

Rapport remis au secrétaire d'État chargé des Transports,
de la Mer et de la Pêche



OUVERTURE DES DONNÉES DE TRANSPORT



Comité présidé par Francis Jutand

Mars 2015

SOMMAIRE

AVANT-PROPOS	3
PARTIE 1 : UNE STRATÉGIE POUR L'OUVERTURE DES DONNÉES TRANSPORT EN FRANCE	5
I. – L'ouverture des données	6
II. – Spécificités des données de transport	7
III. – Les préconisations et recommandations du Comité	9
IV. – Conclusions	12
PARTIE 2 : SYNTHÈSE ET RECOMMANDATIONS	13
I. – Synthèse	14
II. – Recommandations	18
PARTIE 3 : LE RAPPORT DU COMITÉ	27
Introduction	28
I. – Etat des lieux	30
I. 1. – Aspects techniques : données, classes de données et informations	31
I. 2. – État du droit : un cadre juridique inadapté à la réutilisation des données publiques de transport	35
I. 3. – Panorama des initiatives d'ouverture des données en France	42
I. 4. – Enseignements du panorama international	46
II. – Enjeux de l'ouverture des données de l'offre de transport	47
II. 1. – Ouverture des données de l'offre de transport : quels bénéfices ?	47
II. 2. – Ouverture des données de l'offre de transport : quels risques et quels remèdes ?	49
III. – Recommandations	53
III. 1. – Recommandations en matière d'ouverture des données et de conditions permettant de remédier aux risques liés à l'ouverture	54
III. 2. – Recommandations juridiques	62
III. 3. – Recommandations sur les dispositions techniques permettant de faciliter la réutilisation des données	65
III. 4. – Recommandations destinées à favoriser le développement de l'économie numérique dans le domaine du transport	69
III.5. – Recommandations transversales	70
CONCLUSION	75
PARTIE 4 : LES ANNEXES	77
Annexe 1 : Lettre de mission	79
Annexe 2 : Composition du Comité	81
Annexe 3 : Tableau comparatif des licences de réutilisation	85
Annexe 4 : Rappels sur la loi CADA	87
Annexe 5 : Rapport du sous-groupe juridique	89
Annexe 6 : Rapport du sous-groupe technique	95
Annexe 7 : Rapport du sous-groupe modèles économiques	119
Annexe 8 : Le cas des données relatives aux bornes de recharge des véhicules électriques	129
Annexe 9 : Position de la SNCF	131
Annexe 10 : Position de la FNMS	135
Annexe 11 : Position de l'UTP	139

AVANT-PROPOS

Le Ministre en charge des Transports m'a confié en avril dernier la mission de présider les travaux d'un Comité du débat sur l'ouverture des données relatives à l'offre de transport. Ces travaux ont pris place dans un environnement riche et évolutif.

À leur origine, ils s'inscrivaient dans le cadre d'un mouvement puissant pour l'ouverture des données publiques, enclenché par le Secrétariat Général pour la Modernisation de l'Action Publique et bénéficiant de l'activité dynamique de la mission Etalab. Ce mouvement a donné lieu au lancement de débats dans des domaines comme la santé, le logement et les transports ; il a été renforcé entre-temps par la nomination d'un Administrateur Général des Données de l'État.

Le cadre s'est ensuite enrichi avec le lancement d'un débat national sur le numérique, confié au Conseil National du Numérique, pour la préparation d'une loi-cadre sur le numérique, prenant notamment en compte la problématique des données et leur impact sur l'innovation, l'économie et la citoyenneté.

L'ouverture et l'usage des données de transport sont des questions débattues partout dans le monde et la Commission Européenne s'est emparée du sujet avec des actions déjà enclenchées dans le domaine routier et en préparation pour l'ensemble des modes de transport.

La diversification des modes de transport et leur mise en concurrence, facilitées par le développement des services numériques de calcul et d'assistance, peut bouleverser la chaîne de valeur, donner lieu à l'émergence de nouveaux acteurs d'intermédiation et de service, et impacter globalement les modèles économiques des opérateurs de transport.

Par ailleurs les données de transport recèlent des connaissances sur la mobilité, les modes de vie, la consommation, qui sont très riches et sources de valeur pour l'urbanisme, le tourisme, le commerce, les liens sociaux, etc.

Enfin le développement des compétences numériques des utilisateurs et des modes de coopération qu'elles permettent crée des opportunités nouvelles de contribution de la « multitude » (*crowdsourcing*) à l'amélioration de la qualité des transports et au développement de nouveaux services.

Un Comité d'une trentaine de participants a travaillé depuis mai 2014, avec le soutien actif et très compétent de l'équipe de l'AFIMB, pour répondre à la lettre de mission qui nous a été donnée pour guider les travaux.

Il faut remercier l'ensemble des participants, membres du Comité et personnes qualifiées auditionnées, pour l'esprit constructif qui a présidé aux travaux, la qualité d'écoute et la volonté de dialogue, qui ont permis une appropriation approfondie de la problématique, une évolution progressive des points de vue et l'élaboration de consensus ou de quasi consensus sur les points majeurs.

Les données de transport ont un potentiel de valeur d'usage très diversifié. Dans ce Comité, nous nous sommes concentrés sur l'information des voyageurs et les services associés, qui étaient au cœur de la lettre de mission, car ils comportent dans leur problématique les questions essentielles soulevées par l'ouverture des données.

Le rapport qui a été rédigé fournit des recommandations utilisables à court terme au travers de différents vecteurs légaux pour progresser significativement dans l'ouverture des données de transport ; mais il doit être considéré comme une étape, car de nombreuses évolutions se profilent pour l'utilisation des données à l'horizon de quelques années en France, en Europe et dans le monde, qui concernent le droit, les usages, les technologies, mais aussi les modèles économiques des entreprises et le rôle des utilisateurs.

Le rapport a été structuré de façon à répondre à plusieurs besoins et niveaux de lecture.

Dans la première partie, je propose, à la lumière des travaux menés dans le cadre du Comité et dans d'autres cadres sur le numérique et les données, une analyse globale de la problématique de l'ouverture des données de transport. Elle vise à décrire l'esprit des recommandations générales servant de cadre conceptuel pour les recommandations du Comité et à proposer une dynamique pour la poursuite du travail dans les prochaines années pour les acteurs de l'écosystème des transports. Elle se positionne donc en cohérence avec les travaux du Comité, mais va au-delà dans leur lecture et dans les horizons.

La deuxième partie réalise une synthèse des travaux menés dans le cadre du Comité et donne la liste des recommandations précises ayant donné lieu à consensus, qui pourront notamment servir de base aux évolutions légales.

La troisième partie rend compte de l'ensemble des débats et présente les travaux menés par le Comité sur les principaux modes de transport, ainsi que le travail et les conclusions des trois sous-groupes : juridique, technique et économique.

Francis Jutand
Président du Comité du débat



PARTIE 1 : UNE STRATÉGIE POUR L'OUVERTURE ET LA VALORISATION DES DONNÉES TRANSPORT

I. – L'ouverture des données

Le mouvement d'ouverture des données se situe à la confluence de poussée de deux moteurs : l'ouverture des données publiques, c'est-à-dire les données produites dans l'exercice d'une mission de service public, et l'ouverture de données pour des besoins d'intérêt général, que ces données soient de natures publique ou privée.

L'ouverture des données publiques peut être opérée selon deux modes : soit les données sont accessibles pour des usages personnels à but d'information ou pour la réalisation d'études ; soit les données sont réutilisables, ce qui signifie que ces données peuvent être croisées et enrichies afin de créer des produits ou des services valorisables.

La valorisation des données, produits ou services réalisés à partir de données ouvertes peut se faire dans une approche marchande, ou venir enrichir l'espace des biens communs (*Commons*) de données, de contenus, de services ouverts.

L'accès aux données publiques s'effectue généralement dans le cadre d'une licence de réutilisation ouverte garantissant l'indication de l'origine des données ouvertes utilisées et demandant de prendre des engagements de bonne conduite dans les usages de ces données, en termes d'objectifs de complétude et de neutralité. Ceux-ci devraient être précisés dans le cadre de la loi sur le numérique en termes de loyauté et de neutralité.

L'ouverture peut être faite sans restriction autre que celle d'engagement de bonne réutilisation figurant dans la Licence Ouverte (Etalab) ou avec des restrictions liées à la nature ou à l'impact des usages des données.

La valorisation des produits issus de données réutilisables peut s'inscrire dans un cadre de valorisations commerciales, payantes ou gratuites (type *freemium*), respectant les engagements de la Licence Ouverte (indication de source et bon usage), ou être placée dans un environnement *share alike* imposant aux réutilisateurs de partager les données, produits ou services réalisés.

Dans le domaine de l'ouverture de données publiques, le principe est de rendre ces données gratuites d'usage ; toutefois, l'accès à ces données peut donner lieu au paiement de frais de mise à disposition, dûment justifiés, pour répondre à des usages intensifs (par exemple pour des informations temps réel ou pour de très gros volumes de données) et assurer l'équité d'accès entre gros et petits utilisateurs.

Le terme « données brutes publiques » doit être précisé car les fichiers informatiques les contenant peuvent être de nature très variée et spécialisée. Il s'agit parfois de données qui ne permettent pas d'élaborer une application ou une information d'intérêt général, voire peuvent présenter des problèmes liés à la protection de la vie privée, à la sécurité ou à la souveraineté. Il est donc nécessaire d'emblée de déterminer l'accès aux données publiques en fonction d'une finalité d'intérêt général (i.e. d'accès aux informations utiles pour le citoyen ou les entreprises, à partir des données produites dans le cadre d'une mission de service public), pour arriver à la notion d'information publique réutilisable. Un traitement des données brutes par le producteur peut être nécessaire pour produire ces informations publiques ouvertes, que l'on appellera par la suite « données publiques réutilisables ».

Les données brutes sont des données publiques produites par les opérateurs publics ou privés exploitant le service de transport. Les autorités publiques et les opérateurs publics ou privés réalisant des services d'information à partir de ces données publiques, ou les enrichissant, doivent être maîtres de l'usage qu'ils font des données enrichies et des services qu'ils réalisent, dans le respect de leur mission de service public et de considérations économiques de développement de l'innovation, de rentabilité et de meilleure utilisation des fonds publics.

Les données peuvent être source dans leurs usages de problèmes de nature juridique liés aux questions d'usage et de responsabilité sur les conséquences liées à l'usage de données erronées ou d'usage conduisant à des erreurs ou à des biais dans l'information produite. Il convient, dans les licences d'utilisation, de mettre en avant un esprit de responsabilité pour les réutilisateurs sur les conséquences des services qu'ils proposent à partir des données publiques, sans que cela

puisse constituer une garantie de protection absolue vis-à-vis des producteurs de données. Mais on peut considérer qu'en la matière, il y a plus de bénéfices que de risques et que les risques éventuels doivent être traités et suivis, mais ne pas être utilisés comme un moyen de contenir le mouvement d'ouverture ou servir de prétexte pour filtrer ou dégrader des données.

Si certaines données sont considérées d'intérêt général, mais n'existent pas au moment où la nécessité de l'intérêt général est décidée, il est nécessaire que les producteurs potentiels de ces données s'organisent pour les produire. Mais cela peut supposer de mettre sur pied un plan et un calendrier raisonnable de mise en œuvre, globalement pour le jeu de données ou spécifiquement à un opérateur, prenant en compte la complexité et les coûts de réalisation.

II. – Spécificités des données de transport

Par rapport à la problématique générale d'ouverture des données, le secteur des transports présente des spécificités liées à une grande diversité des modes des transports et des données, à leur impact sur la vie des personnes, sur l'économie et sur l'organisation des villes et des territoires. Leur ouverture est donc à la source de risques et d'opportunités économiques et d'usage très importantes :

1. Les activités de transport sont organisées par des maîtrises d'ouvrage publiques ou privées et réalisées par des maîtres d'œuvre publics ou privés ;
2. Les transports publics (ou « conventionnés ») sont réalisés par des entreprises publiques, ou privées, sous régime de régie, délégation ou concession de service public. Les bases de données publiques de données de transport correspondantes sont aujourd'hui hétérogènes quant à leur ouverture et leurs modes d'accès ;
3. Les modes de transport sont très variés en termes de vecteurs : avion, train, tram, bus, voiture, vélo ; d'infrastructures : aéroports, gares, autoroutes, routes, aires de covoiturage, stations de véhicule en libre-service, parking, stationnement de rue ; de modalités de mutualisation : transport collectif, véhicules en libre-service, covoiturage, véhicules privés ;
4. Les données de transport sont, elles aussi, de nature très variée et hétéroclite : horaires théoriques et temps réel, arrêts, infrastructures de transport, positionnement des vecteurs de transport, tarifs, fréquentation et disponibilité, mobilité des voyageurs, incidents, offres de transport... ;
5. Les données de transport peuvent porter dans leurs usages des problématiques de vie privée, de sécurité, notamment pour la gestion de crise suite à des événements, de responsabilité juridique, pour des conséquences dommageables liées à l'impact de l'information de transport, ou de la non réalisation d'un transport ;
6. Les objectifs d'intérêt général dans le domaine du transport sont aussi de natures variées :
 - a) Assistance au voyageur : information « brute » des voyageurs pour planifier leur mobilité, assistance au voyageur pour optimiser leur transport en terme de durée, prix, confort, information des voyageurs en temps réel sur l'accomplissement du transport, assistance au voyageur pour une configuration dynamique de sa mobilité ou information de sécurité ;
 - b) Optimisation des performances des systèmes de transport : adaptation de l'offre de transport, optimisation des coûts de transport, coopération dans la réalisation de transport, gestion globale des trafics, gestion des crises ;
 - c) Information des fabricants de véhicules et des constructeurs des infrastructures et service de transport, route intelligente ;
 - d) Alimentation en données d'autres secteurs intéressés à la mobilité : consommation, tourisme...
7. Les acteurs concernés par les transports ou parties prenantes principales :
 - a) L'État exerçant des prérogatives régaliennes sur l'organisation des systèmes de transport, la sécurité et exerçant la tutelle d'opérateurs de transport nationaux ;

- b) Les Collectivités, autorités organisatrices de transports et de mobilité pour leur territoire et responsables de la qualité de vie de leurs administrés et du développement économique ;
- c) Les opérateurs de transports publics et privés et leurs syndications ;
- d) Les utilisateurs ;
- e) Les acteurs économiques liés au transport ;
- f) L'Europe, qui se substitue aux États.

Dynamique des services de transport et de leurs marchés :

Les besoins de transport s'intensifient du fait de la croissance des agglomérations et du développement des mobilités personnelles et professionnelles de courte et longue distances, et ce, malgré les pressions environnementales et économiques, et malgré, ou à cause, du développement des communications numériques qui jouent à la fois en substitution et à la fois en intensification des besoins de rencontre.

Par ailleurs, les évolutions du commerce électronique posent de nouveaux problèmes de logistique, qui ne font pas partie de l'étude, mais interfèrent cependant avec les autres flux de transport, notamment pour la logistique des derniers kilomètres et les problèmes des trafics de transit.

Concernant les modalités de transport, nous allons vers une diversification des modalités et une concurrence généralisée, sur fond de dérégulation, d'innovation de services, d'optimisation économique, et de recherche d'amélioration de la qualité de vie pour le travail et la vie sociale et individuelle :

1. Dérégulation européenne, séparation des infrastructures et des services pour les monopoles d'État ;
2. Ouverture de la concurrence entre modes dans la longue distance : avion versus train, versus covoiturage et à venir versus autocar ;
3. Diversification de la concurrence dans les territoires : TER versus autocar, versus voiture individuelle, versus covoiturage ;
4. Diversification des modes dans les agglomérations : « versus et complémentaires » métro, tram, bus, vélo, covoiturage... Et pour le stationnement rue versus parking, et location, sous location.

Un mouvement européen et mondial d'ouverture des données de transport :

Le développement du numérique et la dynamique de l'open data, associés aux besoins d'optimisation des mobilités, créent des conditions favorables à l'ouverture des données :

1. Directive européenne sur les systèmes de transport intelligents¹ : sécurité, information des voyageurs, espace de données pour un calculateur d'itinéraire multimodal... ;
2. Pré-positionnement européen vers une ouverture généralisée des données de transport publiques et privées² ;
3. Diversification des expériences d'ouverture dans le monde et prise en compte de leurs apports ;
4. Transformation de la chaîne de valeur et risques de captation : captation et création de chalandise autour d'un service *killer app*, positionnement d'intermédiation de grandes

¹ Directive 2010/40/UE du Parlement Européen et du Conseil du 7 juillet 2010.

² Feuille de route de la Commission Européenne publiée en juin 2014 : « *Towards a roadmap for delivering EU-wide multimodal travel information, planning and ticketing services* »

plates-formes numériques, captation de valeur dans un secteur déjà très tendu, diminution de la rentabilité et des potentiels d'investissement, impact sur la qualité et la sécurité ;

5. Émergence du pouvoir des utilisateurs, de leur capacité de contribution à l'innovation et au développement des services et de leur capacité de pression à l'ouverture ou au contournement en cas de non ouverture.

L'évolution de la chaîne de valeur et des modèles économiques :

La fluidité des données numériques et la croissance des puissances de traitement ont donné lieu à la création de plates-formes mondiales d'intermédiation. Elles se bâtissent autour d'un service phare créant un pôle de chalandise, puis elles valorisent cette chalandise auprès de producteurs de biens ou de services. Ces structures de marchés, dites « bifaces », peuvent conduire très rapidement à une intermédiation forte et l'atteinte de positions de fait dominantes, et la possibilité de capturer ainsi une part croissante de la valeur créée par les opérateurs de transport, contraints de payer pour être référencés par les intermédiaires. Ce phénomène est, bien sûr, attisé par la concurrence entre les modes et les transporteurs.

Par ailleurs, la numérisation du fonctionnement des entreprises (production, marketing, conception, relation client, fonctionnement interne...) fait des données un vecteur clef de leur transformation, de l'amélioration de leurs performances économiques, de progression dans la qualité de service et d'entrée dans un processus d'innovation ouverte. Le mouvement d'ouverture des données de transport est aussi à apprécier dans ce cadre, comme une sorte de partie émergée de l'iceberg de transformation par les données. Il ne doit pas geler les évolutions vers des modèles économiques plus globaux des opérateurs de transport intégrant les données.

III. – Les préconisations et recommandations du Comité

Analyse générale :

Il y a un intérêt fort pour l'ouverture des données de transport afin d'améliorer la qualité de vie des usagers et apporter de l'innovation dans les services, amenant globalement à l'amélioration de la qualité des transports et de la mobilité.

Les villes et agglomérations sont des zones massives de transport avec des services publics de transport prépondérants et se développant en complémentarité face à une concurrence privée encore faible.

Le transport à longue distance est en situation concurrentielle : avion, train, covoiturage, autocar dans un avenir proche.

L'évolution prévue en France consiste à ouvrir les données publiques issues des services conventionnés, mais constituant en réalité un « espace gruyère » du fait de la diversité des modalités de maîtrise d'œuvre, des types de production et de la propriété des données.

L'Europe est en action sur l'ouverture des données de transport avec des décisions déjà en application sur certains modes (route) et d'autres en cours de discussion.

L'impact du coût de production des données et la nature des modèles économiques associés sur la qualité de service et leur valorisation doivent être pris en compte, mais ne doivent pas conduire à pérenniser les « trous du gruyère », quand l'intérêt général est en jeu.

La réalisation d'une information voyageur de qualité, de calculateurs d'itinéraire dynamiques et de services de planification de voyages est d'intérêt général.

Construire des communautés d'acteurs pour améliorer la qualité de vie et la mobilité et contribuer au développement économique est le meilleur chemin pour se mettre en position offensive dans le

paysage mondial, résister aux dangers d'intermédiation et se mettre en situation d'exporter les innovations et solutions.

Inclure dans ces communautés les utilisateurs, usagers, contributeurs est une des clefs pour profiter de la force de la « multitude » des utilisateurs.

Orientation et recommandations générales :

Orientation générale : pour créer un écosystème de transport innovant et générateur de nouvelles sources de valeur économique et d'usage, construire un espace généralisé de données ouvertes de transport, s'appuyant sur une base de données publiques réutilisables alimentée par les opérateurs de transport conventionné et bénéficiant de l'apport de données privées (mises à disposition de façon volontaire dans une première étape), avec des modalités d'ouverture et de réutilisation adaptées à la nature des données et à leurs usages.

