

SYNTHESE DES RESULTATS DE L'ENQUETE AUPRES DE RE-UTILISATEURS DE DONNEES OUVERTES

Françoise Paquienséguy, Valérie Larroche, Marie-France Peyrelong, Martine
Vila-Raimondi, Valentyna Dymytrova

► **To cite this version:**

Françoise Paquienséguy, Valérie Larroche, Marie-France Peyrelong, Martine Vila-Raimondi, Valentyna Dymytrova. SYNTHESE DES RESULTATS DE L'ENQUETE AUPRES DE RE-UTILISATEURS DE DONNEES OUVERTES: LIVRABLE n°1. [Rapport de recherche] SciencesPo Lyon; Ensib; Lyon3. 2016. <hal-01432124>

HAL Id: hal-01432124

<https://hal.archives-ouvertes.fr/hal-01432124>

Submitted on 18 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





elico

Équipe de recherche de Lyon en sciences
de l'information et de la communication

SYNTHESE DES RESULTATS DE L'ENQUETE AUPRES DE RE-UTILISATEURS DE DONNEES OUVERTES

LIVRABLE n°1

Rendu en janvier 2016 et finalisé en septembre 2016

FRANÇOISE PAQUIENSÉGUY (PU)

VALERIE LARROCHE (MCF)

MARIE-FRANCE PEYRELONG (MCF)

MARTINE VILA (MCF)

VALENTYNA DYMYTROVA (Post-doct)

1. INTRODUCTION

Porté par le laboratoire Elico (EA 4147), ce livrable n°1 rend compte de l'enquête menée du septembre 2015 au juin 2016 auprès des personnes-ressources et des ré-utilisateurs des données du portail des données métropolitaines data.grandlyon.com. L'enquête de terrain comportait des entretiens semi-directifs et un questionnaire en ligne (Annexes 1 et 2). Dans un premier temps, en avril 2015, nous avons mené 3 entretiens avec des personnes ressources :

- Nathalie Vernus-Prost, Chief Data Officer de la Métropole de Lyon,
- Laurence Martinez, responsable de l'unité diffusion des données au service Information géographique, et gestionnaire du portail Data Grand Lyon,
- Patrick Gendre, Cérema (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement), Ministère de l'Environnement, de l'Energie et de la Mer.

Dans un deuxième temps, en septembre 2015, Mme Martinez a sollicité les 56 ré-utilisateurs inscrits et déclarés sur le portail Data Grand Lyon. Nous les avons tous contactés. Seuls 9 ont accepté de nous répondre, mais finalement, 2 n'ont pas donné suite malgré trois relances. Dans un troisième temps, de janvier à mars 2016, nous avons interviewé 10 ré-utilisateurs dont les contacts ont été obtenus par des personnes interrogées ou nos relations professionnelles. Enfin, deux derniers entretiens ont été menés en mai et en juin 2016. Ainsi au total, 22 entretiens ont-ils été réalisés (Annexe 3). Ils sont complétés par un questionnaire en ligne visant à recueillir des informations très précises sur les jeux et les formats utilisés. Seuls 13 ré-utilisateurs sur 22 interviewés ont répondu à toutes les questions proposées.

Les pages qui suivent résultent de l'analyse de l'équipe Elico, soutenue par les déclarations, idées et commentaires des personnes interrogées. Afin de préserver l'anonymat des enquêtés, nous avons numéroté les entretiens d'une façon aléatoire. Les citations renvoient à des numéros attribués aux entretiens et précisent, si nécessaire, le positionnement professionnel de ré-utilisateur : développeur, analyste, coordinateur de projets, gestionnaire ou promoteur de la plateforme. Les résultats de l'enquête ont été restitués le 27 juin 2016 à la Métropole de Lyon et ont suscité intérêt et questions de la part de nos interlocuteurs.

2. ETAT DE L'ART

2.1 CONTEXTE

Toutes les décisions prises sur les jeux de données, la structuration du portail métropolitain, l'ouverture de certains jeux, résultent de choix dont la résonance politique est forte et perceptible.

Le caractère stratégique des données d'une part et de l'accès aux données de l'autre prend une ampleur encore plus forte dans le cadre de la réforme territoriale et du transfert de la compétence VOIRIE.

Les portails de données ouvertes fonctionnent au sein d'un mouvement paradoxal. Nous remarquons d'abord une centralisation à outrance puisque toutes les données partent et arrivent de/sur le portail, faisant de lui un acteur à part que, finalement, personne ne maîtrise vraiment : il gagne en autonomie, en même temps, il est le point de convergence d'un très grand nombre d'acteurs qui interviennent dans des registres différents, compartimentés et complémentaires. En effet, ce qui se passe sur le portail relève de plusieurs registres : des dépôts manuels ou automatiques de jeux et des liens vers des jeux offerts par d'autres fournisseurs. De nombreuses directions ou communes de la métropole abondent au portail qui reste, en interne, structurée selon l'administration dont il dépend, en tentant compte des suggestions de la directive européenne *Inspire*. Tout cela compartimente et cloisonne les actions, les jeux de données et les acteurs.

En parallèle, nous constatons une grande diversité du côté des utilisateurs : téléchargements automatiques et anonymes de jeux (licence ouverte), téléchargements identifiés (licence associée) ou liens impossibles à détecter vers des jeux de données de la plateforme. De fait, le poste de relation avec les utilisateurs (Mme Martinez) a été créé en avril 2014, complété par celui de *Chief data officer* pour suivre l'évolution et l'utilisation du portail (Mme Vernus-Prost).

2.2. BESOINS A PROPOS DES DONNEES LIEES A LA MOBILITE

Données anonymes

Les données doivent être anonymes, sinon en croisant des jeux de données, il devient possible d'identifier l'individu par un numéro d'abonné, par exemple, et de tracer son parcours de mobilité [Entretiens 15, 19]. De plus, le croisement des données favorise certains modes de transport et génère des données économiques stratégiques (public captif = augmentation des tarifs) [Entretiens 1, 15, 19]. Autrement dit, l'aspect stratégique des données ne doit pas être minoré.

Données brutes

Déposées, téléchargées en aveugle la donnée brute semble être le niveau de consensus des utilisateurs, internes ou externes du portail. « *Pour permettre la mise en place d'un ensemble de services sur le portail open data, il faudrait que la donnée proposée soit la plus brute possible* » (Chignard, 2013). Certains ré-utilisateurs considèrent les données récupérées sur les portails des données ouvertes comme différentes des données sources et élaborées par le fournisseur : « *Si le fournisseur de données sait exactement ce qu'il fait, s'il est consciencieux et pérenne, on n'est pas*

contre une donnée élaborée et maîtrisée par le fournisseur, s'il ne maîtrise pas ce qu'il fait, que c'est une boîte noire, qu'il met des données à disposition pour dire qu'il met quelque chose à disposition, on préfère les données sources » [Entretien 22]. Dans ce contexte, il est important de comprendre le point de vue du producteur de la donnée : *« la donnée, ce n'est pas suffisant, il faut aussi les moyens de l'interpréter et donc des spécifications sur ces données et éventuellement des rapports sur l'utilité de ces données pour pouvoir vraiment la mettre en contexte »* [Entretien 7]. En effet, les données mises à disposition peuvent présenter certaines limites : *« Il y a des erreurs, des dérives, il y a des anomalies, il y a des pannes, il y a des voitures qui reculent après avoir avancé, qui passent sur la bande du comptage dans le mauvais sens... Il y a tellement d'événements qui font que le comptage d'un parc est ré-calibré sur un parc qui tourne bien toutes les nuits. Toutes les nuits les agents vont compter le nombre de voitures pour vérifier que c'est bien cela qu'on a en comptage. Et il y a toujours des dérives »* [Entretien 18].

Les ré-utilisateurs travaillent les données, souvent manuellement, pour enlever les éléments qui les gênent dans les jeux, ils la « brutifient » (Chignard, 2013). Ensuite, beaucoup la (re)-contextualisent pour lui donner un sens et en permettre la compréhension [Entretien 7]. Certains intègrent les données de façon automatique [Entretien 7], d'autres les déposent dans un intranet après les avoir « brutifiées » [Entretien 4], et d'autres génèrent des API [Entretiens 5, 7, 14, 17].

