



UNIVERSITÉ DE FRIBOURG  
UNIVERSITÄT FREIBURG

Travail de Bachelor

# Gestion du stockage des données de mobilité

Elric Milet

Superviseur : Prof. Marino Widmer

16 juin 2023

Faculté des sciences et de médecine

Département d'informatique



**Abstract :**

Ce document a été réalisé dans le cadre du travail de Bachelor d'Elric Milet, sous la supervision du Prof. Marino Widmer, durant l'année académique 2022-2023 à l'Université de Fribourg.

Il s'appuie principalement sur l'expérience de l'auteur acquise lors d'un stage dans la division mobilité de la Ville de Lausanne, Suisse, où il a pu travailler avec des données de mobilité issues de comptages routiers.

Ce travail présente différents types de données de mobilité utilisées à travers le monde, leurs utilités diverses et les difficultés qu'elles posent à celles et ceux qui travaillent avec. Il se concentre particulièrement sur les données issues de comptages routiers, sur les manières de les stocker et de les représenter. Un regard critique est posé sur leur utilisation, principalement au niveau des problèmes fréquemment rencontrés dans ce domaine et des besoins d'amélioration qui se posent à l'échelle de la Ville de Lausanne mais aussi chez d'autres acteurs du secteur.

Dans le cadre de ce travail, une application a été développée, visant notamment à faciliter la standardisation, le stockage et l'analyse de ces données issues de comptages routiers. Ce document revient sur sa conception et sur les utilisations que l'on peut en faire.

**Mots-clés :**

Mobilité, comptages routiers, données, stockage, amélioration, application.

**Remerciements :**

Je tiens à remercier le Prof. Marino Widmer, qui m'a suivi tout au long de ce travail, pour son investissement à mes côtés, ses retours réguliers, positifs et constructifs, et son soutien constant.

Je remercie également ma famille, et en particulier ma mère, qui m'a toujours soutenu de manière extrêmement enthousiaste et s'est intéressée à ce que je faisais en toute circonstance.

Je remercie enfin la Ville de Lausanne, sa division mobilité, et plus particulièrement Nicolas Vallotton et Daphné Dethier, qui ont su répondre à mes questions et se sont rendus disponibles pour m'aider à réaliser ce travail.



<b>0 Introduction.....</b>	<b>1</b>
<b>1 Ville de Lausanne : Division Mobilité.....</b>	<b>3</b>
1.1 Structure de la Ville.....	3
1.2 Utilisation des données dans le service.....	4
1.3 Difficultés engendrées.....	7
<b>2 Comptages routiers.....</b>	<b>9</b>
2.1 Types de données utilisées.....	9
2.2 Types de comptages.....	11
2.3 Mandants, mandataires et attentes.....	13
<b>3 Autres données de mobilité.....</b>	<b>17</b>
3.1 Catégorisation.....	17
3.2 Utilité de ces données.....	20
3.3 GTFS et GBFS.....	22
3.4 DATEX II.....	28
<b>4 Stockage des données.....</b>	<b>31</b>
4.1 Stockage par fichier.....	31
4.2 Utilité de l'agrégation.....	33
4.3 Bases de données.....	34
4.4 Représentation géographique.....	35
4.5 Conservation à long-terme.....	36
<b>5 Besoins d'amélioration et standardisation.....</b>	<b>39</b>
5.1 Erreurs et biais dans les données.....	39
5.2 Manque d'uniformisation.....	41
5.3 Standardisation à différentes échelles.....	43
5.4 Améliorations internes à l'organisation.....	44
<b>6 Conception de l'application.....</b>	<b>47</b>
6.1 Fondations.....	47
6.2 Base de données.....	49
6.3 Phase d'analyse.....	52
6.4 Structure du programme.....	53
6.5 Introspection.....	55
<b>7 Conclusion.....</b>	<b>59</b>
<b>8 Références.....</b>	<b>61</b>
8.1 Bibliographie.....	61
8.2 Webographie.....	61
<b>9 Annexes.....</b>	<b>65</b>
9.1 GitHub.....	65
9.2 Guide d'utilisation de l'application.....	65



# 0 Introduction

Les routes sont-elles construites au hasard ?

Si la réponse à cette question peut sembler évidente, il est beaucoup plus difficile de déterminer, simplement à l'aide de nos pensées, comment un réseau routier est conçu et construit. En effet, les décisions derrière l'installation de telles infrastructures ne sont pas prises à la légère : de nombreuses données et statistiques viennent les appuyer. Mais quelles sont donc ces données en question, et quelles formes prennent-elles ? C'est la question à laquelle ce travail tente de répondre.

Les données de mobilité sont vastes et leurs champs d'utilisation le sont tout autant.

La planification des transports routiers, qu'ils concernent les véhicules privés, les transports publics ou les deux-roues par exemple, en est totalement dépendante. Dans un monde où la population est en mouvement constant, et dans un pays comme la Suisse où la personne moyenne parcourt environ 30 kilomètres par jour, avoir une infrastructure intelligemment développée est crucial.

Ici, il n'est pas seulement question de l'expansion de l'infrastructure routière, mais aussi des solutions alternatives en termes de mobilité, comme les transports publics, qui se doivent de proposer des options viables pour leurs utilisateurs et utilisatrices. Le développement de tels réseaux repose en grande partie sur les connaissances de leurs besoins. Il est également important de mentionner le développement urbain : une bonne compréhension de l'utilisation de l'infrastructure routière existante et des divers points chauds qui la composent est capitale pour la planification des nouvelles zones résidentielles ou commerciales.

Ainsi, les buts derrière le développement des offres en termes de mobilité sont nombreux : faire gagner du temps à tout le monde, permettre de se déplacer de manière plus efficace mais aussi plus écologique, offrir un environnement de vie plus sain aux résidents et résidentes des alentours des routes... Chacun et chacune peut y trouver son intérêt personnel.

Mais il est possible de voir encore plus loin, et de soulever l'importance des données de mobilité dans des domaines qui n'ont pourtant pas un rapport évident. Par exemple, dans le cadre d'une pandémie comme celle de COVID-19, l'analyse des déplacements et des habitudes de mouvements peut ouvrir des possibilités au niveau de l'identification d'où et quand des individus se sont rencontrés. On peut ainsi contrôler les éventuelles transmissions, afin notamment de mieux cibler les interventions locales. Dans ce même contexte, les données de mobilité ont pu être utilisées pour explorer comment les habitudes des individus ont changé lors des différents confinements et fermetures engendrés par la pandémie.

Dans le cadre d'un stage dans la division mobilité de la Ville de Lausanne, en Suisse, j'ai pu travailler avec ces données de mobilité, et plus particulièrement avec les données issues de comptages routiers. J'ai ainsi eu l'occasion de découvrir les diverses formes qu'elles peuvent prendre, leurs utilités directes et indirectes, mais aussi les nombreuses difficultés que leur utilisation peut poser. Ce travail a été écrit en partie sur la base de mon expérience personnelle issue de ce stage de 8 semaines à plein temps, qui a aussi mené au développement d'une application ayant pour but principal de faciliter la gestion de ces données issues de comptages routiers.

Ce travail commence par revenir sur mon expérience à la Ville de Lausanne, incluant une contextualisation du poste que j'ai occupé, une focalisation sur les données de mobilité utilisées et les difficultés que j'ai pu constater.

Il revient ensuite plus en détail sur les comptages routiers, les types de comptages et les données que l'on en tire, ainsi que les attentes de leurs mandants.

L'horizon est dans la foulée étendu aux données de mobilité de manière plus générale. Une tentative de catégorisation sera présentée, leur utilité sera démontrée et certains formats seront détaillés.

Le travail se concentrera subséquemment à nouveau sur les données issues de comptages routiers plus spécifiquement, les manières de les stocker et de les représenter ainsi que leur conservation à long-terme.

Un regard plus critique est par la suite apporté sur ces données par le biais d'une analyse des besoins d'amélioration et de standardisation. Une attention particulière est mise sur les erreurs et biais dans les données, l'uniformisation, la standardisation aux différentes échelles politiques et les améliorations pouvant être apportées à une organisation telle que celle de la division mobilité.

Enfin, l'application développée dans le cadre de ce travail est décrite, au niveau de son concept, de sa structure et de son processus de développement notamment.

Dans un premier temps, il faut revenir aux origines de ce travail et à l'organisme m'ayant donné l'idée et les clés nécessaires à sa réalisation : la division mobilité de la Ville de Lausanne.

# 1 Ville de Lausanne : Division Mobilité

Avant de se lancer dans des explications sur les données de mobilité, il est crucial d'avoir une compréhension plus accrue de leurs utilisateurs et utilisatrices.

Ce chapitre présente relativement brièvement la division mobilité de la Ville de Lausanne, ses buts, et comment des données de mobilité sont utilisées dans le service qui comporte cette division. Enfin, ce chapitre comporte un aperçu des difficultés engendrées par l'utilisation de ces données.

## 1.1 Structure de la Ville

La Ville de Lausanne comporte une direction des finances et mobilité. Cette direction gère plusieurs services, dont celui de la mobilité et de l'aménagement des espaces publics. Ce service contient lui-même plusieurs divisions, dont celle de la mobilité. Et c'est dans cette division (voir [fig.1](#)) que j'ai eu la chance d'effectuer un stage durant l'été 2022, auprès du gestionnaire de données.

Les missions que s'est donnée la division mobilité sont les suivantes :

- Planifier la mobilité urbaine
- Promouvoir et développer la mobilité douce (piétons et vélos) et les transports publics
- Définir et mettre en application la politique du stationnement sur le domaine public
- Coordonner les chantiers communaux et les interventions sur le domaine public visant à appliquer une stratégie qui garantit l'accessibilité en ville, à minimiser les impacts pour les usagers, les riverains et les activités économiques
- Gérer le trafic par le biais de la programmation de la signalisation lumineuse

Pour résumer, la division mobilité de la Ville de Lausanne se charge à la fois de maintenir les infrastructures servant aux routes et au trafic, et surtout de planifier et co-gérer les projets de développement et d'adaptation de tout ce qui touche de près ou de loin à la mobilité en ville, motorisée ou non. Cette division est donc composée d'une vingtaine de personnes, réparties dans plusieurs sections : accessibilité, stationnement, coordination des chantiers, planification et exploitation.

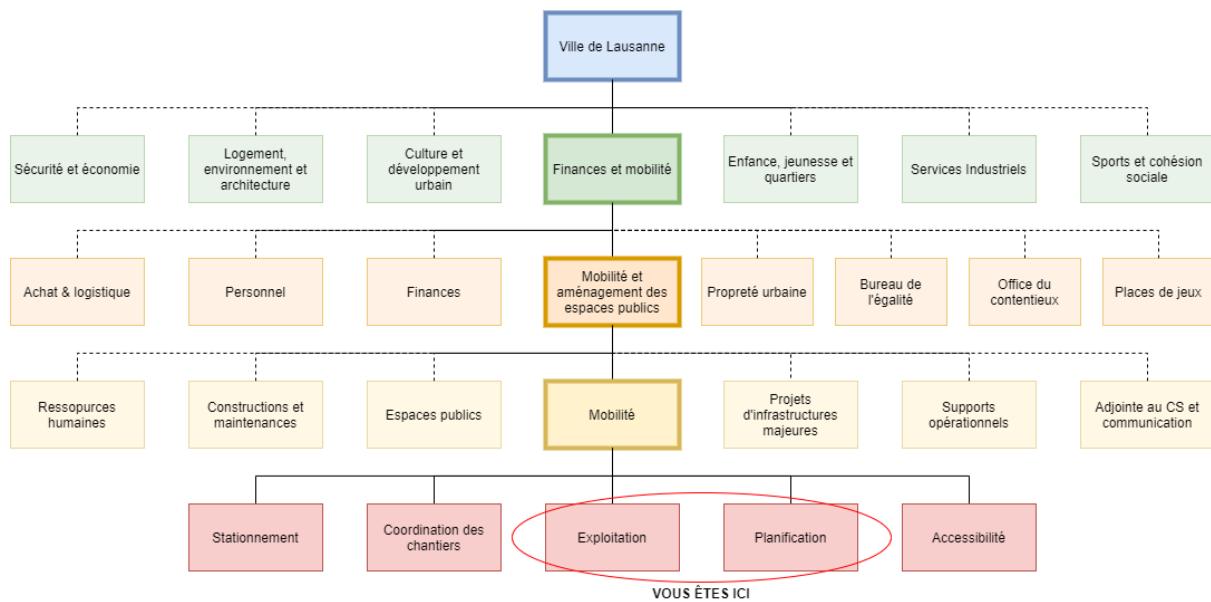


Fig. 1 : Organigramme de la Ville et situation des sections mentionnées (source : conception personnelle)

J'ai ainsi pu travailler notamment en collaboration avec la section planification, ainsi que la section exploitation, le gestionnaire de données de la division se situant dans l'organigramme à cheval entre ces deux sections.

La section planification est en grande partie chargée d'étudier les cas des grands projets et de participer à l'élaboration des stratégies conformes aux besoins de développements durables de la mobilité lausannoise. Par exemple, lors de la mise en place récente de la limitation à 30 km/h de nuit dans la ville, une grande partie des études d'impact et de la gestion du déroulement du projet était supervisée par les collaborateurs de cette section.

La section exploitation est beaucoup plus tournée vers les problématiques de terrain. Là où la section planification est en grande partie composée de chef.fe.s et chargé.e.s de projet, la section exploitation s'occupe en grande partie de la maintenance et de la mise en place de la signalisation lumineuse notamment, comme les feux de circulation ou les indicateurs de vitesse, ainsi que d'autres types d'infrastructures liées à la mobilité urbaine, comme les horodateurs ou les bornes escamotables<sup>1</sup>.

## 1.2 Utilisation des données dans le service

Même si les comptages routiers sont habituellement gérés par la section exploitation et la section planification, ils sont souvent utiles pour l'ensemble du service. Par exemple, dans le cas d'un projet de refonte d'un quartier dans le but de passer ses routes en zones modérées<sup>2</sup>, il se pourrait qu'un.e collaborateur.trice ait besoin de connaître l'utilisation des différentes routes composant le quartier, ceci dans le but d'adapter la requalification pour que les nouvelles routes puissent supporter suffisamment de trafic, ou de trouver des alternatives.

<sup>1</sup> D'autres acteurs sont impliqués dans la mise en place de certaines de ces infrastructures.

<sup>2</sup> Dans une zone à vitesse modérée, la limitation est habituellement fixée à 30 km/h.

Dans ce cas-là, cette personne a deux possibilités :

- aller voir le gestionnaire de données, qui se chargera de lui fournir les données demandées, de préférence traitées au préalable pour n'en ressortir que les informations utiles, ou qui gèrera ou sous-traitera une potentielle nouvelle étude en cas d'absence de données pertinentes,
- ou mandater directement elle-même une entreprise externe, qui sera alors chargée de réaliser une étude sur le quartier en question.

Dans ces deux cas, on en revient normalement au premier type de données stocké, **les études et comptages réalisés par des mandataires externes**. Ces prestations sont commandées auprès d'entreprises spécialisées, qui se chargeront alors d'aller effectuer des comptages à des endroits clés, décidés par la Ville ou par eux suivant les besoins. Ils sont généralement faits automatiquement par des installations temporaires ([fig. 2](#)).

Optionnellement, une analyse des résultats peut être commandée avec l'étude.



*Fig. 2 : exemple d'installation de comptage, ici un radar Doppler (source : [\[swi22\]](#))*

Ces comptages sont parmi les principales données utilisées par le reste du service. Cependant, la section exploitation a d'autres tours dans son sac. En effet, d'autres instruments, installés de manière temporaire ou permanente, sont utilisés dans la Ville pour obtenir des données de mobilité, comptages, vitesse ou autres. Ces instruments sont généralement commandés et installés dans le cadre de projets spécifiques. Reprenons l'exemple du passage aux 30 km/h de nuit : plusieurs dizaines d'indicateurs de vitesse ont alors été installés à des endroits clés de Lausanne. Ces indicateurs de vitesse ont le double avantage de signaler aux conducteurs lorsqu'ils sont en excès de vitesse, et de récolter des données pouvant ensuite servir notamment à la division de la Mobilité.

Ces indicateurs de vitesse se présentent sous plusieurs types. Certains sont directement reliés à l'alimentation électrique sur leur lieu d'installation, d'autres se rechargent de manière solaire, d'autres encore fonctionnent sur batterie qui doivent être changées régulièrement. Suivant leur période de commande et le contexte du projet dans lequel ces indicateurs ont été installés, le fabricant peut également changer, ainsi que le mode de collecte de données.



Fig. 3 : deux exemples d'installation d'indicateurs de vitesse (sources : [\[sma22a\]](#) et [\[rue22\]](#))

Enfin, il existe des compteurs qui n'affichent pas la vitesse aux passants mais envoient tout de même des données à la Ville. Certains types de radars<sup>3</sup> accrochables au mobilier urbain, par exemple, peuvent être employés par la section exploitation pour effectuer des comptages et des relevés de vitesse à des endroits précis sur une certaine période.

Certains de ces appareils ne relèvent d'ailleurs pas la vitesse, ou sont spécialisés dans le comptage des piétons et des vélos. On peut citer aussi les boucles à induction, installations permanentes ne permettant que de compter le nombre de véhicules qui passent. Ou encore le radar bruit, qui depuis mi-2021, relève le bruit émis par chaque véhicule en plus de sa vitesse et de son type.



Fig. 4 : exemples d'appareils de comptage (sources : [\[elw22\]](#), [\[int22\]](#) et [\[sma22b\]](#))

---

<sup>3</sup> On ne parle pas là de radars de police, mais de radars servant uniquement à relever la vitesse et le type de véhicule ; aucun appareil photo n'est intégré au dispositif.

Toutes ces données proviennent donc de sources diverses, n'ont généralement pas le même but (par exemple, les indicateurs de vitesse ne sont pas vraiment prévus pour le comptage de véhicules, mais uniquement pour le relevé de leur vitesse), et pourtant peuvent être utilisées de la même manière par le service.

## 1.3 Difficultés engendrées

Tout cela pose donc plusieurs problèmes au niveau de la gestion des données pour le collaborateur concerné de la section :

- **Inventaire**

Les données provenant de sources souvent différentes, leur représentation numérique varie aussi. Les stocker de manière consistante et organisée pose alors un vrai défi et ralentit considérablement les recherches de données spécifiques. Certains systèmes ont leur propre interface en ligne, d'autres utilisent simplement un stockage sous forme de fichiers sur un serveur interne par exemple.

- **Traitement**

Trouver les données en question n'est probablement pas la difficulté la plus grande : la facilité de traitement de ces données est aussi très influencée par leurs formes diverses. Chaque type de données va demander un procédé particulier pour pouvoir consulter son contenu et l'exploiter, allongeant là aussi significativement le temps nécessaire à la génération de la moindre statistique et du moindre graphique.

D'autant plus que le traitement initial appliqué aux données brutes varie lui aussi. Dans certains cas, les statistiques de base (trafic journalier moyen, vitesse moyenne...) sont calculées automatiquement ou fournies préalablement, tandis que dans d'autres, on se contente de stocker les données brutes.

- **Analyse**

Ceci fait que les failles sont particulièrement difficiles à repérer dans certaines données. Et comme dans tout système reposant sur des instruments utilisés dans des conditions variables, des failles, il y en a.

On parle là de pannes (par exemple causées par un dysfonctionnement ou du vandalisme), de changements de positionnement (par exemple, un radar pédagogique retourné), d'obstacles temporaires empêchant le bon fonctionnement de l'appareil (par exemple, un panneau de chantier positionné devant un radar Doppler), d'activités anormales causées par un événement (par exemple, une route fermée pendant quelques jours)...

Le problème en soit n'est pas qu'il y a des failles, qui sont inhérentes à tout système tentaculaire tel que celui utilisé par la section exploitation, ni leurs résolutions, mais bien comment elles sont intégrées dans le système de gestion de données.

Premièrement, du fait des contraintes mentionnées aux points précédents, il est difficile de repérer les incohérences dans la quantité énorme de données qui est enregistrée quotidiennement. Lorsque ce n'est pas un problème qui a été détecté et répertorié par les collaborateurs de la section exploitation (par exemple, une panne), il n'y a aucune automatisation : la constatation du problème ne se fera qu'à la discréption du gestionnaire des données, et sa réparation ne se fera qu'avec les

maigres outils qui sont à sa disposition. La plupart du temps, les données perdues ou faussées sont purement et simplement rayées du répertoire.

Deuxièmement, même si la section exploitation a à sa disposition une plateforme lui permettant de répertorier chaque opération effectuée sur un des dispositifs dont elle a la responsabilité, il arrive inévitablement qu'il y ait des manquements ou oublis. Par exemple, le cas d'un indicateur de vitesse qui a été retourné sans que personne n'ait noté la date exacte ; il incombe alors au gestionnaire de données de tenter de la retrouver grâce aux données récoltées par l'appareil. L'analyse de ces données a été bien plus longue à effectuer que ce qu'elle aurait pu être avec un programme performant.

- **Demandes**

Enfin, une mauvaise communication et des micro-dysfonctionnements internes à la division posent des problèmes évidents. Il n'y a pas vraiment de processus standard, et le principe de contacter le gestionnaire de données pour s'enquérir de telles ou telles informations est au bon vouloir des collaborateurs et collaboratrices. Il a donc pu arriver que des études soient commandées sur des portions de routes pour lesquelles on a déjà des données, parfois pour la simple raison que le collaborateur ou la collaboratrice ne savait pas à qui s'adresser pour ce faire.

En résumé, beaucoup de données de mobilité sont brassées par la Ville de Lausanne sans que leur utilisation ne soit particulièrement aisée ou efficace.

C'est dans ce contexte qu'une amélioration des processus de collecte, de classification et de stockage de ces comptages peut se révéler utile, ainsi qu'une centralisation des données récoltées. Ce qui va suivre dans ce travail a un but bien précis : répondre aux problèmes rencontrés lors de l'obtention et du traitement de ces données par des pistes d'améliorations ainsi que des solutions concrètes, par la voie du développement d'une application notamment.

## 2 Comptages routiers

Avant de pouvoir parler plus en détail des différents types de comptages utilisés par la Ville de Lausanne, il faut d'abord expliciter les données qui peuvent être mesurées et enregistrées.

Ce chapitre détaille les types de données utilisées en les classant dans trois catégories, présente plusieurs types de comptages routiers employés par la Ville de Lausanne, et explique les attentes des mandants de ces comptages.

