option and have drop down boxes*
- *Détaillé, bien fait; un peu de confusion sur les allers-retours des itinéraires de la journée sélectionnée, mais dans l'ensemble on sent bien la rigueur de la démarche.*
- *Pas mal*
- *Intéressante enquête. Juste la phase des modes de transports un peu rébarbative. Cette étape peut sûrement être améliorée pour devenir plus rapide et intuitive.*
- *C'est bien fait l'interface! J'ai eu un petit problème en ajoutant plusieurs modes de transports, il a fallu appuyer sur ESC pour procéder à la sélection du deuxième mode*

- *It was okay.*
- *It's ok.*
- *Très bien fait. Je suggère d'ajouter l'option même mode pour les déplacements semblables retour.*

ANNEXE B TABLES DE L'ENTREPÔT DE DONNÉES

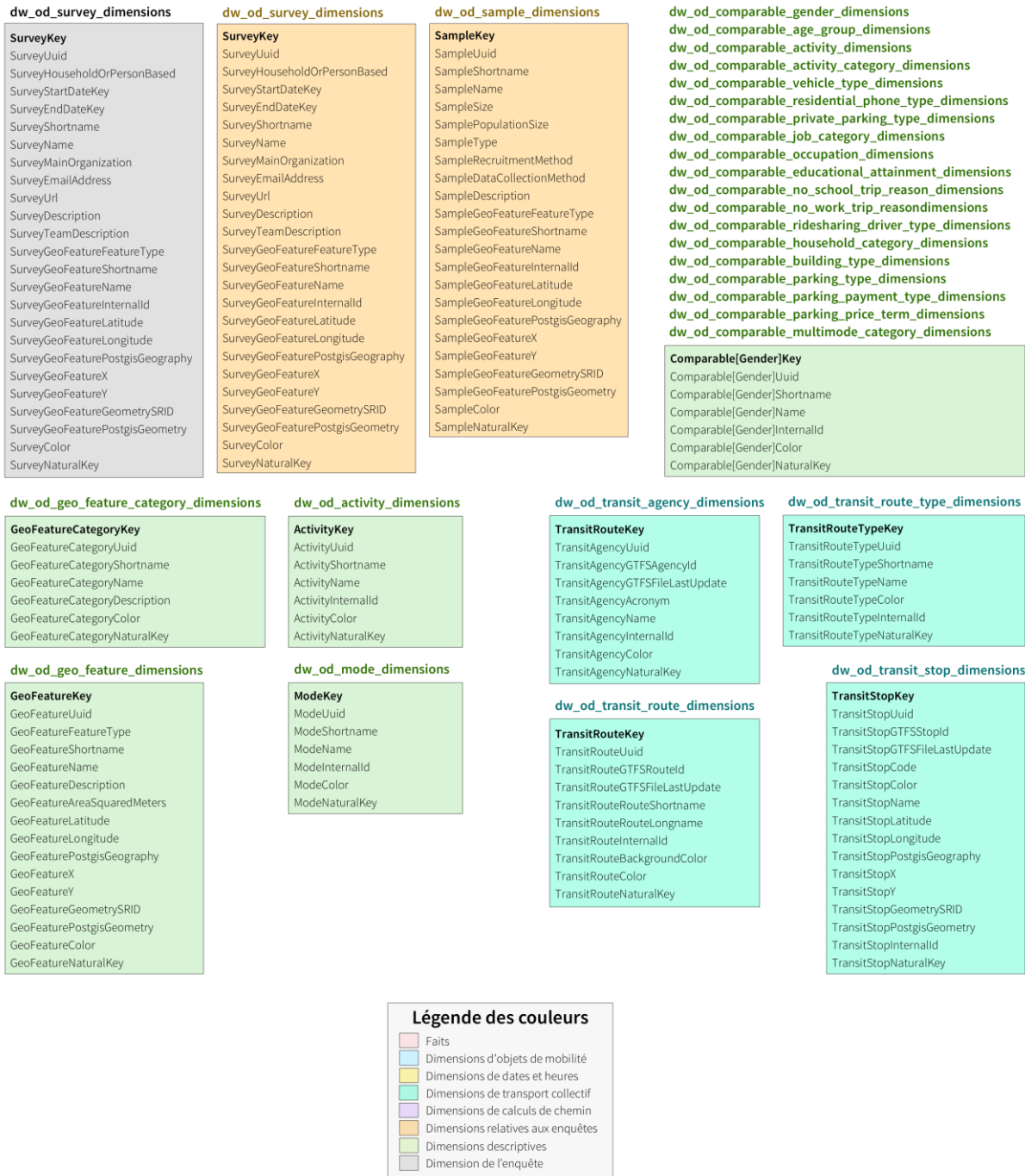


Figure B-1 : Schéma des tables de l'entrepôt de données

dw_od_usual_school_place_dimensions

UsualSchoolPlaceKey

- UsualSchoolPlaceUuid
- UsualSchoolPlaceDeclaredName
- UsualSchoolPlaceDeclaredFullStreetAddress
- UsualSchoolPlaceDeclaredApartmentNumber
- UsualSchoolPlaceDeclaredCivicNumber
- UsualSchoolPlaceDeclaredStreetName
- UsualSchoolPlaceDeclaredStreetType
- UsualSchoolPlaceDeclaredCity
- UsualSchoolPlaceDeclaredRegion
- UsualSchoolPlaceDeclaredCountry
- UsualSchoolPlaceDeclaredPostalCode
- UsualSchoolPlaceCorrectedName
- UsualSchoolPlaceCorrectedFullStreetAddress
- UsualSchoolPlaceCorrectedApartmentNumber
- UsualSchoolPlaceCorrectedCivicNumber
- UsualSchoolPlaceCorrectedStreetName
- UsualSchoolPlaceCorrectedStreetType
- UsualSchoolPlaceCorrectedCity
- UsualSchoolPlaceCorrectedRegion
- UsualSchoolPlaceCorrectedCountry
- UsualSchoolPlaceCorrectedPostalCode
- UsualSchoolPlaceType
- UsualSchoolPlaceTypeShortname
- UsualSchoolPlaceTypeInternalId
- UsualSchoolPlaceNumberOfStudents
- UsualSchoolPlaceLatitude
- UsualSchoolPlaceLongitude
- UsualSchoolPlacePostgisGeography
- UsualSchoolPlaceX
- UsualSchoolPlaceY
- UsualSchoolPlaceGeometrySRID
- UsualSchoolPlacePostgisGeometry
- UsualSchoolPlaceNaturalKey