Recommandation générale n° 1 : mettre en place rapidement l'espace des données publiques réutilisables, en respectant les contraintes spécifiques au secteur des transports

1. Définir trois modes d'ouverture des données publiques fournies par les opérateurs de transport :
 - a) Données ouvertes libres de réutilisation : usage des données gratuit et sans contrôle a priori et sans autre condition que les engagements de bonne conduite ; déclaration optionnelle d'identité, facilitant la mise à jour des données et la création de clubs d'utilisateurs et d'échanges ;
 - b) Données ouvertes avec conditions de réutilisation (du type prise en compte des règles de gestion de crise pour les données en temps réel), donnant lieu, quand cela est pertinent, à déclaration d'identité et de services ;
 - c) Données ouvertes à des fins d'études et de recherche, pour les besoins propres des chercheurs, avec réutilisation conditionnelle des résultats définie par les fournisseurs ou les collecteurs de données.
2. Définir les modalités de sélection des jeux de données ouvertes pour chaque catégorie :
 - a) Données ouvertes libres de réutilisation : gestion par l'État, en concertation avec les autorités organisatrices et les opérateurs de transport, de la définition des finalités d'intérêt général permettant de décider quels jeux de données entrent dans cette catégorie ; plan de mise en œuvre négocié avec l'État et les acteurs du transport pour les données publiques à créer ou mettre en forme ;
 - b) Données ouvertes avec conditions de réutilisation : gestion par l'État, en concertation avec les autorités organisatrices et les opérateurs de transport, d'une liste de jeux de données entrant dans cette catégorie et des conditions éventuelles de réutilisation ;
 - c) Données communiquées à des fins d'études et de recherche avec réutilisation conditionnelle : identification des finalités et décision de l'autorité gestionnaire ou suppléante sur une liste de finalités.

Recommandation générale n° 2 : construire les conditions d'un environnement favorable à la réutilisation des données ouvertes

1. L'utilisation de données donne lieu à une forme de CGU (Conditions générales d'utilisation), incluant les conditions de réutilisation (Licence Ouverte), les modalités de mise en œuvre éventuelles correspondant aux jeux de données concernées, des conditions éventuelles de tarification de l'accès technique, des modalités d'information sur l'évolution des données et la participation éventuelle à un réseau de réutilisateurs ;
2. Pour la licence d'utilisation, il est essentiel de garantir un accès libre et *fair* aux données publiques produites par les opérateurs de transport : à cet effet, l'accès aux données publiques de transport se fait en utilisant la Licence Ouverte, qui établit les bonnes

pratiques pour l'utilisation des données ; cependant, il est également d'intérêt vital que des conditions sur la loyauté des usages soient mises en place pour garantir l'innovation et les conditions de concurrence s'imposant à tous et, au premier chef, aux grands plates-formes d'intermédiation.

Remarques :

- dans ce contexte, les données ouvertes sont celles produites par les opérateurs de transport, qui les rendent directement accessibles ou décident d'en confier l'accessibilité à un tiers qui hérite des conditions de liberté d'accès et de gratuité d'usage ;
 - un réutilisateur produit des données, produits ou services à partir de ces données publiques ouvertes ; il convient de lui laisser ouvert le choix des modalités de réutilisation de sa production, y compris la possibilité d'inscription dans un mode *share alike* de tout ou partie des résultats et ce, que le réutilisateur soit public ou privé, opérateur, autorité ou industriel. Ceci est une condition nécessaire pour développer l'innovation et la création de valeur par les acteurs privés ou publics ;
 - en cas de mise à disposition volontaire de données d'origine privée ou issues d'un domaine concurrentiel, le producteur peut choisir son mode de licence, y compris associer une licence ouverte pour les bonnes pratiques et un cadre coopératif pour l'utilisation de type *share alike* ;
 - de façon conservatoire, pour des données jugées « stratégiques », le producteur peut imposer une réutilisation du type *share alike* dans l'attente que des conditions sérieuses sur la loyauté des usages aient été mises sur pied par les autorités.
3. Définir les conditions de tarification de l'accès en fonction du niveau de service et garantissant la neutralité d'accès : accès simple sans garantie de services, accès avec des garanties de services, prétraitements spécifiques.

Recommandation générale n° 3 : anticiper la possibilité d'ouverture des données en concurrence et l'impact sur leurs usages et les modèles économiques

1. Prendre en compte le mouvement général d'ouverture des données correspondant à un intérêt général : si les données n'étaient pas ouvertes par décision d'intérêt général, l'effet de cisaillement de la concurrence ouvrirait de toute façon la porte à l'entrée de plates-formes d'intermédiation et viendrait contraindre ceux qui n'ouvriraient pas ;
2. Utiliser le temps disponible avant cette ouverture légale ou de fait, pour mettre en place une politique offensive de services aux usagers, s'appuyant sur les données privées et les données publiques, dans un écosystème d'innovation ouvert et/ou coopératif ;
3. Inscrire cette politique dans une action de renforcement de l'écosystème national et européen :
 - a) En créant des espaces de travail avec les opérateurs de transport, entreprises, start-up et utilisateurs pour développer l'innovation de services autour des données publiques et privées ;
 - b) En développant des espaces coopératifs permettant de progresser rapidement vers des services du type calculateurs multimodaux de plus en plus complets et évolutifs.

Recommandation générale n° 4 : penser le rôle du numérique et des données, dans la stratégie d'adaptation et de transformation des opérateurs de transport et d'évolution de leur modèle économique

1. Développer une vision globale du numérique et de la production et l'échange d'information pour chacune des fonctions et de façon holistique pour l'ensemble de l'entreprise ;
2. Dans une vision ouverte du fonctionnement de l'entreprise, prendre en compte le rôle des flux de données dans les interactions avec les voyageurs, les entreprises partenaires et entre les acteurs du transport ;

3. Intégrer le numérique et les données dans les modèles économiques des opérateurs de transport, les capacités de production de valeur interne et externe et les business model de valorisation.

IV. – Conclusions

Les recommandations particulières précises présentées dans la suite de ce rapport se sont concentrées principalement dans le domaine des recommandations générales n° 1 et 2. Elles portent sur les conditions d'ouverture et de réutilisation des données publiques, dans la perspective d'une action législative à court terme.

Les conditions d'ouverture et de valorisation des données en concurrence (recommandation générale n° 3) devraient constituer le deuxième étage de la fusée. Nous recommandons de s'inscrire dans la dynamique européenne et son calendrier. Ces conditions doivent faire l'objet d'une concertation entre tous les acteurs concernés pour contribuer à la réflexion européenne et prendre au mieux en compte les spécificités françaises. Nous recommandons de concevoir et mettre en œuvre dans l'intervalle des stratégies offensives et innovantes, s'appuyant sur des ouvertures volontaires de jeux de données et la création d'un écosystème dynamique d'innovation et de développement de services à partir des données.

La quatrième recommandation générale n'était pas l'objet proprement dit de la saisine et n'a pas été traitée dans le cadre du débat plénier. Elle a en revanche fait l'objet d'échanges avec certains opérateurs, pour lesquels la métamorphose numérique est au cœur de l'évolution de l'entreprise. Elle ne fait donc pas l'objet d'une préconisation précise à ce stade, mais pose le problème de transformation rapide du modèle des entreprises publiques de transport.



PARTIE 2 : SYNTHÈSE ET RECOMMANDATIONS

I. – Synthèse

Méthode de travail du Comité :

La méthode de travail adoptée par le Comité du débat a consisté à croiser deux approches, une approche par thèmes (technique, juridique, modèles économiques), traités dans le cadre de trois groupes de travail spécifiques, et une approche par modes de transport, répartis en quatre ensembles cohérents :

- services publics (notamment bus, tram, métro, RER, Transilien, TER, Intercités, autocars) ;
- services concurrentiels (TGV, covoiturage longue distance, aérien) ;
- nouveaux services de mobilité (vélo, vélo en libre-service, covoiturage du quotidien, autopartage) et stationnement ;
- route.

Pour chacun de ces ensembles de modes, le Comité a analysé le rapport bénéfices / risques d'une ouverture des données au regard de leur modèle économique (conventionné ou concurrentiel).

Problématique juridique :

Le Comité a d'emblée constaté l'inadaptation du cadre juridique actuel à la réutilisation des données publiques liées à l'information du voyageur.

- ce cadre juridique est défini par la loi du 10 juillet 1978 relative à l'accès aux documents administratifs, complétée en 2005 par un chapitre sur la réutilisation des informations ;
- par réutilisation, on entend le fait, pour un tiers, d'utiliser une information à d'autres fins que la mission de service public pour les besoins de laquelle cette information a été produite ; les horaires planifiés par un opérateur de transport, par exemple, sont susceptibles d'être réutilisés par un tiers fournisseur de service d'information ;
- le fait qu'une donnée soit « réutilisable » n'implique pas l'obligation de la produire, mais seulement d'en permettre la réutilisation par les tiers, si elle a été produite ;
- les informations produites dans l'exercice d'une mission de service public administratif (SPA) sont « réutilisables », mais non celles produites dans l'exercice d'une mission de service public à caractère industriel et commercial (SPIC).

Dans le cas du service public de transport, les principaux acteurs à prendre en compte sont :

- l'autorité organisatrice des transports : sa mission d'organisation du service de transport est une mission de service public administratif (SPA) ;
- l'opérateur de transport : sa mission d'exécution du service de transport est une mission de SPIC ;
- les tiers, fournisseurs de service d'information.

Dans ce cas du service public de transport, les données nécessaires à l'information du voyageur sont en règle générale produites en lien avec la mission d'exécution du service de transport, qui est un SPIC. Il en résulte que ces données ne sont pas « réutilisables ».

Le débat a très vite fait apparaître un consensus sur l'opportunité de modifier la loi pour rendre « réutilisables » certaines données des services publics de transport, des nouveaux services de mobilité et du stationnement.

Bénéfices et risques :

Les bénéfices attendus de l'ouverture des données consistent non seulement dans l'amélioration de l'information délivrée à l'usager des transports, mais aussi dans la promotion des alternatives à la voiture individuelle, l'accroissement de la transparence de l'action publique, l'innovation et le développement de l'économie numérique en France.

En matière de risques, deux préoccupations principales ont été exprimées par les acteurs du transport : la première porte sur les coûts engendrés par la mise à disposition des données. C'est

pourquoi le Comité propose que l'accès aux données ouvertes puisse être payant lorsque ce coût est significatif.

La seconde préoccupation porte sur les répercussions de l'apparition, possible, d'une position dominante sur le marché de l'information multimodale.

Si l'ouverture des données débouchait sur le monopole ou le quasi monopole d'un fournisseur d'information, l'objectif de stimulation de l'économie numérique pourrait être remis en question. Dans le but d'éviter cette situation, le Comité a recommandé un ensemble de dispositions, notamment d'ordre technique, destinées à faciliter la réutilisation par les start-up et moyennes entreprises, qui ne disposent pas de moyens importants. Le Comité s'est attaché à trouver un juste équilibre entre les intérêts des réutilisateurs et ceux des producteurs de données.

Certains opérateurs de transport en situation de concurrence ont, pour leur part, fait valoir un risque d'abus de position dominante d'un intermédiaire, lequel pourrait leur imposer le versement de commissions susceptibles de compromettre leur équilibre économique. Ce scénario préoccupe notamment la SNCF (cf. annexe 9). Aussi le Comité recommande-t-il que les données de ces opérateurs ne soient ouvertes qu'à la condition que puissent être adoptées simultanément des mesures robustes permettant de prévenir le risque d'abus de position dominante.

Recommandations sur l'ouverture des données et les conditions permettant de remédier aux risques :

S'appuyant sur la classification élaborée par le sous-groupe technique (cf. figure 1 ci-dessous), le Comité recommande, pour les services publics de transport, d'ouvrir :

- certaines données contenues dans les classes de la tranche supérieure ;
- l'ensemble des données contenues dans les classes des trois tranches inférieures.

Pour remédier aux risques identifiés par les acteurs du transport (cf. annexe 11), le Comité recommande de soumettre cette ouverture à certaines conditions, représentées sur la figure 2 :

- les conditions de complétude et neutralité de l'information et d'exonération de responsabilité du producteur de données font l'objet de clauses équivalentes dans la Licence Ouverte ;
- lorsque la mise à disposition des données entraîne des coûts significatifs, ce qui peut être le cas notamment pour les données « temps réel », l'accès peut être rendu payant (cf. supra) ;
- les données d'usage et billettiques pourraient être mises à disposition sous une forme agrégée (l'ampleur de cette agrégation devant toutefois rester limitée et la mise à disposition des données doit également respecter la réglementation concernant les données personnelles).

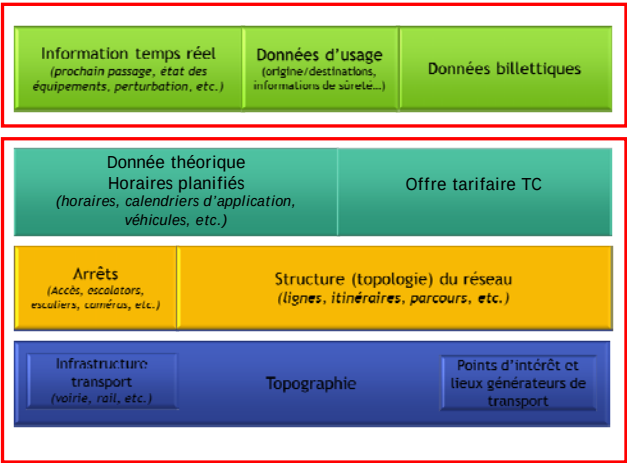


Fig 1. Ouverture des données Services publics de transport

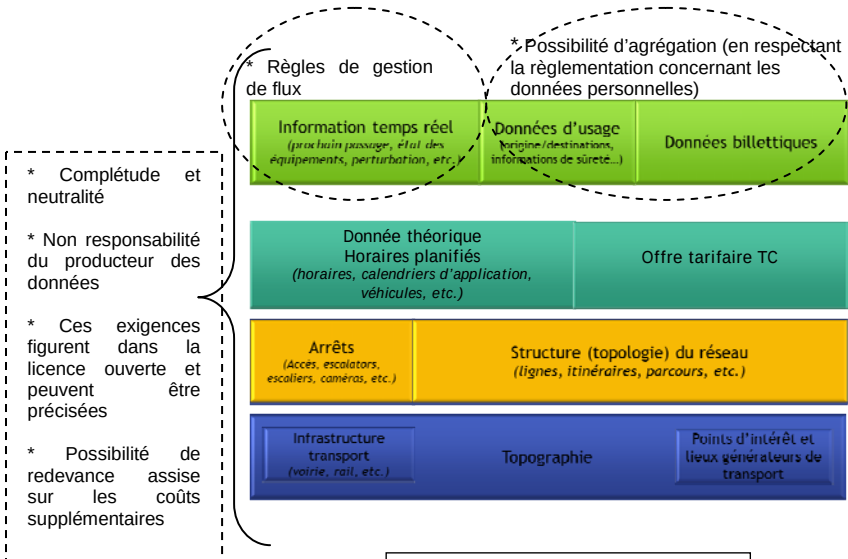


Fig 2. Conditions d'ouverture Services publics de transport

Pour les nouveaux services de mobilité et le stationnement, le Comité formule des recommandations d'ouverture similaires, résumées dans la figure 3 ci-après :

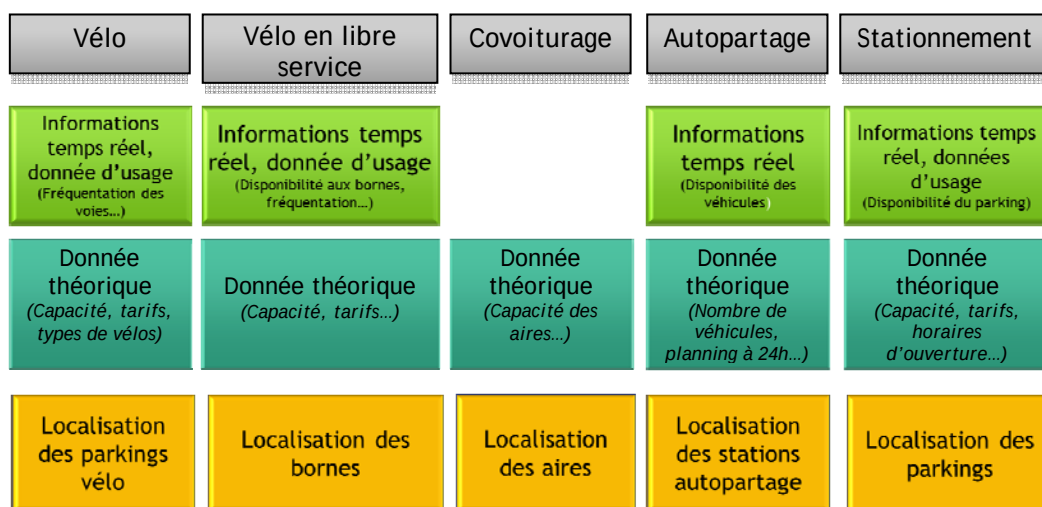


Fig 3. Ouverture des données
Nouveaux services de mobilité

La question se pose de savoir s'il y a lieu de limiter les nouvelles règles en matière d'ouverture des données au strict périmètre du service public. En effet, il n'est pas toujours aisé à déterminer ce qui relève d'une mission de service public: les services de vélo en libre-service (VLS), d'autopartage, de stationnement, les équipements destinés au vélo peuvent, selon les cas, être ou non des services publics. Faut-il alors limiter l'ouverture des données aux seuls services publics ou bien l'étendre à d'autres services ayant le même objet ?

La question de l'ouverture de données de statut privé, qui n'était en toute rigueur pas dans le champ du débat, se pose de façon plus générale :

- la Commission Européenne aborde ce sujet : dans l'information routière, un grand nombre de données sont de source privée ; les obligations de mise à disposition créées par le règlement délégué de l'action prioritaire c) de la directive ITS, relative aux informations de sécurité sur le réseau routier, s'imposent indifféremment aux personnes publiques et privées ;
- le Comité a évoqué, par exemple, la question des données élaborées par les grands opérateurs du numérique sur la description des gares destinées au guidage indoor.

Désireux de formuler des dispositions qui soient d'application simple pour les gestionnaires de service, le Comité recommande que les dispositions en matière d'ouverture des données pour les nouveaux services de mobilité et le stationnement s'appliquent aussi à des services qui, sans constituer des services publics, donnent lieu à une convention avec une personne publique.

La notion d'intérêt général pourra guider la poursuite de la réflexion sur l'ouverture des données.

Les services liés à l'information des voyageurs mis en place par les collectivités territoriales (calculateurs d'itinéraire, calculs d'isochrones, calcul du prix d'un trajet) seraient également rendus réutilisables, sous conditions.

Pour les services concurrentiels de transport et la route, la question de l'ouverture des données est examinée dans le cadre de la mise en œuvre – au niveau européen – de la directive sur les systèmes de transport intelligents.

Concernant les données des services de transport en situation de concurrence, dans l'éventualité d'une ouverture des données, le Comité recommande la mise en place concomitante d'un règlement européen visant à prévenir de possibles abus de position dominante des fournisseurs de service d'information vis-à-vis des opérateurs de transport. S'inspirant du règlement adopté pour les systèmes informatisés de réservation dans le transport aérien, ce règlement pourrait

comprendre, par exemple, un encadrement des commissions d'intermédiation et une obligation de neutralité dans l'affichage des résultats.

Autres recommandations, d'ordre technique, juridique et économique :

Le Comité formule des recommandations d'ordre technique, en vue de favoriser la réutilisation des données. Ces recommandations portent sur les données et leur qualité :

- mettre à disposition les données, de préférence, selon des formats normalisés ou standardisés ;
 - élaborer et utiliser les profils précisant les modalités d'utilisation de la norme ou du standard en France ;
 - pour les nouveaux services de mobilité et le stationnement, élaborer des formats d'échange de données ; participer aux travaux menés dans ce domaine par le Comité européen de normalisation (CEN) ;
 - vérifier la qualité des données et mettre en place l'organisation assurant leur mise à jour ;
- ainsi que sur les outils contribuant à assurer cette qualité des données :
- poursuivre le développement d'outils open source destinés, notamment, à tester la qualité des données et à faciliter la conversion d'un format à un autre.

Les recommandations techniques portent également sur la coordination et la mutualisation des actions à mener :

- élaborer des plates-formes de données multimodales à l'échelle territoriale ;
- élaborer notamment des bases de données d'arrêts à l'échelle territoriale, en assurant leur cohérence globale à l'échelle France entière ;
- décrire dans le schéma régional de l'intermodalité prévu à l'article L. 1213-3-1 du Code des transports l'organisation adoptée en matière d'ouverture des données sur le territoire régional.

et sur la mise en place d'un écosystème favorable aux développeurs :

- animer la communauté des réutilisateurs.