### 2.1 Types de données utilisées

Premièrement, séparons les données récoltées en trois catégories : les données “brutes”, les données “agrégées par temps” et les données “agrégées par vitesse”.

Les données “brutes” correspondent au passage d'un véhicule. On peut enregistrer des informations sur ce passage directement, par exemple :

- La vitesse du véhicule
- Son type
- Le bruit qu'il a émis (en décibels)

La [fig. 5](#) montre un extrait d'un exemple de données brutes issues d'un comptage routier, où pour chaque passage, la date, l'heure, la direction du véhicule et sa vitesse ont été relevés.

```
2022-09-07 11:19:01;1;21
2022-09-07 11:24:06;1;19
2022-09-07 11:27:02;1;26
2022-09-07 11:27:31;1;25
2022-09-07 11:27:37;1;25
2022-09-07 11:29:56;1;26
2022-09-07 11:31:23;0;19
2022-09-07 11:33:48;1;22
```

*Fig. 5 : extrait de données brutes d'un comptage routier (source : conception personnelle)*

Les données “agrégées par temps” correspondent à un certain nombre de passages sur une certaine plage de temps. Souvent, cette plage de temps correspond à 1h ; on enregistre alors des informations sur cette heure, notamment :

- Le nombre de véhicules comptés
- Leur vitesse moyenne
- Leur vitesse maximale
- Leur V85<sup>4</sup>, V50, V30 et V10
- Leur type

La [fig. 6](#) montre un extrait d'un exemple de données agrégées par temps, où pour chaque tranche d'heure, on relève le nombre de véhicules par type.

---

<sup>4</sup> La V85 correspond à la vitesse en-dessous de laquelle 85% des véhicules ont circulé. Ce fonctionnement est le même pour la V50 (équivalent de la médiane), la V30, la V10...

Date	Heure	Moto*	Voiture	PL	PL+	Total
18.05.2021	00:00	0	6	0	0	6
18.05.2021	01:00	0	17	0	0	17
18.05.2021	02:00	0	2	0	0	2
18.05.2021	03:00	0	8	0	0	8
18.05.2021	04:00	1	2	1	0	4
18.05.2021	05:00	1	15	0	0	16
18.05.2021	06:00	2	54	1	0	57
18.05.2021	07:00	16	139	0	1	156
18.05.2021	08:00	15	172	1	0	188

Fig. 6 : extrait de données agrégées par temps d'un comptage routier (source : conception personnelle)

Les données "agrégées par vitesse" fonctionnent de la même manière, mais correspondent à un certain nombre de passages pour une certaine fourchette de vitesse, par exemple de 34 à 36 km/h. On enregistre alors des informations sur cette plage de vitesse, notamment :

- Le nombre de véhicules comptés
- Leur type

La [fig. 7](#) montre un extrait d'un exemple de données agrégées par vitesse, où pour chaque tranche de 5 km/h, on relève le nombre de véhicules.

km/h	Nombre	%
0-5	37	1.4%
5-10	69	2.5%
10-15	152	5.6%
15-20	167	6.1%
20-25	425	15.5%
25-30	746	27.2%
30-35	729	26.6%

Fig. 7 : extrait de données agrégées par vitesse d'un comptage routier (source : conception personnelle)

L'intérêt de ces catégories est de pouvoir plus facilement accepter des données qui ne sont pas toujours reçues de la même manière, certains fabricants préférant fournir aux mandants un résumé des résultats, d'autres préférant fournir directement les données brutes.

Dans le cas où des données agrégées seraient fournies en plus des données brutes, on ne garderait que les données brutes. En effet, les données agrégées ne sont qu'une analyse des données brutes ; les distinguer et les stocker n'est qu'une manière de ne pas perdre d'information lorsque les données brutes ne sont pas disponibles. Plus simplement formulé, il est possible d'obtenir les données agrégées à partir des données brutes, il ne fait donc pas de sens de tout stocker lorsque les données brutes sont disponibles.

On peut encore noter une dernière catégorie de données, un peu différente : celle des données générales. Parfois, les fournisseurs des données incluent une synthèse des données sur la période concernée. Cette synthèse peut comprendre plusieurs informations :

- Le TJM et le TJOM<sup>5</sup>
- L'HPM et l'HPS<sup>6</sup>
- La V85, V50, V30 et V10
- La vitesse moyenne

<sup>5</sup> Le TJM (Trafic Journalier Moyen) désigne le nombre moyen de véhicules par jour. Le TJOM (Trafic de Jour Ouvrable Moyen) est similaire, mais ne comprend que les jours ouvrables.

<sup>6</sup> L'HPM (Heure de Pointe du Matin) désigne le nombre de véhicules durant l'heure de pointe du matin, généralement définie comme étant de 07:00 à 08:00. L'HPS (Heure de Pointe du Soir) est généralement définie de 17:00 à 18:00.

- La vitesse maximum
- Le taux d'excès de vitesse
- L'écart moyen à la vitesse limite
- Le pourcentage de trafic des files d'attente
- Une ou plusieurs photos de l'emplacement du comptage

La fig. 8 montre un exemple de synthèse des données, où sont relevés le TJM, l'HPM et l'HPS pour une journée entière de comptage.

Direction	Avenue de Rumine
TJM [Véh. / Jour]	2530
HPM [Véh. / h]	156
HPS [Véh. / h]	204

Fig. 8 : extrait d'une synthèse des données (source : conception personnelle)

Même si la plupart des informations à disposition dans ces synthèses sont obtenables directement depuis les données brutes ou agrégées, ce n'est pas forcément toujours le cas. Par exemple, la V85 journalière est impossible à calculer en se basant uniquement sur des données agrégées par heure. Il est donc nécessaire, dans une telle situation, d'enregistrer la V85 journalière qui nous est fournie pour ne pas perdre l'information.

## 2.2 Types de comptages

Ce sous-chapitre se focalise sur la description de différents types de comptages routiers qui peuvent être traités par la division mobilité de la Ville de Lausanne, et qui utilisent en principe les types de données définies au [chapitre 2.1](#).

- **Études et comptages réalisés par des mandataires externes**

Comme mentionné au [chapitre 1.2](#), la Ville mandate régulièrement des entreprises externes pour effectuer des comptages à des endroits clés, décidés par la Ville ou par l'entreprise en fonction des besoins.

Le format des données relevées par ces comptages varie d'une entreprise et d'une demande à l'autre. Cependant, les informations principales restent à peu près les mêmes, de manière à ce qu'elles puissent ensuite être enregistrées et exploitées par la Ville. Ces comptages relèvent généralement le nombre de véhicules par heure et leur vitesse, et ce sur une journée ou une semaine entière. Des données générales sur cette période de temps, comme le TJM ou la V50, sont également disponibles.

- **Compteurs piéton/vélo**

Plusieurs de ces compteurs (photo en haut à droite de la [fig. 4](#)), spécialisés dans le passage de piétons et de cyclistes, sont disposés à des endroits stratégiques de Lausanne, en général sur des trottoirs ou des pistes cyclables. Ils sont reliés en temps réel à une plateforme en ligne, qui permet à la Ville d'en tirer facilement des chiffres et des graphiques. Il est aussi possible, depuis cette plateforme, d'exporter sous forme de fichier les données de passage complètes et agrégées par heure sur une période de temps choisie. C'est cette dernière fonctionnalité qui va se révéler utile pour la suite de ce travail.

- **Radar bruit**

Ce type d'installation (photo en bas de la [fig. 4](#)) est le seul incluant des données sur le bruit émis par les véhicules dans ses résultats. Chaque véhicule est identifié, par son type et sa vitesse normalement, permettant d'obtenir des statistiques un peu plus détaillées. Cependant, la fonction de détection de vitesse de l'appareil n'étant pas forcément fiable, l'utilité du radar bruit est généralement volontairement limitée à la détection du bruit et du type de chaque véhicule.

- **Boucles de comptage**

Il existe plusieurs emplacements, en ville, où des boucles à induction sont installées dans la route (photo à la [fig. 9](#)). Leur fonctionnement est relativement simple : elles détectent lorsqu'un véhicule leur passe dessus et l'enregistrent. Il est ensuite possible, via une connexion à distance, d'en tirer le nombre approximatif<sup>7</sup> de passage, agrégé par une certaine période de temps, pour chaque heure par exemple.



Fig. 9 : Boucle à induction de comptage routier (source : [lace231](#))

- **Indicateurs de vitesse**

La Ville de Lausanne utilise plusieurs de ces installations, aussi appelées radars pédagogiques. Ces indicateurs de vitesse proviennent de plusieurs constructeurs différents, et plusieurs modèles différents sont parfois utilisés. Leur mode de fonctionnement est donc assez varié : certains d'entre eux sont directement reliés au réseau électrique, d'autres ont un panneau solaire, et d'autres encore n'ont qu'une batterie qu'il faut régulièrement changer. La collecte des données ne se fait donc pas toujours de la même manière, n'est pas toujours possible, et ces données ne sont pas toujours stockées de la même manière.

Pour certains indicateurs de vitesse, il faut passer par une plateforme en ligne, pour d'autres, par un logiciel qui se chargera de transformer des fichiers illisibles par un humain en données utilisables. Toujours est-il que tous ces cas d'utilisations possèdent un point commun : il est possible d'en tirer les données brutes, et ensuite de travailler avec. C'est sur ce dernier fait que le traitement des données des indicateurs de vitesse va se baser.

---

<sup>7</sup> Les boucles à induction utilisées par la Ville de Lausanne utilisent un calcul particulier pour déterminer le nombre de véhicules qui leur sont réellement passés dessus, ce qui crée une légère marge d'erreur.

Enfin, il me semble utile de préciser que même si les indicateurs de vitesse fournissent pour la plupart une estimation relativement bonne du nombre de passage de véhicules, ces données ne sont pas jugées comme étant suffisamment fiables par la division mobilité, qui se contente donc d'exploiter les données de vitesse.

- **Radar de comptage**

Un dernier appareil (photo en haut à gauche de la [fig. 4](#)) peut être ajouté à cette catégorie. Dans les faits, il produit à peu près les mêmes données et s'utilise à peu près de la même manière qu'un indicateur de vitesse, à la différence près qu'il est capable de différencier les types de véhicules. Il est d'ailleurs commercialisé par la même entreprise. En pratique, cet appareil est principalement utilisé de manière temporaire à des endroits clés, pour évaluer la différence d'utilisation d'un segment de route lors d'une déviation par exemple. À noter également qu'il n'est pas relié à un écran et n'affiche donc aucune vitesse aux véhicules.

Voici finalement un tableau récapitulatif des données disponibles pour chaque type de comptage évoqué ci-dessus :

Comptage	Catégories de données	Distingue les types de véhicule	Utilité(s) supplémentaire(s)
Études réalisées par mandataires externes	- Agrégées par temps - Agrégées par vitesse - Synthèse	Oui	- Vitesse
Compteurs piéton/vélo	- Agrégées par temps	Piétons et vélos uniquement	Non
Radars bruit	- Brutes	Oui	- Vitesse - Bruit
Boucles de comptage	- Agrégées par temps	Non	Non
Indicateurs de vitesse	- Brutes - (Agrégées par temps) - (Agrégées par vitesse) - (Synthèse)	Non	- Vitesse
Radars de comptage	- Brutes - (Agrégées par temps) - (Agrégées par vitesse) - (Synthèse)	Oui	- Vitesse

*Fig. 10 : Tableau récapitulatif des données par type de comptage (source : conception personnelle)*

## 2.3 Mandants, mandataires et attentes

Chaque installation d'un dispositif de comptage est généralement liée à l'exécution d'un projet, et possède un ou plusieurs buts bien définis. Voici plusieurs exemples :

- **Vérification de l'utilité d'aménagements routiers**

Une petite portion de route dans un quartier résidentiel est devenue "zone de rencontre", c'est-à-dire que la vitesse y est limitée à 20 km/h et que des aménagements y sont faits pour améliorer la cohabitation entre les piétons et les véhicules. Au fil des mois, ces aménagements prennent plusieurs formes, allant de simples panneaux et marquages à la surélévation de la route. Ici, un indicateur de vitesse a été posé, dans le but notamment de déterminer l'impact que ces aménagements peuvent avoir sur la vitesse des véhicules.

- **Surveillance des zones particulières**

Avec le déploiement de plusieurs zones où la vitesse est limitée à 30 km/h, la Ville s'est aussi dotée de plusieurs indicateurs de vitesse, commandés auprès d'un nouveau fournisseur. Ces indicateurs de vitesse en particulier sont donc, en principe, uniquement utilisés dans ces zones 30. Leur but est relativement simple : vérifier que la vitesse est suffisamment bien respectée, et constater l'utilité ou non d'une mesure supplémentaire, un nouvel aménagement par exemple.

- **Réponse aux demandes des riverain.e.s**

Il arrive régulièrement que des habitant.e.s de la ville se plaignent du bruit et/ou de la vitesse des véhicules devant leur bâtiment. Les responsables se devant d'agir, dans la plupart des cas, des vérifications sont faites sur les lieux de la plainte afin de vérifier les dires de la personne, par l'installation d'un indicateur de vitesse temporaire par exemple. Si nécessaire, des mesures supplémentaires, telles que des nouveaux aménagements routiers, seront ensuite prises.

- **Comptages réguliers**

Au-delà de projets particuliers, les comptages peuvent simplement servir par la suite aux études communales ou régionales. De manière générale, il est à peu près toujours utile de posséder les chiffres de fréquentation de certains axes. D'où l'installation des boucles à induction à des emplacements clés dans la ville, permettant de tracer avec une certaine précision les flux de déplacement d'un endroit à l'autre, et surtout, la variation en fonction du temps. Des comptages peuvent aussi être commandés à des mandataires externes simplement en cas de déficit d'information dans un quartier en particulier.

- **Préparation d'un projet localisé**

Ces comptages demandés à des entreprises extérieures ne le sont la plupart du temps pas sans raison, et souvent, il s'agit d'une demande d'un ou d'une responsable de projet de la Ville travaillant sur un quartier et ayant besoin de connaître la demande actuelle en termes de transport.

- **Étude sur le bruit**

Dans de nombreuses villes d'Europe, les études sur l'impact du bruit des véhicules motorisés en ville se multiplient. Lausanne n'y fait pas exception et possède depuis 2021 un radar bruit. Déplacé régulièrement, un des premiers buts de ce radar est de déterminer les endroits problématiques afin de pouvoir les étudier plus en détail dans le futur. Les données qu'il génère peuvent aussi ensuite être recoupées avec d'autres sources pour mieux comprendre les causes et solutions du bruit en ville.

- **Vitesse réduite de nuit**

Parallèlement au déploiement d'un radar bruit, la Ville a décidé en 2021 de limiter la vitesse à 30 km/h de nuit, soit à partir de 22h, en grande partie dans le but de réduire ce même bruit causé par les véhicules motorisés. Seulement, sur certains axes, cette limite peine à être respectée, et ce de manière quasiment systématique. Le déploiement de plusieurs dizaines d'indicateurs de vitesse, notamment aux portes de la ville, a permis d'une part de déterminer où se situent ces axes problématiques, et d'autre part de constater l'évolution générale de la vitesse de nuit des véhicules motorisés, ouvrant ainsi la voie aux études sur ce changement majeur : à quel point cette vitesse est-elle respectée, et comment le comportement des automobilistes change-t-il au fil du temps ? L'observation de cette évolution est rendue possible grâce aux données récoltées par ces indicateurs de vitesse.

En résumé, dans la plupart des cas, le mandant de ces comptages est la Ville. Plus particulièrement, quelqu'un au sein du service de la mobilité et de l'aménagement des espaces publics. Les mandataires font généralement partie de la division mobilité, et les exécutants de la section exploitation. Dans le cas des études commandées aux entreprises externes, elles font office de mandataires, mais surtout d'exécutants ; il arrive que les collaborateurs de la division mobilité se chargent eux-mêmes de déterminer les emplacements exacts des comptages, d'analyser les données par la suite, voire d'effectuer les comptages eux-mêmes dans les cas où c'est facilement réalisable.

Tous ces comptages ont des utilités diverses et variées, et l'optimisation du stockage des données qui en sont issues peut se révéler cruciale dans le but de répondre aux demandes des mandants, présentes et futures.



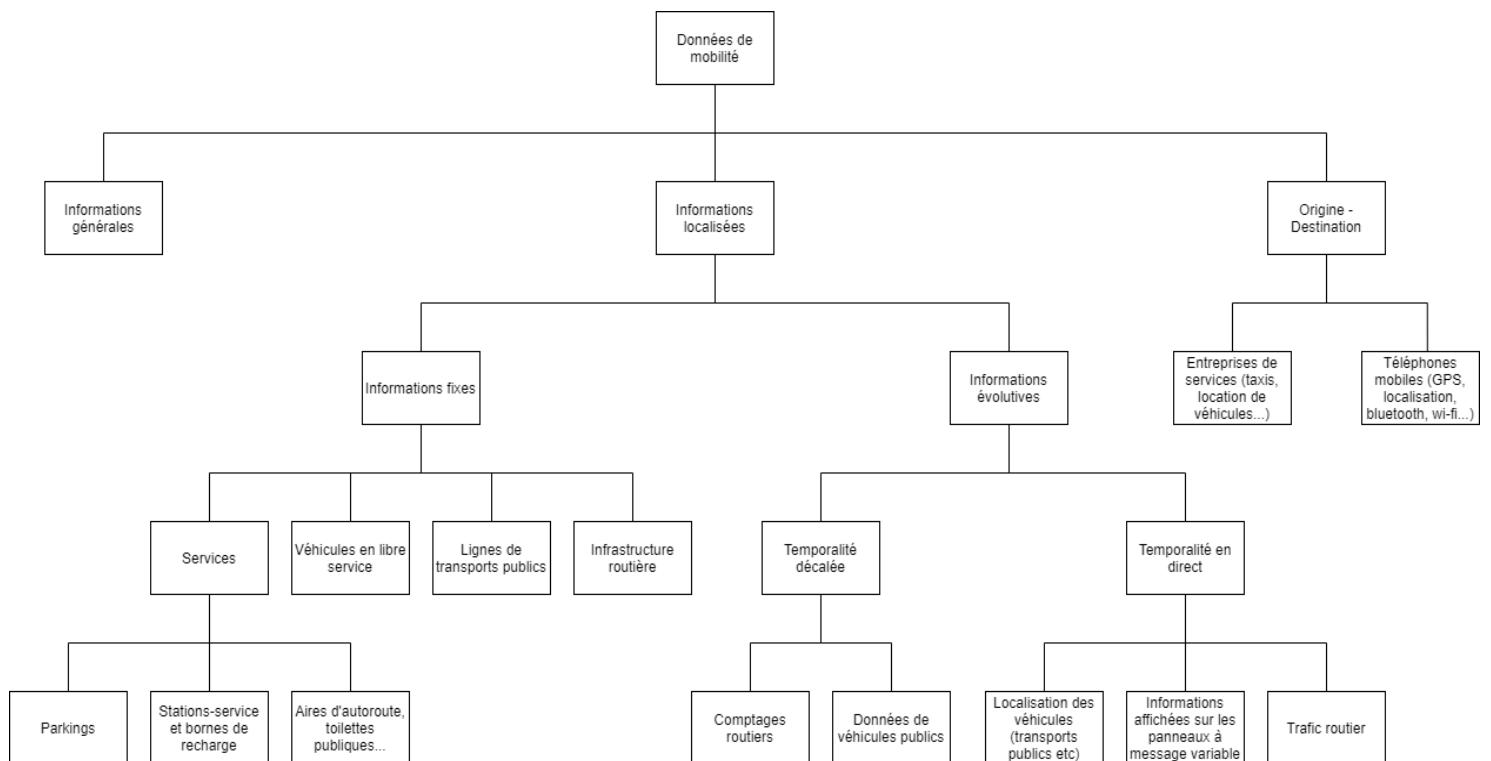
### 3 Autres données de mobilité

Au-delà des comptages routiers traités dans ce travail, il existe de nombreux autres types de données pouvant être qualifiées de “données de mobilité”.

Ce chapitre, tranchant avec le chapitre précédent qui détaillait les données utilisées à la Ville de Lausanne ainsi que dans ce travail, développe les différents types de données de mobilité utilisées dans le monde, les trie dans des catégories classées dans une arborescence, et présente leurs utilités concrètes. Enfin, quelques formats de représentation de ces données, établis et largement utilisés en pratique, sont expliqués.

#### 3.1 Catégorisation

Afin de simplifier et clarifier leur présentation de ces données de mobilité plus générales, elles ont été classées dans une arborescence ([fig. 11](#)). Il est à noter que ce schéma peut être incomplet, et a pour unique but de donner une vue d'ensemble des différents types de données de mobilité existantes et de les hiérarchiser.



*Fig. 11 : Arborescence des données de mobilité (source : conception personnelle)*

- **Informations générales**

Ces données, en principe, concernent le pays entier, un canton ou une zone suffisamment large pour pouvoir en tirer des chiffres significatifs. On peut citer par exemple le microrecensement mobilité et transports effectué par l'Office fédéral du développement territorial (ARE), qui regroupe des informations diverses concernant les habitudes en termes de mobilité de la population suisse (voir [fig. 12](#)).

Ce microrecensement se base sur des données brutes qui ne sont pas nécessairement accessibles à tout un chacun, collectées par l'Office fédéral de la statistique, et dont la finalité est de fournir des chiffres généraux, tels que la distance journalière parcourue en moyenne par personne, ou la part de chaque véhicule dans ces déplacements quotidiens.

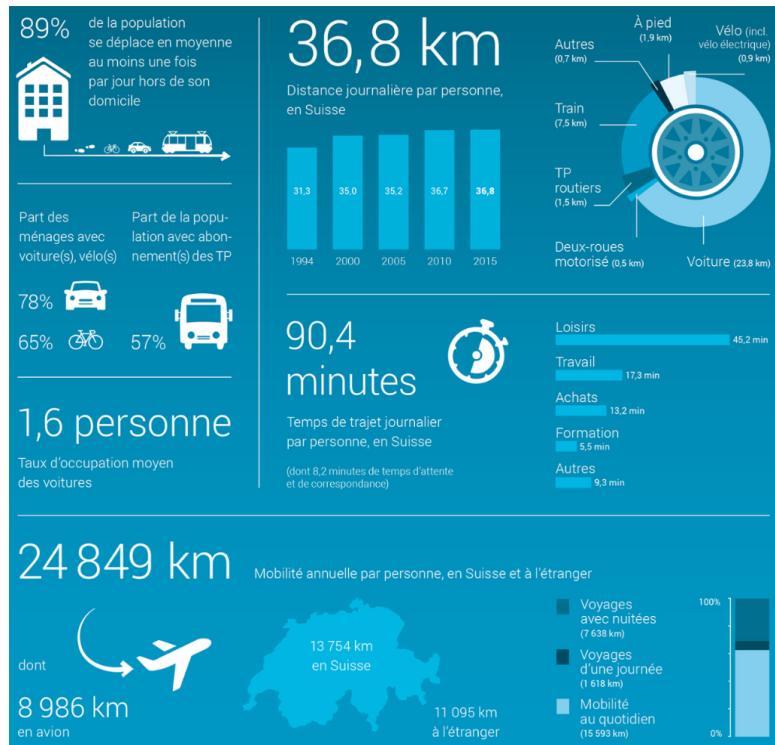


Fig. 12 : Extrait du microrecensement de 2015 (source : [\[are23\]](#))

#### • Informations localisées fixes

Ces informations comprennent une large diversité d'éléments centrés sur la mobilité qui ne varient en principe pas d'un jour à l'autre et concernent une situation géographique particulière, une route par exemple.