L'accès aux données brutes est souvent possible *via* des API. Toutefois, cela impose quelques limites dont les développeurs interrogés sont conscients : *« Une API est une surcouche qui donne des services en plus qui sont les bienvenus, mais du coup il y a des restrictions... Une API, c'est bien pour un développeur qui veut tester des usages, mais si l'on veut faire une entreprise derrière, il faut faire attention parce qu'il peut y avoir des restrictions »* [Entretien 17]. Une autre difficulté consiste dans l'exploitation des données conçues ou prédestinées à une autre utilisation. Cela concerne la plupart des données du portail métropolitain : *« Les données sur la plateforme n'ont pas été produites pour la plateforme, ça c'est bien normal. Ça coûterait trop cher de les produire pour la plateforme. Elles ont été produites en vue de faire quelque chose. Par exemple, les données de capteurs de trafic, c'est en vue de faire de la régulation. Les données emplacement de marchés, les travaux c'est en vue de faire de l'information voyageurs. Toute donnée produite qui est mise en ligne sur la plateforme n'est pas produite pour la plateforme mais dans un autre but. Pour bien interpréter les forces et les limites de cette donnée, il faut comprendre le point de vue du producteur original. »* [Entretien 7].

Licence ouverte

Elle élargit les possibilités et supprime un niveau décisionnel. Bref, c'est plus simple. Cependant, dans le domaine de la mobilité, beaucoup de données intéressantes viennent des

déléataires, d'où l'introduction par la Métropole de Lyon d'une licence associée qui pose d'ailleurs problème pour l'exploitation commerciale de ces données [Entretien 7].

Variété de formats et présentations hétérogènes des données

Nous trouvons une réelle variété de formats des données, des mots-clés et des termes employés pour les présenter : « *Il existe en France 7 fournisseurs des données en temps réel liés aux transports collectifs et 7 formats différents d'API* » [Entretien 11] ou encore « *les données des parkings ne sont pas présentées de la même manière par le STIF, Service des transports en Île de France et par le Grand Lyon* » [Entretien 17]. Les formats des données disponibles viennent souvent des DSI des collectivités qui choisissent des formats informatiquement plus simples mais qui ne sont pas portés par le monde des transports [Entretien 11]. En plus, les collectivités produisent les données en fonction de leurs compétences : « *Chacun met à disposition des données qu'il possède. Par exemple, les données liées aux parkings de covoiturage sont très détaillées chez Nantes Métropole et sur le site du Conseil général il n'y a quasiment pas d'informations associées* » [Entretien 12]. Par ailleurs, certains types de données n'ont pas de standard, comme par exemple, les formats décrivant un bâtiment [Entretien 17].

Si le professionnalisme et la compétence des personnes interrogées permettent la plus part du temps de lever ces difficultés [Entretiens 4, 19], cette hétérogénéité contribue également à « *créer des métiers d'agrégateur des données disparates* » [Entretien 12]. Plus concrètement, les développeurs ne peuvent pas traiter des données hétérogènes d'une manière automatique, ils doivent faire un plus grand travail d'agrégation manuelle et de consolidation pour pouvoir répliquer leurs applications sur d'autres villes [Entretiens 14, 17].

Créer des référentiels multimodaux de données

Selon les coordinateurs de projets, les référentiels multimodaux de données permettraient d'harmoniser des données hétérogènes [Entretien 5] en ménageant l'interopérabilité afin d'avoir une plateforme structurée malgré sa diversité. En effet, plusieurs personnes interrogées qu'ils soient développeurs ou analystes, souhaitent une meilleure structuration et une mise en cohérence des données liées à la mobilité au niveau national, voire européen [Entretiens 11, 17]. En absence d'un tel référentiel commun à l'heure actuelle, les développeurs contribuent à *OpenStreetMap* qui constitue pour eux « *une base vraiment commune des données ouvertes* » [Entretien 14]. En même temps, l'annuaire PASSIM (<http://www.passim.info/>) reste largement méconnu et sous-utilisé alors qu'il a été mis en place par le Ministère de l'Ecologie, du Développement durable et de l'Energie dans l'objectif de recenser les offres de transport et les services d'information liées à la mobilité en France à destination d'un grand public [Entretien 11].

Finalement, nous ne relevons pas de problèmes techniques majeurs. Cependant, les problèmes décisionnels sont assez fréquents, notamment le caractère politique et stratégique de l'ouverture des données : « *C'est moins un frein technique que celui de l'imaginaire des producteurs de données* » [Entretien 1].

3. CADRE D'ANALYSE

3.1. LES PRATIQUES PROFESSIONNELLES

Les postures politiques et l'organisation interne des services des métropoles ont un impact décisif sur les pratiques professionnelles des responsables et informaticiens des portails des données ouvertes [Entretien 12]. Les responsables de services OD et les chefs de projets OD au sein des collectivités territoriales forment tous le poids du politique dans leurs activités, dont ils ne décident que très rarement [Entretiens 4, 6, 12]. En effet, l'ouverture des données se situe à la fois dans l'injonction publique des directives européennes et dans la réforme territoriale ressentie très fortement [Entretien 4]. Les failles structurelles empêchent parfois la remontée des informations du terrain au portail ou la coordination entre les différents services [Entretien 12] ou, pour le moins la complexifient en termes de compétence et de référentiel.

Les enjeux sont politiques et économiques à l'échelle du territoire (le promouvoir, le développer) mais aussi au niveau de la métropole et du portail (vendre certaines données) [Entretiens 1, 1, 21]. Certains interviewés attribuent une place centrale à l'Etat dans les injonctions concernant la publication, la structuration et le contrôle des jeux de données : « *Si l'Etat impose aux transporteurs de s'inscrire obligatoirement dans un registre, il change le fonctionnement du marché et il présente toutes les offres de transport selon les modalités connues et transparentes qui sont celles de proximité par rapport à l'utilisateur. Cela crée des conditions pour une souveraineté numérique et les usagers se tournent toujours vers des moteurs de recherche plus loyaux* » [Entretien 15].

Cela dit, les données liées à la mobilité se révèlent extrêmement stratégiques pour le territoire [Entretiens 2, 12] et les acteurs concernés. L'objectif visé est d'avoir accès aux données en temps réel, d'arriver à passer des horaires théoriques à l'arrêt aux horaires en temps réel. Celui-ci est défini ainsi : « *Une donnée temps réel est une donnée dont la valeur à l'instant T représente la réalité des choses.* » [Entretien 3]. La fréquence dépend de l'information portée par la donnée : un chantier de voirie = 1 fois par jour est suffisant ; un accident sur la rocade = toutes les 10 minutes, etc.... Certaines données en temps réel très fines et très poussées existent déjà pour renseigner des girouettes d'affichages, comme à la SNCF pour les lignes du Francilien, sans être ouvertes pour être basculées dans des API pour autant [Entretien 1]. Toutefois, les données en temps réel « *sont des*

données difficiles à gérer, ça demande de gros volumes de données, des connections entre machine en temps réel, des contraintes de la chaîne technique » [Entretien 6]. En même temps, il faut souligner une réticence des acteurs privés envers la publication des données, particulièrement en temps réel [Entretien 6].

Les chefs de projets OD font de la veille sur les autres portails métropolitains [Entretien 2], mais avec une visée plus concurrentielle que communautaire, car les données sont liées au territoire et souvent difficilement exploitables. La mise en commun et le partage semblent très difficiles car les administrations et les acteurs publics ou en régie des transports ne sont pas les mêmes d'une métropole à une autre et ne sont pas structurés à l'identique [Entretiens 3, 4]. Cependant, ils se retrouvent parfois, à l'instar d'une communauté de pratiques [Entretiens 2, 4]. Au-delà de leurs administrations ou collectivités de référence, la culture du partage d'expériences semble être assez fréquente dans l'OD. Par exemple, la création d'un réseau des géomaticiens de la Métropole de Lyon témoigne d'une ouverture au niveau des pratiques tandis que la localisation des données les segmentent et les rend difficilement partageables, hormis dans certains cas pour data.gouv.fr. Autrement dit,

- la thématique de la mobilité représente un certain enjeu (économique, politique, territorial),
- la marge de manœuvre de ceux qui fournissent ou utilisent des données métropolitaines est très faible,
- celle des communes de 3500 habitants intégrées à la Métro se réduit drastiquement,
- ces communes n'ont aucune visibilité sur l'exploitation des données qu'elles fournissent à la métropole ou à leurs publics,
- ils font tous des métiers différents avec des niveaux de compétence en informatique très variables selon leurs âges et formations,
- la plupart d'entre eux naviguent en permanence entre une API (du service, de la commune) et les jeux de donnée du portail, dans des pratiques plus ou moins artisanales encore.