En matière de licence, le Comité incite les fournisseurs à mettre leurs données à disposition sous la Licence Ouverte (Etalab). Lorsque l'accès aux données est rendu payant, cette licence est complétée d'une clause relative à la tarification ; le prix ne couvre que les coûts supplémentaires engendrés par la mise à disposition.

Toutefois, le temps que soient adoptées des mesures garantissant la loyauté des plates-formes d'intermédiation, il est apparu opportun de laisser aux fournisseurs de données la possibilité d'ouvrir leurs données « stratégiques » sous une licence de type ODbL. Afin d'harmoniser les conditions de réutilisation, les parties concernées poursuivront leur réflexion afin d'identifier, pour chaque ensemble cohérent de modes de transport, quelles données sont susceptibles d'être considérées comme « stratégiques ».

Chacune de ces licences peuvent être complétées, si nécessaire, de clauses relatives aux conditions de réutilisation.

Sur le plan légistique, le Comité recommande d'insérer dans le code des transports une disposition législative déclarant réutilisables les données nécessaires à l'information du voyageur.

Il sera utile d'élaborer un guide des bonnes pratiques comprenant :

- des clauses-types à insérer dans des avenants aux contrats de délégations de service public ;
- des clauses-types à insérer dans les nouveaux contrats (notamment DSP et SIM) détaillant les exigences techniques et juridiques résultant du caractère réutilisable de certaines données.

Enfin, le Comité appelle l'attention sur l'intérêt que pourrait présenter la création d'une notion « d'information d'intérêt général » (IIG), pour définir un statut approprié à ce type de données. Cette notion d'IIG pourrait être utilisée dans de nombreux domaines : santé, éducation, logement,

transport etc. En fonction de l'évolution de la doctrine, telle donnée ou telle classe de donnée serait alors déclarée IIG. Des informations produites hors du cadre strict d'une mission de service public pourraient être déclarées d'intérêt général. Cette hypothèse de travail pourrait être approfondie dans le cadre des travaux préparatoires à la loi sur le numérique.

Sur le plan économique, le Comité invite les autorités publiques à s'appuyer sur l'initiative privée sur les services susceptibles de trouver un modèle économique viable. Il préconise enfin d'offrir aux entreprises innovantes la possibilité de bénéficier des marques des territoires, pour la diffusion de leurs produits, grâce à un dispositif de labellisation.

II. – Liste des recommandations

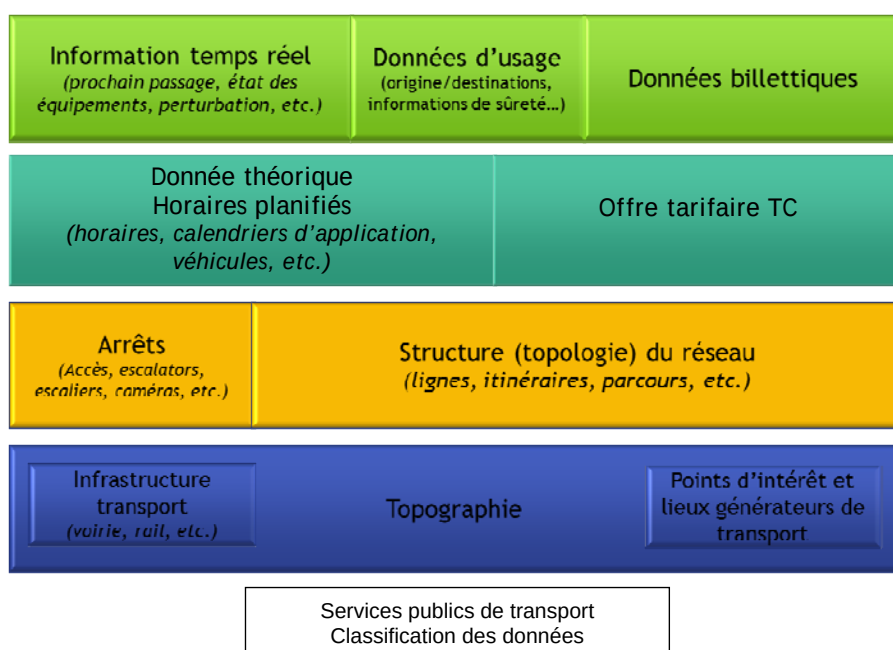
Les recommandations du Comité ont été élaborées, selon la méthode de travail qu'il a retenue, en croisant une approche par modes de transport et une approche par thèmes. Le plan de la présentation reflète cette démarche :

- recommandations par modes :
 - II.1. Recommandations sur les services publics de transport [SPT]
 - II.2. Recommandation sur l'ouverture des données des services de transport en situation de concurrence [SCT]
 - II.3. Recommandations sur les données des nouveaux services de mobilité et le stationnement [NSM]
 - II.4. Recommandations sur les services liés à l'information du voyageur [IV]
- recommandations par thèmes :
 - II.5 Recommandations juridiques [RJ]
 - II.6. Recommandations techniques [RT]
 - II.7. Recommandations destinées à favoriser le développement de l'économie numérique dans le domaine du transport [RE]
 - II.8 Recommandations transversales [RTV]

II. 1. – Recommandations sur les services publics de transport [SPT]

En matière d'ouverture des données :

Les recommandations en matière d'ouverture des données se réfèrent à la classification suivante :



SPT1 : Pour les classes « topographie », « arrêts », « structure du réseau », « données théoriques-horaires planifiés » et « offre tarifaire » :

Les données de ces classes sont réutilisables.

Cas de l'offre tarifaire : à titre exceptionnel, l'ouverture des données de l'offre tarifaire peut ne pas être complète : les informations qui alourdiraient inutilement la présentation ou demanderaient une expertise additionnelle pour être correctement interprétées peuvent être omises.

SPT2 : Pour la classe « information en temps réel » :

Les données suivantes sont réutilisables :

- messages diffusés sur le réseau, y compris messages de perturbation ;
- position des véhicules ;
- prochain passage du véhicule à l'arrêt ;
- état des équipements (escalators, ascenseurs...).

Cette liste pourra faire l'objet de compléments au fil du temps.

SPT3 : Pour les classes « usage » et « billettique » :

Les données suivantes sont réutilisables :

- origine/destination des déplacements ;
- usage des transports (nombre de voyageurs par segment et par période, serpent de charge) ;
- informations de sûreté (vols, agressions) ;
- données de validation.

Ces données peuvent être mises à disposition sous une forme agrégée, l'ampleur de cette agrégation devant toutefois rester limitée. L'ouverture des données doit respecter la réglementation concernant les données personnelles.

Sur les conditions de réutilisation :

SPT4 : Pour toutes les classes de données,

Conditions de réutilisation :

- la réutilisation ne doit pas induire en erreur des tiers quant au contenu de l'information et à sa date de mise à jour ;
- en vue de préciser l'exigence précédente, le fournisseur de données peut demander que l'information réponde à des exigences de complétude et des critères de neutralité ; afin de faciliter les réutilisations, les fournisseurs de données concernés définiront ces exigences de complétude et ces critères de neutralité dans une rédaction commune ;
- le fournisseur des données ne peut garantir l'absence de défaut de la donnée et ne peut être tenu pour responsable des pertes, préjudices ou dommages de quelque sorte causés à des tiers du fait de la réutilisation ;
- prix de l'accès aux données : si, contrairement à ce qui est observé actuellement, les coûts de mise à disposition devaient se révéler significatifs pour des données autres que celles en temps réel, les dispositions prévues en **SPT5** pourraient être mises en œuvre pour ces données.

Remarque : la consultation en cours pour préparer la loi sur le numérique met en avant la notion de « loyauté », qui pourra être utilisée pour préciser le critère de complétude.

SPT5 : Classe « Informations en temps réel » :

Conditions de réutilisation :

- la réutilisation peut être soumise au respect de règles de gestion de flux, notamment en cas d'incident ou de crise. Ces règles sont fixées par l'autorité responsable en cas d'incident ou de crise ;
- prix de l'accès aux données : lorsque les coûts de mise à disposition sont significatifs, cette mise à disposition peut être payante ; le total des recettes ne peut excéder les coûts

supplémentaires engendrés par la mise à disposition des données ; les éléments de calcul permettant de justifier le prix de l'accès aux données à partir des coûts supplémentaires sont publiés ; la mise à disposition reste gratuite pour les petits utilisateurs.

II. 2. – Recommandation sur l'ouverture des données des services concurrentiels de transport [SCT]

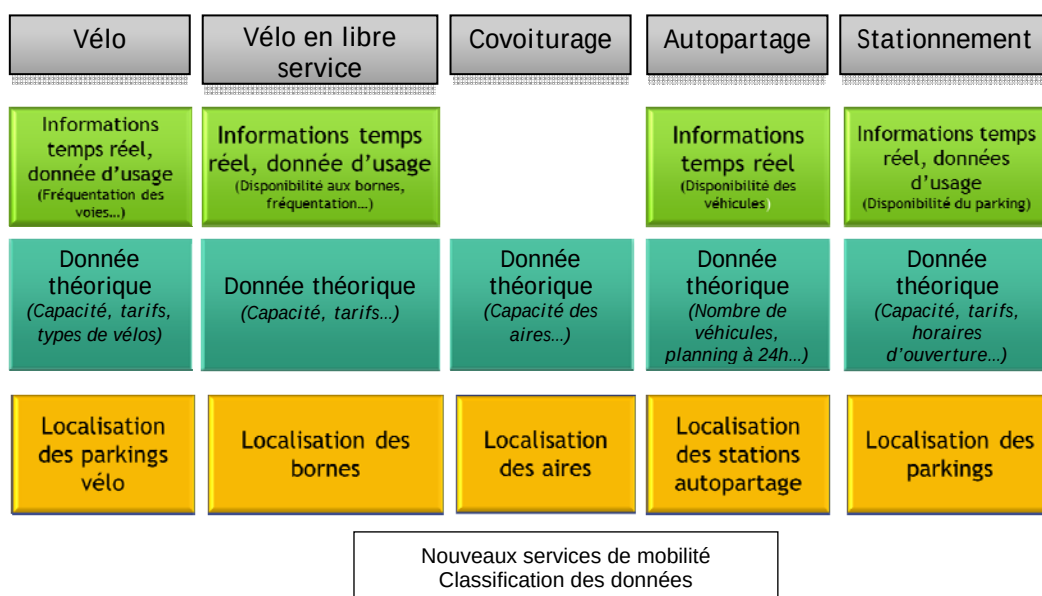
SCT : L'ouverture des données des services de transport en situation de concurrence ne devrait être réalisée que si elle est accompagnée de mesures robustes permettant de prévenir l'abus de position dominante d'un fournisseur de services d'information.

II. 3. – Recommandations sur les données des nouveaux services de mobilité et le stationnement [NSM]

Les recommandations suivantes s'appliquent aux nouveaux services de mobilité et au stationnement, que ces services soient exploités dans le cadre d'une mission de service public ou qu'ils donnent simplement lieu à une convention avec une personne publique.

En matière d'ouverture des données :

Les recommandations en matière d'ouverture des données se réfèrent à la classification suivante :



NSM1 : Vélo en milieu urbain et interurbain :

Les données relatives aux infrastructures ou aux équipements utilisés par les vélos sont réutilisables.

NSM2 : Vélo en libre-service :

- les données des classes « arrêts », « structure du réseau », « données théoriques », « informations en temps réel » et « offre tarifaire » sont réutilisables ;
- les données de la classe « usage » sont réutilisables. Elles peuvent être mises à disposition sous une forme agrégée, l'ampleur de cette agrégation devant toutefois rester limitée et être compatible avec la réglementation concernant les données personnelles.

NSM3 : Covoiturage (courte distance) :

Les données relatives aux aires de covoiturage sont réutilisables.

NSM4 : Autopartage :

Les données sont réutilisables. Il s'agit, notamment, des données suivantes :

- pour la classe « arrêts » : la liste des stations (par une recherche en x, y ou par opérateurs) ;
- pour la classe « données théoriques » :
 - o l'autopartage en boucle, la liste des véhicules à une station, le planning à 24h, le taux d'occupation à 24h ;
 - o pour l'autopartage en trace directe, la capacité de chaque station ;
- pour la classe « informations en temps réel », dans le cas de l'autopartage en trace directe, la disponibilité des stations et des véhicules.

NSM5 : Stationnement :

- les données des classes topographie et « données théoriques », hors « offre tarifaire », sont réutilisables ;
- les données retraitées (places libres, parc saturé, parc en voie de saturation) de la classe « information en temps réel » sont réutilisables ;
- les données de la classe « usage » sont réutilisables ; elles peuvent être mises à disposition sous une forme agrégée, l'ampleur de cette agrégation devant toutefois rester limitée ; l'ouverture des données doit respecter la réglementation concernant les données personnelles.

NSM 6 : Pour l'ensemble des nouveaux services de mobilité et le stationnement – Offre tarifaire :

À titre exceptionnel, l'ouverture des données de l'offre tarifaire peut ne pas être complète : les informations qui alourdiraient inutilement la présentation peuvent être omises.

Pour le stationnement, les données faisant l'objet d'une obligation légale d'affichage en entrée de parc sont réutilisables.

Sur les conditions de réutilisation :

NSM7 : Infrastructures ou équipements relatifs aux vélos, services de vélos en libre-service, plates-formes de covoiturage et d'autopartage, parcs de stationnement :

Conditions de réutilisation :

- la réutilisation ne doit pas induire en erreur des tiers quant au contenu de « l'Information » et sa mise à jour. En vue de préciser cette condition, le fournisseur des données peut demander que l'information réponde à des exigences de complétude et des critères de neutralité qu'il définit ;
- le fournisseur des données ne peut garantir l'absence de défaut de la donnée et ne peut être tenu pour responsable des pertes, préjudices ou dommages de quelque sorte causés à des tiers du fait de la réutilisation ;
- prix de l'accès aux données : voir **SPT 5**.

NSM8 : Inviter les parties concernées à examiner l'opportunité et les modalités d'une ouverture des données relatives aux bornes de recharge des véhicules électriques.

II. 4. – Recommandations sur les services liés à l'information du voyageur [IV]

IV : Les services liés à l'information du voyageur mis en place par les collectivités territoriales sont réutilisables, sous conditions. Il s'agit, par exemple, des services de calculateur d'itinéraire, de calcul d'isochrones ou du prix d'un trajet.

Prix de l'accès aux services : lorsque les coûts de mise à disposition du service sont significatifs, cette mise à disposition peut être payante ; le total des recettes ne peut excéder les coûts supplémentaires engendrés par la mise à disposition. Les éléments de calcul permettant de justifier le prix de l'accès aux services à partir des coûts supplémentaires sont publiés ; la mise à disposition reste gratuite pour les petits utilisateurs.

II. 5. – Recommandations juridiques [RJ]

RJ1 : Modification du cadre législatif

Insérer dans le Code des transports une disposition législative, concernant les services publics de transport, les nouveaux services de mobilité et le stationnement, et ayant pour objet de rendre réutilisables certaines données ou classes de données « nécessaires à l'information du voyageur ».

RJ2 : Élaboration d'un guide de bonne pratique « open data », proposant des clauses-types

- pour les contrats en cours (délégations de service public, notamment), clauses-types à insérer dans des avenants au contrat, pour préciser les modalités de la mise à disposition des données, par la personne publique ou le délégataire ;
- pour les contrats en cours et les nouveaux contrats (notamment DSP et SIM), clauses-types relatives aux exigences techniques résultant du caractère réutilisable des données.

RJ3 : Licences

Les fournisseurs mettent leurs données à disposition sous Licence Ouverte (Etalab).

Lorsque l'accès aux données est payant, la Licence Ouverte est complétée d'une clause relative à la tarification.

La licence peut être complétée, le cas échéant, selon les recommandations de **SPT4** relatives aux conditions de réutilisation.

Toutefois, en attendant que soient adoptées des mesures de nature à garantir la loyauté des plates-formes, il est souhaitable, pour certaines données considérées comme "stratégiques" d'un point de vue économique, de laisser au fournisseur de données le choix du modèle économique de réutilisation qu'il entend promouvoir : modèle "ouvert", autorisant l'appropriation des données et la revente des enrichissements apportés aux bases, ou modèle coopératif, fondé sur le partage des enrichissements apportés à la base de données initiale.

Ces données peuvent être mises à disposition sous une licence alternative, de type ODbL.

Afin d'harmoniser les conditions de réutilisation, les parties concernées veilleront à poursuivre leur réflexion afin d'identifier, pour chaque ensemble cohérent de modes de transport, les données dont le caractère "stratégique" justifie une mise à disposition sous une licence coopérative.

RJ4 : Identification des réutilisateurs

Pour les informations temps réel, les données d'usage, de billettique et les services liés à l'information du voyageur, les producteurs de données peuvent demander l'identification du réutilisateur préalablement à la distribution de clef d'accès à l'API ; ils peuvent exiger, en complément, une déclaration d'usage.

Pour les autres données, les producteurs de données peuvent inviter les réutilisateurs à créer un compte, cette identification ne revêtant toutefois pas un caractère obligatoire.

II. 6. – Recommandations techniques [RT]

Sur la donnée transport :

RT1 : Services publics de transport : utilisation des normes et standards

Dans le domaine des services publics de transports, inciter les producteurs de données à mettre à disposition les données dans les formats d'échange de données suivants, par ordre de priorité :

1. Dans toute la mesure du possible, à la fois aux formats normalisés (NeTEx, SIRI...) ET aux formats orientés « réutilisation *open data* » (GTFS) ;
2. À défaut, seulement aux formats orientés « réutilisation *open data* » ;

3. À défaut, au format propre au producteur des données, accompagné de toute documentation utile sur ce format.

RT2 : Nouveaux services de mobilité et stationnement : élaboration de formats d'échange de données

Inciter les acteurs des autres nouveaux services de mobilité et du stationnement à développer des formats d'échange de données standardisé ou normalisé. Les inciter à participer aux travaux qui vont être engagés par le CEN dans ce domaine.

Les besoins identifiés actuellement portent sur :

- le vélo (en milieu urbain), le covoiturage et l'autopartage ;
- l'information en temps réel sur les parkings.

RT3 : Services publics de transport et nouveaux services de mobilité : élaboration et utilisation de profils précisant la mise en œuvre de la norme en France

Inciter les instances de normalisation (CN03/GT7) :

- à élaborer des profils normalisés au niveau national, permettant une interprétation et une utilisation univoque des formats d'échange de données ;
- à produire, en complément, des règles de tests normalisés au niveau national, permettant de vérifier la conformité des données aux profils.

Inciter les producteurs de données à utiliser les profils disponibles.

RT4 : Services publics de transports et nouveaux services de mobilité : élaboration et utilisation de modèles de données en France

Inciter les acteurs transports (producteurs, réutilisateurs...) à utiliser les modèles de données existants, en particulier Transmodel pour les services publics de transport.

Inciter les instances de normalisation (CN03/GT7) à étendre ces modèles aux nouveaux services de mobilité.

RT5 : Diffusion de données transports de qualité

Inciter les producteurs des données transports à tester la qualité des jeux de données avant leur diffusion et à joindre aux données ouvertes les résultats des tests de qualité qu'elles ont passés.

Inciter les acteurs du transport à utiliser les mêmes données pour la mise à disposition des tiers et pour leurs propres services d'information.

Inciter les producteurs des données transports à mettre en place une organisation pour assurer la mise à jour des données réutilisables. Toute démarche d'ouverture des données devrait inclure un volet « modalités de mise à jour ».

Inciter les producteurs des données transports à publier, pour chaque donnée mise à disposition, les métadonnées décrivant les données (ou l'API).

RT6 : Élaboration de plates-formes de données multimodales à l'échelle territoriale

Inciter les autorités organisatrices des transports et les autorités organisatrices de la mobilité, en lien avec les opérateurs de transport, à mettre en place des plates-formes de données multimodales. Ces plates-formes agrègent les bases de données des différents acteurs du territoire, les mettent en qualité et diffusent des données interopérables et de référence.

Promouvoir l'élaboration de bases de données arrêts à l'échelle territoriale, en assurant leur cohérence globale à l'échelle nationale.

Ajouter dans le Code des transports³ (cf. articles 1213-3-1 et 1213-3-2) une disposition disposant que le schéma régional de l'intermodalité décrit l'organisation adoptée sur le territoire régional, pour la mise à disposition des données liées à l'information du voyageur.