- Les services incluent les larges infrastructures utiles aux voyageurs. Les informations qui y sont liées comprennent notamment la localisation du service, son type, sa capacité (dans le cas d'un parking), les embouts ou types carburants disponibles (dans le cas d'une station-service ou d'une borne de recharge), les genres d'installations utilisables (toilettes publiques, aires de jeux...), le tout dans le but de pouvoir répertorier de manière précise ce qui est proposé aux conducteurs et conductrices.
- De nombreuses villes proposent désormais la possibilité de louer des trottinettes ou des vélos en libre-service (exemple en [fig. 13](#)). Ceci implique tout une démarche afin de répertorier et publier les informations appropriées, telles que la localisation des stations, le prix éventuel ou même les pics d'utilisation à certaines heures. Ces informations de véhicules en libre-service peuvent aussi concerner des entreprises privées, Mobility CarSharing par exemple, qui propose un service de location et de partage de voitures.



*Fig. 13 : Exemple de vélos en libre-service à Fribourg, Suisse (source : conception personnelle)*

- Les transports publics sont aussi une source importante de données de mobilité. On va ici répertorier les différentes lignes exploitées par la compagnie, les arrêts auxquels elles s'arrêtent et leurs horaires notamment.
  - L'infrastructure routière peut comprendre des informations sur les chaussées (leur type, la vitesse limite, leur localisation...), sur les panneaux de signalisation ou sur les voies mises en place, des bandes cyclables par exemple. En bref, il s'agit là de tous les éléments du réseau routier que les véhicules empruntent jour après jour, auxquels on peut aussi inclure les zones piétonnes et les trottoirs.
- **Informations localisées évolutives**

Cette catégorie concerne les informations qui ne sont quasiment jamais identiques d'un jour à l'autre, mais sont suffisamment localisées pour que l'on puisse les identifier sur une carte.

    - Ces informations peuvent avoir une temporalité décalée, c'est-à-dire qu'elles sont récoltées périodiquement et traitées après-coup, afin d'établir des analyses sur les emplacements concernés. C'est dans cette catégorie que seraient placés les comptages routiers qui sont discutés dans ce travail.  
On peut aussi mentionner les données des transports publics, telles que la fréquentation de chaque ligne par tranche d'heures, ou alors la ponctualité des trajets effectués au cours d'une journée.
    - Mais elles peuvent aussi être issues d'un flux en direct, qui seront alors directement disponibles. Par exemple, les transports publics sont en tout temps localisés, d'une part pour que le central puisse les gérer de manière efficace, mais aussi pour pouvoir informer les passagers des temps d'attente.  
On peut aussi citer dans cette catégorie les informations affichées sur les panneaux à message variable, visibles le plus souvent sur l'autoroute (voir [fig. 14](#)).

Enfin, le trafic routier est systématiquement répertorié par les caméras de circulation, satellites et autres moyens de surveillance des automobilistes, et en particulier les événements qui l'affectent : c'est ainsi que les accidents ou les travaux sont enregistrés et immédiatement communiqués aux différents acteurs de l'information routière.



Fig. 14 : Exemple de panneau à message variable (source : [\[sig23\]](#))

- **Origine - Destination**

Ce dernier type d'information est un peu plus particulier. Il est assez similaire aux informations évolutives, à quelques différences près. Le principe est le suivant : on stocke le lieu de départ et le lieu d'arrivée de chaque entité en déplacement. L'information ici n'est donc pas centrée sur une localisation en particulier, mais bien sur les voyageurs.

- Par exemple, les taxis représentent un des cas les plus évidents. On peut ainsi connaître l'emplacement exact de l'origine du client, ainsi que celui de sa destination. Ce procédé peut également fonctionner avec les locations de véhicules, et dans une moindre mesure avec les véhicules en libre-service.
- Une autre source de données Origine - Destination très prisée provient de nos téléphones portables. Dans les limites de ce que la loi du pays en question permet, nos chers compagnons peuvent se révéler être une mine d'or de données de mobilité : dans le cas où les services de géolocalisation sont activés, ils connaissent en permanence notre position et nos déplacements, permettant aux éventuels possesseurs de ces données de retracer avec une certaine précision nos déplacements de tous les jours. De plus, dans le monde ultra-connecté qui est le nôtre, d'autres technologies comme le Bluetooth peuvent servir à tracer les positions relatives des téléphones portables.

## 3.2 Utilité de ces données

Dans ce sous-chapitre, quelques exemples d'utilisation de ces données de mobilité sont présentés, au-delà des comptages routiers qui détaillés dans le reste de ce travail.

- Des chercheurs ont représenté dans une carte ([fig. 15](#)) les positions d'environ 10'000 taxis à Shenzhen (Chine) sur une intervalle de 15 minutes. Ces informations leur ont notamment permis de simuler les conditions du trafic en temps réel sur chaque axe, et ce sans la moindre utilisation de caméra ou d'autre type d'infrastructure potentiellement coûteux.

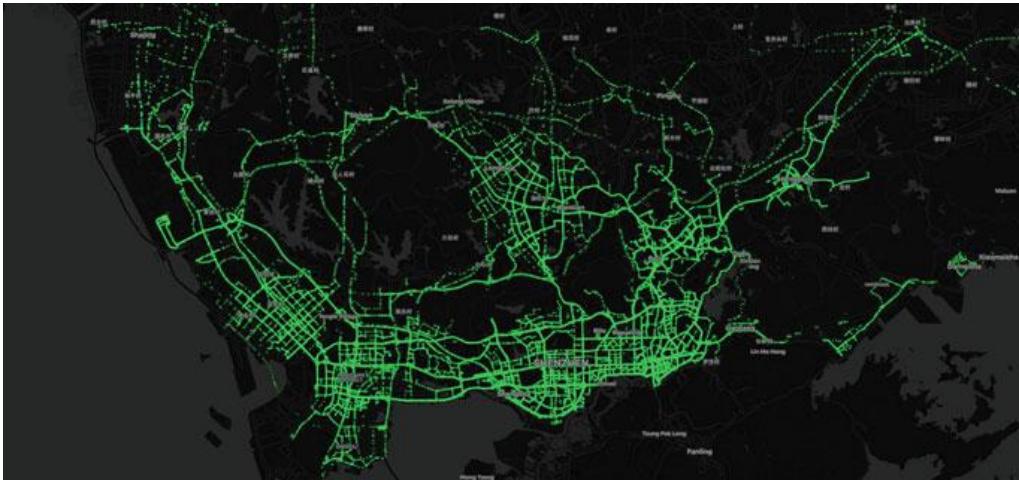


Fig. 15 : Couverture routière fournie par la localisation des taxis (source : [\[liu21\]](#))

- Il est possible de trouver, en accès libre, les données de départ et arrivée de la plupart des compagnies de transports publics en Suisse. À chaque fois qu'un train, un bus ou un autre type de véhicule enregistré passe à un arrêt, plusieurs informations sont enregistrées, telles que la compagnie, la ligne, le nom de l'arrêt, l'heure d'arrivée prévue, l'heure d'arrivée constatée, l'heure de départ prévue et l'heure de départ constatée.
- Ces informations sont échangées de manière directe entre les compagnies concernées, ce qui permet par exemple d'afficher le retard d'un bus TPF<sup>8</sup> sur l'application mobile CFF<sup>9</sup>, ou d'informer les voyageurs d'un changement d'horaires. La disponibilité de ces données on-ne-peut-plus complète ouvre également de nombreuses possibilités. Par exemple, il est possible de déterminer que l'IC1, ligne de train qui relie Genève-Aéroport à Saint-Gall, est arrivé à l'heure en gare de Fribourg dans 63% des cas le 8 février 2023 (voir [fig. 16](#)).

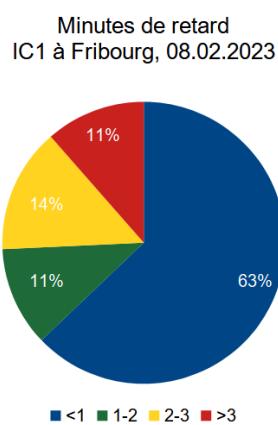


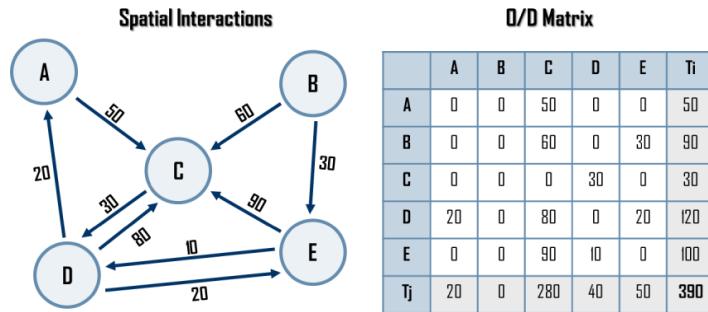
Fig. 16 : Nombre de minutes de retard des trains IC1 à l'arrivée à Fribourg (source : conception personnelle)

- Il existe toute une série de méthodes d'analyse de données de mobilité qui ne seront pas détaillées ici. Par exemple, certaines entreprises spécialisées dans ce genre d'analyse proposent d'établir des matrices origine-destination ; un exemple simple est à la [fig. 17](#).

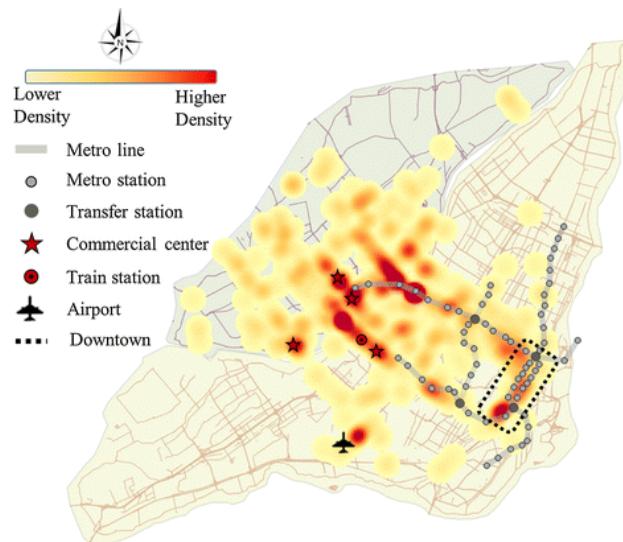
<sup>8</sup> Transports Publics Fribourgeois

<sup>9</sup> Chemins de Fer Fédéraux

Appliqué à l'échelle d'une ville, ce genre de processus peut résulter en une "carte de chaleur" (heatmap en anglais, voir [fig. 18](#)), qui permet de visualiser les points chauds de la ville en termes de transit. Des actions peuvent ensuite être décidées par les responsables de la mobilité de la ville pour pouvoir répondre à la demande.



*Fig. 17 : Exemple simple de matrice origine-destination (source : [\[mit23\]](#))*



*Fig. 18 : Exemple de "heatmap", ici pour les voyages entre Montréal et Laval, Canada (source : [\[ali17\]](#))*

### 3.3 GTFS et GBFS

GTFS<sup>10</sup>, ou Spécification Générale des Flux de Transit, est une spécification de données qui permet aux compagnies de transports publics de publier leurs données de mobilité dans un format qui peut être utilisé par une large variété d'applications.

Plus concrètement, GTFS va fournir aux gestionnaires de données un format à respecter pour qu'elles puissent être représentées d'une manière standardisée par rapport à celles provenant d'autres compagnies de transports publics, et qu'elles puissent ainsi être stockées et utilisées par d'autres entités.

GTFS comprend deux modules :

- Un module contenant des informations fixes, telles que les horaires, les tarifs ou les informations géographiques sur les lignes disponibles au public et leurs arrêts

<sup>10</sup> General Transit Feed Specification en anglais

- Un module contenant des informations en direct sur la localisation des véhicules, ou leur heure d'arrivée estimée par exemple

Le premier module de GTFS est composé d'une série de fichiers texte (donc stockés en .txt), dont voici quelques exemples :

- stops.txt, contient des informations sur les arrêts. La [fig. 19](#) montre un extrait d'un exemple d'utilisation de ce fichier. De gauche à droite sur la figure, chaque arrêt va être représenté par :
  - Son identifiant
  - L'identifiant de l'étage auquel il est
  - Son nom
  - Sa latitude
  - Sa longitude
  - Son type (simple plateforme, station complète, entrée/sortie...)
  - L'arrêt “parent” éventuel, par exemple la station qui contient une plateforme

```
stop_id,level_id,stop_name,stop_lat,stop_lon,location_type,parent_station
F12,,5 Av/53 St,40.760167,-73.975224,1,
E1,L0,5 Av/53 St SW,40.760474,-73.976099,2,F12
E2,L0,5 Av/53 St NE,40.76035,-73.97546,2,F12
E3,L0,5 Av/53 St SE,40.760212,-73.975512,2,F12
E4,L0,Madison/53 St NE,40.759612,-73.973731,2,F12
E5,L0,Madison/53 St SE,40.759491,-73.973820,2,F12
```

*Fig. 19 : Extrait d'exemple d'utilisation de stops.txt (source : [\[gtf23a\]](#))*

- trips.txt, contient des informations sur les trajets effectués par les véhicules de la compagnie. La [fig. 20](#) montre un extrait d'un exemple d'utilisation de ce fichier. De gauche à droite sur la figure, chaque trajet va être représenté par :
  - L'identifiant de la ligne parcourue
  - L'identifiant du service effectué
  - Son identifiant
  - Son en-tête (qui sera souvent affiché sur le véhicule)
  - L'identifiant du bloc (ensemble regroupant plusieurs trajets, par exemple effectués par le même véhicule)

```
route_id,service_id,trip_id,trip_headsign,block_id
A,WE,AWE1,Downtown,1
A,WE,AWE2,Downtown,2
```

*Fig. 20 : Extrait d'exemple d'utilisation de trips.txt (source : [\[gtf23a\]](#))*

- stop\_times.txt, contient des informations sur les horaires. La [fig. 21](#) montre un extrait d'un exemple d'utilisation de ce fichier. De gauche à droite sur la figure, chaque halte d'un véhicule va être représentée par :
  - L'identifiant du trajet concerné
  - L'heure d'arrivée prévue
  - L'heure de départ prévu
  - L'identifiant de l'arrêt
  - La position de l'arrêt par rapport aux autres sur cette ligne (par exemple, celui ayant 5 sera avant celui ayant 12)

- Le type de ramassage (par exemple, 0 indique un ramassage normal, tandis que 1 signifie que les passagers ne peuvent pas entrer à cet arrêt)
- Le type de dépose (fonctionnant avec le même principe)

```
trip_id,arrival_time,departure_time,stop_id,stop_sequence,pickup_type,drop_off_type
AWE1,0:06:10,0:06:10,S1,1,0,0
AWE1,,,S2,2,1,3
AWE1,0:06:20,0:06:30,S3,3,0,0
AWE1,,,S5,4,0,0
AWE1,0:06:45,0:06:45,S6,5,0,0
```

*Fig. 21 : Extrait d'exemple d'utilisation de stop\_times.txt (source : [gtfs23a1](#))*

Il est à noter que dans les exemples ci-dessus, les champs utilisés peuvent varier d'une implémentation à l'autre : certains sont obligatoires, mais d'autres sont optionnels.

Le deuxième module de GTFS comporte 3 types d'informations utilisables :

- Mises à jour sur le trajet (retards, annulations, changement)
- Alerte de service (arrêt déplacé, événement imprévu affectant un endroit particulier, une ligne ou le réseau)
- Position des véhicules

Les informations sont échangées sous forme de flux, et utilisent un format appelé Protocol Buffers, qui est un mécanisme pour sérialiser les données de manière indépendante d'un quelconque langage ou d'une quelconque plateforme, le tout de manière relativement facilement lisible. La [fig. 22](#) montre un extrait d'un exemple d'utilisation de ce module de GTFS, en l'occurrence une mise à jour de trajet.

Plus précisément, la mise à jour concerne un retard de 5 secondes à l'arrivée à l'arrêt numéro 3 dans la séquence du trajet portant l'identifiant "trip-1". Ce retard sera propagé aux arrêts suivant du trajet.

MobilityData, l'entreprise s'occupant de maintenir GTFS, a aussi développé une deuxième spécification pour données de mobilité, cette fois-ci centrée sur les services de véhicules en location ou en libre-service. GBFS<sup>11</sup>, ou Spécification Générale de Flux de Partage de Vélos, a pour but de faciliter le partage d'information par ces services de vélos en libre-accès, et plus largement par toute entreprise proposant un service de mobilité partagée.

Le fonctionnement de GBFS est un peu différent de GTFS. Toutes les informations sont stockées sous forme de fichiers JSON<sup>12</sup>, qui est une notation permettant de représenter les données sous forme de listes d'objets structurés et lisibles. Une douzaine de fichiers peuvent être mis en ligne, mais la plupart sont optionnels, et c'est à l'entreprise de définir lesquels sont pertinents dans son cas. Ils peuvent contenir des informations sur l'entreprise, sur les stations exploitées, sur les véhicules disponibles, sur les horaires du système, sur ses tarifs...

---

<sup>11</sup> General Bikeshare Feed Specification

<sup>12</sup> JavaScript Object Notation

```

# header information
header {
    # version of speed specification. Currently "2.0". Valid versions are "2.0", "1.0".
    gtfs_realtime_version: "2.0"
    # determines whether dataset is incremental or full
    incrementality: FULL_DATASET
    # the moment where this dataset was generated on server
    timestamp: 1284457468
}

# multiple entities can be included in the feed
entity {
    # unique identifier for the entity
    id: "simple-trip"

    # "type" of the entity
    trip_update {
        trip {
            # selects which GTFS entity (trip) will be affected
            trip_id: "trip-1"
        }
        # schedule information update
        stop_time_update {
            # selecting which stop is affected
            stop_sequence: 3
            # for the vehicle's arrival time
            arrival {
                # to be delayed with 5 seconds
                delay: 5
            }
        }
    }
    # ...this vehicle's delay is propagated to its subsequent stops.
}

```

*Fig. 22 : Extrait d'exemple d'une mise à jour de trajet GTFS (source : [\[gtf23b\]](#))*

Voici quelques exemples de fichiers faisant partie de la spécification GBFS :

- `vehicle_types.json`, contient des informations sur les véhicules disponibles à la location. La [fig. 23](#) montre un extrait d'un exemple d'utilisation de ce fichier. Il y est décrit un véhicule dont l'identifiant du type est “abc123”, son facteur de forme est un vélo, à propulsion humaine, nommé “Example Basic Bike”, possédant 2 roues, pouvant être réservé jusqu'à 30 minutes à l'avance et pouvant être retourné à n'importe quelle station. Une icône est également précisée, ainsi que les tarifs de location disponibles.

```

"vehicle_type_id": "abc123",
"form_factor": "bicycle",
"propulsion_type": "human",
"name": "Example Basic Bike",
"wheel_count": 2,
"default_reserve_time": 30,
"return_constraint": "any_station",
"vehicle_assets": {
  "icon_url": "https://www.example.com/assets/icon_bicycle.svg",
  "icon_url_dark": "https://www.example.com/assets/icon_bicycle_dark.svg",
  "icon_last_modified": "2021-06-15"
},
"default_pricing_plan_id": "bike_plan_1",
"pricing_plan_ids": [
  "bike_plan_1",
  "bike_plan_2",
  "bike_plan_3"
]

```

Fig. 23 : Extrait d'exemple d'utilisation de vehicle\_types.json (source : [\[gbf23\]](#))

- station\_information.json, contient des informations sur les véhicules disponibles à la location. La [fig. 24](#) montre un extrait d'un exemple d'utilisation de ce fichier. Il y est décrit une station dont l'identifiant est “pga”, nommée “Parking garage A”, dont la latitude est 12.345678, la longitude est 45.678901, le type de parking dont elle dispose est souterrain, si des arceaux de parkings y sont installés, le numéro de téléphone de contact, si elle a de quoi recharger les véhicules, et enfin sa capacité pour chaque type de véhicule disponible.

```

"station_id": "pga",
"name": "Parking garage A",
"lat": 12.345678,
"lon": 45.678901,
"parking_type": "underground_parking",
"parking_hoop": false,
"contact_phone": "+33109874321",
"is_charging_station": "true",
"vehicle_type_capacity": {
  "abc123": 7,
  "def456": 9
}

```

Fig. 24 : Extrait d'exemple d'utilisation de station\_information.json (source : [\[gbf23\]](#))

- station\_status.json, contient des informations sur les véhicules disponibles à la location. La [fig. 25](#) montre un extrait d'un exemple d'utilisation de ce fichier. Il y est décrit l'état de la station dont l'identifiant est “station1”, qui s'est donc bien installée, est en état de louer des véhicules et de les réceptionner, s'est signalée pour la dernière fois à un temps donné, possède 3 quais disponibles et 1 quai désactivé. Il y figure ensuite une description des quais disponibles avec les véhicules qu'ils peuvent accepter et leur nombre. Enfin, le nombre de véhicules disponibles et désactivés est indiqué, ainsi que le nombre de véhicules disponibles par type.

```

"station_id": "station1",
"is_installed": true,
"is_renting": true,
"is_returning": true,
"last_reported": 1609866125,
"num_docks_available": 3,
"num_docks_disabled": 1,
"vehicle_docks_available": [
  {
    "vehicle_type_ids": [ "abc123", "def456" ],
    "count": 2
  },
  {
    "vehicle_type_ids": [ "def456" ],
    "count": 1
  }
],
"num_bikes_available": 1,
"num_bikes_disabled": 2,
"vehicle_types_available": [
  {
    "vehicle_type_id": "abc123",
    "count": 1
  },
  {
    "vehicle_type_id": "def456",
    "count": 0
  }
]

```

*Fig. 25 : Extrait d'exemple d'utilisation de station\_status.json (source : [\[gbf23\]](#))*

Ces spécifications sont particulièrement avantageuses pour les acteurs de la mobilité, qui peuvent ainsi informer leurs clients et clientes de l'état de leur service de manière simplifiée. En effet, standardiser la manière de partager l'information facilite la tâche des entités qui vont ensuite la récolter pour la distribuer aux utilisateurs et utilisatrices. Les entreprises produisant les données ont un modèle tout fait pour leur organisation et n'ont pas besoin d'en élaborer un eux-mêmes, et ceux qui les exploitent n'ont pas besoin de changer leurs procédés pour chaque nouvelle source de données.