3.2. LES USAGES

Les usages sont à analyser en fonction de la position de l'acteur dans la chaîne. Celle-ci va de la production de la donnée à sa réutilisation par des coordinateurs de projets, développeurs ou analystes des bureaux d'étude, en passant par les gestionnaires des portails OD, même si parfois les frontières entre ces différentes positions restent floues. Par exemple, les développeurs de *Géovélo*, application Internet et mobile qui guide les cyclistes en ville, sont aussi les producteurs au sens où les données générées par application sont exportées vers les collectivités territoriales. De même, la

société *Arcadis* qui est à la base un ré-utilisateur des données pour produire des analyses se positionne aussi en tant que producteur et coproducteur des données.

1. Pour les producteurs de données, en particulier pour le portail Data Grand Lyon, l'échange de données avec d'autres services ou avec des partenaires historiques, se faisait par transfert de fichier *via* FTP [Entretien 3]. Le portail a permis de standardiser les données, de les rendre plus fiables et de gagner du temps. En effet, avant l'ouverture du portail DGL les données étaient structurées selon les formats imposés par les métiers municipaux : un espace n'étant pas répertorié de la même façon selon qu'il est place de parking, arrêt de bus, lot constructible, référence cadastrale ou emplacement de silo à verre. Ces données devaient donc, pour pouvoir passer d'un référentiel métier à un autre, être complétées ou nettoyées à la main. Le portail propose de qualifier les données de façon la plus complète possible et de lier les informations pertinentes, *via* une clé de codage, au référentiel de son producteur, de façon systématique et automatisée.

La qualité de la donnée est une préoccupation importante des producteurs. Malgré une grande homogénéité dans la production des données, celles fournies au portail métropolitain par Lyon Park Auto, par exemple, comportent cependant un irréductible degré d'imprécision : « *Le chiffre que nous envoyons n'est pas mensonger ou frauduleux mais il n'est pas exact...* » [Entretien 18]. En effet, il y a d'une part des erreurs dans le comptage des voitures par des capteurs et de l'autre, ces données issues des capteurs, sont rigoureusement retraitées par des responsables des parkings. C'est une opération complètement manuelle qui dépend des pratiques de chaque agent [Entretien 18]. Par ailleurs, chez certains producteurs des données, un besoin d'outils internes permettant de mieux exploiter et valoriser des données produites se fait sentir : « *On a un très haut potentiel de la production des données et une très faible capacité d'analyse, c'est compliqué par le fait que les facteurs d'analyse sont extrêmement nombreux et complexes* » [Entretien 18].

2. Pour les gestionnaires des portails Open data, la tâche la plus importante est la recherche de la donnée, que ce soit auprès des autres services de la collectivité territoriale ou bien auprès de fournisseurs privés. Elle nécessite un temps, non négligeable, passé à ce que plusieurs interlocuteurs ont appelé un travail « d'évangélisation » [Entretiens 2, 6]. La donnée n'est jamais transformée à ce stade, tout au plus est-elle nettoyée (rendue compréhensible, en particulier pour les en-têtes et déboguée). Le fait que la donnée ait eu, à un moment donné, une valeur pour les services métier de la Métropole suffit : « *donc on la met à disposition sous cette forme-là. La donnée n'est que rarement retouchée* » [Entretien 2]. Ce souci de conserver les données telles que remises par les services satisfait une demande des développeurs, mais reste problématique, nous le verrons ensuite. Tout au plus pour le gestionnaire du portail est-il question de la rendre plus attractive pour améliorer leur consultation et générer du trafic. Le travail du traitement et nettoyage des données, avant leur mise à disposition, sur le portail est compliqué et lourd mais nécessaire : « *Les données ne peuvent pas*

être mises telles qu'elles car elles ne vont pas satisfaire les besoins des ré-utilisateurs. Ces données doivent être complétées et mises en forme » [Entretien 12].

Les gestionnaires des portails OD mènent aussi une réflexion sur les modalités d'accès aux données publiées, notamment sur la mise en place d'une redevance concernant les données en temps réel : *« Parce que d'un côté, il y a l'obligation, la loi sur la mise à disposition des données en temps réel et à la fois, il y a des ressources quand même derrière, ce sont les coûts pour la collectivité. On est obligé de se dire comment je fais ? Soit j'ai des protocoles d'accès et dans ce cas-là je peux faire payer à partir d'un certain niveau de demande ou bien je n'ai pas de protocole d'accès et tout est libre...Ça, ce sont des postures politiques. » [Entretien 12].*

Les formats développés seront *« démultipliés »* pour toucher le plus de gens possibles, que ce soit les citoyens ou les développeurs chevronnés. Par contre, les données ne sont jamais croisées entre elles : faute de temps mais surtout parce que les résultats pourraient être sensibles et ne seraient pas validés ensuite. Ce souci d'ouverture maximale du portail invite alors à transformer certains fichiers géographiques propriétaires, lisibles uniquement par des logiciels spécialisés et généralement coûteux, en des fichiers lisibles, par exemple sur *Google Earth* [Entretien 6]. Il s'agit en effet dans cette optique d'avoir *« des fichiers qui ne créent pas de barrière à l'entrée »*. Les formats les plus cités sont *« ODS et KML pour le commun des mortels, formats PDF et XML et formats plus spécifiques liés à des métiers : XLS, ODS, CSV pour la version à plat, pour les versions géographiques : KML, Shape, Geojson, et des solutions database : DWG pour les plans d'urbanisme ou le cadastre sous Autocad. »* Le choix des formats relève lui aussi des logiques politiques et économiques : *« On le voit dans le cas des données liées aux transports, si tout le monde utilise GTFS, Google bouffe tout... Si on utilise autre chose, chaque développeur doit réinterpréter les données de tous les réseaux. C'est un avantage et un inconvénient » [Entretien 12].*

3. Les analystes et les chefs des projets sont amenés à historiciser, analyser et évaluer les données existantes, à réaliser des comparaisons nationales et à faire des recommandations [Entretiens 11, 16]. De fait, ils accordent une grande importance à la fiabilité et la qualité de la donnée : *« J'aime bien savoir ce qu'il y a dedans et quels ont été les traitements, quelle est la portée de ce qui a été fait » [Entretien 16].* Ayant une expertise historique dans le domaine de la mobilité, les analystes interrogés s'opposent à un autre type de ré-utilisateur des données ouvertes qui sont issus du numérique et abordent les questions de la mobilité par la donnée sans forcément se poser des questions sur les manières dont cette donnée a été produite ou traitée. Or, il existe une difficulté réelle à bien exploiter les données qui ne sont pas conçues ou prédestinées pour une utilisation liée à la mobilité : *« Il y a peu de systématisme dans les données de téléphonie mobile pour pouvoir alimenter des modèles de trafic car il s'agit des données issues de l'exploitation du réseau de téléphonie et non pas des données prédestinées à l'étude de la mobilité » [Entretien 16].* L'utilisation

détournée de ces données pose la question de la fiabilité des résultats : « *On peut très vite faire des nuages des points, on peut très vite faire de belles cartes, mais elles sont probablement assez fausses ou faussées, on ne sait pas encore exactement quels biais sont là, est-ce que le biais est systématiquement sous-compté ou sur-compté etc.* » [Entretien 16].

Les analystes s'adaptent à tout format de données, cependant ils ont besoin de la continuité de celle-ci dans le temps et de la prédictibilité de son évolution, aussi bien par rapport à ses modes de production que par rapport à son prix lorsqu'il s'agit des données privées. D'une manière générale, un meilleur référencement et une meilleure homogénéisation des données liées à la mobilité au niveau national pourraient contribuer à une meilleure exploitation des données ouvertes, surtout lorsqu'il s'agit de croiser la thématique transport/mobilité avec d'autres jeux de données (travaux en cours, interventions d'urgence, horaires des cinémas, etc.). [Entretien 11].

4. Pour les développeurs, le portail constitue un « *magasin de pièces détachées* » [Entretien 5]. En ce sens, les développeurs recherchent et utilisent des données accessibles selon une ergonomie « *simple et limpide* » [Entretien 1]. Ils ont besoin d'éléments clairs sur ce qu'on peut faire avec les données, sur les manières de les télécharger, des informations sur les formats (et leur contexte de création), les métadonnées, les fréquences de mises à jour, une notice explicative de la réutilisation et d'être assuré de la précision de la donnée. Le contexte de la création de la donnée est nécessaire à son interprétation et lui seul permet d'en comprendre les limites: « *L'usage qu'on veut faire de la donnée n'est pas celui pour lequel la donnée a été produite. Toute donnée publique qui est présentée, elle a été produite pour répondre à un problème particulier. Il faut remonter à l'intention originelle du producteur de la donnée pour pouvoir comprendre les contraintes qu'il a eues pour la produire et dans quel esprit il l'a fait. C'est très, très important pour qu'une donnée ait une valeur* » [Entretien 7].