RT7 : Diffusion de données brutes ET de services d'information

En complément à la mise à disposition de données brutes (sous forme de fichiers à plat), inciter les producteurs des données transports à généraliser l'ouverture de services, permettant aux réutilisateurs d'accéder à une information prétraitée et simple d'utilisation. Il peut s'agir des services de recherche d'arrêt, de calcul d'itinéraire, d'affichage de résultat sur une carte, etc.

Sur les outils (logiciels pour manipuler la donnée transport) :

RT8 : Logiciels facilitant la manipulation des données

Inciter les producteurs des données transports, dans leur démarche d'ouverture des données, à utiliser :

- des outils de vérification de la conformité aux profils normalisés. Un outil CHOUETTE open source existe déjà pour la norme NEPTUNE et pourrait être amélioré pour d'autres normes ;
- des outils de conversion d'un format à un autre. CHOUETTE propose également des fonctions de conversion ;
- des outils harmonisés de publication des données et métadonnées.

À des fins de mutualisation, ces outils pourraient être en open source et référencés.

Autres recommandations techniques :

RT9 : Mise en place d'un écosystème *developer-friendly*

Inciter les producteurs des données transport à proposer aux réutilisateurs des lieux de discussion leur permettant d'échanger sur les services développés et les retours d'expériences, ainsi que de s'entraider dans leurs réalisations.

Encourager les opérateurs et les autorités publiques à développer un écosystème public-privé favorisant l'innovation, la qualité et le développement économique.

II. 7. – Recommandations destinées à favoriser le développement de l'économie numérique dans le domaine du transport [RE]

RE1 : Afin de favoriser le décollage de marchés naissants, mettre les données publiques à disposition des tiers de manière gratuite et ce quel que soit le format de rediffusion.

RE2 : Privilégier, pour le secteur public, la fourniture d'un service « de base » d'information voyageur afin de laisser des espaces au secteur privé.

RE3 : Donner gratuitement accès à des services clefs, qui ne sont maîtrisés aujourd'hui que par un faible nombre d'acteurs.

RE4 : Donner de la visibilité aux entreprises innovantes en leur permettant de bénéficier des marques des territoires, pour la diffusion de leurs produits, grâce à un dispositif de labellisation.

II. 8. – Recommandations transversales [RTV]

RTV1 : Mise en œuvre progressive des dispositions en matière d'ouverture

³ http://www.legifrance.gouv.fr/eli/loi/2014/1/27/2014-58/jo/article_6

Mettre en place de manière progressive les dispositions en matière d'ouverture des données : une première étape pourrait concerner, outre les services régionaux et départementaux de voyageurs, les services de mobilité présents dans les agglomérations dont le nombre d'habitants est supérieur à un seuil à définir.

RTV2 : Mise en place d'un répertoire local des données réutilisables ainsi qu'un annuaire national

Inciter les fournisseurs de donnée à mettre en place un répertoire identifiant les données mises à disposition et les conditions d'accès (format, licences, services web...)

Mettre en place un annuaire national des données ouvertes.

RTV3 : Cas où plusieurs personnes détiennent les mêmes données réutilisables

Si plusieurs personnes détiennent les mêmes données réutilisables, elles peuvent décider entre elles laquelle assurera la mise à disposition des données et rendra accessible en ligne le répertoire des informations réutilisables ; elles peuvent également décider de confier ces tâches à un tiers.

RTV4 : Dispositif spécifique pour l'accès aux données aux fins d'études et de recherche

Lorsque le fournisseur fait payer l'accès à ses données, les organismes exerçant une mission de service public administratif ainsi que les organismes de recherche et d'enseignement devraient bénéficier de mesures d'exonération totale.

L'accès aux données non réutilisables pourrait se faire dans le cadre d'une convention pluriannuelle liant l'organisme fournisseur et l'organisme demandeur et comprenant éventuellement une clause de confidentialité afin d'assurer la protection des données personnelles et/ou du secret commercial. Un modèle-type pourrait être proposé. La coopération entre les deux organismes s'inscrirait dans une logique de partenariat gagnant-gagnant, le demandeur devant communiquer au fournisseur le résultat de ses recherches. Après signature de cette convention-cadre, l'organisme demandeur devrait adresser, pour chaque projet de recherche, une demande d'accès à certaines données détenues par le fournisseur ; place serait alors laissée à la négociation.

RTV5 : Information d'intérêt général

Dans le cadre des travaux préparatoires à la loi sur le numérique, examiner l'opportunité de créer une notion « d'information d'intérêt général », applicable à de nombreux domaines (transports, logement, santé etc).

Les « informations d'intérêt général » seraient dotées d'un statut propre, distinct du statut des informations publiques défini par la loi CADA ; ce cadre offrirait toute souplesse pour accueillir différentes informations qui seraient rendues réutilisables au fur et à mesure des évolutions souhaitées ; il pourrait être plus large que le strict champ du service public.



PARTIE 3 : RAPPORT DU COMITÉ

Introduction

Contexte :

L'ouverture des données publiques est un engagement fort du Gouvernement, rappelé dans la feuille de route sur le numérique publiée en février 2013⁴.

La stratégie gouvernementale en faveur de l'ouverture des données stratégiques à fort impact sociétal ou à fort potentiel d'innovation sociale et économique est soutenue par la mission Etalab, créée en 2011, qui a développé une plate-forme nationale d'hébergement et de partage de données publiques, ouvertes et réutilisables : data.gouv.fr. Cette plate-forme peut également héberger les données des collectivités locales, si celles-ci le souhaitent.

En décembre 2012, le premier Comité interministériel pour la modernisation de l'action publique (CIMAP) a réaffirmé le principe de gratuité des données publiques et souhaité l'étendre, en concertation, aux collectivités territoriales et aux délégataires de service public notamment.

À cet effet, il a été décidé d'organiser six débats thématiques, dans les domaines de l'éducation, les risques environnementaux, la transparence du système de santé, l'offre de transport, le logement et la lisibilité des prestations sociales et des dépenses publiques.

Le Secrétaire d'État chargé des Transports, de la Mer et de la Pêche a ainsi annoncé le lancement du débat sur l'offre de transport lors de la journée du 11 février 2014 « Mobilité 2.0 » sur les systèmes de transport intelligents.

Les enjeux d'une ouverture des données liées à l'offre de transport portent non seulement sur une meilleure information des voyageurs, mais aussi sur la promotion des transports publics et des nouvelles formes de mobilité, la transparence de l'action publique et le développement de l'économie numérique.

Objectifs du Comité :

Le Secrétaire d'État chargé des Transports, de la Mer et de la Pêche a confié à Francis Jutand, directeur scientifique de l'Institut Mines-Télécom et membre du Conseil national du numérique, la mission de présider le Comité du débat sur l'ouverture des données liées à l'offre de transport.

Le Comité, qui s'est réuni de mai à décembre 2014, avait pour mission de formuler des recommandations en matière d'ouverture des données utiles à l'information du voyageur pour la préparation et la réalisation de son voyage.

Il convenait de considérer l'ensemble des modes de transport et des services de mobilité, tels que, par exemple, le covoiturage ou le vélo en libre-service. Toutefois, les recommandations devaient principalement porter sur les données « publiques » au sens du CIMAP, c'est-à-dire les données détenues par les collectivités territoriales, les délégataires de service public (notamment les entreprises privées chargées d'une mission de service public industriel et commercial), les organismes exerçant des missions de service public et les entreprises publiques (comme la SNCF).

Les données relatives à un service international ont été exclues du débat. Quant aux données de l'État, leurs modalités de réutilisation sont d'ores et déjà définies par des dispositions réglementaires⁵.

⁴ <http://www.gouvernement.fr/premier-ministre/le-gouvernement-presente-la-feuille-de-route-pour-le-numerique>

⁵ Décret n° 2011-577 du 26 mai 2011 relatif à la réutilisation des informations publiques détenues par l'État et ses établissements publics administratifs et circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'État « data.gouv.fr » par la mission « Etalab » et l'application des dispositions régissant le droit de réutilisation des informations publiques.

Parallèlement à l'ouverture des données proprement dite, le Comité était invité à examiner d'autres modalités de mise à disposition de l'information, par exemple sous forme de services de calcul d'itinéraire.

Le débat devait permettre d'identifier et de publier de nouveaux jeux de données stratégiques, de qualité, en examinant la possibilité d'étendre la politique d'ouverture des données publiques à de nouveaux acteurs dont les données ne sont pas dans le périmètre de la loi de 1978 (dite « loi CADA »), qui prévoit un droit à réutilisation pour les informations publiques.

Le Comité ne s'est pas limité à examiner quelles données ou classes de données rendre réutilisables ; il a également débattu des conditions de la réutilisation. La lettre de mission l'invitait à formuler des recommandations sur les trois questions suivantes :

- convient-il, et dans quelle mesure, de subordonner la réutilisation de certaines données à des conditions liées à l'intérêt général ou au respect des politiques publiques ?
- quelles dispositions devraient être prises en vue d'éviter le risque de position dominante d'un fournisseur de services ?
- quels sont les grands principes d'un modèle de licence permettant, si cela s'avérait nécessaire, de prendre en compte les spécificités du transport et de la mobilité ?

Le Comité devait en outre se prononcer sur la question de savoir si des dispositions particulières sur le plan technique (modèles de données, formats, outils de qualification des données...) devaient être adoptées en vue de faciliter la réutilisation.

Les travaux du Comité ont pris en compte le risque de captation de la valeur pour les services de transport en situation de concurrence, en cas d'émergence d'opérateurs de service d'information acquérant une position dominante.

Dans le domaine de l'information routière, élaborée largement à partir de données privées, le Comité, comme le suggérait la lettre de mission, a estimé opportun de s'en remettre aux travaux menés à l'initiative de la Commission européenne pour la mise en œuvre de la directive sur les systèmes de transport intelligents.

Composition du Comité :

La composition du Comité et des sous-groupes de travail devait permettre aux différents points de vue de s'exprimer afin que les recommandations formulées tiennent compte des objectifs et préoccupations légitimes exprimés par les acteurs du transport.

Par conséquent figuraient parmi les membres et les participants :

- des représentants de producteurs de données :
 - a) gestionnaires d'infrastructures : ADF ;
 - b) autorités organisatrices de transport : GART, ARF, STIF ;
 - c) opérateurs de transport : SNCF, RATP, UTP ;
 - d) fédérations professionnelles ou opérateurs de services de mobilité : CITIZ, CVTC, FEDUCO, FNMS ;
- des représentants du monde de l'open data et de l'ouverture de données : Grand Lyon, Rennes Métropole ;
- des représentants de réutilisateurs de données : start-up JOUL et SNIPS ;
- des représentants des usagers des transports : FNAUT ;
- des experts, chercheurs et universitaires : experts indépendants, cabinet d'études (Chronos), Institut Mines-Télécom, IFSTTAR ;
- des représentants des services de l'État : AFIMB, AQST, CGDD, DST (Ministère de l'écologie, du développement durable et de l'énergie) ; mission Etalab (services du Premier ministre).

Méthode de travail :

Lors de sa première réunion, le Comité a retenu une méthode de travail consistant à croiser :

- une approche par mode de transport pour ses séances plénières ;
- une approche par thème : technique, juridique et modèles économiques.

Le Comité s'est réuni à sept reprises entre mai et décembre 2014, à une fréquence moyenne d'une fois par mois. Chaque séance plénière a fait l'objet d'une ou de plusieurs réunions préparatoires avec les membres du Comité directement concernés et les personnes susceptibles d'être auditionnées.

La première séance plénière a été consacrée à un rappel des éléments de cadrage du débat, à l'adoption de la méthode de travail du Comité ainsi qu'à un panorama des initiatives en matière d'ouverture des données de mobilité en France et dans le monde.

Les quatre séances suivantes ont porté sur des ensembles cohérents de modes de transport voyageur :

1. Les services publics de transport⁶ : notamment bus, tram, car, Intercités, Transilien, TER, autocars, etc. ;
2. Les services concurrentiels de transport longue distance : TGV, covoiturage, aérien ;
3. Les services de mobilité, comprenant les nouveaux services de mobilité (vélo, vélo en libre-service, covoiturage, autopartage) et le stationnement ;
4. La route.

Au cours de chacune de ces séances ont été présentés :

- un état des lieux sur les démarches d'ouverture des données ;
- des propositions concernant :
 - a) les données ou les classes de données à ouvrir ;
 - b) les conditions éventuelles ;
 - c) les caractéristiques techniques ;
- pour les services concurrentiels, le rapport bénéfices / risques d'une ouverture des données.

Les deux dernières séances ont été consacrées à l'élaboration et à l'adoption des recommandations.

Parallèlement au travail réalisé en séance plénière, trois sous-groupes thématiques ont été mis en place :

- un sous-groupe juridique, chargé d'élaborer un nouveau cadre législatif pour la réutilisation des données et d'examiner la validité juridique des orientations adoptées par le Comité ou des propositions des autres sous-groupes ;
- un sous-groupe technique, chargé d'élaborer une typologie des données et de formuler des recommandations sur des dispositions techniques permettant de faciliter leur réutilisation ;
- un sous-groupe « modèles économiques », chargé de réfléchir aux moyens d'assurer un écosystème concurrentiel équitable et de favoriser l'émergence de modèles économiques indépendants des fonds publics.

L'Agence française pour l'information multimodale et la billettique (AFIMB) a assuré le secrétariat du Comité.

Un site dédié a été créé pour rassembler les documents de travail : ordre du jour des séances, comptes rendus ou synthèses des séances plénières, présentations...

I. État des lieux

⁶ Par « services publics de transport », on entend les services de transport soumis au règlement dit OSP (Obligations de service public), CE n° 1370/2007 du Parlement européen et du Conseil du 23 octobre 2007.

I. 1. – Aspects techniques : données, classes de données et informations

Classification des données transport :

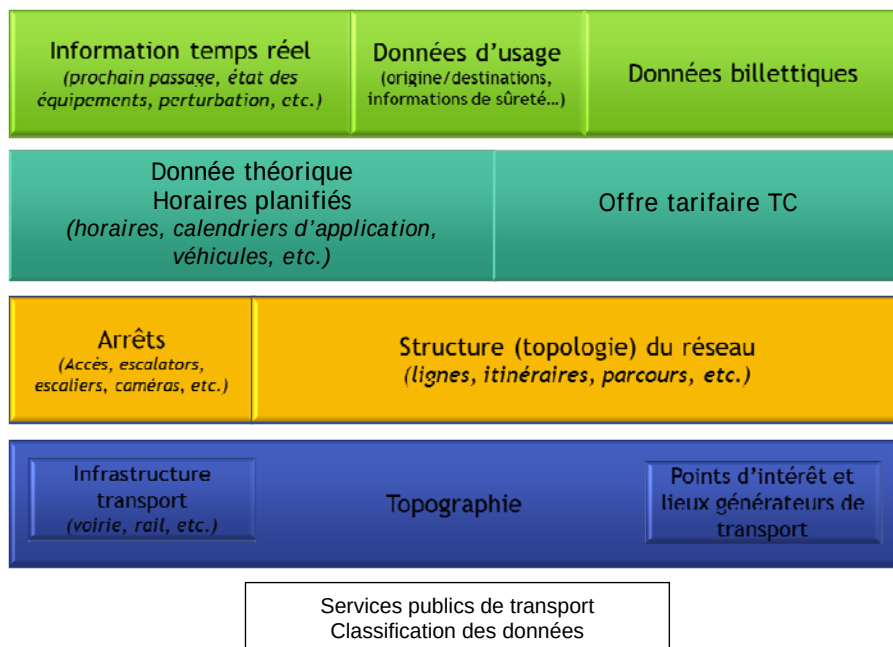
Afin de pouvoir entamer sa réflexion sur la base d'éléments concrets, le Comité a souhaité disposer d'une classification des données et, pour les classes de données les plus sensibles, d'une liste représentative de données.

Il a disposé dans un premier temps de la nomenclature utilisée par Accenture et le Laboratoire d'économie des transports (LET) rattaché à l'Université Lyon 2 et à l'École Nationale des Travaux Publics de l'État, dans leur étude de 2013 sur l'open data dans le monde. Cette étude distinguait quatre catégories de données transports :

- Descriptive : tarifs, noms des opérateurs, descriptif produit...
- Géographique : données géo localisées, rues, arrêts, abri bus ...
- Statistique : fréquentations moyennes, indicateurs de performance...
- Temps réel : données délivrées en temps réel : horaires de passage des bus, images de caméra...

Dans le cadre des travaux du Comité, il est toutefois apparu opportun de disposer d'une classification reflétant de manière plus exhaustive les différentes notions couramment maniées dans les métiers du transport. D'autre part, pour lever toute ambiguïté, il était utile d'établir une correspondance entre les différentes classes de données et les normes existantes, lorsque cela était possible.

Pour les services publics de transport, la figure ci-dessous présente les différentes classes de données, se distinguant chacune par des caractéristiques propres (par exemple : le producteur, le cycle de vie, les systèmes associés) et, le cas échéant, par une norme spécifique :

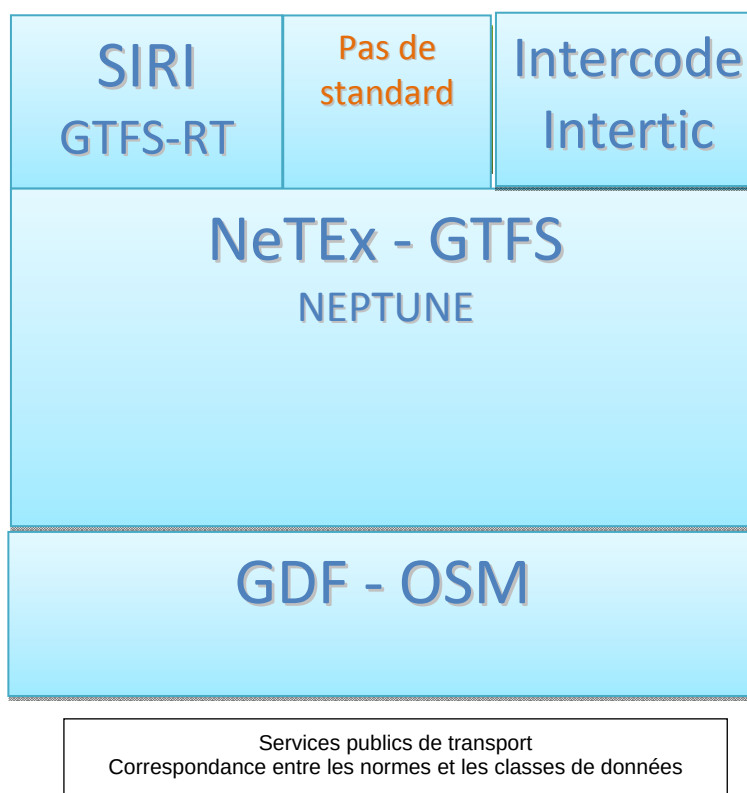


Ces classes peuvent être définies et mises en correspondance avec les normes ou standards de la façon suivante :

- « Topographie » : présentation des formes et détails visibles sur le terrain, qu'ils soient naturels ou artificiels; dans le contexte des données transports, il s'agit de la voirie (voiture, vélo, piéton), des rails, des lieux générateurs de déplacements et des points d'intérêt ;
- « Arrêts » & « Structure du réseau » : toute donnée relevant de la partie 2 de la norme Transmodel v6 et de NeTEx partie 1 ;
- « Donnée théorique – Horaires planifiés » : toute donnée relevant de la partie 3 de la norme Transmodel v6 et de NeTEx partie 2 : course commerciale, heure de passage, jour d'exploitation... ;

- « Information temps réel » : toute information relevant des services définis par la norme SIRI ;
- « Contrôle/régulation » : toute donnée relevant de la partie 5 de la norme Transmodel v6 ;
- « Usage » : toute donnée indiquant comment, quand, pourquoi et dans quelles conditions les voyageurs ont utilisé les réseaux de transport. Cette classe de données comprend les données statistiques liées aux transports (nombre de voyages, nombre de voyageurs * kilomètres, trafic journalier...) ;
- « Billettique » : toute donnée relevant de la partie 6 de la norme Transmodel v6 et des normes Interbob, Intercode, Intertic ;
- « Offre tarifaire » : toute donnée relevant de la partie 6 de la norme Transmodel v6 et de NeTEx partie 3.