dw_od_usual_work_place_dimensions

UsualWorkPlaceKey

- UsualWorkPlaceUuid
- UsualWorkPlaceDeclaredName
- UsualWorkPlaceDeclaredFullStreetAddress
- UsualWorkPlaceDeclaredApartmentNumber
- UsualWorkPlaceDeclaredCivicNumber
- UsualWorkPlaceDeclaredStreetName
- UsualWorkPlaceDeclaredStreetType
- UsualWorkPlaceDeclaredCity
- UsualWorkPlaceDeclaredRegion
- UsualWorkPlaceDeclaredCountry
- UsualWorkPlaceDeclaredPostalCode
- UsualWorkPlaceCorrectedName
- UsualWorkPlaceCorrectedFullStreetAddress
- UsualWorkPlaceCorrectedApartmentNumber
- UsualWorkPlaceCorrectedCivicNumber
- UsualWorkPlaceCorrectedStreetName
- UsualWorkPlaceCorrectedStreetType
- UsualWorkPlaceCorrectedCity
- UsualWorkPlaceCorrectedRegion
- UsualWorkPlaceCorrectedCountry
- UsualWorkPlaceCorrectedPostalCode
- UsualWorkPlaceType
- UsualWorkPlaceTypeShortname
- UsualWorkPlaceTypeInternalId
- UsualWorkPlaceSector
- UsualWorkPlaceSectorShortname
- UsualWorkPlaceSectorInternalId
- UsualWorkPlaceNumberOfEmployees
- UsualWorkPlaceLatitude
- UsualWorkPlaceLongitude
- UsualWorkPlacePostgisGeography
- UsualWorkPlaceX
- UsualWorkPlaceY
- UsualWorkPlaceGeometrySRID
- UsualWorkPlacePostgisGeometry
- UsualWorkPlaceNaturalKey

dw_od_place_dimensions

PlaceKey

- PlaceUuid
- PlaceTypeShortname
- PlaceTypeName
- PlaceCategoryShortname
- PlaceCategoryName
- PlaceDescription
- PlaceDeclaredName
- PlaceDeclaredFullStreetAddress
- PlaceDeclaredApartmentNumber
- PlaceDeclaredCivicNumber
- PlaceDeclaredStreetName
- PlaceDeclaredStreetType
- PlaceDeclaredCity
- PlaceDeclaredRegion
- PlaceDeclaredCountry
- PlaceDeclaredPostalCode
- PlaceCorrectedName
- PlaceCorrectedFullStreetAddress
- PlaceCorrectedApartmentNumber
- PlaceCorrectedCivicNumber
- PlaceCorrectedStreetName
- PlaceCorrectedStreetType
- PlaceCorrectedCity
- PlaceCorrectedRegion
- PlaceCorrectedCountry
- PlaceCorrectedPostalCode
- PlaceLatitude
- PlaceLongitude
- PlacePostgisGeography
- PlaceX
- PlaceY
- PlaceGeometrySRID
- PlacePostgisGeometry
- PlaceNaturalKey

dw_od_home_dimensions

HomeKey

- HomeUuid
- HomeDeclaredFullStreetAddress
- HomeDeclaredApartmentNumber
- HomeDeclaredCivicNumber
- HomeDeclaredStreetName
- HomeDeclaredStreetType
- HomeDeclaredCity
- HomeDeclaredRegion
- HomeDeclaredCountry
- HomeDeclaredPostalCode
- HomeCorrectedFullStreetAddress
- HomeCorrectedApartmentNumber
- HomeCorrectedCivicNumber
- HomeCorrectedStreetName
- HomeCorrectedStreetType
- HomeCorrectedCity
- HomeCorrectedRegion
- HomeCorrectedCountry
- HomeCorrectedPostalCode
- HomeLatitude
- HomeLongitude
- HomePostgisGeography
- HomeX
- HomeY
- HomeGeometrySRID
- HomePostgisGeometry
- HomeNaturalKey

dw_od_date_dimensions

DateKey

- DateFullDate
- DateDayOfWeek
- DateDayNumInMonth
- DateDayNumInYear
- DateDayNumOverall
- DateDayName
- DateDayAbbrev
- DateWeekdayOrWeekend
- DatelsWeekday
- DatelsWeekend
- DateWeekNumInYear
- DateWeekNumOverall
- DateWeekBeginDate
- DateWeekBeginDateKey
- DateMonth
- DateMonthNumOverall
- DateMonthName
- DateMonthAbbrev
- DateQuarter
- DateYear
- DateYearMonth
- DateFiscalMonth
- DateFiscalQuarter
- DateFiscalYear
- DateLastDayInMonthOrNot
- DatelsLastDayInMonth
- DateSameDayAYearAgoDate

dw_od_time_dimensions

TimeKey

- TimeFullTime
- TimeHour24
- TimeHour
- TimeHour12
- TimeHour24ZeroPadded
- TimeHourPad
- TimeHour12ZeroPadded
- TimeHourMin24
- TimeHourMin
- TimeHourMin12
- TimeHourMin24ZeroPadded
- TimeHourMinPad
- TimeHourMin12ZeroPadded
- TimeHourMinSec24
- TimeHourMinSec
- TimeHourMinSec12
- TimeHourMinSec24ZeroPadded
- TimeHourMinSecPad
- TimeHourMinSec12ZeroPadded
- TimeMin
- TimeMinZeroPadded
- TimeMinAfterMidnight
- TimeSec
- TimeSecZeroPadded
- TimeSecAfterMidnight
- TimeAmPm

dw_od_time_extended_dimensions

TimeKey

- TimeHour24
- TimeHour
- TimeHour24ZeroPadded
- TimeHourPad
- TimeHourMin24
- TimeHourMin
- TimeHourMin24ZeroPadded
- TimeHourMinPad
- TimeHourMinSec24
- TimeHourMinSec
- TimeHourMinSec24ZeroPadded
- TimeHourMinSecZeroPadded
- TimeMin
- TimeMinZeroPadded
- TimeMinAfterMidnight
- TimeSec
- TimeSecZeroPadded
- TimeSecAfterMidnight

dw_od_vehicle_dimensions

VehicleKey

- VehicleUuid
- VehicleDeclaredName
- VehicleModelYear
- VehicleModel
- VehicleModelInternalId
- VehicleMake
- VehicleMakeInternalId
- VehicleType
- VehicleTypeInternalId
- VehicleComparableVehicleType
- VehicleComparableVehicleTypeCategory
- VehicleNumberOfSeats
- VehicleOdometerKm
- VehicleOdometerMi
- VehicleFuelEconomyLPer100Km
- VehicleFuelEconomyMPG
- VehicleResidentialParkingPermitOwner
- VehicleResidentialParkingPermitOwnerFlag
- VehicleResidentialParkingPermitName
- VehicleResidentialParkingPermitShortname
- VehicleResidentialParkingPermitInternalId
- VehicleResidentialParkingPermitIdentifier
- VehicleEmployeeParkingPermitOwner
- VehicleEmployeeParkingPermitOwnerFlag
- VehicleEmployeeParkingPermitName
- VehicleEmployeeParkingPermitShortname
- VehicleEmployeeParkingPermitInternalId
- VehicleEmployeeParkingPermitIdentifier
- VehicleNaturalKey