Toutefois, certains développeurs utilisent les données ouvertes sans forcément se poser de questions sur la méthodologie du traitement de la donnée avant sa publication : « *A partir du moment où je peux me dire que ces données sont relativement bonnes à la rigueur je m'en moque. Si un jour on voit que les données ont un écart ou alors certaines données sont supprimées pour des raisons diverses et variées, à ce moment-là on sera intéressé d'aller au plus près de la source. Mais actuellement, ce n'est pas le cas... Ils font leur travail. Pour moi, c'est une sorte de boîte noire, je récupère les données, pour l'instant cela me va* » [Entretien 17].

Le portail des données ouvertes de la SNCF, data. sncf.com, hébergé chez la start-up Opendatasoft est un bon exemple l'horizon d'attente car il propose à la fois les horaires des trains (théoriques et temps réel), les services disponibles en gares, leur localisation, et les comptages de voyageurs. En plus, « *à partir de n'importe quel jeu de données tabulaire, il vous génère une API pour accéder directement au micro service pour accéder directement à la donnée qu'il vous faut* »

[Entretien 1]. En même temps, l'API de SNCF est limité à 1000 appels par jour ce qui amène les développeurs des start-up à surveiller de près leur consommation et à négocier avec les responsables des données s'il y a un dépassement. S'il y avait un accès direct aux fichiers, sans passer par une API, ils pourraient faire un nombre infini de requêtes [Entretien 17]. Certains développeurs restent souvent à une logique de démonstration de ce qu'il est possible de faire avec des données ouvertes, assez semblable aux outils ou services développés dans le cadre des Hackathon. Ces développeurs mettent en place alors « *des petits outils pour masher, revalider, et comparer des données au sens global* » [Entretien 1], principalement dans le domaine de la data géolocalisée ou des transports. Ils ne trouvent pas de données suffisamment précises ou fines pour qu'elles puissent être ensuite croisées avec d'autres données [Entretiens 1, 7]. Ce souci de complétude de la donnée est toutefois assez ambivalent : une donnée incomplète est préférable à l'absence de données. L'idée est alors de renvoyer vers la communauté OD pour avoir des feedbacks et ainsi corriger ou compléter la donnée. Il s'agit là d'une position plus militante, qui n'est pas nécessairement partagée.

Il est rare d'avoir des données qui puissent être réutilisées directement. Par exemple, certaines données tabulaires n'ont pas d'identifiant unique : « *on a des champs dont l'identifiant varie dans le temps, donc on est obligé de remapper tout ça et de faire attention* » [Entretien 7]. Avant de basculer des données sur un portail commun tel que Data Grand Lyon, « *il y a une réflexion pour établir des identifiants inter-service* » [Entretien 4]. Inversement, pour intégrer des données du portail Data Grand Lyon dans l'intranet d'une commune, la compatibilité entre l'application et le portail doit se vérifier, et ce n'est pas forcément le cas car la couche applicative de l'intranet des communes n'est pas toujours faite pour lire des données en temps réel. Ce décalage impose un traitement intermédiaire (manuel pour le moment) afin de régénérer des identifiants compatibles et de pouvoir intégrer les données en temps réel provenant du portail Data Grand Lyon. La solution, pour rendre le système automatisable, serait d'acquérir un ETL (*Extract Transform Load*)¹ [Entretien 4]. Ceci dit, lorsqu'il s'agit du temps réel, obtenir une corrélation entre donnée théorique (la fiche horaire, la planification par exemple pour un transport en commun) et le temps réel reste difficile : « *souvent les deux ne sont pas reliés informatiquement* » [Entretien 5].

Une autre difficulté de l'exploitation des données dynamiques est liée aux décalages entre les flux récupérés et les champs libres concernant les données routières : « *Il faut que le champ libre nous dise la même chose que ce qui est envoyé dans le flux techniquement, que le commentaire corresponde aux données techniques, qu'il n'y ait pas d'équivoque. Aujourd'hui on n'intègre pas les*

¹ Extract-Transform-Load est connu sous le terme ETL, ou extracto-chargeur, (ou parfois : datapumping). Il s'agit d'une technologie informatique intergicielle (comprendre middleware) permettant d'effectuer des synchronisations massives d'information d'une source de données (le plus souvent une base de données) vers une autre. Selon le contexte, on est amené à exploiter différentes fonctions, souvent combinées entre elles : « extraction », « transformation », « constitution » ou « conversion », « alimentation ».

champs textuels, après si on se rend compte demain que le champ textuel correspond plus à la réalité on va chiffrer la complexité, le traiter et si ça apporte une plus-value on fait, mais c'est difficile à quantifier » [Entretien 22]. L'un des développeurs rencontrés conçoit des métadonnées en « *mixant des jeux de données différents qui vont générer d'autres jeux de données pour obtenir quelque chose de relativement pérenne* » [Entretien 1].

En ce qui concerne l'utilisation de jeux de données privées, ils sont superposés et comparés pour générer de nouveau(x) jeu(x) [Entretien 1]. Jeu(x) difficile(s), voire quasi impossibles à obtenir, sauf dans le cadre de licences engagées (sur le portail GL). Ces licences sont le fruit de négociations toujours facilitées lorsque le développeur d'applications propose une application commerciale (par exemple : tournées de livraisons). Mais ces licences engagées posent un problème majeur aux développeurs car elles ne s'inscrivent pas dans l'idéologie portée par l'OD. En effet, elles donnent un droit d'audit des systèmes d'information du ré-utilisateur, ce qui est inacceptable pour certains des développeurs et chefs de projets du secteur privé [Entretiens 1, 7]. D'ailleurs, le problème se (re)-pose ensuite lors de la commercialisation de ces applications [Entretien 7].

Les développeurs expriment un engagement fort envers l'ouverture des données et se soucient beaucoup de la mise à disposition de leurs propres données via *OpenStreetMap* [Entretien 14] ou *Data.gouv.fr* [Entretien 17]. « *Nous ce qu'on essaie de faire, c'est de travailler beaucoup avec OpenStreetMap parce qu'on croit beaucoup à ce projet-là, cela pourrait être finalement une base commune, vraiment commune des données ouvertes. Aujourd'hui la grande difficulté avec tout open data c'est finalement la mise à jour, de quand datent ces données, est-ce que ce sont les dernières, est-ce qu'on va pouvoir les croiser avec d'autres données. Or si tout le monde faisait l'effort aujourd'hui de pousser tout cela vers OpenStreetMap, c'est vrai que cela permettrait d'avoir une base des données commune et cohérente. Nous, en interne, c'est vraiment ce qu'on essaye de faire, garder finalement moins des données chez nous et d'avoir le plus des données possibles sur OpenStreetMap parce que c'est plus facilement maintenable* » [Entretien 14].

En fait, extraite de sa source, la donnée est tellement décontextualisée qu'elle en perd en partie son sens, pour être interprétée, et réutilisée, il faut pouvoir la replacer dans son contexte. La réutilisation dépend de la production amont.

4. DONNES MOBILITES ACCES / USAGES - JEUX DE DONNEES

A ce stade du travail, notre étude est focalisée sur les données mobilité accessibles sur le portail Data du Grand Lyon : data.grandlyon.fr.

4.1. LES DONNEES MOBILITE PRESENTEES DANS LE PROJET EMBLEMATIQUE OPTIMOD LYON

Le schéma ci-dessous (Fig. 1, page suivante) présente l'ensemble des données liées à la mobilité de la Métropole de Lyon qui ont été utiles au projet *Optimod'Lyon*, projet de recherche et projet commercial multi-partenaires permettant à la fois la création d'un référentiel et d'applications en mobilité².

Les données alimentant le référentiel sont de cinq types :

- les données de références topologiques de chacun des services et réseaux ;
- le référentiel géographique, qui est la pièce indispensable permettant de connecter les données de référence de chaque service et réseau ;
- les données temporelles théoriques ;
- les données temps réel ;
- les données prédictives.