C'est ainsi que début février 2023, GTFS est utilisée par plus de 10'000 compagnies de transports publics réparties sur plus de 100 pays. GBFS possède une utilisation plus modeste mais grandissante, avec plus de 700 entreprises inscrites, réparties sur plus de 40 pays. Ces larges bases d'entreprises respectant les spécifications permettent d'exploiter des grandes quantités de données, ce qui peut se révéler très utile notamment pour des applications comme Google Maps, Transit ou Moovit, et ce de manière libre d'accès. C'est précisément grâce à GTFS qu'il est possible pour toutes ces applications de proposer à leurs utilisateurs et utilisatrices de voir en temps réel les prochains départs aux arrêts de bus par exemple.

### 3.4 DATEX II

DATEX II est un langage électronique comprenant une spécification de données de mobilité plus large que GTFS et GBFS, utilisé en Europe, et spécialisé dans l'échange d'informations sur le trafic routier. Ces dernières peuvent notamment couvrir l'état du trafic, les messages affichés sur les panneaux à message variable, les parkings disponibles, les stations-service et bornes de recharge, les régulations du trafic en place...

Le but de DATEX II, comme GTFS et GBFS, est de faciliter l'échange d'informations entre acteurs de la mobilité. Dans les services concernés, on peut citer la communication d'informations de gestion du trafic d'opérateur à opérateur, la situation localisée du trafic et les mesures mises en place par les opérateurs routiers (par exemple en cas d'accident ou de travaux) ou la gestion de données de trafic routier.

La [fig. 26](#) présente les différents composants de DATEX II.

L'opérateur et fournisseur des données doit d'abord déterminer les termes fonctionnels de l'utilisation qu'il veut faire de DATEX II. Par exemple, avec qui va-t-il échanger ? Pour quel réseau routier ? Quel genre d'information, et comment est-elle collectée ?

En fonction des réponses aux questions ci-dessus, certains sous-domaines seront utilisés, d'autres non. Par exemple, il se peut que l'opérateur n'ait pas accès aux informations des parkings du réseau routier, ou qu'il n'y en ait pas. L'encodage et l'organisation, puis la publication des données sont ensuite régulés par les standards définis par DATEX II.

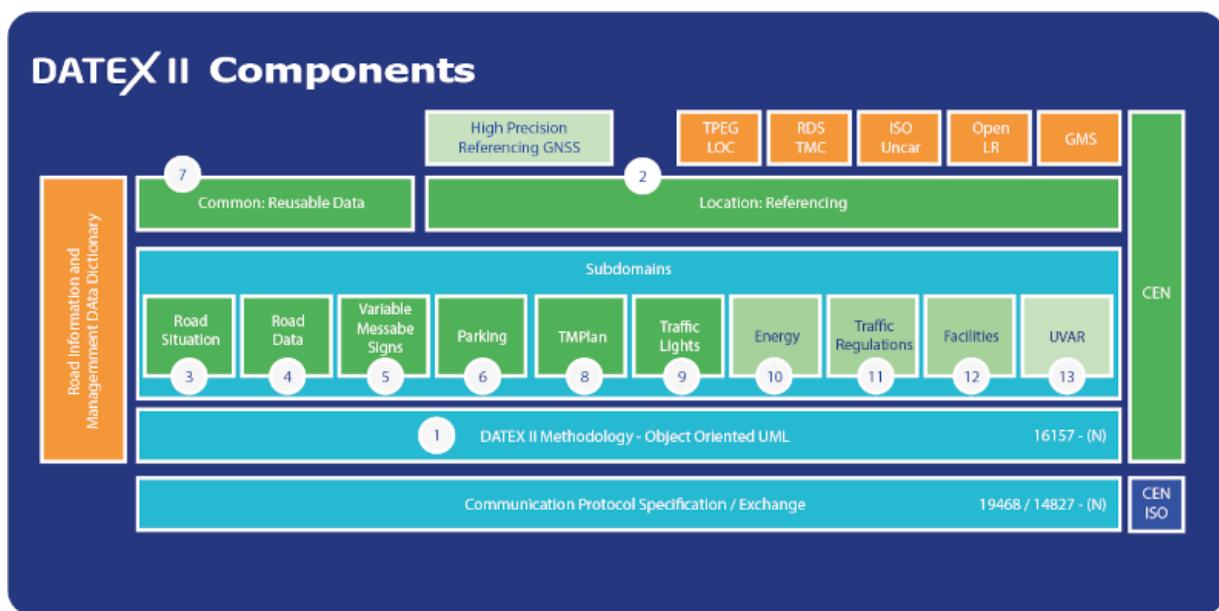


Fig. 26 : Schéma représentant les différents composants de DATEX II (source : [\[dat23\]](#))

Le champ des informations pouvant être traitées par DATEX II est relativement vaste. En voici donc quelques exemples :

- **Location Referencing**

Le référencement de localisation est une collection d'informations liées à un emplacement particulier dans un réseau routier. Cet emplacement peut être enregistré sous forme de point, d'intervalle linéaire entre deux points, ou de zone.

Dans tous ces cas, plusieurs méthodes de représentation de ces données géographiques sont prises en charge par DATEX II, qui ne s'occupe alors que d'enregistrer l'emplacement dans le format choisi par l'utilisateur.

- **Road Traffic Data**

Les données de trafic routier peuvent prendre trois formes différentes : elles peuvent concerner soit un tableau des sites de mesure, soit des données de mesures, soit des données avancées.

Un tableau de sites de mesure est à utiliser en parallèle de données de mesure. Ces informations vont concerner les métadonnées d'une mesure (par exemple un comptage routier), c'est-à-dire l'équipement utilisé, le nombre de voies comptées, la précision de la mesure...

Les données de mesures contiennent les résultats concrets de ces mesures. Elles peuvent concerner les données de mobilité suivantes :

- Temps entre chaque véhicule
- Distance entre chaque véhicule
- Fluidité du trafic
- Vitesse du trafic
- Concentration du trafic
- Mesures individuelles des véhicules

En plus de cela, ces données de mesure peuvent aussi contenir des informations sur le temps de trajet sur l'axe mesuré, ainsi que des données situationnelles telles que les conditions météorologiques, le vent, la pollution ambiante, la condition de la route, la visibilité...

Enfin, les données élaborées ont généralement déjà été traitées et analysées. Elles concernent les mêmes attributs que les données mesurées, mais ne peuvent par nature pas contenir de mesures individuelles de véhicules. La différence entre ces deux types d'information est quelque peu subtile, et s'apparente dans ma classification à la différence entre les données brutes et les données agrégées ou générales.

Les informations échangées avec DATEX II sont généralement sous forme de données mesurées, mais il peut arriver que les données élaborées soient directement utilisées à la place, en fonction du contexte ; par exemple, si le récepteur est connu et a uniquement besoin de certaines informations bien spécifiques. À l'inverse, il peut être bien de publier les données mesurées dans certains cas, par exemple si les données peuvent être utilisées par des applications qui préféreront les traiter elles-mêmes.

- **Traffic Regulation**

Les règles de circulation comprennent plusieurs informations.

- Son origine : elle peut émaner des autorités compétentes, par des acteurs autorisés par les fournisseurs de l'information, ou d'un plan établi au préalable et annoncé. Dans tous ces cas, les informations contiennent différents attributs, tels que la base légale si applicable ou le délai de validité.
- Son champ d'application : il doit contenir au moins une condition et peut en contenir plusieurs. Par exemple, le type de la route concerné, la localisation en question, le type de conducteur.trice concerné...

- Son type : si c'est un avertissement (par exemple en cas de bouchon devant, ou de rétrécissement de la chaussée), une restriction applicable à une voie en particulier, une limitation de la vitesse, une régulation de l'autorisation de se garer...

En résumé, la collecte et l'analyse de données de mobilité de manière générale sont devenues indispensables dans le monde de la mobilité d'aujourd'hui, pour des objectifs de communication, de planification urbaine ou de développement de nouvelles solutions liées au déplacement d'individus par exemple. C'est dans ce contexte que des standards, comme GTFS ou DATEX II, ont vu le jour et continuent à être améliorés et utilisés par de plus en plus d'acteurs de la mobilité.

La parenthèse consacrée à une vue plus générale sur ces données de mobilité étant refermée à l'issue de ce chapitre, la suite de ce travail se recentre sur les données issues des comptages routiers exploitées notamment par la Ville de Lausanne, et dans un premier temps, sur leur stockage.

## 4 Stockage des données

Maintenant que les données utilisées dans ce travail ont été définies, il faut définir la manière dont on veut les stocker, autant en termes de support informatique qu'en termes de techniques employées.

Ce chapitre présente deux manières d'y parvenir, par fichiers individuels et par base de données. Une mise en avant de l'optimisation de l'espace de stockage est effectuée, notamment par l'agrégation de certaines données. Enfin, un point est fait sur les spécificités de la représentation informatique de données géographiques, ainsi que sur la fiabilité sur le long-terme des formats présentés.

### 4.1 Stockage par fichier

Les données issues des comptages routiers, comme présentés au [chapitre 2](#), sont le plus souvent stockées sous forme de fichiers, aux formats divers, comme des .xlsx<sup>13</sup> ou des .csv<sup>14</sup>. À la division mobilité de la Ville de Lausanne, par exemple, chaque source de données utilise sa propre manière de les enregistrer et utilise un de ces deux formats. Ces fichiers sont ensuite stockés dans des dossiers plus ou moins bien organisés, eux-mêmes stockés sur les serveurs internes de la Ville. Il faut ensuite savoir où aller, puis chercher le bon fichier, qui sera tantôt classé par nom de l'endroit où le comptage a eu lieu, tantôt par identifiant du dispositif, tantôt par date...

Au-delà de la perte de temps que peut générer une telle organisation, certaines manières de stocker les données peuvent s'avérer coûteuses en espace disque. La suite de ce chapitre présente une estimation du calcul du besoin actuel d'espace de stockage par jour, pour les comptages traités dans ce travail.

- **Études et comptages réalisés par des mandataires externes**

On peut estimer qu'une feuille .xlsx de comptage ponctuel fait en moyenne 500 kilooctets (abrégés *ko* par la suite), et qu'il y a une centaine de tels sites de comptages par année. La quantité d'espace de stockage nécessaire par jour est donc de  $500 * 100/365 \approx 135 \text{ ko}$  en moyenne.

- **Compteurs piéton/vélo**

Ce système de comptage est un cas un peu plus particulier. En effet, les données sont enregistrées sur la plateforme en ligne du fabricant de ces compteurs, et les seules données éventuellement enregistrées sur les serveurs de la Ville consistent en des données agrégées, téléchargées pour certaines occasions. Aucun espace de stockage n'est donc prévu dans ce calcul pour les compteurs piéton/vélo.

- **Radars bruit**

Les fichiers générés par ce radar, possédé à exemplaire unique par la Ville au moment où ce travail est écrit, sont en format .csv, où chaque ligne représente un passage de véhicule. Sur un axe routier où environ 10'000 passages sont enregistrés en moyenne chaque jour dans les deux sens cumulés, le fichier résultant

---

<sup>13</sup> Extension de nom de fichier pour tableur, utilisée notamment pour les fichiers Excel

<sup>14</sup> Format représentant des données tabulaires sous forme de valeurs séparées par des virgules

est de 450 ko. En sachant que les routes enregistrées par ce radar sont souvent des axes aussi importants que celui-là mais peuvent varier, qu'il y a moins de passage le week-end, et qu'il peut y avoir quelques jours de battements pendant chaque déménagement du dispositif, ce chiffre de 450 ko journaliers est à prendre avec une certaine marge d'erreur.

- **Boucles de comptage**

Ces boucles génèrent des fichiers en .csv sur un modèle similaire aux radars bruits, à la différence près que moins d'informations sont enregistrées et que les fichiers sont donc légèrement moins volumineux ; environ 400 ko pour une journée de comptage sur une route à 10'000 passages, toutes voies cumulées. Ce chiffre est aussi un peu plus stable, car les boucles de comptage ne sont pas déplacées, et fonctionnent (en principe) tous les jours.

Au moment où ce travail est écrit, 7 boucles de comptage sont en service, amenant le total à environ 2'800 ko.

- **Indicateurs de vitesse**

Ces dispositifs comportent des variations.

- Certains indicateurs sont connectés et peuvent transmettre leurs données en direct sur une plateforme en ligne, d'où la section exploitation peut les exporter ponctuellement dans l'année.
- D'autres, fonctionnant sur batterie, nécessitent d'aller télécharger les données sur place.
- D'autres enfin, permettent d'exporter les données directement en ligne aussi, mais sur une plateforme distincte et dans un format différent.

Les données de ce type de dispositif ne sont, en général, pas systématiquement enregistrées sur les serveurs de la Ville. On considérera ici que, sur une année, seulement 5 semaines sont enregistrées en moyenne pour en analyser les données. Chaque indicateur de vitesse produit un fichier pour la semaine enregistrée.

Ceux produits par les deux premières variations définies ci-dessus sont au format .xlsx et pèsent entre 200 et 2'800 ko, en fonction du nombre de passages enregistrés à l'endroit où le dispositif est déployé, pour une moyenne de 1'600 ko par fichier. Au moment où ce travail est écrit, environ 85 indicateurs de vitesse de ces deux variations sont déployés.

Ceux produits par la dernière variation définie ci-dessus sont au format .csv et pèsent entre 20 et 630 ko, pour une moyenne de 400 ko par fichier. Au moment où ce travail est écrit, environ 10 indicateurs de vitesse de cette variation sont déployés. Le calcul du total d'espace requis pour une semaine de comptages est le suivant :

$$85 * 1'600 + 10 * 400 = 140'000 \text{ ko.}$$

Sur une année, on multiplie ce chiffre par 5 comme convenu, ce qui donne un total de 700'000 ko. La moyenne journalière se situe donc aux environs de  $700'000/365 \approx 1'900 \text{ ko.}$

- **Radars de comptage**

L'utilisation de ces radars est très irrégulière. Il est simplement estimé ici que 4 radars seront utilisés dans l'année, pendant 3 semaines chacun au total. Les données récoltées sur ces radars sont dans un format binaire illisible par un humain, mais qui peut être transformé par la voie d'un logiciel en fichiers .xlsx. On estime la

taille d'un fichier à 1'500 ko pour une semaine de comptage. La taille nécessaire pour une année pour tous les appareils combinés est donc de  $1'500 * 12 = 18'000 \text{ ko}$ , soit  $18'000/365 \approx 50 \text{ ko}$  par jour.

Le total d'espace de stockage nécessaire au moment où ce travail est écrit est donc d'environ  $135 + 450 + 2'800 + 1'900 + 50 = 5'335 \text{ ko}$  quotidiennement. Généraliser ce chiffre sur une année augmente inévitablement son imprécision, il est ainsi estimé que l'espace de stockage nécessaire est de près de 2 gigaoctets annuellement.

## 4.2 Utilité de l'agrégation

Si ce chiffre de 2 gigaoctets de données par année peut paraître convenable pour des serveurs comme ceux de la Ville de Lausanne, il représente tout de même une taille assez conséquente à long-terme, taille pouvant poser problème notamment dans le cas d'un stockage centralisé et uniformisé de toutes ces données.

Une question se pose alors : dans quelle mesure est-il possible de réduire cette taille en agrégeant un peu plus les données, et quelles seraient les conséquences pratiques de telles mesures ?

Agréger des données implique de faire des choix et de potentiellement sacrifier certaines informations. La [fig. 27](#) montre un exemple de données fournies par un radar bruit agrégées par jour, ainsi que par tranche de vitesse (note : les données ont été volontairement légèrement tronquées). Si, dans cet exemple, le choix a été fait de garder les informations des types de véhicules en créant une feuille par type (ainsi qu'une feuille "total"), on perd tout de même d'autres informations, le bruit moyen à certaines heures clé notamment. Cependant, le gain d'espace de stockage est énorme : à la place des  $450 * 365 = 164'250 \text{ ko}$  que prendraient les données brutes sur une année complète, le fichier .xlsx fait ici à peine 140 ko.

Jour	Mesures	Moyenne	Max	Db85	Db50	Db30	Db10	Fréquence moyenne	...occurrences	...pourcentage	...nombre par minute	...temps d'intervalle
Total	61031	68,14	99	71	68	67	65	0-59dB	406	0.67%	0.026/min	39 min 1 s
2022_04_01	6748	68,23	96	71	68	67	65	60-64dB	5344	8.76%	0.337/min	2 min 58 s
2022_04_02	5578	67,73	93	70	68	66	64	65-69dB	38892	63.72%	2.455/min	0 min 24 s
2022_04_03	3626	67,76	99	70	68	66	64	70-74dB	13732	22.5%	0.867/min	1 min 9 s
2022_04_04	6192	68,24	96	71	68	67	65	75-79dB	2242	3.67%	0.142/min	7 min 4 s
2022_04_05	6368	67,96	99	71	68	66	64	80-84dB	314	0.51%	0.02/min	50 min 27 s
2022_04_06	6166	68,71	91	72	68	67	65	85-89dB	64	0.1%	0.004/min	247 min 30 s
2022_04_07	6129	68,52	92	71	68	67	65	90-200dB	37	0.06%	0.002/min	428 min 6 s
2022_04_08	6035	68,15	91	71	68	67	65					
2022_04_09	5182	68,00	91	71	68	67	65	0-66dB	17448	28.59%	1.102/min	0 min 54 s
2022_04_10	3451	67,46	82	70	67	66	64	67-71dB	36252	59.4%	2.289/min	0 min 26 s
2022_04_11	5556	68,28	96	71	68	67	65	72-77dB	6390	10.47%	0.403/min	2 min 29 s
								78-200dB	941	1.54%	0.059/min	16 min 50 s

Fig. 27 : Données récoltées par un radar bruit et agrégées (source : conception personnelle)

Dans ce cas, il est également possible d'agréger les données par heure plutôt que par jour, afin de minimiser un peu plus les informations perdues. L'impact d'une telle agrégation est tout de même notable : le fichier fait ici 2'800 ko, soit près de 60 fois moins que le fichier complet de données brutes sur une année complète.

De la même manière, on peut agréger les données obtenues sur un indicateur de vitesse par heure. Pour un fichier contenant les données brutes sur une journée entière et pesant entre 20 et 630 ko, on passe à un fichier de taille à peu près constante d'environ 2 ko. Un exemple est disponible à la [fig. 28](#), où pour chaque heure et pour chaque direction, on enregistre le nombre de véhicules, leur vitesse moyenne, leur V85, leur V50, leur V30, leur V10 et la vitesse maximale mesurée.

Date	dir	count	avg	v85	v50	v30	v10	max
07.09.2022 11:00	1	503	22	25	21	19	18	68
07.09.2022 12:00	1	457	23	26	21	19	17	81
07.09.2022 13:00	1	454	22	25	21	18	18	35
07.09.2022 14:00	1	418	23	26	21	19	18	55
07.09.2022 15:00	1	421	23	26	21	19	18	38
07.09.2022 16:00	1	461	23	26	21	19	18	61
07.09.2022 17:00	1	446	22	25	21	19	17	38
07.09.2022 18:00	1	494	22	24	20	18	18	79
07.09.2022 19:00	1	343	23	26	21	18	18	35
07.09.2022 20:00	1	223	22	24	20	18	18	32

*Fig. 28 : Extrait de données récoltées par un indicateur de vitesse et agrégées (source : conception personnelle)*

Enfin, dans le cas où l'on veut être sûr de ne perdre aucune information pour les données récentes, il y a toujours la possibilité de décider de n'agréger que certaines données au bout d'un certain temps défini, un an par exemple. L'espace de stockage demandé par les années précédentes ne dépasserait vraisemblablement pas, à plus ou moins long-terme, les 2 gigaoctets demandés par les 365 derniers jours.

Ces bons résultats ne sont cependant pas aussi évidents à obtenir pour certaines sources de données. Par exemple, les données sont parfois déjà fournies agrégées, comme dans le cas des comptages ponctuels commandés à des mandataires externes.

Les synthèses des données des comptages représentent aussi souvent au moins un fichier en plus, ainsi que des processus supplémentaires pour les obtenir, potentiellement des scripts ou logiciels pour automatiser l'agrégation, complexifiant encore un système qui n'est déjà pas simple. De plus, la section exploitation emploie déjà dans plusieurs cas des agrégations de données pour pouvoir facilement les analyser sur une période de temps définie.

Ce qu'il faut donc retenir de ces agrégations de données, c'est que c'est un excellent moyen d'économiser de l'espace de stockage, y compris en cas d'agrégation uniquement après un temps défini, mais qu'en pratique il convient de questionner la rentabilité de complexifier un système de fichiers non-uniformisé, en particulier si l'espace de stockage n'est pas un réel problème.

### 4.3 Bases de données

Ce sous-chapitre s'intéresse à une autre option de stockage des données issues des comptages routiers : les bases de données relationnelles. C'est une solution adaptée et très répandue pour répondre à un besoin de centralisation de données, c'est pourquoi le programme présenté plus loin dans ce travail utilise une de ces bases de données.

Le principal intérêt de ces bases de données est d'apporter une certaine facilité de traitement des données du fait de leur rassemblement en un même endroit. Là où des

données stockées sous formes de .xlsx ou de .csv sont séparées en plusieurs fichiers, eux-mêmes pas toujours enregistrés au même endroit dans un dédale de dossiers, et dont le format dépend souvent du fournisseur de ces informations, une base de données a la capacité de tout regrouper en un même endroit et un même format, offrant enfin la possibilité d'effectuer des opérations sur autant de données que l'on le souhaite.

Formulé plus simplement, une base de données permet de traiter des informations, autrefois dispersées et hétérogènes, de manière uniforme et centralisée.