Figure B-1 : Schéma des tables de l'entrepôt de données (suite)

dw_od_routing_dimensions

RoutingKey
RoutingUuid	RoutingTransitDistanceKm	RoutingParkRideTransfersWalkingDistanceKm
RoutingOriginLatitude	RoutingTransitDistanceMi	RoutingParkRideTransfersWalkingDistanceMi
RoutingOriginLongitude	RoutingTransitTortuosityRatio	RoutingParkRideInVehicleTravelTimeSec
RoutingOriginPostgisGeography	RoutingTransitAccessFirstWalkingTravelTimeSec	RoutingParkRideInVehicleTravelTimeMin
RoutingOriginX	RoutingTransitAccessFirstWalkingTravelTimeMin	RoutingParkRideInVehicleDistanceM
RoutingOriginY	RoutingTransitAccessFirstDemoAdjustedWalkingTravelTimeSec	RoutingParkRideInVehicleDistanceKm
RoutingOriginGeometrySRID	RoutingTransitAccessFirstDemoAdjustedWalkingTravelTimeMin	RoutingParkRideInVehicleDistanceMi
RoutingOriginPostgisGeometry	RoutingTransitAccessFirstWalkingDistanceM	RoutingParkRideFirstStopWaitingTimeSec
RoutingDestinationLatitude	RoutingTransitAccessFirstWalkingDistanceKm	RoutingParkRideFirstStopWaitingTimeMin
RoutingDestinationLongitude	RoutingTransitAccessFirstWalkingDistanceMi	RoutingParkRideTransfersWaitingTimeSec
RoutingDestinationPostgisGeography	RoutingTransitAccessLastWalkingTravelTimeSec	RoutingParkRideTransfersWaitingTimeMin
RoutingDestinationX	RoutingTransitAccessLastWalkingTravelTimeMin	RoutingParkRideTotalWaitingTimeSec
RoutingDestinationY	RoutingTransitAccessLastDemoAdjustedWalkingTravelTimeSec	RoutingParkRideTotalWaitingTimeMin
RoutingDestinationGeometrySRID	RoutingTransitAccessLastDemoAdjustedWalkingTravelTimeMin	RoutingParkRideNumberOfTransfers
RoutingDestinationPostgisGeometry	RoutingTransitAccessLastWalkingDistanceM	RoutingParkRideAgencies
RoutingBirdDistanceM	RoutingTransitAccessLastWalkingDistanceKm	RoutingParkRideStops
RoutingBirdDistanceKm	RoutingTransitAccessLastWalkingDistanceMi	RoutingParkRideRoutes
RoutingBirdDistanceMi	RoutingTransitTransfersWalkingTravelTimeSec	RoutingParkRideRouteTypes
RoutingDepartureDateKey	RoutingTransitTransfersWalkingTravelTimeMin	RoutingParkRidePostgisGeography
RoutingDepartureTimeKey	RoutingTransitTransfersDemoAdjustedWalkingTravelTimeSec	RoutingParkRideGeometrySRID
RoutingDepartureDateExtendedKey	RoutingTransitTransfersDemoAdjustedWalkingTravelTimeMin	RoutingParkRidePostgisGeometry
RoutingDepartureTimeExtendedKey	RoutingTransitTransfersWalkingDistanceM	RoutingBikeRideArrivalDateKey
RoutingWalkingArrivalDateKey	RoutingTransitTransfersWalkingDistanceKm	RoutingBikeRideArrivalTimeKey
RoutingWalkingArrivalTimeKey	RoutingTransitTransfersWalkingDistanceMi	RoutingBikeRideArrivalDateExtendedKey
RoutingWalkingArrivalDateExtendedKey	RoutingTransitInVehicleTravelTimeSec	RoutingBikeRideArrivalTimeExtendedKey
RoutingWalkingArrivalTimeExtendedKey	RoutingTransitInVehicleTravelTimeMin	RoutingBikeRideTravelTimeSec
RoutingWalkingTravelTimeSec	RoutingTransitInVehicleDistanceM	RoutingBikeRideTravelTimeMin
RoutingWalkingTravelTimeMin	RoutingTransitInVehicleDistanceKm	RoutingBikeRideDistanceM
RoutingWalkingDemoAdjustedTravelTimeSec	RoutingTransitInVehicleDistanceMi	RoutingBikeRideDistanceKm
RoutingWalkingDemoAdjustedTravelTimeMin	RoutingTransitFirstStopWaitingTimeSec	RoutingBikeRideDistanceMi
RoutingWalkingDistanceM	RoutingTransitFirstStopWaitingTimeMin	RoutingBikeRideTortuosityRatio
RoutingWalkingDistanceKm	RoutingTransitTransfersWaitingTimeSec	RoutingBikeRideAccessFirstWalkingTravelTimeSec
RoutingWalkingDistanceMi	RoutingTransitTransfersWaitingTimeMin	RoutingBikeRideAccessFirstWalkingTravelTimeMin
RoutingWalkingTortuosityRatio	RoutingTransitTotalWaitingTimeSec	RoutingBikeRideAccessFirstDemoAdjustedWalkingTravelTimeSec
RoutingWalkingPostgisGeography	RoutingTransitTotalWaitingTimeMin	RoutingBikeRideAccessFirstDemoAdjustedWalkingTravelTimeMin
RoutingWalkingGeometrySRID	RoutingTransitNumberOfTransfers	RoutingBikeRideAccessFirstWalkingDistanceM
RoutingWalkingPostgisGeometry	RoutingTransitStops	RoutingBikeRideAccessFirstWalkingDistanceKm
RoutingCyclingArrivalDateKey	RoutingTransitAgencies	RoutingBikeRideAccessFirstWalkingDistanceMi
RoutingCyclingArrivalTimeKey	RoutingTransitRoutes	RoutingBikeRideAccessLastWalkingTravelTimeSec
RoutingCyclingArrivalDateExtendedKey	RoutingTransitRouteTypes	RoutingBikeRideAccessLastWalkingTravelTimeMin
RoutingCyclingArrivalTimeExtendedKey	RoutingTransitPostgisGeography	RoutingBikeRideAccessLastDemoAdjustedWalkingTravelTimeSec
RoutingCyclingTravelTimeSec	RoutingTransitGeometrySRID	RoutingBikeRideAccessLastDemoAdjustedWalkingTravelTimeMin
RoutingCyclingTravelTimeMin	RoutingTransitPostgisGeometry	RoutingBikeRideAccessLastWalkingDistanceM
RoutingCyclingDemoAdjustedTravelTimeSec	RoutingParkRideArrivalDateKey	RoutingBikeRideAccessLastWalkingDistanceKm
RoutingCyclingDemoAdjustedTravelTimeMin	RoutingParkRideArrivalTimeKey	RoutingBikeRideAccessLastWalkingDistanceMi
RoutingCyclingDistanceM	RoutingParkRideArrivalDateExtendedKey	RoutingBikeRideTransfersWalkingTravelTimeSec
RoutingCyclingDistanceKm	RoutingParkRideArrivalTimeExtendedKey	RoutingBikeRideTransfersWalkingTravelTimeMin
RoutingCyclingDistanceMi	RoutingParkRideTravelTimeSec	RoutingBikeRideTransfersDemoAdjustedWalkingTravelTimeSec
RoutingCyclingTortuosityRatio	RoutingParkRideTravelTimeMin	RoutingBikeRideTransfersDemoAdjustedWalkingTravelTimeMin
RoutingCyclingPostgisGeography	RoutingParkRideDistanceM	RoutingBikeRideTransfersWalkingDistanceM
RoutingCyclingGeometrySRID	RoutingParkRideDistanceKm	RoutingBikeRideTransfersWalkingDistanceKm
RoutingCyclingPostgisGeometry	RoutingParkRideDistanceMi	RoutingBikeRideTransfersWalkingDistanceMi
RoutingDrivingArrivalDateKey	RoutingParkRideTortuosityRatio	RoutingBikeRideInVehicleTravelTimeSec
RoutingDrivingArrivalTimeKey	RoutingParkRideAccessFirstWalkingTravelTimeSec	RoutingBikeRideInVehicleTravelTimeMin
RoutingDrivingArrivalDateExtendedKey	RoutingParkRideAccessFirstWalkingTravelTimeMin	RoutingBikeRideInVehicleDistanceM
RoutingDrivingArrivalTimeExtendedKey	RoutingParkRideAccessFirstDemoAdjustedWalkingTravelTimeSec	RoutingBikeRideInVehicleDistanceKm
RoutingDrivingTravelTimeSec	RoutingParkRideAccessFirstDemoAdjustedWalkingTravelTimeMin	RoutingBikeRideInVehicleDistanceMi
RoutingDrivingTravelTimeMin	RoutingParkRideAccessFirstWalkingDistanceM	RoutingBikeRideFirstStopWaitingTimeSec
RoutingDrivingDistanceM	RoutingParkRideAccessFirstWalkingDistanceKm	RoutingBikeRideFirstStopWaitingTimeMin
RoutingDrivingDistanceKm	RoutingParkRideAccessFirstWalkingDistanceMi	RoutingBikeRideTransfersWaitingTimeSec
RoutingDrivingDistanceMi	RoutingParkRideAccessLastWalkingTravelTimeSec	RoutingBikeRideTransfersWaitingTimeMin
RoutingDrivingTortuosityRatio	RoutingParkRideAccessLastWalkingTravelTimeMin	RoutingBikeRideTotalWaitingTimeSec
RoutingDrivingPostgisGeography	RoutingParkRideAccessLastDemoAdjustedWalkingTravelTimeSec	RoutingBikeRideTotalWaitingTimeMin
RoutingDrivingGeometrySRID	RoutingParkRideAccessLastDemoAdjustedWalkingTravelTimeMin	RoutingBikeRideNumberOfTransfers
RoutingDrivingPostgisGeometry	RoutingParkRideAccessLastWalkingDistanceM	RoutingBikeRideAgencies
RoutingTransitArrivalDateKey	RoutingParkRideAccessLastWalkingDistanceKm	RoutingBikeRideStops
RoutingTransitArrivalTimeKey	RoutingParkRideAccessLastWalkingDistanceMi	RoutingBikeRideRoutes
RoutingTransitArrivalDateExtendedKey	RoutingParkRideTransfersWalkingTravelTimeSec	RoutingBikeRideRouteTypes
RoutingTransitArrivalTimeExtendedKey	RoutingParkRideTransfersWalkingTravelTimeMin	RoutingBikeRidePostgisGeography
RoutingTransitTravelTimeSec	RoutingParkRideTransfersDemoAdjustedWalkingTravelTimeSec	RoutingBikeRideGeometrySRID
RoutingTransitTravelTimeMin	RoutingParkRideTransfersDemoAdjustedWalkingTravelTimeMin	RoutingBikeRidePostgisGeometry
RoutingTransitDistanceM	RoutingParkRideTransfersWalkingDistanceM	
...	...	