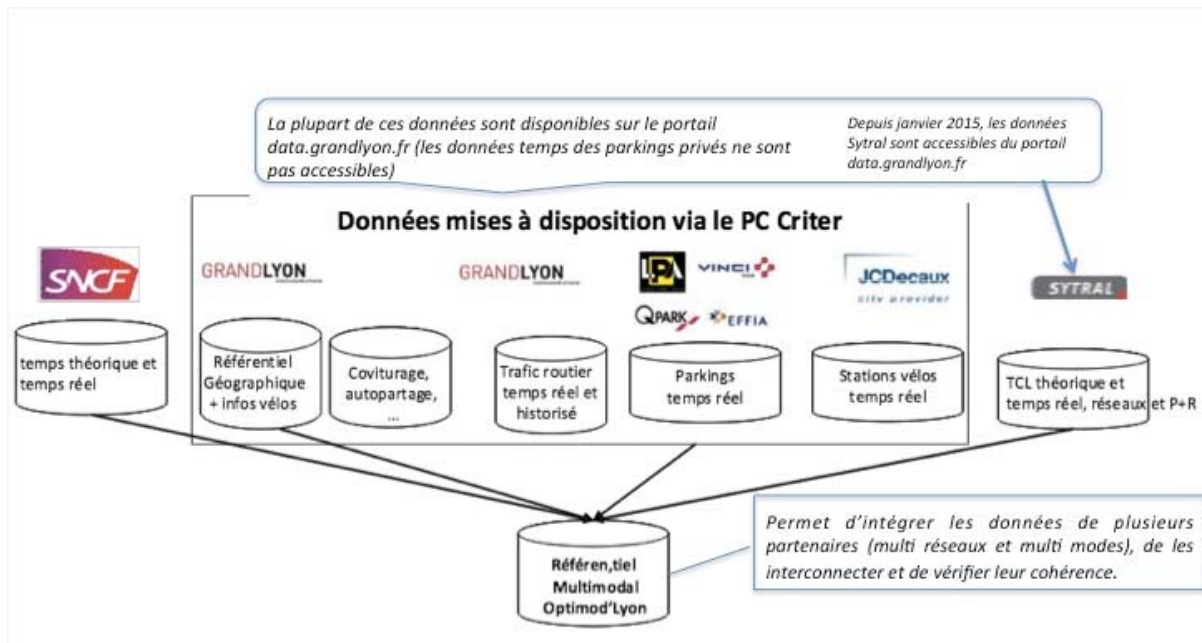


Fig. 1 : Schéma de principe des principales données du référentiel de données mobilité du Grand Lyon reproduit avec l'aimable autorisation de Jean Coldefy (2014).

Ces données mobilité comportent à la fois des données publiques, ouvertes et privées. L'ensemble des jeux de données disponibles à la fois sur la plateforme informatique et sur le portail sont regroupés selon le mode de transport :

- les données de transports collectifs ;
- les données sur les vélos et pistes cyclable ;

² Pour des compléments concernant ces projets : www.optimodlyon.com

- les données routières et de voirie ;
- les données de stationnement ;
- les données relatives aux autres modes de déplacements.

Le référentiel est une base de données intégrée regroupant toutes les informations cartographiques, théoriques, temps réel et prédictives. Il est accompagné des outils de gestion spécifiques (contrôle des flux, visualisation des données en temps réel) qui ont été développés spécialement pour ce projet. Le référentiel vise à intégrer et à mettre en cohérence des données de transport, existantes mais disparates, sur un référentiel cartographique commun dans une optique d'interopérabilité³. Le référentiel multimodal mis en place comporte ainsi 13 flux temps réel et 15 bases de données connectées. Le volume de données dépasse les 20 millions d'enregistrements par jour [Entretien 5].

4.2. LES USAGES ET LES SOUHAITS DES INTERVIEWES EN MATIERE DE MOBILITE

Selon que l'on interroge des producteurs ou des ré-utilisateurs de données les intérêts divergent, malgré des données communes ou partagées. Cependant, le souhait récurrent des développeurs et des analystes des bureaux d'étude porte sur la codification homogène et systématique des données de la mobilité des villes différentes. Il s'agit aussi bien des jeux de données disponibles que des formats ou des descriptions des données [Entretiens 11, 14, 16, 17, 22]. Plusieurs interviewés aspirent à la mise en place d'un outil commun des données ouvertes liées à la mobilité au niveau national, voire européen [Entretiens 11, 17]. D'une manière générale, la continuité et la prédictibilité de l'évolution de la donnée dans le temps sont très importantes pour tous les ré-utilisateurs des données ouvertes. Les différents types de licences proposés par certains portails métropolitains d'OD sont parfois vus comme une réelle contrainte qui obligerait les ré-utilisateurs à se réengager tous les ans [Entretiens 16, 22].

« Les données en temps réels sont des données très intéressantes pour les applications et pour tout ce qui est mobilité. Pour le commun des mortels, c'est le GPS et la météo. Mobilité, c'est transport, parking libre, vélo disponibles et place pour les déposer, le retard des transports, le prochain passage du transport dans la réalité. Ce sont des choses très intéressantes que les gens nous demandent pour pouvoir régler leur vie courante, c'est bien de savoir le temps réel mais aussi l'historique, par exemple que le samedi il y a embouteillage à tel endroit. De l'information en temps réel qui est historicisée, qui permet de prendre des décisions, de prévoir. Ces données-là c'est de la

³ L'interopérabilité est la capacité que possède un produit ou un système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre (source : <http://definition-interoperabilite.info/>)

compétence métropolitaine, en même temps ce sont des données difficiles à gérer, ça demande de gros volumes de données, des connections entre machine en temps réel, des contraintes de la chaîne technique. Il y a un aspect économique très important, c'est que ces services-là sont opérés par des opérateurs dans le cadre d'une délégation du service public, les parkings par Vinci, Eiffage, l'eau par Véolia, etc. Ces acteurs-là n'ont pas du tout envie de publier ces données en OD d'une part parce que on peut observer ce qu'ils font par exemple si le tram a du retard tout le temps, ils n'aiment pas du tout donner de l'information qui risquerait de leur retomber sur la tête, parce que quelque part c'est donner à observer la qualité de service qu'ils devraient assurer, l'autre argument étant que l'on donne des informations pour créer des services concurrents » [Entretien 6] .

4.2.1. LES TYPES DE RE-UTILISATION DES DONNEES MOBILITE

Les données mobilité sont utilisées par nos interviewés dans plusieurs cas :

1. Pour développer des applications.

Nous avons présenté dans la partie précédente *Optimod'Lyon*, une première application utile pour développer d'autres applications. *Optimod'Lyon* rassemble en une seule application les informations théoriques et temps réel sur tous les modes de transport de l'agglomération lyonnaise : bus, car, train, avion, vélo, voiture, parking... Alimentée par tous les opérateurs de transport de l'agglomération, sous l'égide du Grand Lyon, elle permet de trouver des solutions pour tous les trajets des habitants de l'agglomération et des touristes et de bénéficier d'informations mises à jour en temps réel 24h/24h [Entretien 3]. Les données suivantes du portail Data Grand Lyon participent à l'application : état du trafic en temps réel, événement trafic en temps réel, jeu de données orthophoto, fonds de plan d'*OpenStreetMap* et de *Google* [Ibid.]. Développée par Cityway⁴, *Optimod'Lyon* est la première application qui a obtenu l'attestation *Onlymoov*⁵.

L'application *Smart deliveries* développée par le service recherche d'IBM permet d'optimiser la tournée des livreurs. Elle requiert plusieurs jeux de données suivants provenant du portail Data Grand Lyon : données trafic, travaux, carte des détecteurs, carte des marchés et les données temps

⁴ Cityway est une société de services créée en 2001 et spécialisée dans les technologies de l'information appliquées au domaine de la mobilité : sites internet, calculateurs d'itinéraires, suivi des véhicules, systèmes billettiques, centres de relation clients, gestion de pôle d'échanges. Elle est également réputée pour sa contribution aux démarches de normalisation et de partage de données transport à l'échelon européen. (<http://www.cityway.fr/>). L'entreprise est filiale du groupe Transdev, l'un des leaders mondiaux des services de mobilité, membre du groupe Caisse de Dépôts, groupe public œuvrant au service de l'intérêt général et du développement économique.

⁵ Onlymoov' est un service de la Métropole de Lyon, destiné aux grands lyonnais mais également aux personnes en transit sur l'agglomération, empruntant au cours de leurs déplacements le réseau routier, les transports en commun ainsi que les modes doux.

réels du PC Criter [Entretien 7]. Les bases de données des agences d'urbanisme et celles du STIF⁶ sont aussi utilisées par IBM pour d'autres projets de recherche, notamment pour une application *TPT* qui prédit le trafic en fonction des incidents détectés.

À Montpellier, nous avons recensé trois applications : *Mobiliti*, une application d'itinéraire destinée aux handicapés, *Smoove*, qui envoie des alertes transport, *Track'o'matic* qui signale les difficultés de déplacement et qui est destinée aux pouvoirs publics. Celles-ci fonctionnent avec un nombre de jeux de données restreints. Les données qui proviennent du portail Data Grand Lyon concernent les données TCL (horaires théoriques et prochains passages).