Un premier défi quant à l'utilisation d'une base de données réside dans le fait d'établir un schéma relationnel (servant à créer la base de données elle-même) étant capable de contenir toutes les informations utilisables des données sous forme de fichiers, si possible sans la moindre perte.

Ces informations, comme vu au [chapitre 2](#), varient d'une source à l'autre. Cependant, des points communs peuvent être relevés et utilisés. Pour l'élaboration du schéma relationnel, il faut donc se servir de la catégorisation vue précédemment. Ainsi, chaque fichier propose des données brutes, des données agrégées par temps, des données agrégées par vitesse, ou plusieurs de ces options. La base de données doit donc être capable de stocker ces trois catégories de données.

Ces catégories ont tout de même un point commun, puisque toutes les données contiennent la date et l'heure de leur enregistrement, et peuvent spécifier le type de véhicule compté. Enfin, elles proviennent tous d'une certaine source, qui devrait pouvoir spécifier le lieu du comptage.

Cette base de réflexion permet donc d'établir un schéma relationnel et une base de données qui sera utilisée pour l'application, dont le processus sera détaillé au [chapitre 6.1](#).

Pour une base de données telle que décrite dans ce travail, l'espace de stockage nécessaire est comparable à celui occupé par un stockage par fichiers pour des données brutes : pour 10'000 passages de véhicules enregistrés individuellement dans une base de données SQLite<sup>15</sup>, on demande 400 à 500 ko, ce qui est comparable à un fichier généré par un radar bruit ou une boucle de comptage, comme vu au [chapitre 4.1](#).

## 4.4 Représentation géographique

Une des informations les plus importantes enregistrées dans un comptage routier, si ce n'est la plus importante, est son emplacement géographique. Sans cela, le comptage n'a plus aucun intérêt.

La représentation de cette information de géolocalisation est assez variable lorsque les données sont stockées sous forme de fichiers. Le plus souvent, le nom de la rue ou de l'endroit où a été effectué le comptage est mentionné dans le nom du fichier, ou dans le nom du dossier le contenant.

Utiliser les coordonnées GPS est un moyen d'enregistrer l'emplacement du comptage de manière précise, standardisée et simple : il suffit de stocker deux valeurs numériques, la latitude et la longitude. La difficulté réside alors dans le fait de passer d'un format imprécis et hétérogène, tel que celui utilisé dans le cas d'un stockage des données dans des fichiers indépendants, à des coordonnées universelles, déterminées, et surtout très précises.

---

<sup>15</sup> SQLite est un moteur de base de données relationnelle accessible par le langage SQL, dont le produit est stocké dans un fichier indépendant plutôt que dans un serveur.

Cette transition, pour passer d'un système de stockage par fichiers vers une base de données par exemple, doit se faire par un être humain dans la plupart des cas. De multiples solutions existent pour permettre de définir des coordonnées en ayant l'emplacement en tête, une carte sur laquelle on peut créer un pointeur par exemple.

Il existe également des logiciels de SIG<sup>16</sup> qui permettent d'afficher et d'interagir avec une telle carte. On peut ainsi mentionner par exemple QGIS, qui permet à la Ville de Lausanne de représenter de nombreuses informations sur une carte de la ville, dont les comptages routiers.

Dans une telle situation, on représente chaque croisement routier par un "nœud", soit une position géographique. Chaque morceau de route reliant chaque nœud forme alors une arête. Un comptage peut ensuite être associé à une arête, ce qui facilite la visualisation des routes sur lesquelles un comptage a déjà été effectué, mais aussi la sélection des comptages intéressants, ainsi que l'insertion de nouveaux comptages dans la base de données associée à l'interface QGIS.

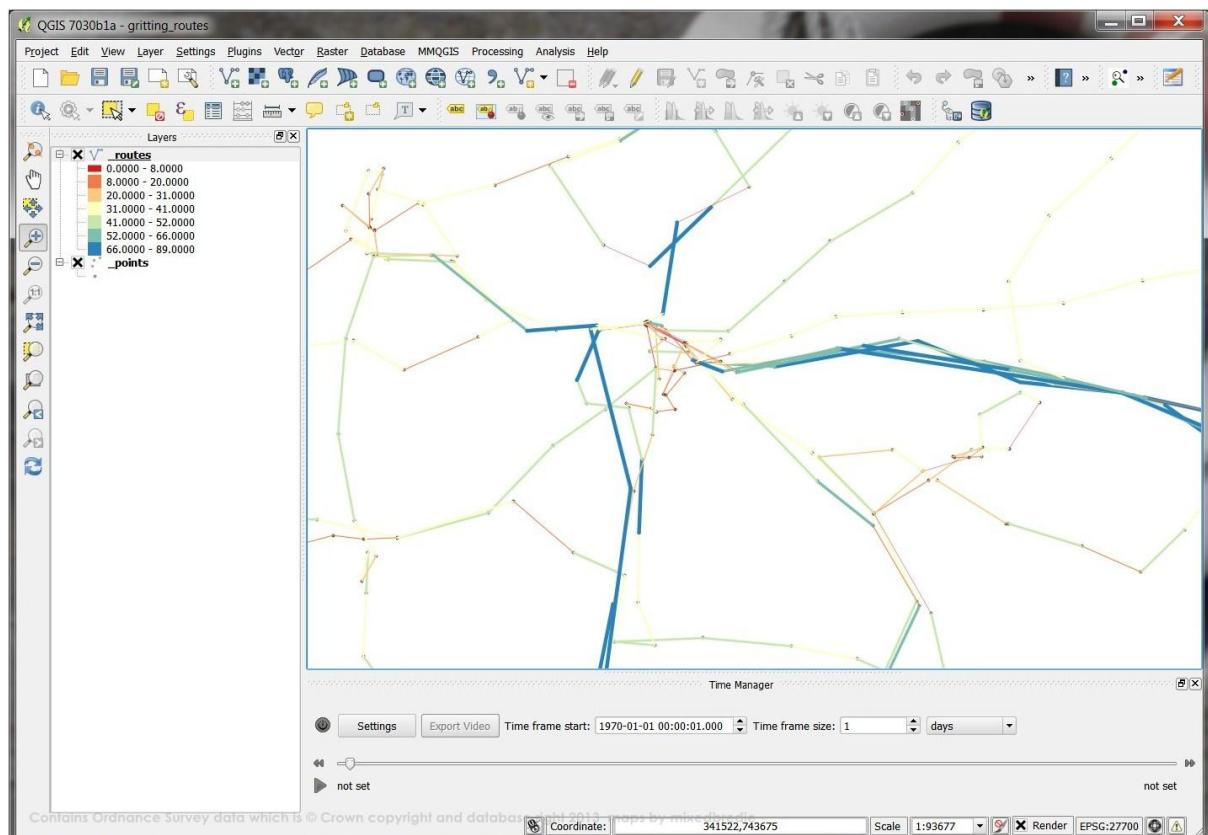


Fig. 29 : Exemple d'utilisation de QGIS avec un réseau de routes et des données associées (source : [\[mix13\]](#))

## 4.5 Conservation à long-terme

Avoir la capacité de stocker ces données sur le long-terme, soit sur plusieurs décennies est important. Mais encore faut-il avoir la certitude qu'il est utile de le faire, et que les formats dans lequel elles sont représentées garantissent leur utilisabilité sur ce même long-terme.

<sup>16</sup> Système d'Information Géographique

Dans l'état actuel des choses, certaines données, parfois datant de 10 ou 20 ans en arrière, sont bien utilisées. Elles servent principalement à établir des comparatifs avec les données plus récentes. Par exemple, on peut très facilement constater l'évolution annuelle du trafic motorisé sur certains axes ou dans certains quartiers en comparant le Trafic Moyen Journalier (TJM). Ces chiffres sont souvent utiles pour des études à plus large échelle, communale voire cantonale. Cependant, la ville étant un immense maillage de routes inter-connectées, il est toujours important d'avoir des informations sur les axes voisins, ou dans certains quartiers spécifiques. Et même si les projets d'aménagements et de mobilité d'une Ville comme celle de Lausanne sont souvent prévus plusieurs années en avance, il est impossible de prédire avec une certitude absolue les points chauds du trafic routier dans 10 ans, rendant nécessaire le fait de garder des informations sur toutes les zones disponibles plutôt que de réduire aux axes majeurs.

La conclusion à retenir est donc la suivante : il est tout à fait possible d'agréger, voire de supprimer en cas de nécessité certaines informations plus détaillées datant de plus de 5-10 ans en arrière, la plus importante restant le TJM. Cependant, il a été vu au chapitre [4.2](#) que l'espace de stockage gagné en agrégeant est déjà suffisamment conséquent pour ne pas forcément avoir besoin d'aller plus loin. Par contre, agréger géospatialement les informations n'est pas une alternative intéressante, du fait de l'impossibilité de prédire lesquelles seront utiles dans un futur un peu plus lointain.

Une question subsiste alors : comment faire pour s'assurer que les données stockées soient toujours lisibles à long-terme ? Sept critères peuvent être étudiés par rapport au format informatique utilisé :

- Il doit comporter une documentation complète et libre d'utilisation.
- Il ne doit pas dépendre d'une plateforme en particulier.
- Il ne doit pas appartenir à un propriétaire quelconque.
- Si certaines informations sont compressées, rien ne doit être perdu ou verrouillé.
- Il ne doit pas dépendre de l'exécution d'un script ou d'un programme.
- Les informations ne doivent pas être encryptées.
- Il ne doit pas y avoir de protection par mot de passe.

Les formats présentés dans ce travail, .csv et .xlsx, répondent à respectivement tous ou presque tous les critères. Le risque se situerait surtout au niveau des fonctionnalités spécifiques à Excel, comme des macros ou des formatages spéciaux, qui dépendent alors du logiciel en question et pourraient être modifiées ou disparaître au fil des versions.

Cependant, les informations contenues par le biais de ces formats devraient rester. Ils sont universellement utilisés, et le développement de logiciels pouvant les lire, les modifier et les produire permet d'assurer une fiabilité sur le long-terme d'un stockage par ces formats.

Un dernier format reste à évaluer : celui utilisé par les bases de données.

Ces dernières peuvent être établies sur des DBMS<sup>17</sup> différents. Par exemple, le code source de PostgreSQL est libre d'accès, offrant ainsi une sécurité quant à son utilisation puisqu'aucune autorité n'a la compétence de rendre inutilisables toutes les données stockées par ce système.

Dans ce travail, SQLite est utilisé. Et si, comme pour tout projet en cours de développement, il se peut que certaines fonctionnalités changent et évoluent, une rétro-compatibilité est

---

<sup>17</sup> DataBase Management System (système de gestion de base de données)

systématiquement et explicitement assurée par les développeurs du système. Des trillions de bases de données SQLite sont utilisées dans le monde, et leurs possesseurs ne peuvent pas se permettre de perdre l'accès aux informations qui y sont contenues, y compris dans un futur lointain. De plus, son code source est également libre d'accès et d'utilisation.

Le risque d'une impossibilité d'accès aux données stockées par la voie d'une base de données ne se pose donc pas pour les prochaines décennies, tout comme pour celles stockées sous forme de fichiers tels que .csv et .xlsx.

En résumé, stocker les données de comptages routiers dans des fichiers indépendants est un moyen simple de procéder, mais très inefficace, lourd et vite inorganisé sur le long-terme. Un stockage dans une base de données centrale permettrait de fluidifier et simplifier le traitement de ces données grâce à leur uniformisation et à leur regroupement, ainsi que d'ouvrir la voie vers leur agrégation généralisée. Cette agrégation permettrait d'économiser un espace de stockage non négligeable, et pourrait être mise en place pour les données les plus anciennes par exemple. Ces solutions sont tout à fait compatibles avec le besoin de représenter des données de géolocalisation, et leur accessibilité est garantie sur le long-terme.

Maintenant que les données, leur exploitation et leurs moyens de stockage sont définis, il va falloir s'intéresser à la détection, la réparation et la discussion des problèmes sous-jacents à leur utilisation.

## 5 Besoins d'amélioration et standardisation

Les différents types de données de mobilité ayant été présentés, ainsi que les difficultés liées à leur exploitation comme dans le cadre de la division mobilité de la Ville de Lausanne, il est temps de formaliser les sources des besoins d'amélioration, leur localisation précise, en quoi la situation actuelle pose problème et de quelles solutions on dispose pour y remédier.

Ce chapitre s'intéresse de plus près aux problèmes que rencontre tout exploitant de données de mobilité dans un premier temps, la Ville de Lausanne dans un second temps, et propose des pistes de solutions en se basant sur ce qui a été présenté plus haut dans ce travail.

### 5.1 Erreurs et biais dans les données

- **Sources**

Un des problèmes phares que rencontrent les exploitants de données de mobilité, c'est les incohérences qui viennent se glisser dans ces données. Elles peuvent provenir d'un certain nombre de facteurs, les plus courants concernant des défauts matériels : travailler avec plusieurs dizaines, voire centaines d'instruments de comptages actifs en simultané implique qu'il y en aura quasiment toujours qui auront des problèmes. Perte de connexion, défaut de batterie, défaut d'un autre composant, dégradation volontaire ou involontaire, altération de l'environnement...

Les causes ne manquent pas, et comme pour tout travail avec des instruments mesurant la réalité du terrain, il est strictement impossible de garantir un fonctionnement exempt de tout défaut. Le problème est donc que ces défauts se retrouvent dans les données générées.

- **Identification**

Avant de pouvoir y remédier, encore faut-il pouvoir localiser les erreurs dans les données à disposition. Des données absentes sont relativement faciles à repérer par quelqu'un dont la tâche est de les analyser, en particulier lorsque, comme c'est souvent le cas, des jours entiers manquent à l'appel. Une absence de données sur une période plus courte peut s'avérer plus problématique, car plus difficile à apercevoir par un œil humain. Cependant, fort heureusement, ce cas de figure est bien plus rare, un dysfonctionnement empêchant généralement complètement la machine de prendre des mesures jusqu'à sa réparation.

L'affaire se corse encore lorsque des erreurs de mesure se glissent dans les données : il n'y a alors aucune garantie que le gestionnaire ou la gestionnaire en charge de l'analyse remarque les défauts. Heureusement, là aussi, ce genre de problèmes est relativement rare : les instruments utilisés sont généralement très fiables, et les erreurs viennent alors de facteurs externes. Une erreur de positionnement de l'instrument de comptage, ou une modification de la configuration routière comme la fermeture au trafic de la voie comptabilisée, par exemple. Dans la plupart des cas, ces facteurs externes sont reconnus et enregistrés, puis transmis au responsable ou à la responsable de l'analyse des données qui peut alors prendre en compte le dérangement dans ses calculs.

Certains de ces biais ne provoquent d'ailleurs pas de réelle erreur dans les données, simplement un changement temporaire de paradigme. Il peut s'agir de travaux, entraînant une déviation et donc un flux de véhicules différent, de vacances, d'une manifestation... Par exemple, lorsque le Tour de France est passé à Lausanne et que de nombreuses routes ont été totalement fermées à la circulation pendant toute une journée, il a fallu prendre en compte ce facteur lorsque des données couvrant cette période ont été analysées. Dans ce genre de cas, il est à la discrétion de la personne en charge de l'analyse des données de penser aux influences et facteurs pouvant contribuer à un biais dans les données, sans quoi ils ne seront probablement pas pris en compte.

- **Problèmes engendrés**

L'absence de données crée surtout une difficulté de traitement. Par exemple, établir un comparatif de chiffres sur une certaine période entre tous les indicateurs de vitesse de la Ville relève de l'impossible, puisqu'il n'existe généralement aucune période durant laquelle tous les indicateurs étaient fonctionnels simultanément. Il faut alors séparer la période étudiée en deux tranches comparables en termes de fréquentation et de trafic routier, en admettant que tous les indicateurs de vitesse ayant subi un défaut durant la première période n'en ont pas subi durant la deuxième.

Cette absence de données, ainsi que les biais mentionnés plus haut, peuvent induire une falsification des chiffres. Et si cette falsification n'est souvent pas remarquable et ne causera pas forcément d'autre problème, elle pourrait remettre en question la véracité scientifique des statistiques produites par la Ville. Leur fiabilité en serait amoindrie, et en cas de découverte d'erreurs par des sources internes ou externes à la Ville, son image en serait entachée. Les problèmes qu'impliquent un manque de rigourosité en statistique ne seront pas détaillés ici, mais il est de mise de considérer que tout chiffre produit au moyen des données de mobilité se doit d'être le plus juste possible, en particulier lorsque des décisions importantes peuvent être prises en se fondant dessus.

- **Solutions**

En ce qui concerne la détection des biais dans les données de mobilité, il n'y a pas de réponse miraculeuse et fiable à 100%. Cependant, la centralisation de ces données peut faciliter le développement d'une solution informatique donnant la possibilité de mettre en évidence plus simplement les absences de données dans un premier temps, et même les erreurs dans un second temps.

Ce dernier élément est plus compliqué à mettre en place, car contrairement à des données absentes, des données biaisées ne sont pas toujours aisément repérables, autant pour un être humain que pour un ordinateur. Cependant, des solutions d'apprentissage automatique existent et pourraient contribuer à souligner les irrégularités dans les données, qui seraient alors laissées au jugement de la personne chargée de leur analyse.

Des méthodes informatiques pour réparer les données sont aussi disponibles. Elles impliquent également une certaine centralisation et uniformisation des données, sans quoi il est impossible de les traiter proprement, mais elles pourraient permettre de simuler les données absentes ou biaisées au moyen de modèles d'apprentissage automatique. Et s'il est impossible de répliquer la réalité avec une certitude absolue,

ces modèles se veulent suffisamment perfectionnés pour pouvoir être utilisés dans le contexte d'une réparation de données de mobilité.

## 5.2 Manque d'uniformisation

- **Sources**

Un autre gros problème souvent relevé dans le domaine, c'est un manque de pratiques et de terminologies communes dans la collecte et l'étude de données de mobilité. Si on retrouve certaines grandes lignes un peu partout, telles que le Trafic Journalier Moyen ou la vitesse médiane, le format des données de mobilité que l'on collecte diffère quasi-systématiquement d'un support à l'autre.

La raison est simple : dans une infrastructure comme celle de la Ville de Lausanne, de nombreux outils sont utilisés, améliorés, changés ou supprimés au fil des années et des projets. Il a été vu précédemment dans ce travail que les outils utilisés pour collecter les données employées par la division mobilité sont multiples et variés, ce qui pose inévitablement un réel problème de standardisation. Il est impossible d'exiger des fabricants, qui ont souvent des dizaines, voire des centaines de clients différents, qu'ils adaptent leur solution spécifiquement aux besoins de la Ville. C'est ainsi que chaque nouvel instrument vient avec sa procédure qui lui est propre, et c'est aux utilisateurs et utilisatrices de s'adapter au format proposé.

- **Identification**

Le problème peut se poser même dans le cas de données brutes : il suffit que certains appareils fassent une distinction suivant le sens du véhicule enregistré, que d'autres enregistrent leurs informations dans un fichier CSV séparé par des points-virgules plutôt que des virgules, et que d'autres encore aient une ligne d'en-tête avant les valeurs, et on se retrouve très vite avec une multitude de combinaisons possibles pour quelque chose d'aussi basique que des informations brutes sur le passage de chaque véhicule sur une route.

Date	Vitesse	Direction de l'évaluation	Geschwindigkeit	Zeit	Datum	Spur	Laenge [m]	Klasse	Klassenname	Achsen	Schall [dB]	Abstand [m]	Name	Time	Q@KFZ
05/09/2022 00:00:55	22	Arrivant	03504:01:2816.07.2214.2007Pkvw2644.700										M024E40	02.09.2021 00:00	313,77426
05/09/2022 00:00:57	23	Arrivant	03404:02:2716.07.2223.6007Pkvw2012.300										M024S10	02.09.2021 00:00	313,4298
05/09/2022 00:00:58	24	Arrivant	03404:02:3016.07.2214.8007Pkvw2704.300										M024E40	03.09.2021 00:00	308,93195
05/09/2022 00:01:00	24	Arrivant	03404:03:1716.07.2214.2007Pkvw2724.600										M024S10	03.09.2021 00:00	305,62292
05/09/2022 00:01:01	26	Arrivant	05004:04:5416.07.2216.00011Lfw2714.700										M024E40	04.09.2021 00:00	316,895
05/09/2022 00:01:04	26	Arrivant	03104:05:0316.07.2214.6007Pkvw2644.100										M024S10	04.09.2021 00:00	312,9367
05/09/2022 00:01:06	26	Arrivant	02304:05:1616.07.2214.8007Pkvw2614.400										M024E40	05.09.2021 00:00	264,80832
05/09/2022 00:06:48	44	Arrivant	03804:05:3016.07.2225.7003Lkw4012.200										M024S10	05.09.2021 00:00	293,8868
05/09/2022 00:06:49	46	Arrivant	03804:05:3216.07.2225.1007Pkvw2012.200										M024E40	06.09.2021 00:00	168,95253
05/09/2022 00:06:51	46	Arrivant	04004:05:3516.07.2225.1003Lkw4012.100										M024S10	06.09.2021 00:00	178,34296
05/09/2022 00:06:52	45	Arrivant	03404:05:4116.07.2225.1003Lkw3012.100										M024E40	07.09.2021 00:00	279,28983
05/09/2022 00:06:54	43	Arrivant	03504:06:3316.07.2214.7007Pkvw2664.200										M024S10	07.09.2021 00:00	280,2001
05/09/2022 00:07:01	28	Arrivant	02504:06:4816.07.2215.40011Lfw2644.300										M024E40	08.09.2021 00:00	303,51596
05/09/2022 00:07:03	28	Arrivant	04304:07:1816.07.2211.60010Krad1724.300										M024S10	08.09.2021 00:00	302,44446
05/09/2022 00:07:04	30	Arrivant	03304:07:2816.07.2224.50010Krad3012.400										M024E40	09.09.2021 00:00	304,6389
05/09/2022 00:07:06	31	Arrivant	03204:07:3116.07.2225.1003Lkw3012.300										M024S10	09.09.2021 00:00	310,70624
05/09/2022 00:07:09	34	Arrivant	03404:07:3316.07.2225.3002PkvwA4012.100										M024E40	10.09.2021 00:00	313,3597

Fig. 30 : Exemple de formats différents de données brutes (source : conception personnelle)

Ces combinaisons sont donc d'autant plus nombreuses lorsque l'on parle de données agrégées, où il faut prendre en compte les différentes manières d'agrégérer des données et de les présenter dans un fichier. Et elles sont quasi-infinies lorsque l'on parle de synthèses produites par une entreprise externe dans le cadre d'une

étude, ou par un logiciel utilisé par la division mobilité par exemple. Dans ce dernier cas, les informations présentées et leur mise en page ne seront jamais similaires d'un fournisseur de données à l'autre.