Figure B-1 : Schéma des tables de l'entrepôt de données (suite)

dw_od_household_facts

HouseholdKey
HouseholdUuid
HouseholdDimensionKey
HouseholdSingleWeight
HouseholdSurveyKey
HouseholdSampleKey
HouseholdHomeKey
HouseholdTripsDateKey
HouseholdComparableHouseholdCategoryKey
HouseholdComparableBuildingTypeKey
HouseholdComparableResidentialPhoneTypeKey
HouseholdComparablePrivateParkingTypeKey
HouseholdIsSingleHouseholdFlag
HouseholdOwnershipFlag
HouseholdInternetAccessFlag
HouseholdHighSpeedInternetAccessFlag
HouseholdResidentialPhoneOwnerFlag
HouseholdResidentialPhoneInPhonebookFlag
HouseholdWouldLikeToParticipateToOtherSurveysFlag
HouseholdPrivateParkingAvailabilityFlag
HouseholdNaturalKey

dw_od_household_dimensions

HouseholdKey	...
HouseholdSingleWeight	HouseholdNumberOfPartTimeWorkers
HouseholdSize	HouseholdNumberOfPartTimeStudents
HouseholdCarNumber	HouseholdNumberOfRetired
HouseholdIsSingleHouseholdFlag	HouseholdNumberOfInterviewable
HouseholdIsSingleHousehold	HouseholdNumberOfInterviewed
HouseholdTripsWeekday	HouseholdNumberOfNonInterviewable
HouseholdCategory	HouseholdNumberOfNonInterviewed
HouseholdCategoryShortname	HouseholdNumberOfMobileInterviewed
HouseholdCategoryInternalId	HouseholdNumberOfNonMobileInterviewed
HouseholdComparableHouseholdCategory	HouseholdNumberOfYoungerThan5
HouseholdIncomeLevel	HouseholdNumberOfYoungerThan16
HouseholdIncomeLevelInternalId	HouseholdNumberOfYoungerThan18
HouseholdOwnershipFlag	HouseholdNumberOfYoungerThan21
HouseholdOwnership	HouseholdNumberOfYoungerThan65
HouseholdBuildingType	HouseholdNumberOfYoungerThan75
HouseholdBuildingTypeInternalId	HouseholdNumberOfOlderThanOrEqualTo5
HouseholdComparableBuildingType	HouseholdNumberOfOlderThanOrEqualTo16
HouseholdInternetAccess	HouseholdNumberOfOlderThanOrEqualTo18
HouseholdInternetAccessFlag	HouseholdNumberOfOlderThanOrEqualTo21
HouseholdHighSpeedInternetAccess	HouseholdNumberOfOlderThanOrEqualTo65
HouseholdHighSpeedInternetAccessFlag	HouseholdNumberOfOlderThanOrEqualTo75
HouseholdResidentialPhoneOwnerFlag	HouseholdTotalNumberOfTripsForInterviewed
HouseholdResidentialPhoneOwner	HouseholdTotalTripsBirdDistanceForInterviewedM
HouseholdResidentialPhoneOwnerInternalId	HouseholdTotalTripsBirdDistanceForInterviewedKm
HouseholdComparableResidentialPhoneOwner	HouseholdTotalTripsBirdDistanceForInterviewedMi
HouseholdResidentialPhoneType	HouseholdAverageNumberOfTripsForInterviewed
HouseholdResidentialPhoneTypeInternalId	HouseholdAverageTripsBirdDistanceForInterviewedM
HouseholdComparableResidentialPhoneType	HouseholdAverageTripsBirdDistanceForInterviewedKm
HouseholdResidentialPhoneInPhonebookFlag	HouseholdAverageTripsBirdDistanceForInterviewedMi
HouseholdResidentialPhoneInPhonebook	HouseholdNumberOfInterviewedStartingAwayFromHome
HouseholdResidentialPhoneInPhonebookInternalId	HouseholdNumberOfInterviewedEndingAwayFromHome
HouseholdComparableResidentialPhoneInPhonebook	HouseholdNumberOfAvailableVehicles
HouseholdResidentialPhoneProvider	HouseholdNumberOfOwnedVehicles
HouseholdResidentialPhoneProviderInternalId	HouseholdMinimumNumberOfNeededVehiclesAtHome
HouseholdWouldLikeToParticipateToOtherSurveysFlag	HouseholdMinimumNumberOfMobileVehicles
HouseholdWouldLikeToParticipateToOtherSurveys	HouseholdMaximumNumberOfMobileVehicles
HouseholdEmailAddress	HouseholdNumberofMobileVehicles
HouseholdNumberOfFemales	HouseholdMinimumNumberOfNonMobileVehicles
HouseholdNumberOfMales	HouseholdMaximumNumberOfNonMobileVehicles
HouseholdAverageAge	HouseholdNumberofNonMobileVehicles
HouseholdAgeOfYoungest	HouseholdPrivateParkingAvailabilityFlag
HouseholdAgeOfOldest	HouseholdPrivateParkingAvailability
HouseholdAgeDifferenceBetweenYoungestAndOldest	HouseholdPrivateParkingType
HouseholdNumberOfWorkers	HouseholdPrivateParkingTypeInternalId
HouseholdNumberOfStudents	HouseholdComparablePrivateParkingType
HouseholdNumberOfFullTimeWorkers	HouseholdNumberOfPrivateParkingSpaces
HouseholdNumberOfFullTimeStudents	HouseholdCommentOnSurvey
...	HouseholdNaturalKey