La figure ci-dessous, reprenant les éléments de la figure précédente, montre la correspondance entre classes et normes :

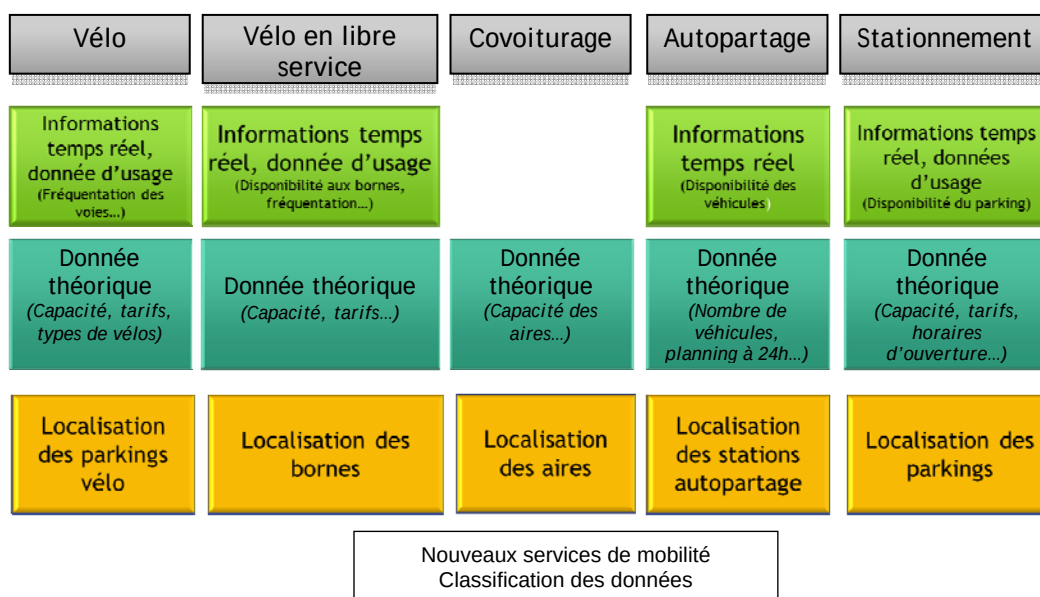


Pour les classes de données les plus sensibles dans le cadre du débat, le sous-groupe technique a fourni une liste indicative des données. Ainsi, pour les services publics de transport, la classe de données « information en temps réel » comprend notamment :

Heure de passage estimée et réalisée et informations associées (quai, etc.)	Position des véhicules	Messages diffusés sur le réseau	État des équipements (aux arrêts et dans les véhicules)	Événements, incidents et leurs conséquences	Taux de remplissage des véhicules et fréquentation
---	------------------------	---------------------------------	---	---	--

Note : le rapport du sous-groupe technique donne des compléments similaires pour les autres classes de données (cf. annexe 6 du présent rapport).

Pour les nouveaux services de mobilité, un tableau similaire aux services publics de transport a été élaboré et présente les principales classes de données pour chaque mode de déplacement.



Données brutes et informations :

En complément de cette classification des données qui permet d'avoir une vision d'ensemble du point de vue des différents métiers du transport, il est utile de connaître la distinction, couramment introduite dans le monde de l'open data, entre données brutes et informations.

Simon Chignard explique concrètement cette distinction sur son site internet⁷ :

« Une donnée est un fait brut, qui n'a pas encore été interprété. Par exemple, si en consultant un thermomètre je lis "18°C" il s'agit bien d'une donnée. Si je dis qu'il fait "plutôt doux pour la saison", on a affaire à une information. L'open data ne s'intéresse qu'aux données brutes. Les acteurs publics (et les acteurs privés) manipulent un grand nombre de données et d'informations, qu'ils mettent parfois en ligne. »

La distinction introduite porte sur le contenu : soit un « fait » objectif, soit une « information » qualitative, incertaine et qui peut être évolutive. Rappelons toutefois que, si l'on s'en tient à la définition usuelle du terme « information », une donnée brute est bien aussi une information.

Dans le cas du transport, si l'on essaie d'appliquer la distinction proposée par Simon Chignard, il apparaît que non seulement des données brutes, mais aussi des informations paraissent susceptibles d'être ouvertes de façon pertinente. Revenons à la classification des données présentée plus haut :

- la localisation d'un arrêt, l'horaire de départ d'un bus, le prix d'un abonnement mensuel sont des données brutes, des faits dans la définition ci-dessus ;
- l'heure de passage estimée du prochain train est une information et non un fait; c'est une information obtenue à partir de la donnée brute « localisation du train à telle heure » ; l'information « prochain train dans 5 mn » peut évoluer ;
- un message de perturbation tel que « à la suite de l'incident survenu à 8h15, le trafic sur la ligne se rétablit progressivement et devrait redevenir normal vers 9h30 » est une information ;
- enfin, un service de calculateur d'itinéraire fournit une information sous la forme d'une feuille de route. Cette information est élaborée en prenant en compte à la fois des données brutes (horaires), des critères retenus par l'utilisateur (temps le plus court...), des informations (temps de correspondance...) et divers algorithmes.

Dans tous ces cas, qu'il s'agisse de données brutes ou d'informations, la question de l'ouverture de la donnée ou information paraît se poser.

⁷ <http://donneesouvertes.info/>

On verra également au § 1.2 ci-dessous que le droit actuel de la réutilisation ne connaît que la notion « *d'information contenue dans un document* », « *quel que soit le support* ». Une donnée est, juridiquement, une information.

Mise à disposition statique ou dynamique :

Au paragraphe précédent, on analysait le contenu de l'information susceptible d'être ouverte à la réutilisation et il apparaissait qu'aussi bien un fait objectif qu'une information qualitative était susceptible d'être rendue réutilisable.

Si l'on se place maintenant du point de vue de la méthode utilisée en informatique, on observe deux principaux modes de mise à disposition.

Une donnée peut être diffusée de manière statique, généralement dans un fichier à plat (pdf, xls, csv, xml, shp) à l'instant t. La fréquence de mise à jour ou bien la fraîcheur de cette donnée dépendent donc du producteur. La notion usuelle d'open data concerne plutôt ce cas.

Une donnée (ou bien une information) peut également être mise à disposition de façon dynamique, sous la forme d'un flux, par l'intermédiaire d'une « interface de programmation ». Dans le langage informatique, on parle d'API, pour *Application Programming Interface*. Pour créer sa propre application, un développeur peut ainsi se connecter au flux grâce à l'API et utiliser la donnée aussi souvent qu'il le souhaite.

Ainsi, dans le domaine du transport :

- la mise à disposition statique concerne, en général, plutôt les données brutes. Dans la plupart des sites open data, on peut ainsi télécharger des arrêts de transport ou des horaires théoriques fournis dans un fichier à plat ;
- la mise à disposition dynamique, sous forme d'API, concerne plutôt le flux des informations : l'horaire du prochain passage du bus par exemple.

Pour avoir la garantie d'avoir des données à jour, il peut également être utile de proposer une API pour fournir des données brutes.

À titre d'illustration, il est intéressant de présenter une démarche menée en Allemagne à l'initiative du VDV (équivalent de l'Union des transports publics). Le VDV a défini, sous le nom de TRIAS (*Traveller Real-time Information & Advisory Services*) les spécifications d'un ensemble d'API. L'approche retenue consiste à définir des API permettant d'accéder à des données ou informations. Ces API standardisées pourront ainsi être utilisées de manière homogène par des développeurs, qui mettront en place des services d'information à l'intention des usagers.

La démarche a associé des universités et des industriels. Elle s'est appuyée sur une étude fonctionnelle approfondie, faisant intervenir des personnes-types impliquées dans des scénarios. À ce stade, chaque territoire décide d'utiliser ou non l'interface et de fournir tout ou partie des données ou informations correspondantes.

Le standard TRIAS comprend 17 services, parmi lesquels les opérateurs de transport sont invités à sélectionner ceux qu'ils souhaitent (ou peuvent) alimenter. Parmi ces services, on trouve ainsi :

Ainsi, dans le domaine du transport :

- les départs et arrivées à tel arrêt ;
- l'information en temps réel sur un déplacement donné ;
- les informations sur les correspondances pendant le déplacement (avec ré-annonce en cas de rupture) ;
- les informations et calculs tarifaires ;
- les informations pour les réservations et achat de titres ;
- les informations sur les parties d'itinéraire en transport individuel (piéton – dont prise en compte des escalators et escaliers, suivant les préférences exprimées – vélo, véhicule particulier) ;
- le signalement par les usagers des anomalies sur les équipements ;

- le service d'annonces de perturbation et de re-routage proactif ;
- la personnalisation (préférences exprimées par l'utilisateur, administration des données personnelles) ;
- les informations fournies par le véhicule (bus...) ;
- la demande au véhicule de stopper au prochain arrêt ;
- le calculateur d'itinéraire.

I. 2. – État du droit : un cadre juridique inadapté à la réutilisation des données publiques de transport

Dans le cadre juridique actuel, la majorité des données relatives à l'offre de transport dans le périmètre défini par le CIMAP, c'est-à-dire essentiellement les données des transports publics, ne sont pas réutilisables, et ce quel qu'en soit le producteur : collectivités (lorsque le service public de transport est exécuté en régie), délégataires de services publics ou entreprises publiques.

Les données publiques de transport en droit communautaire :

Le cadre juridique européen est constitué par trois directives :

1. La directive PSI (*Public Service Information*)

La directive 2003/98/CE du 17 novembre 2003 concernant la réutilisation des informations du secteur public, révisée par la directive 2013/37/UE, « *fixe un ensemble minimal de règles concernant la réutilisation et les moyens pratiques destinés à faciliter la réutilisation de documents existants détenus par des organismes du secteur public* », c'est-à-dire des organismes « *créé(s) pour satisfaire spécifiquement des besoins d'intérêt général ayant un caractère autre qu'industriel ou commercial* ».

Aux termes de l'article 2, le terme de « *document* » peut renvoyer à « *tout contenu quel que soit son support (écrit sur support papier ou stocké sous forme électronique, enregistrement sonore, visuel ou audiovisuel)* ».

L'article 1er précise que la directive ne s'applique pas « *aux documents dont la fourniture est une activité qui ne relève pas de la mission de service public dévolue aux organismes du secteur public concernés telle qu'elle est définie par la loi ou d'autres règles contraignantes en vigueur dans l'État membre* » ainsi qu' « *aux documents dont l'accès est exclu [...] pour des motifs de [...] confidentialité des informations commerciales (par exemple secret d'affaires [...]) ou secret d'entreprise* ».

Il en résulte que la directive PSI ne concerne que très peu le domaine du transport. En effet, les données détenues par les EPIC (SNCF, RATP...) ou les opérateurs de transport privés délégataires de service public n'entrent pas dans le champ d'application de la directive. Il en est de même des données couvertes par le secret commercial, quand bien même celles-ci seraient transmises à un organisme public par un opérateur de transport. Enfin, si les calculateurs d'itinéraire mis en place par les collectivités locales relèvent bien d'une mission de service public administratif, ces webservices, ne sauraient être considérés comme des « documents » au sens de la directive PSI. Par conséquent, ils n'entrent pas non plus dans son champ d'application.

L'article 6 de la directive dispose que « *lorsque la réutilisation de documents est soumise à des redevances prélevées par des organismes du secteur public, ces redevances devraient, en principe, être limitées aux coûts marginaux* ». Seuls certains organismes sont autorisés à pratiquer des redevances supérieures à ce coût (les services culturels et des établissements tels que l'INSEE qui sont tenus de générer des recettes pour couvrir leurs coûts de fonctionnement). Le principe est de ne faire payer que les coûts spécifiquement liés à la mise à disposition des données, celles-ci étant détenues par les organismes publics dans le cadre de l'exécution de leur mission de service public, pour laquelle les contribuables ont déjà consenti un effort.

La directive PSI ne définit pas ce que recouvre au juste ce « *coût marginal* ». Toutefois, une communication de la Commission⁸ précise que « *les coûts marginaux pourraient être définis comme les coûts directement liés et nécessaires à la reproduction d'un exemplaire supplémentaire d'un document et à sa mise à la disposition des réutilisateurs* ».

La communication indique que « *dans le calcul des redevances, les coûts suivants pourraient être considérés comme éligibles* :

- *infrastructure* : coût du développement, de la maintenance logicielle, de la maintenance matérielle et de la connectivité, dans les limites de ce qui est nécessaire pour rendre les documents accessibles et réutilisables ;
- *duplication* : coût d'un exemplaire supplémentaire de DVD, clé USB, carte SD, etc. ;
- *gestion* : matériel d'emballage, préparation de la commande ;
- *consultation* : communications téléphoniques et échange de messages électroniques avec les réutilisateurs, coûts du service à la clientèle ;
- *livraison* : coûts du port, qu'il s'agisse d'affranchissement standard ou de messagerie rapide ;
- *et demandes spéciales* : coûts de la préparation et du formatage des données à la demande ».

Il est encore précisé : « *Lorsque des documents non numériques sont diffusés physiquement, la redevance peut être calculée sur la base de toutes les catégories de coûts ci-dessus. Toutefois, dans un environnement en ligne, les redevances totales pourraient être limitées aux coûts directement liés à la maintenance et au fonctionnement de l'infrastructure (base de données électronique), dans les limites de ce qui est nécessaire pour reproduire les documents et les mettre à disposition d'un réutilisateur supplémentaire* ».

La Commission conclut : « *Comme les coûts de fonctionnement moyens d'une base de données sont faibles et ont tendance à baisser, le chiffre sera probablement proche de zéro. Il est donc recommandé que les organismes du secteur public évaluent régulièrement les coûts et avantages potentiels d'une politique de coût nul et d'une politique de coûts marginaux, en gardant à l'esprit que l'application de redevances a aussi un coût (gestion des factures, suivi et encadrement, paiements, etc.). En conclusion, il est possible d'appliquer la méthode des coûts marginaux pour recouvrer les coûts liés à la reproduction de documents non numériques supplémentaires et à leur diffusion physique tandis que, lorsque des documents numériques (fichiers) sont diffusés par voie électronique (téléchargement), une méthode de coût nul pourrait être recommandée* ».

Toutefois, l'emploi du conditionnel (« *pourraient* ») indique qu'il est permis de ne pas limiter la redevance aux « *coûts directement liés à la maintenance et au fonctionnement de l'infrastructure* » ; il est donc possible d'intégrer dans l'assiette de la redevance les coûts de développement d'une infrastructure ou les coûts de formatage des données lorsque celui-ci s'effectue à la demande du réutilisateur et non pour répondre aux besoins propres du fournisseur.

Très concrètement, la directive PSI nous semble autoriser de faire supporter aux réutilisateurs les coûts de mise en place d'un nouveau serveur ou de mise au format des données engendrés par la mise à disposition des données à des tiers.

La directive PSI, dans sa version révisée de 2013, n'a pas encore fait l'objet d'une transposition en droit interne. Bien qu'a priori cette directive n'impacte que très marginalement le domaine du transport, il convient toutefois de veiller à la cohérence de sa transposition avec les modifications législatives résultant des propositions du présent rapport.

C'est pourquoi, dans un souci de lisibilité, le Comité propose d'appliquer à l'ensemble des données publiques de transport la méthode de tarification au coût marginal, que ces données n'entrent ou n'entrent pas dans le champ de la directive PSI.

2. La directive ITS (*Intelligent Transport System*)

⁸ Communication de la Commission, *Orientations sur les licences types recommandées, les ensembles de données et la tarification de la réutilisation des documents*, 2014/C 240/01, 24 juillet 2014 : <http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52014XC0724%2801%29&from=FR>

La directive 2010/40/UE du 7 juillet 2010 sur les systèmes de transports intelligents définit plusieurs actions prioritaires destinées à favoriser l'émergence de services d'information dans l'Union européenne. Les trois premières actions prioritaires traitent des services d'information à destination de l'utilisateur des transports :

- « a) la mise à disposition, dans l'ensemble de l'Union, de services d'information sur les déplacements multimodaux ;
- b) la mise à disposition, dans l'ensemble de l'Union, de services d'information en temps réel sur la circulation ;
- c) les données et procédures pour la fourniture, dans la mesure du possible, d'informations minimales universelles sur la circulation liées à la sécurité routière gratuites pour les usagers ».

Chacune de ces actions prioritaires donne lieu à l'adoption par la Commission Européenne d'un règlement délégué :

- le règlement de l'action prioritaire c) concernant les informations de sécurité routière a été publié. Il comporte des obligations à la charge d'acteurs publics et privés : tout acteur disposant de données de sécurité, qu'il soit public ou privé, doit les rendre accessibles par l'intermédiaire d'un « *point d'accès national* » (TIPI en France). Les données doivent être mises à la disposition des usagers finaux gratuitement ou, plus exactement, sans surcoût par rapport au coût des services dont ils disposent. Dans ces conditions, l'utilisateur final ne dispose pas nécessairement de l'ensemble des données de sécurité. La diffusion des données s'effectue selon les modèles économiques existants. Si un service payant est déployé par un acteur détenteur de données de sécurité, celui-ci doit fournir les informations de sécurité sans surcoût ;
- le règlement concernant l'action prioritaire b) relative à l'information routière en temps réel est paru le 18 décembre 2014. Il comporte des obligations à la charge de l'autorité publique gestionnaire de l'infrastructure routière et de l'opérateur, public ou privé, chargé de la maintenance et de l'exploitation.

Les actions prioritaires b) comme c) portent sur le réseau principal, c'est-à-dire sur le réseau routier transeuropéen, les autoroutes non comprises dans ce réseau et des zones prioritaires. Il ne s'agit pas d'obliger les acteurs à créer les données, mais, lorsqu'elles existent, à les rendre réutilisables par les tiers. La mise à disposition des données doit être effectuée de façon non discriminatoire et par l'intermédiaire d'un point d'accès national unique. Les données doivent être mises à jour. Pour l'action c), les événements susceptibles d'être couverts par le service sont limitativement énumérés (travaux routiers de courte durée, conducteur en contresens...) Pour l'action b) est indiquée une liste de données de type statique, dynamique ou relatives au trafic.

Dans ces deux actions prioritaires b) et c), l'ouverture des données obéit à des approches différentes : l'action c) portant sur les données de sécurité impose des obligations aux détenteurs de ces données, qu'ils soient publics ou privés, tandis que l'action b) portant sur les services d'information en temps réel n'impose des obligations qu'aux personnes chargées d'une mission de service public.

- pour l'action prioritaire a) relative à la mise en place de services d'information multimodale dans l'Union européenne, la Commission vient de mettre en place le Comité d'experts chargé d'accompagner l'élaboration du projet de règlement délégué. L'objectif est de publier ce règlement en 2015.

Enfin, dans la feuille de route sur les services d'information multimodale à l'échelle européenne publiée en juin 2014⁹, la Commission annonce le lancement d'une « *Initiative pour l'accès aux données* », qui consisterait en une nouvelle directive.

Par analogie avec la démarche mise en œuvre pour les actions b) et c), on peut supposer que le règlement délégué relatif à l'action a) aura notamment pour objet de rendre réutilisables les données nécessaires à la mise en place de services d'information multimodale.

⁹ Commission Staff working document, Towards a roadmap for delivering EU wide multimodal travel information, planning and ticketing services : <http://ec.europa.eu/transport/themes/its/doc/swd%282014%29194.pdf>

Il en résulte que les recommandations du présent rapport sont destinées à préparer aussi bien l'adoption de mesures législatives en France que la position française dans les travaux européens. Il conviendra d'être très attentif aux calendriers respectifs de mesures législatives françaises et des textes européens.

3. La directive INSPIRE

La directive 2007/2/CE du 14 mars 2007 impose des obligations de publication et de partage des données environnementales et géographiques, notamment pour les données relatives aux réseaux de transport (cf. annexe I de la directive). Toutefois, cette directive ne traite pas de la réutilisation de ces données.

Les données publiques de transport en droit interne :

Introduction

Le fondement juridique de l'ouverture des informations publiques est constitué par la loi n° 78-753 du 17 juillet 1978, dite « loi CADA ». Le chapitre Ier du titre Ier crée un droit à communication des documents administratifs définis comme les « *documents produits ou reçus, dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission* » (article 1er).

Le chapitre II, de rédaction plus récente (2005), transpose la directive PSI de 2003 en créant un droit à réutilisation des informations « *figurant dans des documents produits ou reçus par les administrations mentionnées à l'article 1er, quel que soit le support [...]* ». La réutilisation s'entend comme une utilisation « *à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus* » (article 10).

Ce même article 10 exclut du droit à réutilisation les informations contenues dans des documents produits ou reçus dans l'exercice d'une mission de service public à caractère industriel (SPIC).