- **Problèmes engendrés**

Ici aussi, le principal problème engendré par le manque d'uniformisation des données de mobilité est une difficulté de traitement élevée. S'il n'est, en général, pas trop dur de passer d'un format à l'autre (en remplaçant les points-virgules par des virgules, en supprimant les colonnes superflues...), cela reste un processus usant et inefficace. Plus grande est la différence entre les formats de données à disposition, plus long sera le traitement, et ce à chaque itération d'analyse des données.

Un autre cas peut être encore plus problématique : si les données fournies ne suffisent pas à recouper les informations nécessaires. Il se peut, par exemple, que l'entreprise externe ait réalisé des choix par rapport aux données envoyées au mandant, n'incluant pas dans le produit certaines informations qui n'étaient pas explicitement demandées. Comparer les données avec d'autres sources relève alors de l'impossible.

Si le processus de commande du mandataire envers le mandant peut être sujet à reproche dans un tel cas, c'est bien le manque de standards dans le domaine qui est à blâmer ici. Au-delà des quelques informations clés mentionnées plus haut, les mandants ont en général une grande liberté quant au format et à la sélection des informations qu'ils rendront aux mandataires. Et ces derniers, en l'absence de standards établis et unilatéralement reconnus, ne peuvent que se contenter de ce que les fournisseurs de données ont à leur proposer et se débrouiller avec ce qu'ils peuvent en tirer.

- **Solutions**

La recherche sur une uniformisation des données de mobilité et l'élaboration de standards utilisables par les différents acteurs du milieu est en plein essor. Le but affiché est clair : converger vers un cadre de travail plus uniforme de pré-traitement des données de mobilité, pour que tout fournisseur de données puisse s'y plier plutôt que de procéder comme il l'entend.

Pour ce faire, le développement en open-source et favorisant la reproductibilité est largement reconnu comme étant un facteur d'avancement, de par son accessibilité. Un bon standard doit être ouvert à tout acteur souhaitant l'utiliser ou y contribuer.

Cependant, les options s'offrant actuellement à un organisme comme la Ville de Lausanne ne sont pour le moment pas suffisantes pour améliorer la situation à son échelle. Si des standards spécifiques à certains milieux, comme GTFS ou DATEX, sont en plein développement, il n'existe pas de solution suffisamment adaptée à un cas comme celui des comptages routiers effectués en ville de Lausanne. Il faut donc composer avec ce que l'on a à notre disposition : avoir des prérequis clairs lors du lancement de mandats impliquant des solutions proposées par des entreprises externes, et avoir à disposition des solutions internes apportant une flexibilité dans les possibilités de traitement des données, sous formes de logiciels par exemple. Un tel logiciel est un des buts de ce travail, et est présenté plus en détail au [chapitre 6](#).

## 5.3 Standardisation à différentes échelles

- **Sources**

Tout comme la standardisation des informations communiquées entre mandataires et mandants, celle entre les différents échelons des groupes de décideurs (au niveau municipal, cantonal, fédéral...) est quasiment inexistante. La situation dans le domaine des données des postes de comptage en Suisse est très hétérogène.

En l'absence de lignes directrices ou de standards nationaux, aucune procédure fixe n'est établie pour la communication d'informations de comptages routiers entre le niveau communal, le niveau cantonal et le niveau fédéral. Et sans communication établie entre ces différents niveaux, difficile d'aller au-delà, au niveau international.

- **Identification**

Si l'OFROU<sup>18</sup> a bien publié une directive, ASTRA 13012 ([\[ofr09\]](#)), afin d'harmoniser l'installation de nouveaux postes de comptages, celle-ci ne s'applique en principe qu'aux nouvelles réalisations et renouvellements de postes de comptages exploités par la Confédération. Sans base légale appropriée, l'OFROU n'a pas la compétence d'imposer un standard aux cantons et communes.

Autrement, il n'existe pas de standard en termes de formatage de données issues de comptages routiers utilisable simplement par un acteur comme la Ville de Lausanne, à l'échelle suisse du moins.

- **Problèmes engendrés**

Une mauvaise communication entre instances de décideurs mène inévitablement à des procédures plus longues et compliquées, dans le cadre d'études de trafic ou d'établissements de nouveaux projets d'infrastructures routières par exemple.

Plus globalement, une situation où chaque producteur de données les garde pour lui, par manque d'un moyen simple de les partager, va à l'encontre du principe d'ouverture intrinsèque à la recherche et au développement de nouvelles solutions de mobilité. Pour que ces dernières émergent et que le monde avance dans ce domaine, il est nécessaire que les données soient accessibles.

- **Solutions**

En l'absence d'une base légale permettant à l'OFROU d'imposer des standards aux communes et aux cantons, c'est leur volonté propre qui est mise à l'épreuve ici. La directive ASTRA 13012 s'adresse à tout organisme mandaté par l'OFROU pour l'entretien, l'exploitation ou l'étude des routes fédérales, mais elle peut également être utilisée par des cantons ou des communes directement. Cette directive établit « les exigences fonctionnelles et techniques de base posées aux postes de comptages, et définit les standards devant être respectés aux stades de la planification, de l'établissement des projets, de la réalisation et de l'exploitation ». Elle contient donc tout ce qu'il faut pour que les communes et les cantons aient une base sur laquelle se mettre d'accord. Reste à ce qu'ils passent le cap, et que cette directive devienne un standard dans le milieu. En l'absence d'autorité centrale, un tel changement est relativement difficile à appliquer.

---

<sup>18</sup> Office Fédéral des Routes

Il convient aussi de prendre en compte le processus de standardisation pour l'échange de données au niveau international. Ce standard, c'est DATEX II. Mais là aussi, en l'absence d'une obligation ainsi que d'un organisme chargé d'aider les communes et les cantons à traduire leurs données de comptage dans le format proposé par DATEX II, c'est un changement compliqué à mettre en place.

## 5.4 Améliorations internes à l'organisation

- **Sources et identification**

Une autre grande cause de dysfonctionnements dans le processus de collecte et de traitement de données de comptages routiers est interne à l'organisme ou à l'entreprise qui les supervise, la division mobilité de la Ville de Lausanne par exemple.

Sans aller jusqu'à analyser en profondeur les rouages de l'organisation, certains problèmes peuvent être relevés. Par exemple, en l'absence d'une documentation bien établie ou d'une sorte de grand registre, l'accès aux connaissances liées aux données est très limité : on ne peut alors que compter sur ce que les collaborateurs impliqués en savent. Il se peut aussi, dans certains cas, que certaines informations soient trouvables dans les méandres des fichiers appartenant à la division mobilité. Quoi qu'il en soit, savoir qui a l'information recherchée est souvent un défi, et il peut même arriver que plus personne ne l'ait dans les personnes actuellement employées.

Il en va de même pour les études commandées à des mandataires externes. En l'absence de processus bien établi ou de personne ressource, toute personne ayant besoin d'un certain comptage a la liberté de procéder à peu près comme bon lui semble, dans la limite de certaines consignes et du budget à disposition.

- **Problèmes engendrés**

Tout cela apporte une certaine inefficacité dans le traitement des données à disposition. Cette inefficacité va d'une légère perte de temps sur les aspects où la personne chargée de traiter les données sait à peu près quoi faire, à une perte totale d'informations lorsque les lignes directrices des explications d'un comptage ont été perdues.

Un traitement inefficace des données peut toucher aussi les autres collaborateurs et collaboratrices de l'organisme. S'il leur est impossible ou trop compliqué de trouver les informations dont ils et elles auraient besoin pour un projet quelconque, il ne leur reste plus qu'à perdre du temps et de l'énergie à chercher qui pourrait leur venir en aide dans leur quête. Et le plus probable dans un tel cas, c'est plutôt que l'on ne va pas s'embêter, et une nouvelle commande auprès de la section exploitation ou d'une entreprise externe va être effectuée, engendrant ainsi une perte de moyens.

- **Solutions**

Les solutions à de tels problèmes sont multiples, et chacune d'entre elles vise à améliorer un peu la situation.

Premièrement, une personne responsable des données, de superviser leur collecte, puis de gérer leur stockage et leur traitement, est un must. Elle peut aussi servir de

personne de référence pour les autres collaborateurs et collaboratrices du service, faisant d'une pierre deux coups : d'une part, la gestion des données est effectuée à un seul et unique endroit et supervisée par une personne ayant une vue globale et complète de ce que le service a à sa disposition, et d'autre part elle offre un point de contact clair pour les autres collaborateurs et collaboratrices qui n'ont ainsi pas à se démener pour trouver les informations dont ils et elles ont besoin. Dans une structure comme celle de la division mobilité de la Ville de Lausanne, ce poste peut aisément être occupé par une personne à temps plein, comme cela a pu être le cas par le passé.

Deuxièmement, des processus clairs et établis devraient être mis en place. Ceci concerne les mandats à des entreprises externes pour des études de comptages aussi bien que la collecte et le traitement des données utilisées par la division mobilité ou encore les demandes d'accès à des informations internes au service.

Troisièmement, un répertoire d'informations devrait être construit. La section exploitation a déjà à sa disposition un outil lui permettant de répertorier la position dans la ville d'une bonne partie de ses dispositifs, ainsi que les opérations qui sont effectuées, mais il serait utile d'aller plus loin. De nombreuses informations peuvent être enregistrées, telles que les procédures de collecte des données sur les différents appareils, les procédures actuellement mises en place de traitement des données, où trouver les sources, les résultats, les archives...

En résumé, le domaine des données de mobilité doit faire face à un certain nombre de difficultés, et la division mobilité de la Ville de Lausanne n'y fait pas exception. Certains problèmes concernent tout acteur travaillant avec des comptages routiers, d'autres sont inhérents au système fédéraliste suisse, d'autres enfin sont internes à l'organisation de la section.

Dans tous les cas, même si un certain nombre de problèmes ne peuvent pas être résolus à l'échelle d'un "petit" acteur comme la Ville, la recherche et l'innovation évoluent à toute vitesse dans ce milieu, et une grosse partie des améliorations à plus large échelle passe par une standardisation des données de mobilité. Certains problèmes, pour finir, peuvent être corrigés plus facilement, ou en partie. C'est notamment aux erreurs dans les données et au manque d'uniformisation que s'attaque le programme présenté dans le chapitre suivant.



## 6 Conception de l'application

Dans le cadre de ce travail de Bachelor, une application a été développée. Ce chapitre présente son concept, mais aussi son développement et les détails techniques qui vont avec.

### 6.1 Fondations

Le principe de l'application peut être résumé par le schéma suivant :

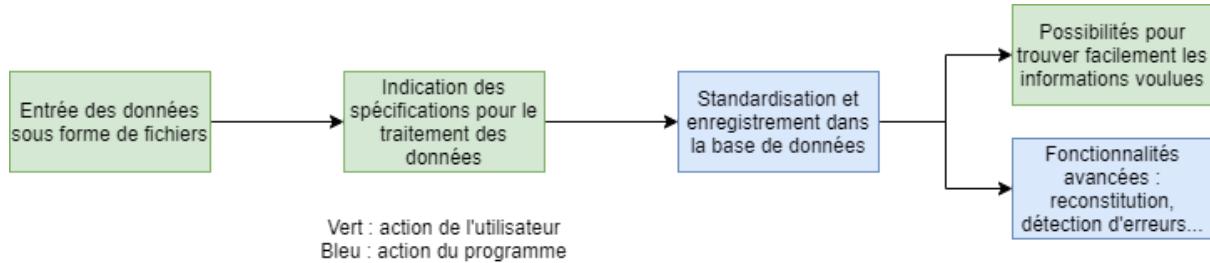


Fig. 31 : Résumé du principe de l'application (source : conception personnelle)

Pour les deux premières étapes, il faut donc fournir aux utilisateurs et utilisatrices un moyen de déposer leurs données, qui sont alors encore sous forme de fichiers dont le format peut varier grandement d'un exemplaire à l'autre.

Le procédé de réflexion utilisé ici est grandement basé sur des exemples de comptages mis à disposition dans le cadre du développement de l'information. Ces exemples couvrent les différentes catégories de comptages routiers décrites précédemment dans ce travail, il est donc considéré que si l'application peut tous les prendre en charge, elle pourrait prendre en charge la grande majorité des formats qui se présenteraient à elle.

L'étape suivante dans l'élaboration de l'application consiste alors à déterminer les similarités et les différences entre ces différents exemples de comptages. Plus concrètement, la question posée est la suivante : "comment standardiser toutes ces données pour les ranger dans une seule et même base de données ?"

La machine ne pouvant que très difficilement deviner par elle-même quelles informations sont contenues dans le fichier qui lui est donné et où les trouver, l'étape de saisie va inévitablement comporter une spécification manuelle, à demander à quiconque tente d'insérer des nouvelles données.

En se basant sur les types de données établis précédemment ainsi que sur les similarités et différences relevées entre les fichiers exemples, il est possible d'établir un diagramme BPMN<sup>19</sup> comportant les différents passages qui composent le processus d'insertion des informations dans le programme, et ultimement dans la base de données, en distinguant ce qui est fait par un être humain et ce qui est fait par le programme.

<sup>19</sup> Business Process Modeling Notation

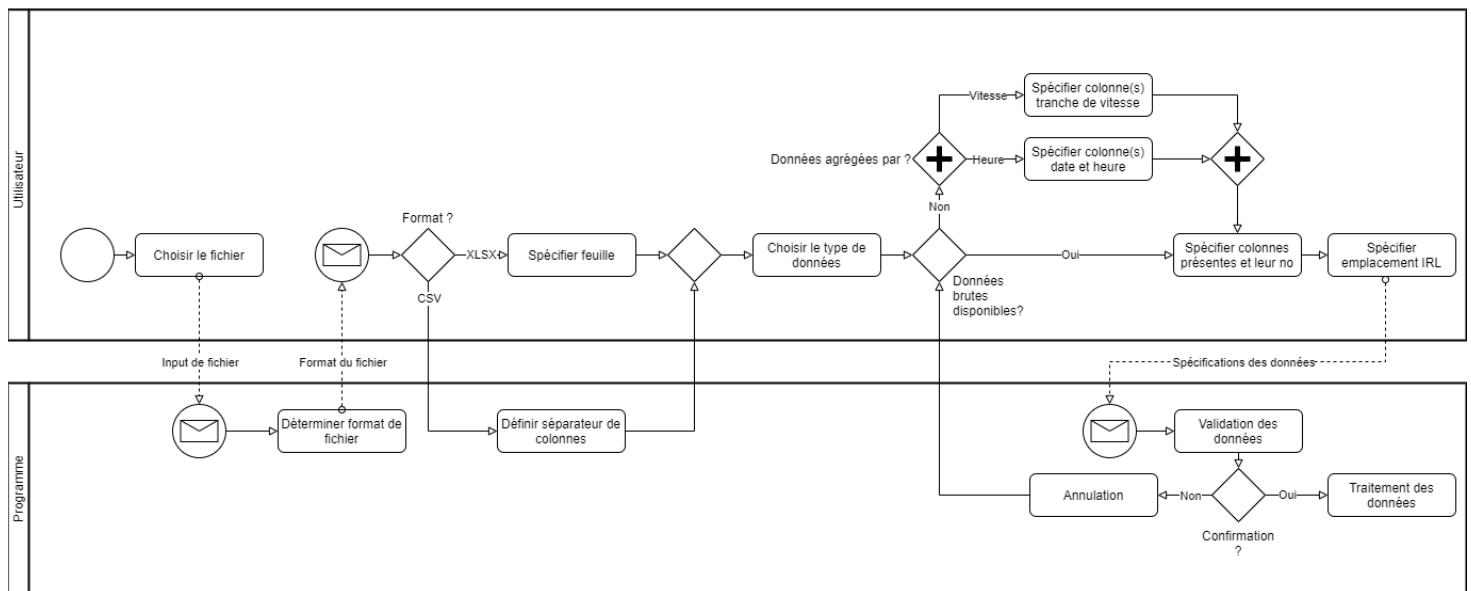


Fig. 32 : Diagramme BPMN du processus de traitement de l'information (source : conception personnelle)

Les différentes étapes présentées ici forment une ligne directrice pour le développement du programme. Ainsi, ce dernier va d'abord demander à l'utilisateur ou l'utilisatrice de choisir le fichier à importer, puis va en fonction du format du fichier demander à spécifier un séparateur de colonnes ou le nom de la feuille Excel à utiliser. Les colonnes à spécifier (par exemple la date, la vitesse, le bruit...) vont varier en fonction du type de données disponibles, mais d'autres informations comme la localisation du comptage seront toujours demandées.

C'est ensuite la machine qui prendra le relais et qui tentera de traiter l'information qu'on lui a transmise. Si elle y parvient, les entrées sont insérées dans la base de données. Sinon, une erreur parvient à l'utilisateur ou l'utilisatrice et le processus doit être en partie recommandé avec d'autres spécifications.

On en vient maintenant aux informations techniques utilisées dans le cadre de la réalisation de ce programme.

Deux options étaient sur la table au départ : ce serait soit un logiciel exécutable, soit une solution basée sur une interface web. C'est finalement la première qui a été choisie, pour des raisons d'utilisabilité au sein de la Ville de Lausanne notamment.

Le programme est développé en Python. Ce choix a été fait en fonction de ma connaissance du langage, ainsi que pour les nombreuses possibilités qu'offre Python en termes de traitement de données par ses librairies. Le programme utilise ainsi les librairies NumPy, Pandas et Matplotlib notamment, mais aussi Tkinter pour l'interface utilisateur, openpyxl pour la prise en charge de fichiers Excel, PyInstaller pour la distribution du programme, ou encore sqlite3 pour le raccord avec la base de données.

Cette base de données a ainsi été créée en SQLite, qui est un moteur de base de données relationnelle fonctionnant par fichiers. La base de données créée et utilisée par l'application est donc enregistrée sous la forme d'un fichier unique avec une extension .db.

Le choix de ce moteur plutôt qu'un autre s'est effectué, là aussi, en partie sur mon expérience personnelle. Le fait que la base de données soit enregistrée sous forme de fichier individuel et facilement déplaçable et accessible est aussi plutôt intéressant, puisqu'elle n'implique pas l'utilisation d'un serveur comme pour d'autres systèmes de gestion de bases de données comme PostgreSQL ou MySQL.

Les quelques limitations que SQLite implique ne sont pas non plus un problème pour son utilisation dans notre contexte. Par exemple, il n'y a pas de risque de conflit puisque l'application n'est pas prévue pour une utilisation sur plus d'un ordinateur en simultané. La vitesse de traitement des données par SQLite est également jugée comme suffisante pour le volume de données qui sera traité par le programme.

Finalement, il est bon de préciser que le code du programme est open-source et est entièrement disponible sur GitHub. Le but, ici, est que toute personne voulant utiliser le programme et disposant de compétences en programmation puisse le modifier et l'adapter pour qu'il convienne à ses besoins, notamment pour que l'application puisse inclure d'éventuels autres formats de données pas pris en compte lors de son développement initial. De plus, aucun revenu n'étant généré par l'application, il ne fait pas de sens de vouloir garder le code-source secret.

Les principes et caractéristiques clés de l'application étant posés, un dernier détail d'une certaine importance reste à préciser. Ainsi, le nom choisi pour l'application est KallaxComptages. Ce nom provient des étagères Kallax, qui fournissent un moyen simple de ranger toute une panoplie d'objets n'ayant souvent que la taille en commun, offrent la possibilité de les organiser d'une manière satisfaisante, et surtout sont relativement modulables. Un parallèle avec les fondements du programme développé dans le cadre de ce travail peut alors être fait.



Fig. 33 : Exemple d'étagères Kallax (source : [Ike23](#))

## 6.2 Base de données

La base de données constitue le cœur du programme. Toutes les informations entrées par l'utilisateur ou l'utilisatrice y sont stockées, et c'est donc vers elle aussi que le programme va se tourner au moment de devoir afficher des données traitées à l'utilisateur ou l'utilisatrice. Elle dépend énormément des principes de standardisation mentionnés plus haut. En effet, elle doit être capable de répertorier les données diverses tout en ne perdant aucune information importante. Le modèle relationnel présenté ci-dessous est une version prenant

en compte les exemples qui ont été mis à disposition de la réalisation de ce travail par la Ville de Lausanne et qui couvre une large variété de formats de données de comptages routiers.

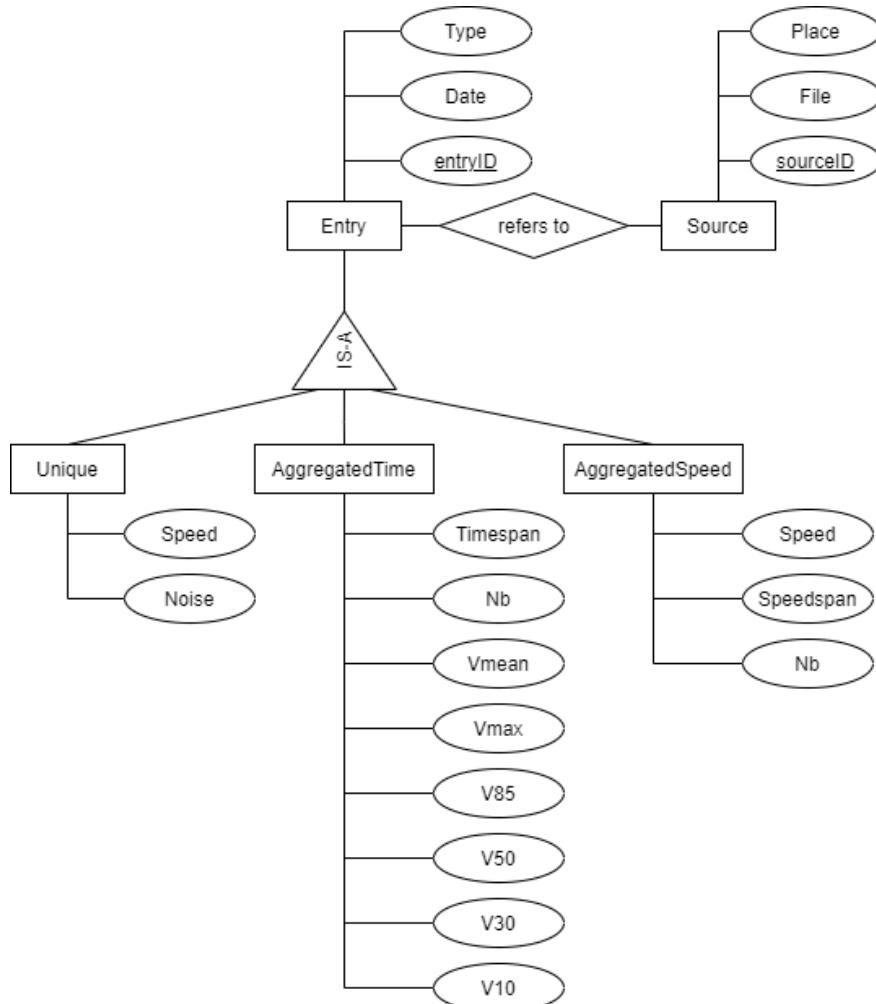


Fig. 34 : Modèle Entité-Association de la base de données de l'application (source : conception personnelle)

À chaque fois que l'utilisateur ou l'utilisatrice entre un fichier dans l'application, un objet Source est créé, qui comprend la localisation du comptage renseigné, ainsi que le nom du fichier fourni, dans le but de garder une trace de la provenance des informations contenues dans la base de données.