Figure B-1 : Schéma des tables de l'entrepôt de données (suite)

dw_od_person_facts

PersonKey
PersonUuid
PersonDimensionKey
PersonSingleWeight
PersonSurveyKey
PersonSampleKey
PersonHouseholdKey
PersonHomeKey
PersonUsualWorkPlaceKey
PersonUsualSchoolPlaceKey
PersonRespondentPersonKey
PersonHomeToUsualSchoolPlaceRoutingKey
PersonHomeToUsualWorkPlaceRoutingKey
PersonUsualSchoolPlaceToHomeRoutingKey
PersonUsualWorkPlaceToHomeRoutingKey
PersonComparableAgeGroupKey
PersonComparableGenderKey
PersonComparableOccupationKey
PersonComparableJobCategoryKey
PersonComparableAgeGroupOfFirstDrivingLicenseKey
PersonComparableAgeGroupOfFirstFullTimeJobKey
PersonComparableNoSchoolTripReasonKey
PersonComparableNoWorkTripReasonKey
PersonComparableEducationalAttainmentKey
PersonComparableAgeGroupOfEduAttainmentObtentionKey
PersonInterviewableFlag
PersonSelfReportingFlag
PersonWorkerFlag
PersonStudentFlag
PersonFullTimeWorkerFlag
PersonFullTimeStudentFlag
PersonPartTimeWorkerFlag
PersonPartTimeStudentFlag
PersonOnTheRoadWorkerFlag
PersonWorkAtHomeFlag
PersonDrivingLicenseOwnerFlag
PersonTransitPassOwnerFlag
PersonBikesharingMemberFlag
PersonCarsharingMemberFlag
PersonRidesharingMemberFlag
PersonCellphoneOwnerFlag
PersonIntelligentPhoneOwnerFlag
PersonTabletOwnerFlag
PersonLaptopOwnerFlag
PersonStartingTripsAwayFromHomeFlag
PersonEndingTripsAwayFromHomeFlag
PersonWouldLikeToParticipateToOtherSurveysFlag
PersonNaturalKey

dw_od_person_dimensions

PersonKey	...
PersonSingleWeight	PersonLaptopOwner
PersonInterviewableFlag	PersonStartingTripsAwayFromHomeFlag
PersonInterviewable	PersonStartingTripsAwayFromHome
PersonInterviewLanguage	PersonEndingTripsAwayFromHomeFlag
PersonSelfReportingFlag	PersonEndingTripsAwayFromHome
PersonSelfReportingOrProxy	PersonComparableNoTripReason
PersonAge	PersonNoSchoolTripReason
PersonAgeGroup	PersonNoSchoolTripReasonInternalId
PersonAgeGroupInternalId	PersonComparableNoSchoolTripReason
PersonComparableAgeGroup	PersonNoWorkTripReason
PersonGenderInternalId	PersonNoWorkTripReasonInternalId
PersonComparableGender	PersonComparableNoWorkTripReason
PersonOccupation	PersonBirdDistanceHomeToUsualSchoolPlaceM
PersonOccupationInternalId	PersonBirdDistanceHomeToUsualSchoolPlaceKm
PersonComparableOccupation	PersonBirdDistanceHomeToUsualSchoolPlaceMi
PersonJobCategory	PersonBirdDistanceHomeToUsualWorkPlaceM
PersonJobCategoryInternalId	PersonBirdDistanceHomeToUsualWorkPlaceKm
PersonComparableJobCategory	PersonBirdDistanceHomeToUsualWorkPlaceMi
PersonJobName	PersonEducationalAttainment
PersonWorkerFlag	PersonEducationalAttainmentInternalId
PersonWorker	PersonComparableEducationalAttainment
PersonStudentFlag	PersonAgeOfEducationalAttainmentObtention
PersonStudent	PersonComparableAgeGroupOfEduAttainmentObtention
PersonFullTimeWorkerFlag	PersonNumberOfOwnedVehicles
PersonFullTimeWorker	PersonNumberOfAvailableVehicles
PersonFullTimeStudentFlag	PersonNumberOfTripChains
PersonFullTimeStudent	PersonNumberOfJourneys
PersonPartTimeWorkerFlag	PersonNumberOfVisitedPlaces
PersonPartTimeWorker	PersonNumberOfSegments
PersonPartTimeStudentFlag	PersonNumberOfTrips
PersonPartTimeStudent	PersonNumberOfTransitTrips
PersonOnTheRoadWorkerFlag	PersonNumberOfCarDrivingTrips
PersonOnTheRoadWorker	PersonNumberOfCarPassengerTrips
PersonWorkAtHomeFlag	PersonNumberOfCyclingTrips
PersonWorkAtHome	PersonNumberOfWalkingTrips
PersonDrivingLicenseOwnerFlag	PersonNumberOfParkAndRideTrips
PersonDrivingLicenseOwner	PersonNumberOfKissAndRideTrips
PersonAgeOfFirstDrivingLicense	PersonNumberOfBikeAndRideTrips
PersonComparableAgeGroupOfFirstDrivingLicense	PersonNumberOfMotorcycleTrips
PersonAgeOfFirstFullTimeJob	PersonNumberOfTaxiTrips
PersonComparableAgeGroupOfFirstFullTimeJob	PersonNumberOfParatransitTrips
PersonTransitPassOwnerFlag	PersonNumberOfSchoolBusTrips
PersonTransitPassOwner	PersonNumberOfIntercityTransitTrips
PersonTransitPasses	PersonNumberOfPlaneTrips
PersonTransitPassesInternalIds	PersonTotalTripsBirdDistanceM
PersonTransitPassesIdentifiers	PersonTotalTripsBirdDistanceKm
PersonBikesharingMemberFlag	PersonTotalTripsBirdDistanceMi
PersonBikesharingMember	PersonAverageTripsBirdDistanceM
PersonBikesharingMemberships	PersonAverageTripsBirdDistanceKm
PersonBikesharingMembershipsInternalIds	PersonAverageTripsBirdDistanceMi
PersonBikesharingMembershipsIdentifiers	PersonTotalTripsDeclaredDurationSec
PersonCarsharingMemberFlag	PersonTotalTripsDeclaredDurationMin
PersonCarsharingMember	PersonAverageTripsDeclaredDurationSec
PersonCarsharingMemberships	PersonAverageTripsDeclaredDurationMin
PersonCarsharingMembershipsInternalIds	PersonMinimumWalkedDistanceDayM
PersonCarsharingMembershipsIdentifiers	PersonMinimumWalkedDistanceDayKm
PersonRidesharingMemberFlag	PersonMinimumWalkedDistanceDayMi
PersonRidesharingMember	PersonMinimumWalkedDurationDaySec
PersonRidesharingMemberships	PersonMinimumWalkedDurationDayMin
PersonRidesharingMembershipsInternalIds	PersonNickname
PersonRidesharingMembershipsIdentifiers	PersonFirstname
PersonPersonalIncomeLevel	PersonLastname
PersonPersonalIncomeLevelInternalId	PersonFullname
PersonCellphoneOwnerFlag	PersonBirthdateDateKey
PersonCellphoneOwner	PersonEmailAddress
PersonIntelligentPhoneOwnerFlag	PersonWouldLikeToParticipateToOtherSurveysFlag
PersonIntelligentPhoneOwner	PersonWouldLikeToParticipateToOtherSurveys
PersonTabletOwnerFlag	PersonPhoneNumber
PersonTabletOwner	PersonCellphoneNumber
PersonLaptopOwnerFlag	PersonCommentOnSurvey
...	PersonNaturalKey