La société Here cite l'application *Heremaps*, logiciel de cartographie sur smartphone (*Android, Windows phone, ios*), elle utilise les données suivantes de Data Grand Lyon : horaires théoriques du réseau TCL, tronçon de la trame viaire (voies et adresses) et le jeu de données aménagement cyclable. Cette application va chercher des données sur le site du SITRA (informations touristique en Rhône-Alpes) et sur les sites open data régionaux.

La Compagnie des Mobilités a développé *Géovélo*, une application web et mobile d'aide à la navigation qui offre aux cyclistes urbains plusieurs services : recherche d'un itinéraire, information sur les aménagements cyclables, stations vélo, disponibilités des places. L'application utilise les données du portail Data Grand Lyon concernent les aménagements cyclables et les données en temps réel de la disponibilité des places dans les stations Vélo'V (une API de JC Decaux). *Géovélo* fonctionne à Tours, à Paris, à Lyon, à Caen, à Toulouse, à Rennes et à Nantes.

La start-up Tayo développe *Mytayo*, une application multimodale et collaborative d'aide à la mobilité à destination des professionnels. Elle permet de créer un plan d'accès multimodal pour un événement professionnel et de le partager auprès des invités *via* un *Doodle*, de planifier un trajet et de mettre en relation les invités. L'application utilise les données TCL issues du portail Data Grand Lyon, notamment les jeux de données concernant bus, métro, tram et parkings, et les croise avec les données issues d'*OpenStreetMap*.

L'application *Nantes dans ma poche*, réalisée en mai 2015 dans le cadre du programme Smart City, vise à faciliter le quotidien des habitants de Nantes Métropole en les informant de la qualité de l'air, des transports en commun, des horaires des piscines, du menu des écoliers... Pour cela, l'application utilise les données statiques et les données en temps réel issues de la plateforme des données ouvertes de la ville.

Afin de faciliter l'accès des clients aux taxis, le registre des données *LeTaxi.fr*, développé au sein d'Etalab permet de connaître la disponibilité et géolocaliser les taxis à l'échelle nationale. Il est

⁶ Le STIF fédère tous les acteurs (voyageurs, élus, constructeurs, transporteurs, gestionnaires d'infrastructures...) et travaille à améliorer le service rendu aux voyageurs.

disponible *via* tous les navigateurs Internet et plusieurs applications liées à la mobilité aussi bien pour les chauffeurs des taxis que pour les clients.

Filiale de Michelin, ViaMichelin conçoit, développe et commercialise des produits et services numériques d'aide au déplacement et à la mobilité, notamment le site web et l'application mobile *ViaMichelin*, le premier destiné à la préparation au voyage, la deuxième étant un *GPS* avec des alertes trafics et des fonctionnalités communautaires (une personne en mobilité peut déclarer un incident sur la route).

2. Pour intégrer les données dans un intranet.

C'est le cas de la ville de Chassieu qui intègre des données qui ne sont pas orientées sur la mobilité [Entretien 4]. En effet, Chassieu a développé son propre intranet pour les interventions sur les espaces verts municipaux, ce malgré l'unique utilisateur (le chef de services des espaces verts) auquel elle est destinée.

3. Pour intégrer les données dans une base des données nationales.

Le Cérema utilise les données ouvertes liées aux transports collectifs (horaires de TCL) pour nourrir une base nationale des arrêts des transports collectifs (ATC) qui vise à recenser et rediffuser l'ensemble de l'offre des transports collectifs disponibles avec les arrêts dans un format homogène et standard. Cette base offre des données concernant les populations autour des arrêts, nombre de passage des transports, premier et dernier passage des transports. Les données sont disponibles à un nombre restreint des usagers professionnels travaillant avec le Ministère des Transports et les différents services publics. Par ailleurs, le Cérema contribue à la constitution de l'annuaire des offres de transport et services d'informations liées à la mobilité, à l'instar du site web PASSIM (<http://www.passim.info/>). Au départ destiné à un large public (citoyens, journalistes, étudiants, professionnels), ce site web est aujourd'hui largement méconnu et sous-utilisé.

4. Pour réaliser les études liées à la mobilité.

Certains analystes dans le domaine de la mobilité recherche, par exemple, des données sur les stations TCL, les parkings relais, les stations Velo'v, les plans mode doux, les aménagements cyclables. Ils sont aussi amenés à utiliser les données d'IAU IDF⁷, les jeux SDRIF⁸, les projets de transport, les projets d'aménagement. Certaines études nécessitent de recourir aux données de comptage routier du portail Data du Grand Lyon, même si elles ne font pas partie de données ouvertes. D'autres analystes étudient les données des différents portails métropolitains pour relever les données en lien avec le transport.

⁷ IAU-IDF : Institut d'urbanisme d'Ile de France

⁸ SDRIF : Schémas directeur d'Ile de France

5. Pour s'informer et expérimenter.

Les chargés de mission et les analystes font un travail important de veille. Ils consultent tous les jeux de données généraux et ceux orientés transport temps réel du portail Data du Grand Lyon ; ainsi que les jeux de données d'*OpenStreetMap*, ceux d'acteurs publics comme Etalab ou privés comme la SNCF.

4.2.2. LES QUALITES ATTENDUES DES DONNEES DYNAMIQUES EN MOBILITE

Les données de mobilité et notamment les données dynamiques sont considérées comme stratégiques par l'ensemble des interviewés, mais elles n'ont pas encore les qualités attendues par ces ré-utilisateurs.

Les données mobilité attendues sont multimodales [Entretien 3] et concernent la circulation (encombrement/travaux), la disponibilité (vélo, place de parking...), les historiques et les prévisions pour anticiper les problèmes [Entretien 6]. Les historiques supposent un archivage des données qui n'est pas toujours possible du fait des capacités de stockage. La mise à jour des données de flux doit être garantie [Entretien 3]. Les données dynamiques sont produites avec une fréquence élevée qui varie en fonction de type de la donnée. Cela peut aller de quelques minutes pour la disponibilité des équipements jusqu'à peu près 15 secondes pour les prochains passages des véhicules. Il est important de réduire le laps de temps entre la collecte ou l'observation de l'information sur la route et sa diffusion à l'utilisateur : « *Typiquement : un véhicule à contre-sens, il faut vite le remonter aux personnes pour l'indiquer..., c'est là justement qu'on dit que la donnée doit être dynamique pour les cas critiques...alors que 10 minutes pour la mise en place d'un chantier, ça ne nous dérange pas* » [Entretien 22].

Nos interviewés aimeraient que les données de mobilité en temps réel soient accessibles gratuitement, ce qui est rarement le cas. Elles sont le plus souvent accessibles à l'aide d'une licence associée [Entretien 6]. Le coût élevé de certaines données, qu'elles soient ouvertes ou privées est considérée comme une restriction du potentiel créatif [Entretien 16]. Les interviewés aimeraient que les données de la mobilité en temps réel soient associées à une localisation d'où la nécessité de les fournir dans un format compatible avec des formats SIG (Shapefile) et dans des formats GTFS [Entretien 9], Datex CSV [Entretien 3] et Neptune [Entretien 19]. D'une manière générale, les fichiers aux formats GTFS ou Neptune sont considérés comme proposant des données plus complètes [Entretien 13]. « *La complexité de certains formats et la structuration métier de certaines données les rendent bien souvent inintelligibles au citoyen* » [Entretien 5].

Pour le moment, il manque des restitutions intuitives de données dynamiques en temps réel. Un accès aux données dynamiques sur toutes les plateformes serait souhaitable, ainsi qu'un accès simplifié aux API.

Un focus sur les formats en mobilité pour les données statiques et temps réels

Concernant les transports publics urbains, nous pouvons rencontrer des données dans les formats les plus simples (tel que CSV, XLS), ou dans des formats utilisés sous système d'informations géographiques (SHP, KML/ KMZ, TAB, ce qui requiert déjà un minimum de connaissance sur les SIG). Les données d'horaires de transport collectif nécessitent des formats plus complexes. Elles sont diffusées selon le format standard GTFS (qui peut être utilisé pour développer des applications pour smartphone) ou selon la norme Neptune. (Chevereau, 2015, p. 33).