Chaque Source va ensuite générer une multitude d'objets Entry. Une entrée correspond au passage d'un véhicule unique dans le cas de données brutes, ou aux données récoltées sur une tranche d'heure ou de vitesse dans le cas de données agrégées. Pour chaque entrée, un deuxième objet va être créé, soit de type Unique, soit de type AggregatedTime, soit de type AggregatedSpeed, en fonction du type de données fourni par le fichier.

Prenons deux exemples concrets.

	A	B	C
1	Date	Direction	Vitesse
20	07.09.2022 11:47	1	21
21	07.09.2022 11:51	1	29
22	07.09.2022 11:53	1	19
23	07.09.2022 11:53	1	25
24	07.09.2022 11:59	1	22

Fig. 35 : Extrait de données brutes issues d'un comptage routier (source : conception personnelle)

En premier lieu, la [fig. 35](#) montre un court extrait de données brutes issues d'un comptage routier. Voici donc ce qui va être introduit dans la base de données :

- Un objet Source contenant...
  - Un attribut Place : la localisation du comptage en question (sous forme de coordonnées géographiques)
  - Un attribut File : le nom du fichier d'où proviennent les données
  - Un attribut sourceID : un identifiant unique permettant de se référer à l'objet Source créé
- Plusieurs objets Entry (un pour chaque passage de véhicule compté) contenant chacun...
  - Un attribut Type : la catégorie à laquelle appartient le véhicule compté. Dans notre cas, cette catégorie n'étant pas notée, l'attribut sera systématiquement défini comme "null", soit sans valeur
  - Un attribut Date : la date et l'heure à laquelle le véhicule a été enregistré. Dans notre cas, la valeur contenue dans la première colonne
  - Un attribut entryID : un identifiant unique permettant de se référer à l'objet Entry créé
  - Un attribut sourceID : une clé étrangère<sup>20</sup> contenant l'identifiant de l'objet Source auquel est associé l'objet Entry créé
- Plusieurs objets Unique (un pour chaque objet Entry créé) contenant chacun...
  - Un attribut Speed : la vitesse à laquelle le véhicule compté a été enregistré. Dans notre cas, la valeur contenue dans la troisième colonne
  - Un attribut Noise : le bruit émis par le véhicule compté. Dans notre cas, les données ne notant pas cette valeur, l'attribut sera systématiquement défini comme "null"
  - Un attribut entryID : une clé étrangère contenant l'identifiant de l'objet Entry auquel est associé l'objet Unique créé

Note : la deuxième colonne, dans cet exemple, contient le sens de circulation du véhicule. Cette information n'est pas enregistrée dans la base de données, plus de détails sont disponibles au [chapitre 6.5](#).

---

<sup>20</sup> Terme utilisé en bases de données relationnelles pour représenter un attribut se référant à l'attribut d'un autre objet

Date	Heure	Moto*	Voiture	PL	PL+	Total
18.05.2021	06:00	2	54	1	0	57
18.05.2021	07:00	16	139	0	1	156
18.05.2021	08:00	15	172	1	0	188
18.05.2021	09:00	7	144	2	0	153
18.05.2021	10:00	3	150	2	0	155

Fig. 36 : Extrait de données agrégées par temps issues d'un comptage routier (source : conception personnelle)

En second lieu, la [fig. 36](#) montre un court extrait de données agrégées par tranches d'une durée fixe définie, issues d'un comptage routier. Voici donc ce qui va être introduit dans la base de données :

- Un objet Source créé de la même manière que pour l'exemple précédent.
- Plusieurs objets Entry créés de la même manière que pour l'exemple précédent. À noter que l'attribut Date contiendra l'heure notée pour chaque entrée, soit généralement l'heure de début de la tranche de temps. En l'occurrence, il s'agit des valeurs notées dans la deuxième colonne.
- Plusieurs objets AggregatedTime (un pour chaque objet Entry créé) contenant chacun...
  - Un attribut Timespan : l'étendue de la tranche de temps. Dans notre cas, cette valeur va systématiquement être 1h00
  - Un attribut Nb : le nombre de véhicules comptés dans cette entrée. Dans notre cas, cette valeur est donnée par la troisième colonne
  - Un attribut Vmean : la vitesse moyenne des véhicules enregistrés dans cette entrée. Dans notre cas, cette valeur est donnée par la sixième colonne
  - Un attribut Vmax : la vitesse maximale des véhicules enregistrés dans cette entrée. Dans notre cas, cette valeur est donnée par la cinquième colonne
  - Un attribut V85 : la vitesse en-dessous de laquelle ont été enregistrés 85% des véhicules comptés dans cette entrée. Dans notre cas, cette valeur est donnée par la quatrième colonne
  - Un attribut V50, un attribut V30, un attribut V10 : la vitesse en-dessous de laquelle ont été enregistrés respectivement 50%, 30% et 10% des véhicules comptés dans cette entrée. Dans notre cas, les données ne présentent pas ces valeurs, la valeur de ces attributs sera donc systématiquement de "null"

La base de données est ensuite utilisée dans le cadre de la phase d'analyse, explicitée au sous-chapitre suivant.

### 6.3 Phase d'analyse

Cette phase, bien que capitale pour l'utilité du programme, n'a pas pu être réalisée dans le cadre de ce travail, pour des raisons de temps à disposition et de volume de travail nécessaire principalement. Le contenu de ce sous-chapitre se base donc sur les principes voulus pour cette phase et sur ce qui pourrait être implémenté dans le cadre de futurs ajouts au programme.

La première étape de la phase d'analyse consiste à sélectionner les données avec lesquelles on veut travailler. Cette fonctionnalité repose sur un principe similaire à un moteur

de recherche avancé, dans le sens où l'utilisateur ou l'utilisatrice doit spécifier toute une série de critères visant à ne sélectionner que certaines données en particulier.

Les critères proposés sont les suivants :

- Emplacement des données sous forme de zone à dessiner sur une carte géographique
- Dates minimum et maximum auxquelles les comptages ont été effectués
- Jours de la semaine et heures particuliers (par exemple, seulement les nuits de week-ends)
- Catégories de véhicules à prendre en compte

En fonction des critères spécifiés, le programme va ensuite chercher dans la base de données les informations correspondantes. Il offre alors plusieurs fonctionnalités à l'utilisateur ou l'utilisatrice, dont voici des exemples :

- Affichage des données telles qu'elles sont stockées dans la base, pas forcément très lisibles ou utilisables mais complètes
- Affichage du TJM<sup>21</sup>, TJOM<sup>22</sup> et autres informations générales sur les données sélectionnées
- Génération de graphiques dont l'abscisse et l'ordonnée peuvent être spécifiées, par exemple le nombre de véhicules en fonction de l'heure de la journée
- Prédiction d'informations et de données futures basées sur les données à disposition
- Détection, et éventuellement réparation, des anomalies dans les données : variations anormales, pertes de données...

## 6.4 Structure du programme

Le programme est divisé en cinq fichiers Python.

- *main.py* est le fichier central, gérant la fenêtre de départ ainsi que les différentes fonctionnalités du programme, les coordonnant les unes avec les autres et leur passant les informations dont elles ont besoin.  
Dans le cas où l'utilisateur ou l'utilisatrice désire insérer des nouvelles données dans la base, le code va lancer le contenu de *input.py*, puis celui de *database.py*.  
Dans le cas où il ou elle désire travailler avec les données déjà présentes dans la base, le code va lancer le contenu de *selection.py*, puis celui de *analyze.py*.

---

<sup>21</sup> Trafic Journalier Moyen

<sup>22</sup> Trafic des Jours Ouvrables Moyen

C'est également dans ce code que sont contenues certaines variables communes à plusieurs phases du programme, telles que le texte "Pas de valeur" utilisé dans certains menus et permettant de ne sélectionner aucune colonne lorsque le comptage ne possède pas les valeurs demandées (par exemple, le bruit dans l'exemple de données brutes à la [fig. 35](#)).

- *input.py* est le fichier qui s'occupe de la phase de spécifications des données.  
À chaque étape de spécification, le code va afficher la fenêtre qui contient les menus permettant à l'utilisateur ou à l'utilisatrice de donner les informations nécessaires au programme. Il va ensuite les enregistrer, et à la fin de son exécution, va renvoyer au fichier central les informations récoltées pour qu'il puisse les passer plus loin.  
Les étapes distinguées lors de la spécification des données sont les suivantes :
  - Choisir le fichier
  - Définir le séparateur de données (dans le cas d'un fichier .csv, l'ordinateur se charge alors de l'étape) ou la feuille à utiliser (dans le cas d'un fichier .xlsx)
  - Définir le type de données : brutes, ou alors agrégées par durée et/ou agrégées par vitesse
  - Définir la portée des lignes utilisées. Cette étape permet de s'affranchir des en-têtes éventuellement contenus dans le fichier au-dessus des données, ainsi que des lignes superflues au-dessus ou en-dessous
  - Définir les colonnes contenant les tranches de temps (date et heure) et/ou les tranches de vitesse. Cette étape est passée si les données sont brutes plutôt qu'agrégées
  - Définir les colonnes à utiliser. Les informations à donner comprennent les différents attributs explicités plus hauts, par exemple la vitesse et le bruit en cas de données brutes, mais aussi la date, l'heure et le type de véhicule. Une option est prévue au cas où la date et l'heure sont contenus dans la même colonne
  - Définir la localisation du comptage enregistré à l'aide d'une carte géographique
  - Faire une dernière vérification avec l'utilisateur ou l'utilisatrice que les informations entrées sont les bonnes
- *database.py* est le fichier qui s'occupe de l'insertion des données dans la base. Le fichier central s'occupe de lui fournir toutes les informations récoltées par *input.py*, et ce code va alors les utiliser pour prendre les données spécifiées, les traiter, et enfin les insérer dans la base de données.
- *selection.py* est le fichier qui s'occupe de la sélection des données insérées dans la base. Le principe de la fenêtre gérée par ce code est similaire à celui d'un moteur de recherche avancé. L'utilisateur ou l'utilisatrice peut sélectionner des données en fonction de critères spécifiés par le biais des différents menus affichés dans la fenêtre, et le programme va lui proposer une série de données avec laquelle il va pouvoir travailler.
- *analyze.py* est le fichier qui s'occupe de l'étape d'analyse et de visualisation. Il permet à l'utilisateur ou l'utilisatrice de tirer toute une série d'informations des données sélectionnées par *selection.py*, ainsi que de les réparer et nettoyer.

Ce code gère alors une nouvelle fenêtre, toujours activée par le biais du fichier central, affichant plusieurs boutons, chacun reliés à une fonction qui offre une fonctionnalité différente : générer un graphique avec les données, donner la moyenne de vitesse, analyser les potentielles irrégularités contenues dans les données sélectionnées...

## 6.5 Introspection

Ce dernier sous-chapitre est consacré à une réflexion sur l'état du programme au moment où ce travail est écrit, sur ce qui a été fait et ce qu'il resterait à faire.

Comme mentionné plus haut, le programme n'est malheureusement pas complet, et n'est en fait réalisé qu'à moitié : si la première partie, qui consiste à remplir la base de données, est en grande partie terminée, ce n'est absolument pas le cas de la partie analyse qui est quasiment inexistante.

Le gros du travail restant consisterait donc à fournir aux utilisateurs et utilisatrices une interface pour sélectionner les données avec lesquelles ils et elles veulent travailler, puis une interface leur permettant de les analyser comme cela a été décrit plus haut. La base du programme est, cependant, complètement faite, puisqu'il est d'ores et déjà possible de lui fournir nombre de fichiers de comptages routiers sous des formats divers et variés et il se chargera bel et bien de les traiter et de remplir la base de données.

Cependant, il y a aussi des améliorations qui pourraient être apportées au programme actuel, c'est-à-dire à la phase de spécification des données :

- Le sens de circulation n'est pour le moment pas du tout pris en compte. Or, cette information peut être importante dans certains cas. Par exemple, savoir qu'une rue est beaucoup plus fréquentée dans un sens le matin et dans l'autre le soir peut permettre une analyse à plus large échelle sur les besoins en termes de mobilité dans la zone concernée.
- Les catégories de véhicules comptés, bien qu'enregistrées dans la base de données, ne sont que difficilement utilisables en l'état actuel des choses. La raison à ceci est simple : chaque source de données possède sa propre manière de noter le type de véhicule. Par exemple, un poids-lourd peut être noté "poids lourd", "heavyweight", "pl", "5"...

Cette différence de notation pour une même catégorie pose un problème au niveau de la sélection des données par la suite. Cela demande ainsi à l'utilisateur ou à l'utilisatrice de connaître tous les types de notation et de tous les sélectionner au moment de sa requête, ce qui peut s'avérer quasi-impossible si la notation est abstraite (par exemple dans le cas d'une classification par chiffre, comme "5" pour "poids lourd").

Il n'y a malheureusement pas de réponse parfaite à ce problème. Cependant, une solution pourrait consister à demander, au moment de la spécification des données, à ce que la catégorie des véhicules comptés soit sélectionnée dans une liste de

catégories pré-définies.

- Les informations générales sur les comptages, bien que pas forcément présentes dans les fichiers à disposition, sont perdues lors de l'enregistrement dans la base de données. Or, certaines informations comme l'écart moyen à la vitesse limite ou le pourcentage de trafic des files d'attente sont impossibles à obtenir si les données brutes ne sont pas disponibles.

Il faudrait donc étendre l'objet Source de la base de données avec de nouveaux attributs, et demander à l'utilisateur ou l'utilisatrice de définir les informations générales fournies par le fichier de données au moment de l'exécution du programme.

- De manière générale, l'application se concentre sur son aspect fonctionnel ; l'aspect esthétique est quasi-entièrement passé à la trappe. Ce point relève plus de détails de définitions, ce qui explique que l'esthétique n'aît pas été considérée comme une priorité lors du développement du programme. Cependant, il devrait être possible, sans partir dans un nouveau design graphique complet, d'améliorer l'apparence visuelle de l'application qui se contente pour le moment de l'habillage visuel de base proposé par la librairie Tkinter.
- Sur une note relativement similaire, l'expérience utilisateur pourrait être améliorée. À noter qu'une grande partie de cet aspect a déjà été prise en compte durant le développement de l'application. Par exemple, la sélection de lignes que le programme doit intégrer lors du traitement des données se fait en partie automatiquement et prend par défaut l'intégralité des lignes non-vides du document. Dans la même optique de facilitation, dans certains cas, la première valeur de chaque colonne est affichée à côté de son numéro dans les menus déroulants, permettant ainsi de vérifier que l'on sélectionne bien les colonnes voulues. Cependant, l'expérience utilisateur repose en général en grande partie sur le test du programme par des personnes extérieures à son développement, ce qui n'a pas été fait. Il est alors probable que certains cas d'utilisation n'aient pas été pris en compte dans la réflexion de l'expérience utilisateur, ou que certaines données censées être prises en charge génèrent des erreurs imprévues par le programme.

KallaxComptages se veut extensible et utilisable par toute personne voulant contribuer à son développement. Si le principe de l'open-source offre cette possibilité-là, une attention a également été apportée à la lisibilité du code et à l'écriture de commentaires ayant pour but de faciliter sa compréhension.

Le code pourra donc, dans le futur, être complété pour que le programme offre les fonctionnalités qu'il est censé offrir. Mais il est possible d'aller plus loin, puisqu'il devrait être également possible d'étendre le code afin de prendre en charge d'éventuels autres formats qui n'auraient pas été considérés dans le cadre de ce travail. Ou encore, le programme devrait pouvoir être étendu pour proposer des solutions plus poussées en termes d'apprentissage automatique, aspirant ainsi à offrir des fonctionnalités de traitement de données bien plus avancées que ce qui était prévu au départ.

Il est bon, enfin, de préciser que le programme vient avec une documentation relativement simple, qui explicite son fonctionnement et clarifie les étapes du processus pour toute personne souhaitant l'utiliser.

En résumé, si le programme est encore relativement loin d'être terminé, une base solide a déjà été élaborée et devrait permettre, dans le futur, d'accomplir tout ce pour quoi KallaxComptages a été imaginé, voire plus.

Les réflexions autour de la conception du programme ont déjà pour la plupart été faites. Les différentes étapes de son utilisation ont été établies et schématisées, la base de données a été créée et testée, la structure du programme a été construite, la phase de spécifications et d'introduction des données ont été codées, et les idées pour la phase d'analyse ont été listées. Les ingrédients sont donc réunis pour que KallaxComptages puisse, dans le futur, continuer son développement et, un jour peut-être, être utilisé en production.



## 7 Conclusion

Les données de mobilité regroupent une large variété d'informations, et les stocker peut poser un véritable défi. Les principales difficultés qui peuvent être rencontrées lors du traitement de données de mobilité, et plus particulièrement celles issues de comptages routiers, ont été détaillées dans ce travail, et des solutions partielles ou complètes ont été proposées.

Les processus de collecte, de classification et de traitement des données de mobilité ont encore une marge d'amélioration non négligeable, à la Ville de Lausanne comme ailleurs. Les recherches dans le domaine sont en plein essor mais leurs avancées sont encore trop faibles. C'est notamment par la voie d'une standardisation et d'une mise à disposition des informations au public que le domaine des données de mobilité avancera.

Or, ces avancements sont nécessaires. La collecte et l'analyse des données de mobilité sont devenues indispensables dans le monde de la mobilité d'aujourd'hui. C'est dans ce contexte que des standards comme GTFS ou DATEX II sont développés, mais restent relativement inadaptés pour une utilisation à l'échelle d'un organisme comme la division mobilité de la Ville de Lausanne.

Les données issues de comptages routiers plus spécifiquement apportent leur lot de spécificités à prendre en compte au moment du stockage. Plusieurs décisions doivent alors être prises, portant sur le format des données enregistrées ou sur l'agrégation des passages de véhicules enregistrés par exemple.

Ces données posent aussi certaines difficultés, au niveau de leur traitement, de leur uniformisation et du processus de collecte notamment, mais des solutions existent. Le programme développé dans le cadre de ce travail s'attaque en particulier aux erreurs potentielles contenues dans les données et au manque de standardisation, en offrant une solution de stockage plus homogène à l'aide de sa base de données.

Si KallaxComptages est encore relativement loin d'être fini, une base solide a déjà été construite et ouvre la voie vers la suite de son développement. À titre personnel, je suis plutôt satisfait de l'état du programme. Il est vrai que j'avais pour ambition d'aller plus loin et de proposer directement les différentes fonctionnalités liées à l'analyse des données, mais je me suis assez vite rendu compte que cela n'allait pas pouvoir rentrer dans le cadre d'un travail de Bachelor : le développement s'est effectué en parallèle de l'écriture de ce document, limitant grandement le temps que je pouvais y allouer. J'ai donc choisi de me concentrer sur les fondements de l'application et de faire en sorte que l'entrée des informations dans la base de données soit fonctionnelle. La plupart des questions autour des principes clés de l'application auraient ainsi une réponse, et la suite du développement concernerait surtout l'ajout de fonctionnalités de traitement de données relativement généralistes.

J'estime donc que je n'ai pas particulièrement fait d'erreur au niveau de la gestion de mon temps, et que j'ai pris les bonnes décisions. Il n'était pas vraiment envisageable de réduire les fonctionnalités prévues pour les premières parties du programme, sans quoi la base de données aurait été trop incomplète ou inutilisable. Avoir connaissance du temps que me prendrait le développement du projet dès son initialisation ne m'aurait pas permis de structurer mon travail autrement. De plus, la phase de planification du développement pour laquelle j'ai pris du temps au début s'est révélée elle aussi utile, puisque c'est elle qui m'a permis de concevoir le schéma du traitement de l'information ([fig. 32](#)). J'avais ainsi une

vision claire des différentes étapes que je voulais développer, et je n'ai quasiment pas eu besoin de revenir en arrière pour modifier le processus imaginé.

Cependant, même s'il ne m'aurait pas permis d'achever le programme, du temps aurait pu être gagné à certaines étapes du développement, notamment en raison de certaines difficultés auxquelles j'ai été confronté et que je n'avais pas vraiment vu venir.

Par exemple, durant le développement, je me suis rendu compte que la standardisation des données pose un défi encore plus grand que ce que j'imaginais. Ce n'est pas tellement que certains formats de données peuvent être totalement différents, mais plutôt que certains détails, qui peuvent paraître minuscules, m'ont en fait forcé à revoir certains principes et certaines fonctionnalités. Par exemple, les données agrégées par tranches de vitesse comportent une information cruciale, mais qui n'est pas sous forme de colonnes : la date du comptage. J'ai donc dû repenser mon étape 6, qui était initialement prévue comme la définition de colonnes uniquement, pour qu'elle intègre un calendrier permettant de sélectionner une date fixe de manière pratique.

Et encore maintenant, certaines questions attendent encore une réponse définitive. Par exemple, la standardisation des modèles de véhicules disponibles : je partirais sur une solution impliquant un menu donnant le choix entre plusieurs modèles intégrés dans l'application, mais la mise en œuvre d'une telle solution n'est pas totalement triviale au niveau de l'implémentation. Cependant, je suis confiant quant au fait que ces défis, tout comme ceux auxquels j'ai été confronté durant le développement de KallaxComptages, peuvent être résolus.