Les données des services publics à caractère industriel et commercial ne sont pas réutilisables

Parmi les exceptions au droit à réutilisation énumérées à l'article 10 figurent les informations « *produites ou reçues dans l'exercice d'une mission de service public à caractère industriel et commercial* ». Or, l'exécution du service de transport public de personnes constitue un service public à caractère industriel et commercial (SPIC) aux termes de l'article L. 1221-3 du code des transports.

Quant aux données nécessaires à l'information du voyageur, objet des recommandations du présent rapport, elles apparaissent bien d'une manière générale liée à l'exécution du service et non pas à l'organisation du transport public (article L. 1211-4), qui constitue un service public à caractère administratif (SPA). La chose est évidente pour les données en temps réel. Pour les données horaires théoriques, même en admettant que l'autorité organisatrice définisse une fréquence sur la ligne, c'est bien l'opérateur de transport qui fixe les horaires de passage. De même, pour les arrêts, même en admettant que l'autorité organisatrice en définisse le principe, c'est bien aussi l'opérateur de transport qui en définit précisément l'emplacement.

Il en résulte que les données nécessaires à l'information du voyageur, lorsqu'elles concernent un service public de transport, ne sont en général pas réutilisables. Peut-être pourrait-on démontrer que certaines données le sont, mais l'exercice paraît tellement ardu et incertain qu'il paraît préférable de trancher le nœud gordien en adoptant des dispositions législatives claires pour les données que l'on entend rendre réutilisables.

Pour éviter toute ambiguïté, il convient à ce stade de souligner la distinction entre les notions « *d'accès aux documents administratifs* » et de « *réutilisation des informations publiques* » visées respectivement par les chapitres Ier et II du titre Ier de la loi CADA, et qui se recoupent sans coïncider parfaitement (cf. annexe 4). Les données relatives à l'offre de transport détenues par les

collectivités, les délégataires de service public ou les entreprises publiques de transport sont ainsi communicables, mais pas réutilisables de droit.

Dans ce contexte, le débat avait fondamentalement pour objet d'examiner l'opportunité de mettre fin à l'exclusion des données des SPIC du champ de la réutilisation, dans le domaine du transport.

Un récent avis de la CADA affirme que toute donnée transport devient réutilisable dès lors qu'elle est collectée par une administration aux fins d'information du voyageur

L'avis *Agence française pour l'information multimodale et la billettique* rendu par la CADA en 2011¹⁰ apporte un tempérament au caractère non-réutilisable des données des SPIC et, plus généralement, des données privées, lorsqu'elles sont collectées par l'administration.

Dans cet avis, en effet, la CADA considère que les informations sur les transports revêtent le caractère d'informations publiques réutilisables, « *quelle que soit leur provenance* », dès lors elles sont collectées dans l'exercice d'une mission de service public à caractère administratif. Constitue une telle mission l'instauration d'un service d'information multimodale « *dans les périmètres de transports urbains inclus dans des agglomérations de plus de 100000 habitants* » (article L. 1231-8 du code des transports). La mission d'information dévolue aux autorités organisatrices du transport public de personnes par cet article se rattache à leurs compétences en matière d'organisation du transport public (SPA) et non à l'exécution du service de transport public (SPIC).

Les données des SIM seraient donc réutilisables de droit. Les seules données non réutilisables seraient celles protégées par le secret en matière commerciale et industrielle (art. 6 de la loi CADA) ou grevées d'un droit de propriété intellectuelle (art. 10).

Les conclusions de la CADA ne se limitent pas aux SIM mis en place dans les périmètres de transports urbains inclus dans des agglomérations de plus de 100000 habitants, mais peuvent être étendues à tout système d'information sur les transports mis en place par l'État, les collectivités territoriales et leurs établissements publics, en liaison avec les entreprises privées ou publiques. En effet, si, dans son avis, la CADA ne fait mention que de l'article L. 1231-8, elle s'est également appuyée¹¹, sur l'article L. 1211-4 du code des transports qui fait du « *développement de l'information sur le système des transports* » une mission de service public. C'est ainsi que « *l'ensemble des données collectées dans ce cadre [...] sont a priori des informations publiques* », telles les données du système d'information multimodale Destineo de la Région des Pays de la Loire.

Une analyse équivalente permet de conclure que les données collectées par l'AQST¹² sont également réutilisables.

La validité de l'avis de la CADA a été discutée par le sous-groupe juridique, dont certains membres pointent le risque de retrait des données versées dans les SIM par les opérateurs de transport soucieux de se prémunir d'une réutilisation par des tiers. Les interrogations formulées portent sur le fait que la CADA considère que la provenance des informations est sans incidence sur leur caractère réutilisable. Le caractère réutilisable d'une information devrait se comprendre comme une notion relative : une information produite dans le cadre d'un SPIC n'est pas réutilisable, mais peut l'être si elle est reçue dans le cadre d'un SPA.

La question de l'articulation de cette interprétation avec l'exception SPIC posée par l'article 10 de la loi a été soulevée. En effet, si la première phrase du premier alinéa de l'article 10 pose bien le principe général selon lequel les informations « *produites ou reçues* » dans l'exercice d'une mission de service public sont réutilisables, la seconde précise que cette réutilisation comprend des « *limites et conditions* », dont notamment l'exception b) des informations « *produites ou reçues* » dans l'exercice d'une mission de SPIC.

¹⁰ CADA, Conseil n° 20104559 du 17 février 2011, *Agence française pour l'information multimodale et la billettique*.

¹¹ Comme la CADA le précise dans un échange informel, cf. courriel du 19 avril 2011 adressé à l'AFIMB.

¹² L'Agence est chargée de « *diffuser [...] des informations statistiques relatives à la qualité dans les services de transport relevant de sa compétence* », à savoir, « *les transports terrestres, aériens et maritimes, publics et réguliers de voyageurs* » (cf. décret n° 2012-211 du 14 février 2012, art. 2).

À la lumière de ce texte, comment analyser le cas d'une information produite par un SPIC et reçue par un SPA ? Faut-il considérer que l'exception joue au motif que cette information a été produite par un SPIC ? Selon la CADA, dès lors que cette information est reçue par un organisme chargé d'une mission de SPA, elle n'est pas dans l'exception du b). Cette interprétation a paru correcte au Comité, qui estime néanmoins qu'elle est porteuse d'insécurité juridique.

Plus généralement, toute information, qu'elle soit produite dans l'exercice d'une mission de SPIC ou par un opérateur privé en dehors de toute mission de service public, devient réutilisable dès lors qu'elle est reçue par un SPA.

Le Comité observe l'absence d'enjeu sur cette fragilité éventuelle de l'avis de la CADA. En effet, ses recommandations conduisent bien à rendre réutilisables les données des SPIC dans le domaine du transport. Quant à l'activité TGV, le Comité n'a pas considéré qu'il s'agissait d'un service public, du moins sur le plan strictement juridique.

Le statut juridique des services de transport est parfois difficile à déterminer

Selon la lettre de mission, elle-même élaborée à partir des décisions prises par le CIMAP au sujet du contenu des débats thématiques, le Comité a tout d'abord examiné les possibilités :

- d'élargissement de la stratégie d'ouverture des données publiques aux délégataires de service public et aux collectivités territoriales ;
- d'extension de la politique d'ouverture des données publiques aux missions de service public à caractère industriel et commercial et aux entreprises publiques.

Toutefois, le périmètre assigné au débat, circonscrit aux données produites dans l'exercice d'une mission de service public, s'est révélé d'application délicate au domaine des transports : il n'est pas toujours aisé de déterminer si un service de transport relève d'un SPIC et SPA, voire s'il relève seulement d'une mission de service public. Pour ces raisons, appliquer la loi CADA au cas du transport est particulièrement délicat.

Examinons les quatre ensembles de modes de transport définis lors de la mise au point de la méthode de travail du Comité :

1. Les services publics de transport comprennent les services urbains (bus, tram, métro...), départementaux (autocars...) ou régionaux (TER) de voyageurs, ainsi que les trains Intercités (trains d'équilibre du territoire). Ce sont des services publics, parfois exécutés en régie, et, majoritairement, des services publics à caractère industriel et commercial. Précisons que les trains Intercités, qui circulent sur la moyenne et longue distance ferroviaire, assurent une mission de service public sur 35 liaisons non assurées par la grande vitesse. Ces liaisons ont fait l'objet d'une convention avec l'État, notamment du fait de leur absence de rentabilité.
2. les services concurrentiels de transport longue distance comprennent notamment le TGV, le covoiturage et les transports aérien et maritime.
Le TGV, exploité par l'EPIC SNCF « selon les principes du service public¹³ », et non pas par l'une de ses filiales, relève, à ce titre, du débat ; la formulation « selon les principes... » entretient toutefois l'incertitude sur la question de savoir s'il s'agit d'un service public ou d'une activité commerciale. En tout état de cause, sur les lignes dénuées d'alternative, le TGV assure de fait, aussi, une mission de service public. Toutefois, la tarification en yield management est caractéristique d'une activité commerciale.
Les modes aérien et maritime, ainsi que le covoiturage longue distance relèvent quant à eux de sociétés privées ; les compagnies aériennes et maritimes ne sont délégataires de service public que dans un nombre très limité de cas¹⁴.
TGV, air et covoiturage longue distance étant en concurrence, il paraît justifié que les obligations en termes d'ouverture des données qui s'imposent à eux soient similaires.

¹³ Article L. 2141-1 du Code des transports.

¹⁴ 28 liaisons aériennes et une poignée de dessertes maritimes (Corse, îles Bretonnes etc.) sont sous obligation de service public.

3. les services de mobilité comprennent le stationnement, le vélo, ainsi que de nouveaux services comme le vélo en libre-service (VLS), le covoiturage et l'autopartage. La situation pour ces différents services apparaît contrastée : si le statut juridique des données vélo et stationnement est globalement aisé à déterminer, tel n'est pas le cas pour les données des autres services.
- Les données vélo sont en grande partie réutilisables : les voies cyclables sont des infrastructures routières ou des voies propres gérées par des collectivités territoriales dans le cadre d'une mission de SPA. En revanche, les données relatives aux équipements (enclos sécurisés, gonfleur automatique...) et aux services à destination des cyclistes (consignes, location longue durée, prêt de casques, location de remorques enfants¹⁵...) sont a priori non-réutilisables, puisque majoritairement produites dans le cadre d'une délégation de service public, ce qui fait peser une présomption de SPIC. Ainsi, l'espace de consigne aménagé par la ville de Bourg-en-Bresse est géré en partenariat avec une société privée (SAGS). De même, la gestion du service de location longue durée V'Lille (Lille) a été confié à Keolis. De nombreux services sont gérés par des sociétés d'économie mixte, comme NGE à Nantes (Métropole à Vélo) ou Sémitag à Grenoble (Métrovélo). Il existe aussi des services de location purement privés : associatifs sur les campus (Vélocampus à Nantes) ou commerciaux dans les zones touristiques (Astel au château de Versailles). d'extension de la politique d'ouverture des données publiques aux missions de service public à caractère industriel et commercial et aux entreprises publiques.
 - La grande majorité des données stationnement sont des données SPIC non-réutilisables. Rappelons que 80% des parcs de stationnement sont gérés par des sociétés privées dans le cadre d'une délégation de service public (cf. annexe 10).

Pour les données des autres services de mobilité, l'analyse juridique se révèle plus ardue, d'autant que la loi n° 2014-58 du 27 janvier 2014, dite loi « MAPAM » permet aux AOM d'intervenir en cas d'inadaptation ou d'insuffisance de l'offre privée en matière de VLS ou de covoiturage¹⁶.

- Ainsi, un service de VLS peut relever soit d'une activité privée, soit, dans l'écrasante majorité des cas, d'un service public, géré en régie comme à Belfort (Optymo) ou délégué à une entreprise privée (Vélodi à Dijon, géré par JCDecaux). Il est toutefois impossible de conclure au caractère réutilisable des données produites dans le cadre d'une mission de service public de VLS, puisque la loi MAPAM ne qualifie pas ce service (SPA ? SPIC ?).
- Le statut juridique des données covoiturage est encore plus délicat à déterminer, dans la mesure où la loi MAPAM précitée n'octroie pas la qualification de service public aux plates-formes dématérialisées facilitant la rencontre des offres et demandes de covoiturage que les AOT ont la possibilité de mettre en place. Les données de ces plates-formes ne seraient donc pas réutilisables, ce que semble indiquer les conditions générales d'utilisation du site covoiturage.orne.fr créé par le Conseil Général de l'Orne, qui donne compétence aux tribunaux (judiciaires) d'Alençon (et non au tribunal administratif de Caen) pour trancher les litiges.
- La situation des données autopartage est particulièrement complexe. Si les données produites dans le cadre de délégations de service public (Autolib' Lyon géré par la SEM Lyon Parc Auto) sont vraisemblablement non-réutilisables, on ne saurait aboutir à une conclusion aussi tranchée pour les données produites dans le cadre d'un service régional d'autopartage tel que Regionlib dont la gestion est assurée par une société publique locale, la SA Poitou-Charentes AutoPartage, qui a vocation à intervenir pour le compte de ses actionnaires (sept collectivités locales) dans le cadre de prestations intégrées (quasi-régie ou « *in house* ») ; on ne saurait donc totalement exclure que Regionlib puisse recevoir la qualification de SPA¹⁷. En revanche, il paraît acquis que les données produites

¹⁵ À l'exclusion du VLS, qui fait l'objet d'une rubrique distincte.

¹⁶ Article 52 de la loi MAPAM, codifié aux articles L. 1231-1-14 à L. 1231-1 du code des transports.

¹⁷ Cf. circulaire n° COT/B/11/08052/C du 29 avril 2011 : « Si l'alinéa 2 de l'article L. 1531-1 du CGCT ne vise expressément que l'exploitation des services publics à caractère industriel ou commercial, il n'en demeure pas moins que les SPL pourront aussi se voir confier l'exploitation de services publics à caractère administratif (« Ces sociétés sont compétentes pour [...] »).

dans le cadre du service d'autopartage géré par Autocool (devenue Citiz Bordeaux en 2013), une société coopérative d'intérêt collectif, ne sont pas non-réutilisables ; en effet, bien que ce service soit subventionné par la Communauté urbaine de Bordeaux¹⁸, il s'inscrit hors du cadre strict d'une mission de service public.

Pour les nouveaux services de mobilité, la difficulté est donc tout autant de trancher ce qui relève d'un SPIC ou d'un SPA que de déterminer où commence et où s'arrête le service public. D'où l'opportunité d'adopter pour ces services une approche différente de celle retenue pour les services publics de transport.

4. La route. Les infrastructures routières qui composent le domaine public routier sont gérées dans le cadre d'une mission de service public administratif ; les données correspondantes sont réutilisables sans qu'il soit besoin de modifier le cadre législatif actuel.

I. 3. – Panorama des initiatives d'ouverture des données en France

Un certain nombre d'acteurs du transport, en France, ont pris l'initiative d'ouvrir des données de l'offre de transport. Il s'agit notamment :

- de collectivités territoriales (Rennes Métropole¹⁹, Communauté Urbaine de Bordeaux²⁰, conseil général de Loire-Atlantique²¹) ou bien de syndicats mixtes de transports comme le SMT13²² ;
- d'opérateurs de transport en charge d'une mission de service public (RATP²³, SNCF Proximités²⁴).

Il convient également de mentionner des initiatives privées : données d'OpenStreetMap ou bien de la plate-forme « navitia.io », réutilisant des données publiques.

De premières études réalisées par l'AFIMB²⁵ et le CEREMA²⁶ se sont attachées à analyser ces initiatives d'ouverture de données « publiques » en France. Les présentations en réunions plénières du Comité, ainsi que les travaux du sous-groupe technique ont permis d'enrichir ce panorama.

Les principales caractéristiques des initiatives d'ouverture des données publiques en France sont présentées ci-après selon deux axes :

- les données transports ou mobilité mises à disposition du réutilisateur ;
- les conditions de réutilisation.

Les données publiques mises à disposition :

En s'appuyant sur la classification des données des services publics de transport, le tableau ci-dessous présente, de manière non exhaustive, les données ou informations mises à disposition des réutilisateurs sur différentes plates-formes open data.

Quelques tendances se dégagent :

- Rennes et SNCF ouvrent non seulement des données théoriques, mais aussi temps réel ;
- le STIF ouvre un calculateur d'itinéraires ;
- une faible ouverture des données d'usage et d'offre tarifaire ;
- l'absence d'ouverture des données billettiques.

exploiter des services publics à caractère industriel ou commercial ou toutes autres activités d'intérêt général ») ; document téléchargeable à l'adresse : http://circulaire.legifrance.gouv.fr/pdf/2011/05/cir_33089.pdf

¹⁸ http://www.lacub.fr/sites/default/files/synchro_docs/ConseildeCommunaute_2/2012/2012_07_13/P0EEP.pdf

¹⁹ <http://www.data.rennes-metropole.fr/>

²⁰ <http://data.lacub.fr/>

²¹ <http://data.loire-atlantique.fr/>

²² http://www.lepilote.com/presentation/?rub_code=92

²³ <http://data.ratp.fr/fr/les-donnees.html>

²⁴ <https://data.sncf.com/>

²⁵ http://www.developpement-durable.gouv.fr/IMG/130408_OpenData_Rapport%20final.pdf

²⁶ <http://www.certu.fr/open-data-les-travaux-du-cerema-a1219.html>

<u>Classes de données</u>	<u>STIF</u>	<u>Rennes</u>	<u>RATP</u>	<u>SNCF Proximité (TER, Transilien, Intercités)</u>
Topographie		Voies, tronçons de voies Adresses	Coordonnées (x ; y) des stations ferrées (barycentrées) et des 12000 arrêts de bus	Voies et tronçons de voies (sur OSM) Adresses sur SNCF
Arrêt	Lieu d'arrêt, Correspondances	Lieu d'arrêt	Lieu d'arrêt Correspondance	Lieu d'arrêt Entrées/ sorties du Transilien (OSM)
Structure du réseau	Lignes	Lignes	Lignes Plan du réseau Accessibilité arrêts, lignes & véhicules	Lignes Équipements accessibles en gare et sur quai (limité au Transilien)
Données théoriques – Horaires planifiés	Horaires planifiés	Horaires planifiés, Position des points de ventes	Horaires planifiés, Position des points de vente	Horaires planifiés Position des points de ventes (sur OSM)
Information en temps réel		Prochain départ/arrivée à l'arrêt Perturbation État des équipements		Prochain départ/arrivée à l'arrêt
Usage			Fréquentation (limité au réseau ferré)	Fréquentation (limité au Transilien)
Offre tarifaire				Tarif des abonnements
Services liés à l'information du voyageur	Calculateur d'itinéraires, prix d'un trajet, horaires à l'arrêt et à la destination			

Services publics de transport
Exemples de données mises à disposition sur des portails internet

S'agissant des nouveaux services de mobilité, le tableau ci-dessous présente des exemples d'initiative d'ouverture de données pour l'autopartage et le vélo :

<u>Classes de données</u>	<u>Autopartage (Autolib²⁷)</u>	<u>Autopartage (Citiz²⁸)</u>	<u>Vélo (ON3V²⁹)</u>	<u>Vélo en libre-Service (JC Decaux³⁰)</u>
Topographie			Véloroute et voies vertes	
Arrêt	Localisation des stations Nom de la station Nombre de places	Localisation des stations Liste des stations gérées par un opérateur Liste des véhicules à une station Photo d'un véhicule		Localisation des stations VLS
Information en temps réel				État de la station Nombre de points d'attache opérationnels Nombre de points d'attache disponibles pour y ranger un vélo Nombre de vélos disponibles et opérationnels
Services liés à l'information du voyageur		Calcul du prix prévisionnel à partir de critères (date, opérateur)		

Nouveaux services de mobilité
Exemples de données mises à disposition sur des portails internet

²⁷ http://opendata.paris.fr/explore/dataset/stations_et_espaces_autolib_de_la_metropole_parisienne/?tab=table

²⁸ <http://production.citiz.fr/doc/>

²⁹ <http://www.departements-regions-cyclables.org/page/telecharger--p-73.html>

³⁰ <https://developer.jcdecaux.com/#/opendata/vls?page=getstarted>

Ce tableau ne prétend pas à l'exhaustivité et d'autres démarches pourraient être mentionnées, telles que l'élaboration d'un registre national recensant les données de disponibilité et de géolocalisation des taxis, conformément aux dispositions de la loi n° 2014-1104 du 1er octobre 2014 relative aux taxis et aux voitures de transport avec chauffeur.