Dans la mesure du possible, il est recommandé de doubler ce format GTFS par un format plus complet que sont les normes Neptune et NeTeX9, permettant des fonctionnalités plus larges (Chevereau, 2015, p. 44)

Pour des services vraiment innovants utilisant des données en temps réels, une standardisation des formats est nécessaire. Cette standardisation peut se faire en utilisant la norme SIRI-Lite, ou un autre format unique. La première solution nécessite un travail de communication et d'accompagnement auprès des collectivités sur les possibilités offertes par SIRI-Lite et sur la manière concrète de l'utiliser. La deuxième solution nécessite une mise en synergie des collectivités impliquées dans la diffusion de ce type d'information en temps réel, afin de choisir un format standard, sur la base des potentialités et de la facilité pour les collectivités d'exporter leurs données dans ce format. Dans ce cas, le choix d'un API est recommandé, plutôt qu'un fichier complet qui serait mis à jour très régulièrement (toutes les minutes par exemple). (Chevereau, 2015, p. 48).

Les pratiques professionnelles sont très fortes et diversifiées. « *Chaque communauté a ses formats favoris. C'est pour ça qu'on démultiplie les formats pour pouvoir répondre à divers types de réutilisations. Chaque communauté a ses formats favoris. On a toujours un fichier Excel, un fichier ODS pour le commun des mortels, des fichiers KML pour qu'ils puissent le glisser dans Google Maps ou Google Earth, et après des formats plus spécifiques liés à des métiers particuliers. En gros pour la version à plat XLS, ODS, CSV. Pour la version géographique : KML, Shape, Geojson, et après on a parfois des solutions de database : DWG pour du autocad (pour le plan d'urbanisme, ou le cadastre)*

⁹ NeTeX (Network Exchange) est un format de référence pour échanger des données d'offre théorique du transport collectif, défini au niveau européen.

on a même du PDF texte pour les délibérations du conseil municipal. On essaie au maximum d'éviter le PDF image. On doit avoir aussi du XML [Entretien 2].

Lancé en 2013, le projet *Opticities* est le prolongement à l'échelle européenne d'*Optimod'Lyon*. Réunissant 25 partenaires¹⁰ de 8 pays différents, ce projet vise à améliorer en 3 ans des transports urbains à travers l'expérimentation des solutions ITS (Intelligent Transport System) interopérables dans 6 villes européennes. Le projet vise les objectifs suivants :

- Industrialisation des process,
- Intégration du covoiturage,
- Continuité GPS voiture – smartphone,
- Régulation de trafic urbain,
- Coordination des modes de transport,
- Diffusion d'informations ciblées, temps réel, tous modes,
- À destination des particuliers et des acteurs du fret urbain.

Toutes les villes ont plus ou moins un portail. Les données sont très éclatées. Il faut « *recréer les connecteurs d'une ville à l'autre, ça peut être très compliqué* » [Entretien 5].

5. PISTES POUR DES SERVICES PERTINENTS

Nos interviewés n'avaient pas de réponse à cette question, et la majorité d'entre eux l'a contournée en abordant ses propres applications, celles qui ont été développées et celles à développer. « *Avec la plateforme on a des données, donc de la matière pour travailler. La plus grosse difficulté des projets innovants, ce n'est pas le développement lui-même que d'avoir des utilisateurs d'un côté et des données à exploiter de l'autre. Si vous supposez à l'avance, que quelqu'un va vous fournir des données de parkings, l'emplacement de parking, que vous développez un système et qu'ensuite vous vous demandez qui va vous fournir les données, là il va y avoir un décalage. Les hypothèses que vous avez faites ne sont pas nécessairement égales à ce que l'on sera capable de vous fournir. La recommandation du projet réussi. D'abord on a les données, et ensuite on voit ce que l'on fait avec. Enfin, c'est une bonne façon !* » [Entretien 7].

Les services liés à la mobilité développés par nos interlocuteurs sont les suivants :

- Réalisation de navigateur multimodal en temps réel (c'est déjà effectif pour le Grand Lyon avec *Optimod'Lyon*) : calcul intermodal en temps réel couplant des modes individuels et

¹⁰ Parmi les partenaires : Lyon, Madrid, Turin, Göteborg, Wrocław, Birmingham, Spie, Cityway, Hacon, Institut Vedecom, UITP, Ertico Eurocities, etc.

collectifs, avec assistance au passager durant tout son trajet [Entretiens 1, 5]. Ce projet sera déployé par Cityway sur plusieurs villes en Europe dans le cadre d'*Opticities* (déploiement sur 6 villes européennes d'*Optimod'Lyon*). D'autres services liés à l'intermodalité sont également à l'étude [Entretien 1],

- Interfaçage avec des navigateurs embarqués automobile (suivi en situation de conduite) [Entretien 5],
- Planification d'un trajet multimodal et collaboratif à destination des professionnels [Entretien 16],
- Aide à la navigation destinée aux cyclistes [Entretien 14],
- Optimisation des parcours de livreurs. Il s'agit d'une application à visée de commercialisation auprès des fournisseurs de ce type de service [Entretien 7],
- Parking : savoir si on peut ou non se garer (parking de surface) [Entretien 18].
- Registre des données pour faciliter l'accès des clients aux taxis [Entretien 15].

BIBLIOGRAPHIE DU LIVRABLE

- Asdourian, B., Zimmerli, V. (2015), « *Open data*, usagers du numérique et entreprises publiques », *Revue française des sciences de l'information et de la communication*, n°6. En ligne : <http://rfsic.revues.org/1316>.
- Chevereau, L. (2015), (Dir.), *L'Open Data en collectivité à la lumière des données de mobilité*, Cérema, coll. Connaissances.
- Chignard, S. (2013), *Données brutes ou données contextualisées*. En ligne : <http://donneesouvertes.info/2013/05/02/donnee-brute-ou-donnee-contextualisee/>.
- Coldefy, J. (2014), « Optimod'Lyons ou la mise en synergie des acteurs publics et privés pour des villes plus intelligentes », *les Ecosystèmes de l'innovation, TEC*, n° 221, mars-avril 2014.
- Courmont, A. (2015), « Open data et recomposition du gouvernement urbain : de la donnée comme instrument à la donnée comme enjeu politique », *Informations sociales*, 5, n° 191, p. 40-50.
- Denis, J. & Goëta S. (2016), "*Brutification*" et instauration des données. *La fabrique attentionnée de l'open data*. i3 Working Papers Series, 16 – CSI – 01. En ligne : <https://hal.archives-ouvertes.fr/hal-01347301/document>.
- Paquienséguy, F. (Dir.) (2016), *Open data : Accès, territoires, citoyenneté : des problématiques info-communicationnelles*, Paris, Editions des Archives contemporaines.

Contexte :

Entreprise – domaine de compétence – depuis quand...

Fonction de la personne interrogée

1. Le portail métropolitain :

Quels jeux de données utilisez-vous ?

Pourquoi ces données – Data du Grand Lyon / généalogie de la pratique / de leur projet Combien de données dans le jeu ?

Depuis quand ?

Avec quel type de licence ?

Quels formats de données utilisez-vous en entrée ?

Quels formats de données utilisez-vous en interne ? (format pivot ?)

Quels formats de données utilisez-vous en sortie ?

Utilisez-vous de données en temps réel, si oui lesquelles ?

Que représentent pour vous les OD dynamiques ?

Autres sources :

Quelles données autres que l'open data utilisez-vous ? Lesquelles / quoi ?

Comment avez-vous choisi/trouvé les jeux de données ?

(Accès – Fournitures/dépôt aux données)

Manques / Soucis / freins du portail Data du Grand Lyon - données

Souhaits / attentes / vos projets à venir

2. Les applications développées :

Développez-vous des applications ?

Combien – lesquelles ?

Comment l'idée vous en est-elle venue ?

Quelles sont les fonctionnalités de l'application ? Quels objectifs ?

Pourquoi/comment développez-vous ces applications avec de l'open data ?

A quoi vous servent les données si vous ne produisez pas d'application ?

Quelle cible visez-vous ?

Démarche :

Que manquait-il dans les données ?

Quels traitements faites-vous avec les données ?

Qu'en est-il de la partie nettoyage ?

Notions :

Qu'entendez-vous par données consolidées ?

Que savez-vous des technologies du Web sémantique et du Web de données ?

Comment pourrait-on simplifier l'utilisation des données ouvertes ?

3. Les usagers :

Quels sont les usagers de votre/vos application(s) ?

Profils / caractéristiques ? Les connaître ?

En tenez-vous compte ? Quelle place ont-ils (expérimentations – tests)

Attentes identifiées ? connues ?

Interactions avec les usagers ? client ? (retour/échange)

1. Identité professionnelle

Nom :

Prénom :

Fonction :

Employeur :

Service de rattachement :

Profil de votre poste :

Depuis quelle date assurez-vous ces fonctions :

Mail :

Quelles sources de données ou d'applications utilisez-vous ?