Figure B-1 : Schéma des tables de l'entrepôt de données (suite)

dw_od_visited_place_facts

VisitedPlaceKey
 VisitedPlaceUuid
 VisitedPlaceDimensionKey
 VisitedPlaceSingleWeight
 VisitedPlaceSequenceInInterviewDay
 VisitedPlaceSurveyKey
 VisitedPlaceSampleKey
 VisitedPlaceHouseholdKey
 VisitedPlaceHomeKey
 VisitedPlacePersonKey
 VisitedPlacePersonUsualWorkPlaceKey
 VisitedPlacePersonUsualSchoolPlaceKey
 VisitedPlaceActivityKey
 VisitedPlaceComparableActivityKey
 VisitedPlaceComparableActivityCategoryKey
 VisitedPlaceArrivalDateKey
 VisitedPlaceArrivalTimeKey
 VisitedPlaceArrivalDateExtendedKey
 VisitedPlaceArrivalTimeExtendedKey
 VisitedPlaceDepartureDateKey
 VisitedPlaceDepartureTimeKey
 VisitedPlaceDepartureDateExtendedKey
 VisitedPlaceDepartureTimeExtendedKey
 VisitedPlacePreviousDepartureDateKey
 VisitedPlacePreviousDepartureTimeKey
 VisitedPlacePreviousDepartureDateExtendedKey
 VisitedPlacePreviousDepartureTimeExtendedKey
 VisitedPlaceNaturalKey

dw_od_visited_place_dimensions

VisitedPlaceKey
 VisitedPlaceSingleWeight
 VisitedPlaceSequenceInInterviewDay
 VisitedPlaceArrivalWeekday
 VisitedPlaceArrivalWeekdayExtended
 VisitedPlaceDepartureWeekday
 VisitedPlaceDepartureWeekdayExtended
 VisitedPlaceActivity
 VisitedPlaceActivityInternalId
 VisitedPlaceComparableActivity
 VisitedPlaceComparableActivityCategory
 VisitedPlaceDeclaredActivityDurationSec
 VisitedPlaceDeclaredActivityDurationMin
 VisitedPlaceDeclaredName
 VisitedPlaceDeclaredFullStreetAddress
 VisitedPlaceDeclaredApartmentNumber
 VisitedPlaceDeclaredCivicNumber
 VisitedPlaceDeclaredStreetName
 VisitedPlaceDeclaredStreetType
 VisitedPlaceDeclaredCity
 VisitedPlaceDeclaredRegion
 VisitedPlaceDeclaredCountry
 VisitedPlaceDeclaredPostalCode
 VisitedPlaceCorrectedName
 VisitedPlaceCorrectedFullStreetAddress
 VisitedPlaceCorrectedApartmentNumber
 VisitedPlaceCorrectedCivicNumber
 VisitedPlaceCorrectedStreetName
 VisitedPlaceCorrectedStreetType
 VisitedPlaceCorrectedCity
 VisitedPlaceCorrectedRegion
 VisitedPlaceCorrectedCountry
 VisitedPlaceCorrectedPostalCode
 VisitedPlaceLatitude
 VisitedPlaceLongitude
 VisitedPlacePostgisGeography
 VisitedPlaceX
 VisitedPlaceY
 VisitedPlaceGeometrySRID
 VisitedPlacePostgisGeometry
 VisitedPlaceNaturalKey

dw_od_trip_chain_facts

TripChainKey
 TripChainUuid
 TripChainDimensionKey
 TripChainSingleWeight
 TripChainSequenceInJourney
 TripChainSequenceInInterviewDay
 TripChainSurveyKey
 TripChainSampleKey
 TripChainJourneyKey
 TripChainPersonKey
 TripChainHouseholdKey
 TripChainHomeKey
 TripChainPersonUsualWorkPlaceKey
 TripChainPersonUsualSchoolPlaceKey
 TripChainAnchorVisitedPlaceKey
 TripChainMainActivityKey
 TripChainMainComparableActivityKey
 TripChainMainComparableActivityCategoryKey
 TripChainStartDateKey
 TripChainStartTimeKey
 TripChainStartDateExtendedKey
 TripChainStartTimeExtendedKey
 TripChainEndDateKey
 TripChainEndTimeKey
 TripChainEndDateExtendedKey
 TripChainEndTimeExtendedKey
 TripChainSimpleFlag
 TripChainOpenFlag
 TripChainComplexFlag
 TripChainSingleLoopFlag
 TripChainMultiLoopFlag
 TripChainConstrainedFlag
 TripChainUnconstrainedFlag
 TripChainFirstVisitedPlaceHomeFlag
 TripChainLastVisitedPlaceHomeFlag
 TripChainNaturalKey