La partie écrite du travail s'est aussi révélée très instructive. Par rapport aux compétences acquises d'une part, puisque j'ai dû effectuer une analyse en profondeur de ce que j'avais appris et constaté lors de mon stage à la Ville de Lausanne. Par rapport à des faits très concrets d'autre part, puisque j'ai dû me renseigner sur des sujets comme l'utilisation et le partage d'informations par les acteurs de la mobilité. J'ai ainsi eu l'occasion de mettre en pratique des connaissances sur les outils de standardisation existants dans le domaine, et j'en ai beaucoup appris sur des sujets qui me tiennent à cœur.

Cependant, le manque de ressources scientifiques accessibles sur le thème des données de mobilité et leur stockage m'a posé une certaine difficulté. Le développement de ce domaine est encore relativement récent, et surtout est très niche. La littérature portant dessus est donc encore trop limitée.

Je ne compte pas abandonner KallaxComptages à la fin de ce travail. Les plans sont posés pour la suite du développement, et je suis tout à fait partant de continuer l'avancement de l'application, dans le but qu'elle serve un jour, à la Ville de Lausanne ou ailleurs, ou qu'elle puisse servir de base au développement d'une solution plus complète. Je crois avoir fait le nécessaire à cet effet.

Les recherches autour de la collecte, du stockage, du traitement et de la standardisation des données de mobilité ont encore une certaine distance à parcourir. Mais le progrès dans ce milieu est en accélération constante, et semble être sur la bonne voie. De plus en plus de décideurs reconnaissent l'importance des données de mobilité pour le développement urbain de manière générale, ce qui laisse présager de belles avancées pour l'avenir des comptages routiers et des données de mobilité et tout ce qu'elles peuvent déclencher dans notre société connectée. Il n'y a plus qu'à espérer que les décideurs et autres facilitateurs politiques prennent le train en marche.

## 8 Références

### 8.1 Bibliographie

- [ali17] Alizadeh, Hamzeh & Farooq, Bilal & Morency, Catherine & Saunier, Nicolas. (2018). On the role of bridges as anchor points in route choice modeling. *Transportation*. 45. 10.1007/s11116-017-9761-7.
- [asi13] Asif, Muhammad Tayyab & Dauwels, Justin & Goh, Chong & Oran, Ali & Fathi, Esmail & Xu, Muye & Mohan, Dhanya & Mitrovic, Nikola & Jaillet, Patrick. (2013). Spatiotemporal Patterns in Large-Scale Traffic Speed Prediction. *IEEE Transactions on ITS*. 15. 1-11. 10.1109/TITS.2013.2290285.
- [asl12] Aslam, Javed & Lim, Sejoon & Pan, Xinghao & Rus, Daniela. (2012). City-scale traffic estimation from a roving sensor network. *SenSys 2012 - Proceedings of the 10th ACM Conference on Embedded Networked Sensor Systems*. 141-154. 10.1145/2426656.2426671.
- [cho11] Cho, Eunjoon & Myers, Seth & Leskovec, Jure. (2011). Friendship and Mobility: User Movement In Location-Based Social Networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1082-1090. 10.1145/2020408.2020579.
- [liu21] Liu, Zidan & Wu, Kaishun. (2021). Mobility Data-Driven Urban Traffic Monitoring. 10.1007/978-981-16-2241-0.
- [row22] Rowe, Francisco & Arribas-Bel, Daniel & Calafiore, Alessia & MacDonald, Jacob & Samardzhiev, Krasen & Fleischmann, Martin. (2022). Mobility data in urban science. 10.13140/RG.2.2.16202.24009.

### 8.2 Webographie

- [ace23] Atelier de Construction & Etudes Techniques SA : Boucles à induction.  
<https://acet.ch/comptage-et-statistiques/recueil-de-donn%C3%A9es-du-trafic>  
(dernière consultation le 22.01.2023)
- [are23] Office fédéral du développement territorial ARE : Microrecensement mobilité et transports.  
<https://www.are.admin.ch/are/fr/home/mobilite/bases-et-donnees/mrmt.html>  
(dernière consultation le 13.02.2023)
- [dat23] DATEX II Docs : The components of the DATEX II suite of standards.  
<https://docs.datex2.eu/basics/basicstructure.html>  
(dernière consultation le 13.02.2023)

- [elw22] ELWC : ViaCount II - Traffic Data Collection.  
<https://www.elwc.com.au/viacount-ii-traffic-data-collection/>  
(dernière consultation le 16.12.2022)
- [gbf23] General Bikeshare Feed Specification : Current Version.  
<https://gbfs.mobilitydata.org/specification/reference/>  
(dernière consultation le 13.02.2023)
- [geo23] Georgia Southern University : Data Management Services.  
<https://georgiasouthern.libguides.com/c.php?g=833713&p=5953146>  
(dernière consultation le 25.03.2023)
- [gtf23a] General Transit Feed Specification : Example GTFS Feed.  
<https://gtfs.org/schedule/example-feed/>  
(dernière consultation le 13.02.2023)
- [gtf23b] General Transit Feed Specification : Trip Updates.  
<https://gtfs.org/realtime/feed-examples/trip-updates/>  
(dernière consultation le 13.02.2023)
- [ike23] IKEA : Kallax.  
<https://www.ikea.com/ch/fr/p/kallax-etagere-effet-chene-blanchi-00324518/>  
(dernière consultation le 30.04.2023)
- [int22] InterTraffic : Eco-Counter.  
<https://company.intertraffic.com/Eco-Counter?Language=EN&eventid=22096&account=00018077-0>  
(dernière consultation le 16.12.2022)
- [lau22] Ville de Lausanne : Missions du Service de la mobilité et de l'aménagement des espaces publics.  
<https://www.lausanne.ch/officiel/administration/finances-et-mobilite/mobilite-et-amenagements-publics/a-propos/missions.html>  
(dernière consultation le 27.11.2022)
- [mit23] Ira Winder : Computational Ants: Agent Based Visualization Technique with CDR OD Matrix.  
<https://ira.mit.edu/blog/agent-based-visualization>  
(dernière consultation le 13.02.2023)
- [mix13] mixedbredie : Time Manager plugin in QGIS 2.1.0 master.  
<https://www.flickr.com/photos/mixedbredie/9737909855/in/pool-qgis-screenshots/>  
(dernière consultation le 12.03.2023)
- [mob23] MobilityData : What We Do.  
<https://mobilitydata.org/what-we-do/>  
(dernière consultation le 12.02.2023)

- [ofr09] Office fédéral des routes OFROU : Directive Postes de Comptage du Trafic.  
[https://www.astra.admin.ch/dam/astra/fr/dokumente/standards\\_fuer\\_nationalstrassen/astra\\_13012\\_verkehrszaehler2009v105.pdf.download.pdf/astra\\_13012\\_postesdecomptagedutrafic.pdf](https://www.astra.admin.ch/dam/astra/fr/dokumente/standards_fuer_nationalstrassen/astra_13012_verkehrszaehler2009v105.pdf.download.pdf/astra_13012_postesdecomptagedutrafic.pdf)  
(dernière consultation le 15.04.2023)
- [ope23] Open Data Plattform öV Schweiz : Open data platform mobility Switzerland.  
<https://opentransportdata.swiss/en/group>  
(dernière consultation le 12.02.2023)
- [rts23] Radio Télévision Suisse : Chaque Helvète parcourt en moyenne 30 kilomètres par jour, le plus souvent en voiture.  
<https://www.rts.ch/info/suisse/13925929-chaque-helvete-parcourt-en-moyenne-30-kilometres-par-jour-le-plus-souvent-en-voiture.html>  
(dernière consultation le 25.05.2023)
- [rue22] Rue de l'Avenir : Lausanne - Essai de 30 km/h la nuit.  
<https://rue-avenir.ch/themes/generalites-zones-30-2/mise-en-oeuvre/zones-30-nocturnes/>  
(dernière consultation le 16.12.2022)
- [sig23] Signal : Panneaux à messages variables LED pour routes à grand débit.  
<https://www.signal.ch/fr/assortiment/panneaux-afficheurs-a-messages-variables-led/>  
(dernière consultation le 16.12.2022)
- [sma22a] SmartRoad : Radar pédagogique.  
<https://smartroad.ch/fr/solutions/radar-pedagogique>  
(dernière consultation le 16.12.2022)
- [sma22b] SmartRoad : Radar anti-bruit.  
<https://smartroad.ch/fr/product/radar-anti-bruit>  
(dernière consultation le 16.12.2022)
- [sql23] SQLite : File Format Changes in SQLite.  
<https://sqlite.org/formatchng.html>  
(dernière consultation le 16.04.2023)
- [swi22] SwissTraffic : Produits  
<https://www.swisstraffic.ch/fr/produkte>  
(dernière consultation le 30.05.2023)
- [una22] Unacast : What is mobility data, and why is it important?  
<https://www.unacast.com/post/mobility-data>  
(dernière consultation le 16.12.2022)



## 9 Annexes

### 9.1 GitHub

Le code-source de KallaxComptages est disponible sur GitHub au lien suivant :

<https://github.com/Elric02/KallaxComptages>

Le GitHub contient également des informations résumées sur le programme et son développement (se trouvant de manière plus complète dans ce rapport), des informations sur comment y contribuer, ainsi que le guide d'utilisation retranscrit au [chapitre 9.2](#).

### 9.2 Guide d'utilisation de l'application

KallaxComptages ne nécessite pas vraiment d'installation. Il suffit de copier le dossier "out" sur votre machine, puis de lancer l'exécutable (.exe).

Une fois l'exécutable lancé, la fenêtre suivante apparaît :



#### Ajout de données

##### Étape 1 : Choix du fichier



La première étape consiste à choisir le fichier contenant les informations à introduire dans la base de données.

Le fichier peut être au format .xlsx ou .csv.

## Étape 2 : Choix de la feuille à utiliser



Si le fichier choisi est au format .xlsx, cette étape consiste à choisir la feuille à utiliser parmi celles contenues dans le fichier.

Si le fichier est au format .csv, le programme détermine automatiquement le séparateur des colonnes et passe à l'étape suivante sans intervention de l'utilisateur.trice.

## Étape 3 : Choix du type de données présentes

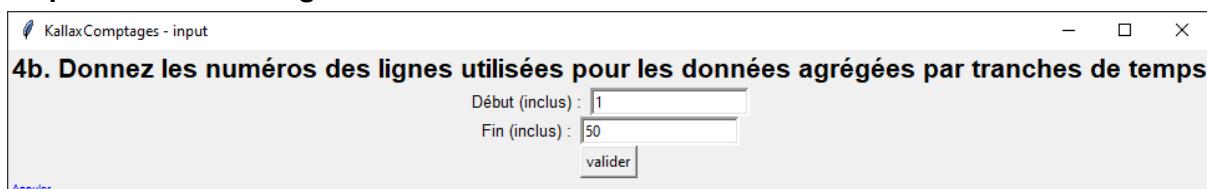


Cette étape consiste à spécifier le type de données présent dans le fichier. Il y a 3 types de données possibles :

- Données brutes : chaque passage de véhicule est enregistré sur sa propre ligne.
- Données agrégées par temps : chaque ligne représente les passages de véhicule sur une certaine période de temps, par exemple par tranches d'1h.
- Données agrégées par vitesse : chaque ligne représente les passages de véhicule ayant eu une certaine vitesse, par exemple par tranches de 5 km/h.

Si des données brutes sont disponibles, le programme n'a pas besoin de données agrégées, d'où la sélection unique dans le menu. Autrement, il faut spécifier si les données à disposition sont agrégées par temps, par heure, ou si les deux types sont présents.

## Étape 4 : Choix des lignes à utiliser



Cette étape consiste à spécifier les numéros de ligne du fichier à utiliser. Elle permet d'éviter de prendre en compte les éventuels en-têtes et lignes superflues contenues dans le fichier.

Par défaut, les lignes non-vides du fichier sont prises en compte.

Le contenu de la fenêtre varie en fonction des types de données spécifiées à l'étape 3.

## Étape 5 : Choix des colonnes de temps et/ou de vitesse



Cette étape consiste à spécifier les numéros/lettres des colonnes contenant l'information des tranches de temps et/ou de vitesse (en fonction du type de données agrégées sélectionnées).

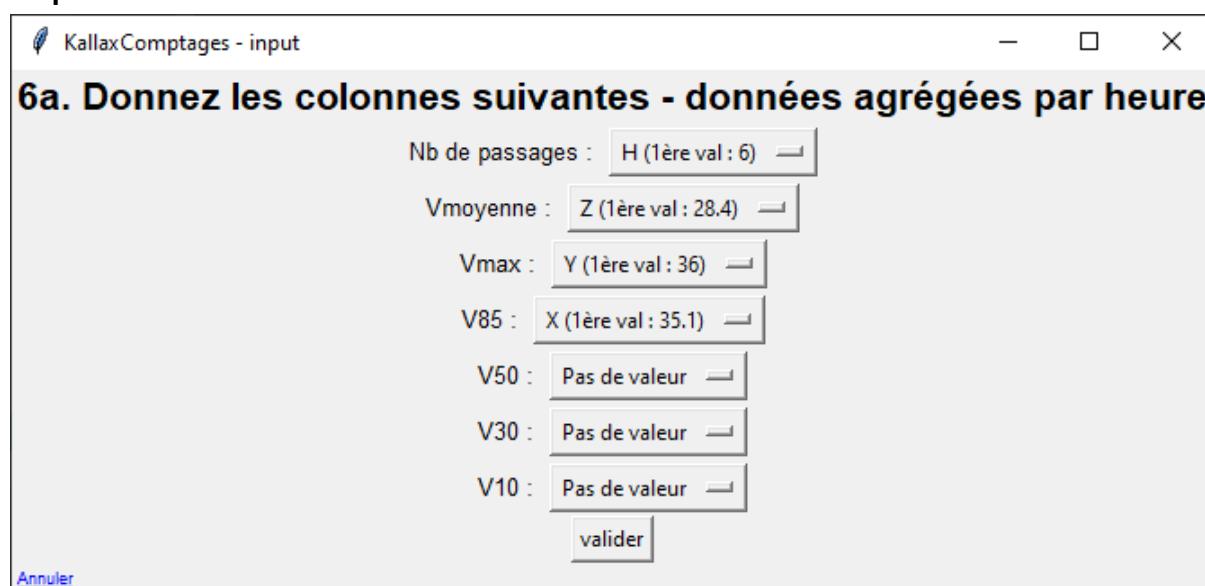
Elle est automatiquement passée si le type de données a été défini sur "brutes" à l'étape 3. La première valeur de chaque colonne (déterminée à partir de la première ligne indiquée à l'étape 4) est aussi affichée, dans le but de faciliter la sélection de la bonne colonne.

Si la cellule à cet endroit est vide, "vide~" sera affiché.

Si la date et l'heure sont affichées dans la même colonne, il est possible de sélectionner "Déjà présent dans "Date"" pour la colonne d'heure.

Il en va de même pour le début et la fin de la tranche de vitesse pour les données agrégées par vitesse.

## Étape 6 : Choix des colonnes à utiliser



Cette étape consiste à spécifier les numéros/lettres des colonnes contenant les informations présentes dans le fichier.

Le contenu de la fenêtre varie en fonction des types de données spécifiées à l'étape 3.

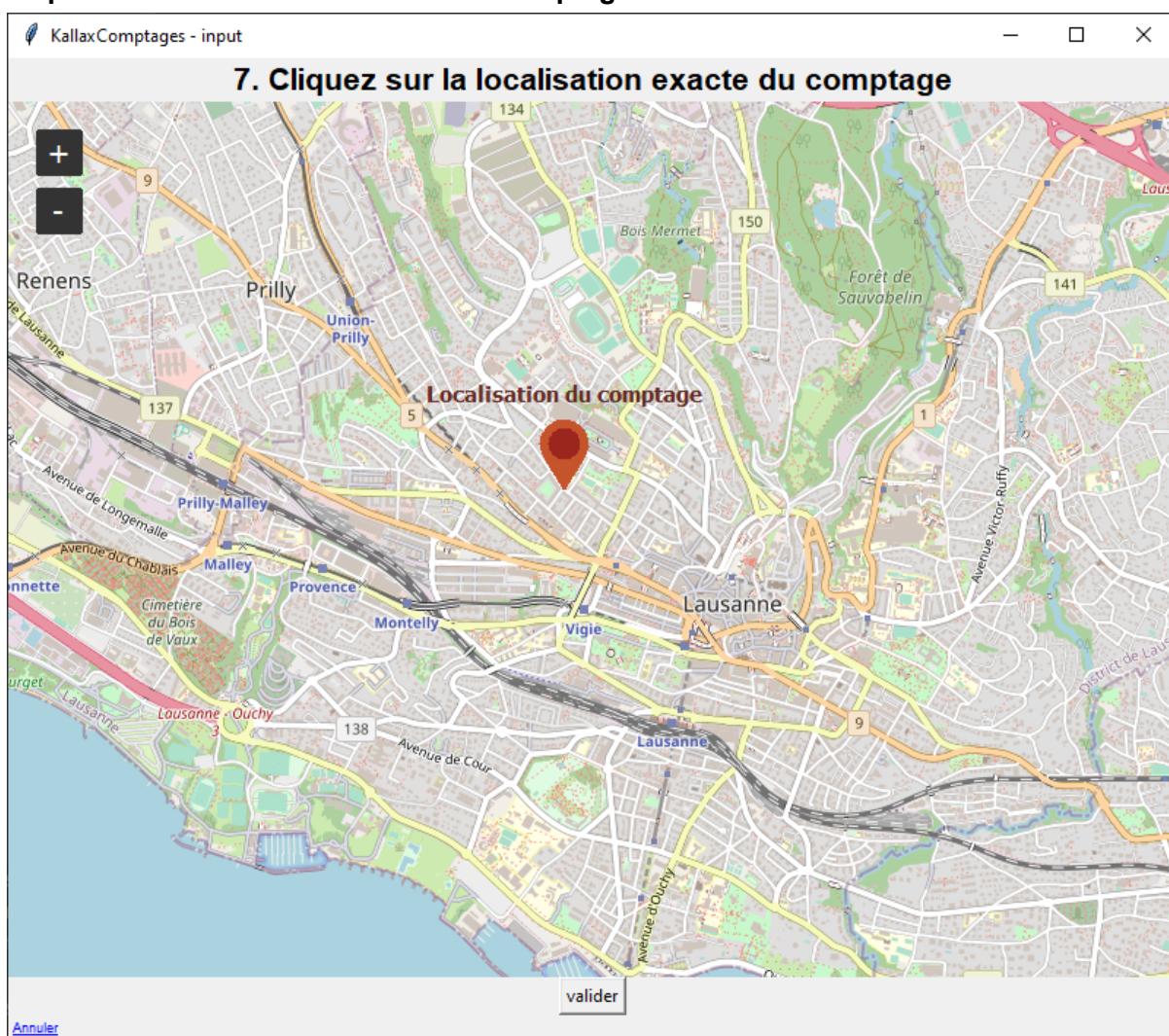
Si des données agrégées par tranches de vitesse sont présentes, il est également demandé d'entrer le jour où le comptage a été effectué.

La première valeur de chaque colonne est affichée de la même manière qu'à l'étape 5.

Il est possible de spécifier lorsque des données sont déjà présentes dans d'autres colonnes (par exemple l'heure et la date), ainsi que lorsque des données ne sont pas présentes.

Il n'est pas possible de définir certaines données comme absentes, comme le nombre de passages de véhicules par exemple.

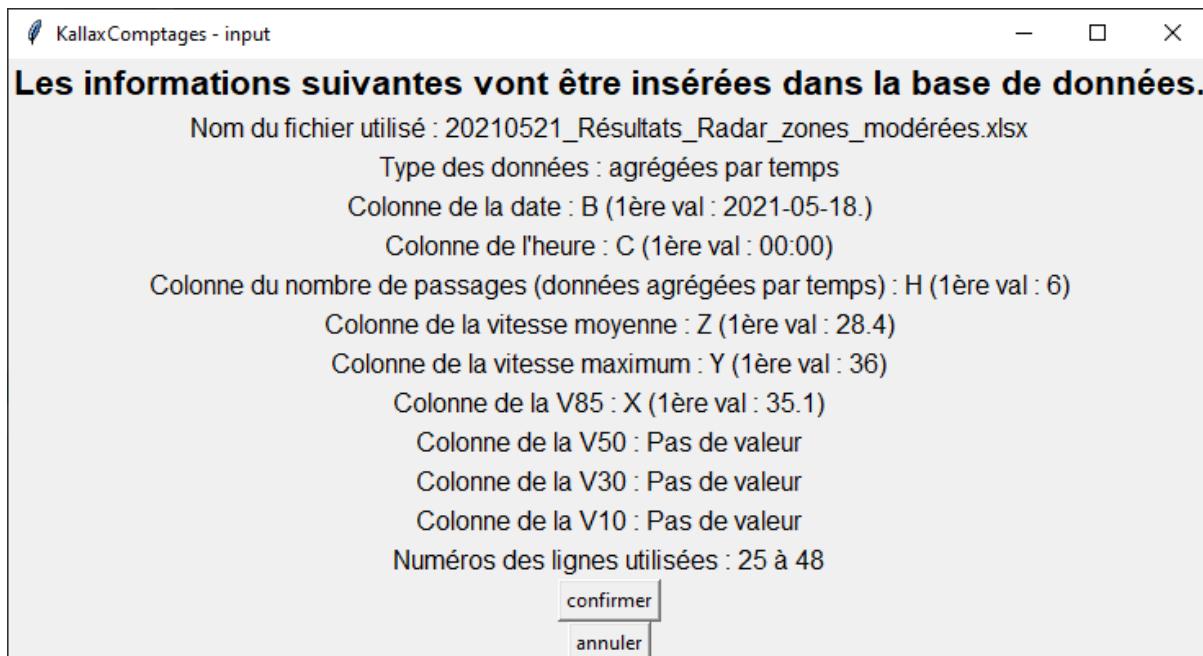
## Étape 7 : Choix de la localisation du comptage routier



Cette étape consiste à préciser la localisation exacte du comptage routier, sous forme de pointeur sur une carte.

Le pointeur est ensuite automatiquement converti en coordonnées géographiques au moment de l'enregistrement dans la base de données.

## Étape 8 : Récapitulatif des données à entrer



Cette étape consiste simplement à vérifier si les informations entrées précédemment sont correctes. Pour ce faire, le programme génère un récapitulatif contenant tout ce qui a été spécifié aux étapes précédentes.

Une fois l'étape 8 validée, les données sont envoyées dans la base de données. Ceci peut prendre un peu de temps, en particulier si le comptage contient une grande quantité de données.

Il est recommandé de ne pas fermer le programme tant que la fenêtre "Données envoyées à la base de données !" n'est pas apparue.