Veillez choisir toutes les réponses qui conviennent :

- data.grandlyon.com ?
- autres sources ?

2. A propos de la plateforme data.grandlyon.com

Comment avez-vous connaissance des données de cette plateforme ?

Quel(s) jeu(x) de données utilisez-vous prioritairement et de façon récurrente ?

Pour le jeu de données 1 depuis quand ? (mois-année)

Combien de données exploitez-vous dans le jeu de données n°1 ?

Avec quel type de licence (jeu de données n°1)?

Veillez choisir toutes les réponses qui conviennent :

- Aucune licence
- Licence associée
- Licence engagée

Autre jeu de données ?

- Oui
- Non

Quel autre jeu de données utilisez-vous ?

Pour le jeu de données 2 depuis quand ? (mois-année)

Combien de données exploitez-vous dans ce jeu (jeu de données 2) ?

Avec quel type de licence (jeu de données 2) ?

Veillez choisir toutes les réponses qui conviennent :

- Aucune licence

- Licence associée
- Licence engagée

Autre jeu de données ?

- Oui
- Non

Quel autre jeu de données utilisez-vous ?

Pour le jeu de données 3 depuis quand ? (mois-année)

Combien de données exploitez-vous dans ce jeu (Jeu de données 3) ?

Avec quel type de licence (jeu de données 3) ?

Veillez choisir toutes les réponses qui conviennent :

- Aucune licence
- Licence associée
- Licence engagée

Quels formats de données utilisez-vous en entrée ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Format privilégié :
- Format fréquemment utilisé :
- Format parfois utilisé :

Quels formats de données utilisez-vous en interne (format pivot) ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Format privilégié :
- Format fréquemment utilisé :
- Format parfois utilisé :

Quels formats de données utilisez-vous en sortie ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Format privilégié :
- Format fréquemment utilisé :
- Format parfois utilisé :

Utilisez-vous des données en temps réel ?

- Oui
- Non

Lesquelles ?

Quels sont les 3 freins majeurs que vous rencontrez avec les données de la plateforme data.grandlyon.com ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Frein n°1 :
- Frein n°2 :
- Frein n°3 :
- Aucune difficulté

Comment vous informez-vous des nouvelles données déposées sur la plateforme data.grandlyon.com ?

Veillez choisir toutes les réponses qui conviennent :

- Flux RSS
- Visites régulières sur la plateforme
- Autre :

3. Autres sources de données ouvertes

Quelles sont les principales sources de données ouvertes que vous utilisez ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Source 1
- Source 2
- Source 3

Pour la source 1 quels sont les jeux utilisés ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Jeu le plus utilisé :
- Jeux parfois utilisés :

Pour la source 2 quels sont les jeux utilisés ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Jeu le plus utilisé :
- Jeux parfois utilisés :

Pour la source 3 quels sont les jeux utilisés ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Jeu le plus utilisé :
- Jeux parfois utilisés :

Que faites-vous avec ces données ?

Veillez choisir toutes les réponses qui conviennent :

- Vous les utilisez
- Vous les intégrez
- Vous les produisez (données propriétaires)

Les données que vous utilisez sont-elles dynamiques ?

- Oui
- Non

Toutes ?

- Oui
- Non

Les données que vous intégrez sont-elles dynamiques ?

- Oui
- Non

Toutes ?

- Oui
- Non

Les données que vous produisez sont-elles dynamiques ?

- Oui
- Non

Toutes ?

- Oui
- Non

Quelles sont les trois difficultés les plus fréquentes que vous rencontrez lors de l'utilisation des données citées précédemment ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Difficulté n°1 :
- Difficulté n°2 :
- Difficulté n°3 :

Quelles sont les trois difficultés que vous rencontrez lors de la production de vos propres données ?

Veillez choisir toutes les réponses qui conviennent et laissez un commentaire :

- Difficulté n°1 :
- Difficulté n°2 :
- Difficulté n°3 :

Comment avez-vous eu connaissance de ces données ouvertes ?

4. Applications développées exploitant des données ouvertes dynamiques

Développez-vous des applications exploitant des données ouvertes dynamiques ?

- Oui
- Non

Combien d'applications avez-vous développées ?

- 1
- 2
- 3
- 4
- Autre valeur :

Nom de l'application n°1

Veillez écrire votre (vos) réponse(s) ici :

- nom :
- fonctionnalité centrale :
- cible(s) visée(s) :

Nom de l'application n°2

Veillez écrire votre (vos) réponse(s) ici :

- nom :
- fonctionnalité centrale :
- cible(s) visée(s) :

Nom de l'application n°3

Veillez écrire votre (vos) réponse(s) ici :

- nom :
- fonctionnalité centrale :
- cible(s) visée(s) :

Nom de l'application n°4

Veillez écrire votre (vos) réponse(s) ici :

- Nom :
- fonctionnalité centrale :
- cible(s) visée(s) :

Vous diriez que vous connaissez les attentes de vos utilisateurs/clients :

Veillez sélectionner une seule des propositions suivantes :

- Pas du tout
- Peu
- Assez bien
- Bien
- Très précisément

5. Données liées à la mobilité

Les données liées à la mobilité vous paraissent-elles stratégiques pour votre activité/mission ?

- Non
- Rarement
- Parfois
- Souvent
- Oui

Donnez trois mots-clés caractérisant les données ouvertes dynamiques :

Veillez écrire votre (vos) réponse(s) ici :

- Mot-clé n°1 :
- Mot-clé n°2 :
- Mot-clé n°3 :

Donnez deux mots-clés liés à la mobilité :

Veillez écrire votre (vos) réponse(s) ici :

- Mot-clé n°4 :
- Mot-clé n°5 :

Quelles qualités attendez-vous des données liées à la mobilité ? (trois réponses maximum)

Veillez choisir toutes les réponses qui conviennent :

- Mise à jour en temps réel
- Quantité croissante
- Localisation
- Fiabilité technique
- Format particulier
- Gratuité
- Relation avec le fournisseur
- Autre qualité :

Quel format particulier ?

Selon vous quelles sont les trois caractéristiques dominantes des données dynamiques liées à la mobilité ?

Veillez écrire votre (vos) réponse(s) ici :

- Caractéristique n°1 :
- Caractéristique n°2 :
- Caractéristique n°3 :

6. Dépôt de données sur la plateforme

6.1 Déposez-vous ou envisagez-vous de déposer des données sur la plateforme data.grandlyon.com ?

Veillez sélectionner une seule des propositions suivantes :

- Oui, j'en dépose
- Oui, j'envisage d'en déposer
- Non

6.2 Quel type de données déposez-vous ?

6.3 De quelle manière privilégiez-vous l'accès aux données ?

Veillez choisir toutes les réponses qui conviennent :

- Flux
- Téléchargements
- Autres :

Quel type de données envisagez-vous de déposer ?

ANNEXE 3. LISTE DES PERSONNES INTERROGÉES

Personnes ressources :

- Nathalie Vernus-Prost, Chief Data Officer de la Métropole de Lyon,
- Laurence Martinez, responsable de l'unité diffusion des données au service Information géographique, et gestionnaire du portail Data Grand Lyon,
- Patrick Gendre, Cérema, (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement), Ministère de l'Environnement, de l'Energie et de la Mer.

Producteurs :

- Directeur des parkings d'une métropole,
- Responsable du service trafic d'une métropole,
- Chef de projet dans une entreprise privée.

Gestionnaires des portails OD :

- Chef de projet OD dans une collectivité territoriale (300 000 habitants),
- Chef de projet dans une collectivité territoriale (300 000 habitants),
- Géomaticien/informaticien dans une collectivité territoriale (10 000 habitants),
- Responsable du service de promotion de la mobilité durable dans une collectivité territoriale (600 000 habitants).

Coordinateurs de projets :

- Chef de projet dans une entreprise privée,
- Directeur de la recherche dans une entreprise privée,
- Chargé de mission Innovation numérique dans une structure publique métropolitaine,
- Chef de projet dans une entreprise privée.

Développeurs :

- Développeur dans une entreprise privée,
- Développeur dans une entreprise privée,
- Développeur dans une start-up,
- Développeur dans une start-up,
- Développeur dans une structure publique nationale.

Analystes :

- Chargé de projet Information Multimodale et analyste dans une structure publique nationale,
- Chargé de mission et expert-analyste dans une structure publique nationale,
- Analyste dans un bureau d'études,
- Responsable Développement et Innovation en Economie des Transports et analyste dans un bureau d'études.