Les conditions de réutilisation :

Que ce soit sur le plan technique ou juridique, l'analyse des initiatives d'ouverture des données met en évidence des démarches différentes.

Données statiques et dynamiques

La mise à disposition statique des données est la plus fréquente. Il s'agit de la possibilité pour les réutilisateurs de télécharger sur un portail internet des fichiers informatiques à plat : .xls, .csv ou encore .xml.

Des dispositifs d'accès dynamique, au travers de flux, commencent à apparaître pour les informations en temps réel (prochain passage du bus, état des équipements). L'accès aux informations temps réel de la STAR (Rennes-Métropole) ou de Tisséo est ainsi rendu possible sous la forme d'une connexion à un serveur, grâce à une API.

Relève également de la catégorie accès dynamique l'ouverture d'un calculateur d'itinéraires, pratiquée par exemple par le STIF : dans une logique d'*open service*, les réutilisateurs peuvent adresser des requêtes au système Vianavigo et en recevoir des réponses. Une dizaine d'entreprises ont signé la convention proposée par le STIF : Mappy, Urban Pulse, ADP...

Format et qualité des données

Trois types de format sont principalement utilisés dans la mise à disposition des données transports :

- les formats correspondant à une norme (Neptune) ou à un standard (GTFS) ;
- les formats de type tabulaire (.xls ou .csv) ;
- les formats orientés « SIG » (.shp ou .kml).

S'agissant des API (données temps réel ou calculateur d'itinéraires), le format proposé est systématiquement propre au fournisseur. Pourtant, des normes ou standards existent.

L'étude de 2013 de l'AFIMB sur l'open data propose une analyse du contenu des données transports mises à disposition en open data, de leur qualité et, le cas échéant, de leur conformité à la norme. Les tests ont été effectués à l'aide de l'outil open source CHOUETTE.

Les constats suivants se dégagent :

- la position géographique des arrêts s'avère parfois inexacte : la simple visualisation des arrêts sur la carte de l'outil CHOUETTE permet ce constat (arrêt placé sur un cours d'eau...) ;
- les arrêts ne se voient pas toujours attribuer un identifiant ou celui-ci n'est pas unique ;
- sur la sémantique des données : les données ne sont pas toujours définies. Ainsi, pour les arrêts, on ne sait pas toujours s'il s'agit d'arrêts commerciaux, de points d'embarquement, de quais ou de points d'accès ;
- des doublons peuvent subsister : par exemple, deux correspondances au sein d'un même arrêt commercial avec des temps de correspondance différents, pour un même type d'utilisateur.

Plates-formes de données

On observe l'apparition de plates-formes de données territoriales (transport ou mobilité), qui permettent de faciliter la réutilisation des données en offrant un point d'entrée unique sur un territoire.

Dans le cadre du projet Optimod'Lyon, une plate-forme de données théoriques et temps réel a ainsi agrégé les données de différents opérateurs : routier, parking, vélo, vélo en libre service...

De même, un projet de l'agglomération de Montpellier vise à créer une plate-forme de données multimodales urbaine, en agrégeant des données statiques, temps réel et *analytics* (temps réel retraité) pour différents modes : vélo, vélo en libre service, parking...

À l'échelle de la région Île-de-France, le STIF développe actuellement une base de données d'arrêts avec l'ensemble des opérateurs de transport (base dite REFLEX), dont la mise à disposition des tiers est envisagée. De manière similaire, sous l'impulsion du SMIRT, une base de données « Points de Mobilité » est également en cours de création dans le Nord-Pas-de-Calais.

Portails internet permettant d'accéder aux données

La structuration et le contenu des portails internet d'accès aux données ne sont que faiblement harmonisés.

Les données transport peuvent, par exemple, se retrouver dans des rubriques variables d'une plate-forme à l'autre ou d'un mode de transport à l'autre : ex. « transport », « mobilité », « déplacement », « cadre de vie », « localisation », « équipements ».

Par ailleurs, les intitulés pour une même donnée peuvent eux aussi varier. Pour une donnée telle que les « horaires de transports », on trouve différentes terminologies : « Offre », « Description complète des lignes », « Données théoriques », « Réseau », « Horaires »...

Labellisation

En complément à l'ouverture de leurs données, certains acteurs ont souhaité procéder à la labellisation des services les réutilisant. Pour Rennes, pionnier dans l'open data, la labellisation est apparue comme une alternative au contrôle a priori de l'usage de la donnée. De son côté, le Grand Lyon a créé un label (Onlymoov) pour donner de la visibilité à des PME innovantes, répertoriées sur son site. De même, la plate-forme open data de la RATP référence les applications les plus innovantes et un label est à l'étude.

Les licences de réutilisation

Les fournisseurs de données optent pour l'*Open Database License* (ODbL), la Licence Ouverte ou encore des licences « maison », notamment pour les données temps réel (cf. annexe 3). Les données mises à disposition sur une même plate-forme peuvent être soumises à des licences différentes, en fonction de leurs spécificités.

Les collectivités territoriales concernées expliquent leurs choix de la façon suivante :

- le Grand Lyon propose trois licences, adaptées chacune à une certaine catégorie de données. Pour la majorité des données, il s'agit de la Licence Ouverte. Le Grand Lyon considère en effet la licence ODbL comme faisant obstacle à la réutilisation commerciale des données, qu'il entend précisément promouvoir. Par ailleurs, le Grand Lyon a voulu réserver un traitement spécifique à des données identifiées comme sensibles, notamment les données en temps réel. Les deux licences destinées à ces données imposent une identification du réutilisateur et le respect des politiques publiques de la collectivité territoriale. Ces deux licences se distinguent par le fait que l'une (licence dite « engagée ») est gratuite, tandis que l'autre (licence dite « associée ») prévoit une redevance (voir à ce sujet le paragraphe **Redevance versus principe de gratuité**, ci-dessous).
- A contrario, Rennes Métropole n'a pas estimé nécessaire de prendre de précautions particulières concernant ses données transport. Son représentant déclare n'avoir pas observé, en plusieurs années, de cas de mauvais usage de ses données, mais reconnaît que les enjeux pourraient être autres dans des agglomérations de plus grande taille. La licence ODbL a été adoptée en raison de son modèle coopératif, plus conforme, selon Rennes, aux attentes de la communauté open data.

- Enfin, l'accès au service de calcul d'itinéraires du STIF donne lieu à la conclusion d'un contrat de service, dans lequel le fournisseur des données prend des engagements sur le niveau de performance de son service (tenue à la charge, disponibilité...)

On observe parfois une certaine confusion entre plusieurs documents destinés à définir les conditions d'utilisation des données, notamment les conditions générales d'utilisation (CGU) du portail et la licence de réutilisation des données.

Redevance versus principe de gratuité

Le principe de gratuité prévaut pour la majorité des initiatives d'ouverture des données en France. Ce principe peut s'appliquer à la fois pour les données transport théoriques, mais également pour le temps réel (Rennes, Toulouse, SNCF Proximité...).

Le Grand Lyon part également du principe de gratuité, mais se fixe l'objectif supplémentaire de « *garantir un écosystème concurrentiel équitable* ». Il s'agit d'encourager les initiatives de PME innovantes et d'éviter la formation d'une situation de monopole. À cet effet, dans la licence dite « associée », la réutilisation des données à fort potentiel commercial est soumise au paiement d'une redevance d'un montant élevé pour les acteurs ayant une position concurrentielle forte. Une enquête réalisée sur l'agglomération de Lyon a montré que deux services internet avaient une part de marché dominante, celui du Sytral pour les transports publics et Google Maps pour la route.

Le STIF offre, quant à lui, un service assorti d'engagements de performance. Il justifie la redevance demandée par le coût de mise à disposition des informations.

I. 4. – Enseignements du panorama international

Sybille Berjoan, directrice Monde des Études Transports Accenture Research, et Catherine Bouteiller, doctorante au Laboratoire d'Economie des Transports de Lyon (LET), ont présenté une étude qualitative de la stratégie de six villes (Paris, Londres, Singapour, Toronto, Madrid et New York) en matière d'ouverture de données dans le domaine de la mobilité urbaine.

Il existe deux types de données : les données de production et les données de clients, ces dernières étant de sources très hétérogènes (par exemple, données des réseaux sociaux). Les données brutes ont peu de valeur. Le flux des données est extrêmement important, ce qui pose la question de savoir si les acteurs traditionnels de la mobilité urbaine ont les capacités suffisantes pour les collecter, les analyser et leur donner du sens.

Les données sont mises à disposition sur les sites web des AOT, des territoires urbains ou des institutions. Certaines villes, comme Singapour, ont mis en place un portail d'accès unique, d'autres, comme Paris, une constellation de sites (RATP, Vélib', SNCF, Paris data...)

L'analyse croise plusieurs approches : par types de données, par modes et par formats.

En 2013, 150 jeux de données ont été recensés et 252 au début de 2014. Singapour a publié 80 jeux de données, Londres 44, New York 39, Paris une trentaine. La répartition entre catégories de données est la suivante : près de 40% de données statistiques et 13% de données temps réel, le reste se répartissant équitablement entre données informatives et données géographiques. Une forte disparité s'observe entre les villes : Londres met à disposition 30% de données temps réel quand Singapour et New York sont les plus grands pourvoyeurs de données statistiques. Paris est le champion des données informatives et géographiques, mais ne met à disposition que 6% de données temps réel. Toronto a un profil similaire, avec 9% de données temps réel.

La donnée temps réel présente un intérêt particulièrement fort en raison de sa valeur marchande. Si elle est gratuite, elle n'est pas toujours facilement accessible. À Londres, Toronto et Singapour, il est désormais nécessaire de s'authentifier et de déclarer le type d'utilisation qu'on compte faire des données avant de pouvoir les télécharger. Singapour n'ouvre des données temps réel que dans le domaine routier, alors que Madrid se concentre sur le bus.

Les formats les plus utilisés sont excel et pdf, peu intéressants pour les développeurs. Le format GTFS est de plus en plus répandu, notamment à New York et Toronto. 18% des données sont disponibles via une API (60% des données temps réel), le plus souvent avec des clefs d'identification.

De nombreuses applications ont été tirées de ces données : *Moovit*, qui donne de l'information en temps réel ; *Get Me A Lot* (Singapour), qui géolocalise les parkings et permet de connaître les places disponibles ; *London Jamcams* qui informe sur l'état du trafic routier ; *Tripease London* qui indique le départ des prochains bus. Les informations produites par le réseau et enrichies de données locales sont réutilisables à la fois par les gestionnaires d'infrastructures et par les opérateurs.

En cinq ans, TfL (Londres) a fédéré autour de ses données près de 5000 utilisateurs permettant ainsi le développement de près de 200 applications. La composante open data est prise en compte pour chaque projet digital. On observe une volonté de mettre en valeur les productions des développeurs, voire d'inciter au développement de certaines fonctionnalités, tout en contrôlant les usages. Il reste à explorer les attentes des concepteurs en matière de cohérence, de formats, de qualité de la donnée, ainsi que la question de savoir quels sont les bénéfices induits et les implications organisationnelles d'une politique plus systématique d'open data pour les acteurs de la mobilité.

Si sur les sites internet français, la licence est bien visible, à New York, c'est *No License*. A Singapour figurent des *terms of agreement* très contraignants en termes de réutilisation et de monétarisation des données. Toutefois les données sont gratuites et accessibles. On relève également des adaptations de licences standards comme à Londres ou encore à Toronto, dont l'*Open government license* est adapté de l'ODbL. À la différence de ce qu'on observe en France, la licence est toujours unique. Sur le site de Paris, en revanche, on relève la licence Open Data Paris, l'ODbL, la Licence Ouverte.

L'autorité organisatrice de Londres a réalisé un travail en amont avec les opérateurs, privés pour l'essentiel : la mise à disposition des données est ainsi prévue dans les contrats.

II. Enjeux de l'ouverture des données de l'offre de transport

II. 1. – Ouverture des données de l'offre de transport : quels bénéfices ?

Amélioration de l'information délivrée à l'utilisateur des transports :

Comme l'observe la charte du G8, « *les données publiques librement accessibles et gratuitement réutilisables peuvent être à la source de services et des produits innovants susceptibles d'aider les individus à faire face plus facilement à la vie moderne* ».

L'ouverture des données de l'offre de transport crée un environnement propice à l'apparition de services répondant de manière plus performante aux attentes des usagers des transports en matière d'information. À ce titre, l'open data s'inscrit pleinement dans la « *mission de développement de l'information sur les transports* » mentionnée par l'article L. 1211-4 du code des transports.

L'émergence de services d'information combinant une large palette de modes de transport (collectif + vélo en libre service + autopartage + covoiturage...) pourrait ainsi être rendu possible. La combinaison d'information transport (ou mobilité) avec des informations touristiques, culturelles etc. serait également facilitée.

L'inaccessibilité des données apparaît comme un frein au développement de services innovants, notamment ceux élaborés à partir de données en temps réel – les données théoriques étant plus aisément reconstituables (l'application comme Paris ci la Sortie du Métro, qui indique à l'utilisateur dans quelle voiture monter pour descendre en face de la bonne sortie, n'a pas été créée avec les données de la RATP).

Si ZenBus, une application mobile créée par la start-up Joul qui permet de géolocaliser en temps réel les bus du réseau de transports urbains d'Issy-les-Moulineaux (TUVIM), a pu voir le jour, c'est parce que ce projet s'inscrivait dans le cadre d'une expérimentation cofinancée par la Région Île-de-France, en collaboration avec Grand Paris Seine Ouest, la Ville d'Issy-les-Moulineaux, la RATP et l'ENSTA ParisTech. En effet, faute d'accès aux données du SAE de l'opérateur de transport, Joul a dû les créer par ses propres moyens.

Promotion des alternatives à la voiture individuelle :

Certains modes de transports seraient davantage utilisés, notamment dans le cadre de déplacements occasionnels, si les usagers pouvaient avoir plus facilement accès à l'information qu'ils recherchent, via des calculateurs d'itinéraires ou des services de géolocalisation, par exemple.

En améliorant la circulation de l'information (théorique et temps réel), l'ouverture des données pourrait ainsi favoriser le report modal sur les transports publics, mais aussi les mobilités actives (vélo en libre service, notamment) et les services de mobilité proposant une alternative à l'usage individuel de la voiture, comme le covoiturage et l'autopartage. Dans ce dernier cas, l'ouverture des données pourrait également être un accélérateur pour l'électro-mobilité, la flotte de nombreux services d'autopartage étant composée de véhicules décarbonés³¹.

La mise en avant des alternatives à la voiture individuelle permettrait à la fois d'améliorer la fluidité du trafic et de diminuer les émissions polluantes.

Stimulation de la croissance économique :

Comme le rappelle une récente communication de la commission européenne, « *les données générées par le secteur public peuvent servir de matière première à des services et produits innovants à valeur ajoutée qui, en favorisant la création d'emplois et l'investissement dans des secteurs fondés sur les données, dopent l'économie*³² ».

Si un rapport du cabinet McKinsey³³ évalue les gains pour l'économie mondiale entre 3 220 et 5 390 milliards de dollars annuels, l'impact économique d'une politique d'ouverture des données n'est pas aisé à mesurer, d'autant que les initiatives d'ouverture des données sont encore récentes.

L'une des premières initiatives en France, celle de Rennes, conduit, après quelques années de recul, à un bilan nuancé : seule une poignée de la quarantaine d'applications ayant concouru au concours organisé par la métropole a survécu. Cette expérience en demi-teintes montre qu'une initiative isolée d'ouverture des données ne permet pas d'apprécier pleinement les bénéfices de l'open data sur l'économie.

Le débat a ainsi fait apparaître que la valeur d'une donnée se construisait en association avec d'autres données. En d'autres termes, la valeur économique réside moins dans la donnée elle-même qu'elle ne s'élabore dans la réutilisation qui en est faite. Le Comité a souligné à plusieurs reprises l'importance de ne pas préjuger des réutilisations en excluant de l'ouverture des données jugées a priori sans intérêt.

Par conséquent, pour révéler tout le potentiel économique de l'open data, il est nécessaire de remédier à l'isolement et au disparate des initiatives, et de mettre en place une politique d'ouverture à la fois globale et harmonisée.

³¹ Auto Bleue (Nice), Auto'trement (Strasbourg), AutoCité (Besançon), Autolib (Paris), Wattmobile Paris (Paris gare de Lyon), Moebius (Rueil-Malmaison), Blueclub (Bordeaux), Cité VU (Antibes), Citélib by Ha:Mo (Grenoble), [Mobili'Volt](#) (Angoulême), [Mopeasy](#) (Neuilly), [Plurial'Move](#) (Reims), [SunMoov](#) (Lyon), Twizy Way (Saint-Quentin-en-Yvelines), Wattmobile Marseille (Marseille Gare Saint Charles), Yelomobile (La Rochelle) etc.

³² Communication de la Commission, *Orientations sur les licences types recommandées, les ensembles de données et la tarification de la réutilisation des documents*, 2014/C 240/01, 24 juillet 2014 : <http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52014XC0724%2801%29&from=FR>

³³ http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

Amélioration de la transparence de l'action publique :

Enfin, l'open data permet d'accroître la transparence de l'action publique en permettant à chaque citoyen et contribuable d'accéder facilement aux données produites ou reçues par les personnes publiques dans l'exercice de leurs missions.

Dans le domaine des transports, l'ouverture des données de régularité, de ponctualité ou bien de qualité de l'information diffusée aux voyageurs permet, par exemple, de tendre vers une plus grande transparence sur la qualité des services de transport en France.

II. 2. – Ouverture des données de l'offre de transport : quels risques et quels remèdes ?

La lettre de mission invitait le Comité à examiner s'il convenait, et dans quelle mesure, de subordonner la réutilisation de certaines données à des conditions liées à l'intérêt général ou au respect des politiques publiques. Sur ces questions, le GART et l'UTP ont été invités à présenter leur analyse métier au Comité.

La lettre de mission demandait également d'examiner quelles dispositions devraient être prises pour éviter le risque de position dominante d'un fournisseur de service. Elle précisait que les travaux du Comité devraient prendre en compte le risque de captation de la valeur pour les services de transport en situation de concurrence, en cas d'émergence d'opérateurs de service d'information acquérant une position dominante.

Inexactitude de l'information fournie aux usagers :

Les AOT et les opérateurs de services publics de transport font valoir le risque qu'un réutilisateur propose une information :

- incomplète, c'est-à-dire qui, dans un calculateur d'itinéraire, écarterait une solution existante par exemple, ou bien, n'afficherait pas l'ensemble des horaires (ex. horaires du soir pour privilégier une offre en taxi) ;
- biaisée, c'est-à-dire qui favoriserait (sans que cette préférence résulte des critères de choix annoncés) un mode de transport par rapport à un autre (par exemple, mise en avant de la route au détriment des transports publics), voire un opérateur de transport par rapport à un autre, ce dernier risque étant également souligné par les opérateurs de transport craignant une distorsion de concurrence ;
- erronée, c'est-à-dire fausse ou caduque.

C'est pourquoi les AOT et les opérateurs de transport jugent nécessaire de formuler des exigences contractuelles en termes de complétude et de neutralité de l'information fournie à l'usager des transports.

Dans le cas du temps réel, l'information fournie par le détenteur des données prendrait la forme d'une API (par ex. : une API fournissant les horaires de départ et d'arrivée aux arrêts). Si le fournisseur de service élabore son service à partir de cette API, la fraîcheur de l'information est alors assurée.

Le détenteur de données peut également proposer une API délivrant une information conforme à la politique de mobilité de l'AOT. Si le fournisseur de service la réutilise telle quelle, cela assure, par construction du service, que les données ne soient pas altérées ni leur sens dénaturé par le réutilisateur.

La Licence Ouverte a entendu prévenir le risque d'inexactitude de l'information en stipulant que « la réutilisation ne doit pas induire en erreur des tiers quant au contenu de « l'Information », sa source et sa date de mise à jour ». Cette clause, dont il convient de noter qu'elle ne figure pas dans l'ODbL, garantit la complétude et la neutralité de l'information délivrée et permet d'apprécier son niveau de fraîcheur.

Risque pour la sécurité des voyageurs :

Ce risque concerne tout particulièrement la classe « informations temps réel ». Dans des situations de crise ou de perturbations, la gestion des flux revêt en effet une importance cruciale, afin de prévenir engorgements et bousculades. Si un train est en panne, un fournisseur d'information réutilisant des données publiques ne doit pas inciter les usagers à se reporter sur un bus d'une capacité inférieure.