dw_od_trip_chain_dimensions

TripChainKey
 TripChainSingleWeight
 TripChainSequenceInJourney
 TripChainSequenceInInterviewDay
 TripChainShortname
 TripChainCategory
 TripChainSimpleFlag
 TripChainOpenFlag
 TripChainComplexFlag
 TripChainSimpleComplexOrOpen
 TripChainSingleLoopFlag
 TripChainMultiLoopFlag
 TripChainSingleOrMultiLoop
 TripChainConstrainedFlag
 TripChainUnconstrainedFlag
 TripChainConstrainedOrUnconstrained
 TripChainMainActivity
 TripChainMainActivityInternalId
 TripChainMainComparableActivity
 TripChainMainComparableActivityCategory
 TripChainFirstVisitedPlaceHomeFlag
 TripChainLastVisitedPlaceHomeFlag
 TripChainLastVisitedPlaceHomeOrNot
 TripChainNumberOfLoops
 TripChainNumberOfTrips
 TripChainNumberOfVisitedPlaces
 TripChainNumberOfSegments
 TripChainNumberOfUniqueModes
 TripChainStartWeekday
 TripChainStartWeekdayExtended
 TripChainEndWeekday
 TripChainEndWeekdayExtended
 TripChainTotalBirdDistanceM
 TripChainTotalBirdDistanceKm
 TripChainTotalBirdDistanceMi
 TripChainTotalDeclaredDurationSec
 TripChainTotalDeclaredDurationMin
 TripChainMainActivityDeclaredDurationSec
 TripChainMainActivityDeclaredDurationMin
 TripChainNaturalKey

dw_od_segment_facts

SegmentKey
 SegmentUuid
 SegmentDimensionKey
 SegmentSingleWeight
 SegmentSequenceInTrip
 SegmentSequenceInJourney
 SegmentSequenceInInterviewDay
 SegmentSequenceInTripChain
 SegmentSequenceInJourney
 SegmentSequenceInInterviewDay
 SegmentFirstSegmentOfTripFlag
 SegmentLastSegmentOfTripFlag
 SegmentFirstOrLastSegmentOfTrip
 SegmentVehicleOccupancy
 SegmentNumberOfSegmentPartners
 SegmentNumberOfTripPartners
 SegmentStartWeekday
 SegmentStartWeekdayExtended
 SegmentEndWeekday
 SegmentEndWeekdayExtended
 SegmentStartParkingType
 SegmentStartParkingTypeInternalId
 SegmentStartComparableParkingType
 SegmentStartParkingPaymentType
 SegmentStartParkingPaymentTypeInternalId
 SegmentStartComparableParkingPaymentType
 SegmentStartParkingPriceTerm
 SegmentStartParkingPriceTermInternalId
 SegmentStartComparableParkingPriceTerm
 SegmentStartParkingPricePaidAmount
 SegmentStartParkingPricePaidCurrency
 SegmentEndParkingType
 SegmentEndParkingTypeInternalId
 SegmentEndComparableParkingType
 SegmentEndParkingPaymentType
 SegmentEndParkingPaymentTypeInternalId
 SegmentEndComparableParkingPaymentType
 SegmentEndParkingPriceTerm
 SegmentEndParkingPriceTermInternalId
 SegmentEndComparableParkingPriceTerm
 SegmentEndParkingPricePaidAmount
 SegmentEndParkingPricePaidCurrency
 SegmentPassengerRidesharingDriverType
 SegmentStartDateKey
 SegmentStartTimeKey
 SegmentStartDateExtendedKey
 SegmentStartTimeExtendedKey
 SegmentEndDateKey
 SegmentEndTimeKey
 SegmentEndDateExtendedKey
 SegmentEndTimeExtendedKey
 SegmentNaturalKey

dw_od_segment_dimensions

SegmentKey
 SegmentSingleWeight
 SegmentSequenceInTrip
 SegmentSequenceInJourney
 SegmentSequenceInInterviewDay
 SegmentFirstSegmentOfTripFlag
 SegmentLastSegmentOfTripFlag
 SegmentFirstOrLastSegmentOfTrip
 SegmentVehicleOccupancy
 SegmentNumberOfSegmentPartners
 SegmentNumberOfTripPartners
 SegmentStartWeekday
 SegmentStartWeekdayExtended
 SegmentEndWeekday
 SegmentEndWeekdayExtended
 SegmentStartParkingType
 SegmentStartParkingTypeInternalId
 SegmentStartComparableParkingType
 SegmentStartParkingPaymentType
 SegmentStartParkingPaymentTypeInternalId
 SegmentStartComparableParkingPaymentType
 SegmentStartParkingPriceTerm
 SegmentStartParkingPriceTermInternalId
 SegmentStartComparableParkingPriceTerm
 SegmentStartParkingPricePaidAmount
 SegmentStartParkingPricePaidCurrency
 SegmentEndParkingType
 SegmentEndParkingTypeInternalId
 SegmentEndComparableParkingType
 SegmentEndParkingPaymentType
 SegmentEndParkingPaymentTypeInternalId
 SegmentEndComparableParkingPaymentType
 SegmentEndParkingPriceTerm
 SegmentEndParkingPriceTermInternalId
 SegmentEndComparableParkingPriceTerm
 SegmentEndParkingPricePaidAmount
 SegmentEndParkingPricePaidCurrency
 SegmentPassengerRidesharingDriverType
 SegmentNaturalKey

dw_od_journey_facts

JourneyKey
 JourneyUuid
 JourneyDimensionKey
 JourneySingleWeight
 JourneySequenceInInterview
 JourneySurveyKey
 JourneySampleKey
 JourneyPersonKey
 JourneyHouseholdKey
 JourneyHomeKey
 JourneyPersonUsualWorkPlaceKey
 JourneyPersonUsualSchoolPlaceKey
 JourneyVisitedPlaceFirstKey
 JourneyVisitedPlaceLastKey
 JourneyStartDateKey
 JourneyStartTimeKey
 JourneyStartDateExtendedKey
 JourneyStartTimeExtendedKey
 JourneyEndDateKey
 JourneyEndTimeKey
 JourneyEndDateExtendedKey
 JourneyEndTimeExtendedKey
 JourneyNaturalKey

dw_od_journey_dimensions

JourneyKey
 JourneySingleWeight
 JourneySequenceInInterview
 JourneyShortname
 JourneyName
 JourneyDescription
 JourneyNumberOfTrips
 JourneyNumberOfVisitedPlaces
 JourneyNumberOfTripChains
 JourneyNumberOfSegments
 JourneyNumberOfUniqueModes
 JourneyStartWeekday
 JourneyStartWeekdayExtended
 JourneyEndWeekday
 JourneyEndWeekdayExtended
 JourneyTotalBirdDistanceM
 JourneyTotalBirdDistanceKm
 JourneyTotalDeclaredDurationSec
 JourneyTotalDeclaredDurationMin
 JourneyNaturalKey

Figure B-1 : Schéma des tables de l'entrepôt de données (suite)

dw_od_trip_facts

TripKey
 TripUuid
 TripDimensionKey
 TripSingleWeight
 TripSequenceInTripChain
 TripSequenceInJourney
 TripSequenceInInterviewDay
 TripSurveyKey
 TripSampleKey
 TripTripChainKey
 TripJourneyKey
 TripPersonKey
 TripHouseholdKey
 TripHomeKey
 TripPersonUsualWorkPlaceKey
 TripPersonUsualSchoolPlaceKey
 TripVisitedPlaceOriginKey
 TripVisitedPlaceDestinationKey
 TripActivityOriginKey
 TripActivityDestinationKey
 TripComparableActivityOriginKey
 TripComparableActivityDestinationKey
 TripComparableActivityCategoryOriginKey
 TripComparableActivityCategoryDestinationKey
 TripComparableMultimodeCategoryKey
 TripFirstParkAndRideJunctionParkingLocationKey
 TripFirstComparableParkingTypeKey
 TripFirstComparableParkingPaymentTypeKey
 TripFirstComparableParkingPriceTermKey
 TripLastComparableParkingTypeKey
 TripLastComparableParkingPaymentTypeKey
 TripLastComparableParkingPriceTermKey
 TripFirstPassengerSegmentComparableDriverTypeKey
 TripRoutingKey
 TripStartDateKey
 TripStartTimeKey
 TripStartDateExtendedKey
 TripStartTimeExtendedKey
 TripEndDateKey
 TripEndTimeKey
 TripEndDateExtendedKey
 TripEndTimeExtendedKey
 TripNextStartDateKey
 TripNextStartTimeKey
 TripNextStartDateExtendedKey
 TripNextStartTimeExtendedKey
 TripFirstTriplnTripChainFlag
 TripLastTriplnTripChainFlag
 TripFirstTriplnJourneyFlag
 TripLastTriplnJourneyFlag
 TripFirstTriplnInterviewDayFlag
 TripLastTriplnInterviewDayFlag
 TripSinglemodeFlag
 TripMultimodeFlag
 TripNaturalKey

dw_od_trip_dimensions

TripKey
 TripSingleWeight
 TripSequenceInTripChain
 TripSequenceInJourney
 TripSequenceInInterviewDay
 TripNumberOfSegments
 TripNumberOfUniqueModes
 TripNumberOfDrivingSegments
 TripNumberOfCyclingSegments
 TripFirstOrLastTrip
 TripFirstTriplnTripChainFlag
 TripLastTriplnTripChainFlag
 TripFirstOrLastTriplnTripChain
 TripFirstTriplnJourneyFlag
 TripLastTriplnJourneyFlag
 TripFirstOrLastTriplnJourney
 TripFirstTriplnInterviewDayFlag
 TripLastTriplnInterviewDayFlag
 TripFirstOrLastTriplnInterviewDay
 TripSinglemodeFlag
 TripMultimodeFlag
 TripSinglemodeOrMultimode
 TripComparableMultimodeCategory
 TripModes
 TripStartWeekday
 TripStartWeekdayExtended
 TripEndWeekday
 TripEndWeekdayExtended
 TripActivityAtOrigin
 TripActivityAtOriginInternalId
 TripComparableActivityAtOrigin
 TripComparableActivityCategoryAtOrigin
 TripActivityAtDestination
 TripActivityAtDestinationInternalId
 TripComparableActivityAtDestination
 TripComparableActivityCategoryAtDestination
 TripDeclaredActivityDurationAtOriginSec
 TripDeclaredActivityDurationAtOriginMin
 TripDeclaredActivityDurationAtDestinationSec
 TripDeclaredActivityDurationAtDestinationMin
 TripFirstParkingType
 TripFirstParkingTypeInternalId
 TripFirstComparableParkingType
 TripFirstParkingPaymentType
 TripFirstParkingPaymentTypeInternalId
 TripFirstComparableParkingPaymentType
 TripFirstParkingPriceTerm
 TripFirstParkingPriceTermInternalId
 TripFirstComparableParkingPriceTerm
 TripFirstParkingPricePaidAmount
 TripFirstParkingPricePaidCurrency
 TripLastParkingType
 TripLastParkingTypeShortname
 TripLastParkingTypeInternalId
 TripLastComparableParkingType
 TripLastParkingPaymentType
 TripLastParkingPaymentTypeInternalId
 TripLastComparableParkingPaymentType
 TripLastParkingPriceTerm
 TripLastParkingPriceTermInternalId
 TripLastComparableParkingPriceTerm
 TripLastParkingPricePaidAmount
 TripLastParkingPricePaidCurrency
 TripFirstPassengerSegmentRidesharingDriverType
 TripFirstPassengerSegmentRidesharingDriverTypeInternalId
 TripFirstPassengerSegmentComparableDriverType
 TripFirstDrivingOrPassengerSegmentVehicleOccupancy
 TripBirdDistanceM
 TripBirdDistanceKm
 TripBirdDistanceMi
 TripDeclaredTravelTimeSec
 TripDeclaredTravelTimeMin
 TripStraightLinePostgisGeography
 TripStraightLineGeometrySRID
 TripStraightLinePostgisGeometry
 TripNaturalKey

Figure B-1 : Schéma des tables de l'entrepôt de données (suite)

ANNEXE C CALCULATEURS DE CHEMIN ET DONNÉES CARTOGRAPHIQUES LIBRES

Calculateur de chemin routier • Marche, vélo, voiture

Un calculateur de chemin routier conçu pour les déplacements à pied, à vélo ou en voiture est utilisé lors de l'enrichissement de données vers l'entrepôt de données, notamment pour les trajets entre le domicile et les lieux habituels, ainsi que pour chaque déplacement déclaré. Le calculateur de chemin utilisé est appelé OSRM (Open Source Routing Machine) (Luxen & Vetter, 2011). Développé au Karlsruher Institut für Technologie (Karlsruher Institut für Technologie, n.d.), OSRM est un moteur de calcul de chemin performant utilisant les données d'OpenStreetMap ("OpenStreetMap," n.d.). Il met en œuvre l'algorithme « Contraction Hierarchies ».

Calculateur de chemin de transport collectif

Pour obtenir des calculs de chemin en transport collectif, un outil a été développé en parallèle de la plateforme et est basé sur l'algorithme CSA (Connection Scan Algorithm) (Dibbelt, Pajor, Strasser, & Wagner, 2013).

Cartes collaboratives libres et données ouvertes • OpenStreetmap

Les deux calculateurs de chemins utilisent les données d'OpenStreetMap (OSM) pour représenter le réseau routier, cyclable et piétonnier. OSM est créé et mis à jour en mode collaboratif et peut être visualisé et modifié par n'importe quel utilisateur enregistré. Le réseau de transport routier et cyclable d'OSM est disponible pour la planète entière et fait l'objet de mises à jour fréquentes dans de nombreuses régions du monde.

L'utilisation de logiciels libres comme OSRM et de données libres comme OSM assurent une grande flexibilité lors de leur intégration dans un outil comme celui qui a été développé dans le

cadre de la thèse. Bien que certains défis accompagnent leur utilisation (qualité et précision variables), les données libres facilitent l'enrichissement à un moindre coût et les rendent davantage compatibles avec d'autres applications, puisque la documentation qui leur est associée est de plus en plus étendue et complète.