Pour prévenir ces risques, le GART et l'UTP proposent – par analogie à ce qui se pratique dans l'information routière, où la hiérarchie du réseau est pris en compte par les concepteurs de GPS – de définir des règles de gestion de flux, de mettre à disposition des réutilisateurs les données pertinentes dans les situations de crise ou de perturbations (comme des plans de transport adaptés) et, éventuellement, de proposer au réutilisateur une API délivrant une information conforme à la politique de gestion des flux de l'AOT.

Mise en cause de la responsabilité du fournisseur des données :

Les AOT et les opérateurs de transport pointent le risque que leur responsabilité, civile ou pénale, soit engagée dans le cas où :

- les données fournies par eux-mêmes se révéleraient défectueuses ;
- une mauvaise information serait fournie à un usager des transports par un tiers réutilisant des données mises en open data.

Par conséquent, les AOT et les opérateurs de transport jugent nécessaire d'introduire dans la licence une clause les exonérant de toute responsabilité au regard du préjudice que pourrait subir le réutilisateur des données ou l'utilisateur d'un service créé à partir de ces mêmes données.

Il pourrait être pris modèle sur la clause figurant dans la Licence Ouverte : « le « Producteur » [...] ne peut garantir l'absence de défauts ou d'irrégularités éventuellement contenues dans « l'Information » [...], ne garantit pas la fourniture continue de « l'Information » [et] ne peut être tenu pour responsable de toute perte, préjudice ou dommage de quelque sorte causé à des tiers du fait de la réutilisation ».

Coût de la mise à disposition des données, notamment en temps réel :

Les AOT et les opérateurs de transport mettent l'accent sur le coût qu'est susceptible d'entraîner la mise à disposition des données à des tiers, notamment pour les données en temps réel.

Pour amortir ces éventuels coûts supplémentaires, les AOT et les opérateurs de transport proposent d'instaurer une redevance prenant en compte les coûts supplémentaires engendrés par la mise à disposition des données.

En outre, avant qu'il soit procédé aux investissements nécessaires à la mise à disposition des données à des tiers, il convient de s'assurer qu'une vraie demande existe.

Apparition d'une position dominante d'un fournisseur d'information :

De nombreux membres du Comité ont estimé l'ouverture des données pouvait conduire à l'acquisition par un acteur de l'économie numérique tel que Google d'une position dominante sur le marché de l'information multimodale. Un rapport récemment publié par la Commission européenne envisage ce scénario³⁴.

Sur ce point, il convient d'emblée de rappeler que l'existence d'une position dominante d'un acteur sur un marché n'est pas contraire au droit de la concurrence. C'est l'abus de position dominante qui est sanctionnable.

³⁴ *All Ways Travelling Final Report*, contract MOVE/C2/SER/2012 489/SI2.646722 To develop and validate a European passenger transport information and booking system across transport modes, cf. p. 178 : « [...] as long as Google Transit is a free service for operators who are willing to share their data with Google, and if GTFS is updated for real-time utilisation, Google might seize the market for a potential MMITS, include almost every transport in future. » <http://ec.europa.eu/transport/themes/its/studies/doc/20140812-july9thversion-awtfinalreport.pdf>

Google est déjà présent sur le marché de l'information transport français avec des services tels que *Flight*, un comparateur de vols lancé en 2013, et surtout *Maps*, un service de cartographie en ligne qui intègre un calculateur d'itinéraire multimodal (piéton, vélo, routier). En 2012, Google a ajouté à *Maps* la fonctionnalité *Transit* qui permet de calculer son itinéraire en transport public, service qui peine toutefois à décoller en raison de la réticence des AOT³⁵ à signer des partenariats avec Google. Suite à l'entrée du fond de capital-risque, Google Venture, en août 2013, au capital d'Uber, l'application mobile *Maps* offre désormais, pour une trentaine de villes dans le monde, un comparateur de trajet « transport en commun / VTC »³⁶.

Flight comme *Maps* sont des services offerts gratuitement aux internautes, le business model de Google consistant à se faire rémunérer par les fournisseurs de biens ou de services au titre de l'apport d'affaires, dans le cadre du modèle général de l'affiliation³⁷ – en d'autres termes, à vendre de la publicité. L'achat de liens sponsorisés permet à un annonceur de bénéficier d'un référencement préférentiel.

Or, l'existence de pratiques anti-concurrentielles de la part de Google est avérée : une enquête de la Commission européenne a fait apparaître que Google abaissait dans ses résultats de recherche le rang de services concurrents et accordait à ses propres services un placement préférentiel, profitant de son monopole sur le marché de la recherche en ligne³⁸. En janvier 2012, le Tribunal de Commerce de Paris a, par ailleurs, jugé Google coupable d'un abus de position dominante sur les marchés des liens sponsorisés et de la cartographie en ligne. Pour le tribunal, « *le comportement [de Google, qui offre gratuitement son service de cartographie en ligne], aboutit à l'éviction de tout concurrent [...] mais en outre s'inscrit à l'évidence dans le cadre d'une stratégie générale d'élimination*³⁹ ».

L'apparition d'une position dominante de Google sur le marché de l'information multimodale serait préjudiciable à la fois pour les opérateurs de transport en situation de concurrence, mais aussi pour les autres entreprises opérant sur ce marché (cf. II.2).

Pour les opérateurs de transport en situation de concurrence : un risque d'abus de position dominante d'un fournisseur d'information

Lors du débat sur les services concurrentiels de transport longue distance, a été décrit un scénario dans lequel certains intermédiaires de l'économie numérique proposant des moteurs de voyages pourraient abuser de leur position dominante pour prélever des commissions de plus en plus importantes. Citant le précédent de l'hôtellerie, les opérateurs évoquent un prélèvement sur leurs marges susceptible de compromettre leur équilibre économique. Ce risque d'un abus de position dominante ne concerne pas, en revanche, les opérateurs gérant des services de transport « conventionnés ». Faiblement soumis à la pression de la concurrence, ces services ne se verraient pas contraints d'accepter de payer des commissions élevées.

Pour prévenir un abus de position dominante, la SNCF a émis le souhait que les opérateurs de transport en situation de concurrence puissent disposer de la liberté de conclure des accords de gré à gré, ce qui implique d'une part de pouvoir choisir les entreprises auxquelles ils souhaitent ouvrir leurs données et, d'autre part, de négocier de manière bilatérale les conditions de leur mise à disposition. Dans la situation actuelle, la SNCF met ses données TGV à la disposition des entreprises ferroviaires européennes, des agences de voyage selon des conditions qui ne doivent pas être discriminatoires, ainsi que de certains infomédiateurs.

³⁵ Faute d'avoir trouvé un accord avec le STIF, Google a recréé, avec beaucoup d'approximations, les données théoriques des métros et RER parisiens : tracés des lignes, localisation des stations, horaires de passage ou fréquence...

³⁶ Ce service est disponible depuis le 5 novembre 2014 : <http://blog.uber.com/googlemaps>

³⁷ La part de marché mondiale de Google sur la publicité en ligne devrait atteindre 31,45 % en 2014 : http://www.lemonde.fr/economie/article/2014/08/23/publicites-en-ligne-amazon-s-appreterait-a-concurrencer-google_4475623_3234.html

³⁸ La part de marché mondiale de Google sur la publicité en ligne devrait atteindre 31,45 % en 2014 : http://www.lemonde.fr/economie/article/2014/08/23/publicites-en-ligne-amazon-s-appreterait-a-concurrencer-google_4475623_3234.html

³⁹ Condamnée, Google a fait appel. Saisie par la cour d'appel de Paris, l'Autorité de la concurrence devrait prochainement rendre un avis sur cette affaire.

Les dispositions susceptibles d'être adoptées en cas d'ouverture des données TGV ont été discutées :

1. La mise à disposition des données sous une licence de type Open Database License (ODbL) ; cette licence comporte une clause de *share alike* (« partage à l'identique »), qui contraint le réutilisateur à mettre gratuitement à la disposition de la communauté des développeurs la base de données qu'il a enrichie ;
2. L'instauration d'une redevance garantissant « *un écosystème concurrentiel équitable* », sur le modèle de celle du Grand Lyon, c'est-à-dire modulée en fonction du « *volume d'activité* » du service fourni. D'un montant nul pour les petits acteurs, la redevance s'élève à un montant dissuasif pour les entreprises en position dominante ;
3. L'interdiction d'accès aux données pour les entreprises ayant été condamnées par la CNIL ou ne s'étant pas acquittées de leurs obligations fiscales et sociales ; cette disposition s'inspire de celles existantes en matière d'accès aux marchés publics (article 43 du code des marchés publics).

Ces trois solutions ont toutefois été déconseillées par le sous-groupe juridique, aux motifs suivants :

1. Imposer l'ODbL, plus exactement une obligation de partage à l'identique, pourrait constituer une restriction à la liberté de réutilisation en rendant impossibles les modèles économiques fondés sur la revente de bases de données enrichies. En outre, on ne saurait exclure que pour un opérateur de transport en position de monopole, imposer une obligation de cette sorte puisse s'assimiler à un abus de position dominante.
2. Si la redevance pour service rendu permet d'amortir les coûts de mise à disposition des données, voire de récupérer une partie de la valeur créée à partir de celles-ci, elle n'apparaît pas comme un outil juridique pertinent pour prévenir un abus de position dominante (mission qui, du reste, ne relève pas de la compétence des collectivités territoriales). Si une jurisprudence récente⁴⁰ permet, en effet, de prendre en compte, pour le calcul du montant de la redevance, « l'avantage économique » susceptible d'être retiré par le bénéficiaire, la redevance doit toutefois rester proportionnée au service rendu et respecter le principe d'égalité devant les charges publiques. Un mode de calcul conçu pour mettre l'essentiel du produit de la redevance à la charge d'un acteur dominant, en exonérant toutes les autres entreprises, pourrait être qualifié de discriminatoire, la différence de traitement entre les bénéficiaires n'apparaissant pas, en l'espèce, fondée sur une différence objective de situation au regard du service rendu. De plus, la mise en œuvre du critère « part de marché » pourrait prêter à contestation, compte tenu des difficultés méthodologiques auxquelles se heurte la quantification d'une part de marché ou d'activité, surtout au niveau local. Sur un plan technique, ce type de critère paraît peu adapté pour définir un seuil de déclenchement d'une obligation de payer : le montant d'une redevance ne saurait reposer sur des critères potentiellement variables dans le temps et dépendants des autres acteurs présents sur le marché.
3. Interdire l'accès aux données aux entreprises condamnées par la CNIL paraît peu pertinent en raison de l'absence de rapport entre l'objet de la licence et la nature de l'infraction qui ferait obstacle à sa souscription. En outre, il convient de rappeler que seule une condamnation définitive pourrait être prise en compte. De même, interdire l'accès aux données aux entreprises ne s'étant pas acquittées de leurs obligations fiscales et sociales est sans effet sur les intermédiaires de l'économie numérique pratiquant l'optimisation fiscale⁴¹.

Une autre solution, étudiée par les sous-groupes juridiques et modèles économiques, serait que les rapports entre opérateurs de transport en situation de concurrence et fournisseurs d'information fassent l'objet d'une régulation, sur le modèle de ce qui a été fait dès 1989 (règlement

⁴⁰ CE, *Société d'équipement de Tahiti et des îles*, 7 octobre 2009, n° 309499 ; la redevance peut prendre la forme d'un pourcentage du chiffre d'affaires réalisé.

⁴¹ Le redressement fiscal notifié à Google en mars 2014 ne constitue pas une condamnation.

n° 2299/1989) pour les Systèmes Informatisés de Réservation (SIR) dans le transport aérien. C'est, du reste, ce que préconise le rapport final d'All Ways travelling récemment publié par la DG Move⁴².

Un règlement européen pourrait ainsi encadrer le montant des commissions perçues par l'intermédiaire et imposerait aux moteurs de voyage de fournir une information neutre. Ces dispositions empêcheraient qu'une meilleure visibilité ne soit accordée au plus offrant.

Sur le marché de l'information multimodale : monopole d'un fournisseur d'information au détriment du développement de l'économie numérique en France

Le sous-groupe modèles économiques estime que la position dominante d'un acteur aurait des répercussions négatives sur le développement de l'économie numérique en France. Quand bien même de nombreuses start-up se développeraient dans l'écosystème des géants du numérique, il n'en reste pas moins que l'apparition d'un monopole sur le marché de l'information multimodale se traduirait par une contraction de la concurrence, de nature à freiner l'innovation.

Une domination des intermédiaires (tels que Google) atténuerait l'effet bénéfique de l'ouverture des données pour l'économie française. Ces grands groupes pratiquant l'optimisation fiscale, une grande part de la valeur créée à partir des données serait transférée hors de France.

C'est pourquoi le sous-groupe s'est attaché à proposer des mesures permettant de rééquilibrer le jeu de la concurrence et à promouvoir l'innovation. Le sous-groupe recommande ainsi de lever les barrières à l'entrée pour les acteurs de taille modeste en mettant gratuitement les données à disposition et en fournissant des services complémentaires susceptibles de les aider à se développer.

Le sous-groupe souligne également l'importance de ne pas créer d'entrave à la réutilisation commerciale des données et de laisser une place à l'initiative privée. Enfin, le sous-groupe préconise de permettre aux entreprises innovantes de bénéficier des marques des territoires pour la diffusion de leurs produits.

S'inspirant d'une décision du conseil de la concurrence de 2001, qui a interdit pendant trois ans à Vinci Park de répondre à des appels d'offre dans les villes du sud de la France où il était en position ultra-dominante, le sous-groupe a également proposé d'interdire l'accès aux données aux entreprises en position de monopole. Cette solution a toutefois été déconseillée par le sous-groupe juridique, qui a fait valoir qu'on ne pouvait transposer au cas d'espèce une décision prise dans le contexte d'un litige portant sur l'attribution de marchés publics.

III. Recommandations

Les travaux du Comité ont été menés selon une double approche, par modes de transport, d'une part, et par thèmes (technique, juridique, modèles économiques), d'autre part.

Les différents modes de transport ont été répartis en quatre ensembles cohérents :

- Les services publics de transport : notamment bus, tram, Intercités, Transilien, TER, autocars, etc. ;
- les services concurrentiels de transport longue distance : TGV, covoiturage, aérien ;
- les nouveaux services de mobilité (vélo, vélo en libre-service, covoiturage, autopartage) et le stationnement ;
- la route.

Pour chacun de ces ensembles de modes de transport, les membres du Comité ont examiné les deux questions suivantes :

- quelles données ou classes de données ouvrir ?
- Le cas échéant, quelles conditions imposer pour remédier à des risques identifiés ?

⁴² <http://ec.europa.eu/transport/themes/its/studies/doc/20140812-july9thversion-awtfinalreport.pdf>

La section III.1. du rapport présente les recommandations relatives à ces deux questions, en abordant successivement chacun des ensembles de modes. Cette section se termine par une recommandation transversale relative aux services liés à l'information du voyageur.

Les sections III. 2. à III. 4. présentent les autres recommandations du Comité, issues des travaux des sous-groupes juridique, technique et modèles économiques, en évitant les redites lorsque ces sous-groupes ont formulé des recommandations voisines.

Des recommandations à caractère général figurent au III.5.

III. 1. – Recommandations en matière d'ouverture des données et de conditions permettant de remédier aux risques liés à l'ouverture

Dans les recommandations qui suivent, nous employons la notion de « donnée réutilisable ». Par réutilisation, nous entendons le fait, pour un tiers, d'utiliser une information à d'autres fins que celles pour les besoins de laquelle elle a été produite; les « horaires planifiés » par un opérateur de transport, par exemple, sont susceptibles d'être réutilisés par un tiers fournisseur de service d'information.

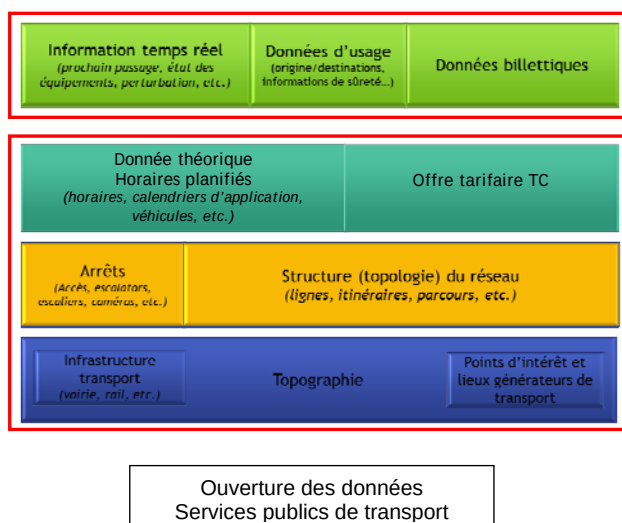
Le fait qu'une donnée soit « réutilisable » n'implique pas l'obligation de la produire, mais seulement d'en permettre la réutilisation par les tiers, si elle a été produite.

Recommandations relatives aux services publics de transport :

Lors de la réunion consacrée aux services publics de transport, les quatre intervenants ayant présenté leurs initiatives d'ouverture des données, qu'ils soient issus d'une collectivité locale, d'une autorité organisatrice ou d'un opérateur de transport, ont tous conclu en faveur d'une ouverture des données de l'offre de transport public. De son côté, le sous-groupe modèle économique a proposé que « *tout jeu de données relatives à l'information voyageur et aux flux de mobilité (données statiques, théoriques, temps réel, historiques...) issu de ce type de contrat reviennent de droit au cocontractant public et qu'il en ait toute liberté d'usage. Ceci reviendrait à dire que ces données aient un statut public* ».

Le Comité, en s'appuyant sur la classification des données élaborée par le sous-groupe technique (cf. figure ci-dessous), a adopté une approche plus nuancée :

- pour les classes des trois tranches inférieures (« arrêts » et « structure du réseau », « données théoriques, horaires planifiés », « offre tarifaire ») : ouverture des données de ces classes, éventuellement sous certaines conditions ;
- pour les classes de la tranche supérieure (« temps réel », « usage », « billettiques ») : ouverture de certaines données de ces classes, sous certaines conditions.



Pour examiner les conditions à prendre en compte, le Comité a tout d'abord entendu le GART et l'UTP, qui ont présenté leur analyse des risques entraînés par l'ouverture des données et des remèdes susceptibles de les prévenir.

D'une manière générale, il apparaît opportun de subordonner la réutilisation des données à des conditions :

- de « complétude et neutralité » ;
- de « non responsabilité du producteur de données ».

Ces deux conditions correspondant à des dispositions présentes dans la Licence Ouverte (Etalab).

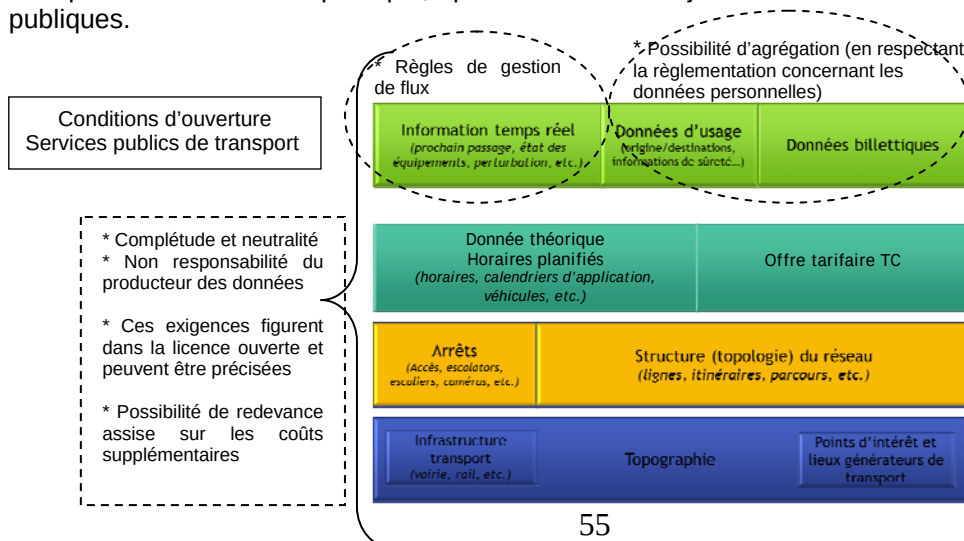
Pour la classe « informations en temps réel » :

- un consensus se dégage sur l'ouverture de certaines données :
« prochain passage du bus »
position du véhicule
« état des équipements » (escalators, ascenseurs)
données de perturbations
- pour les données de perturbations, les risques liés aux reports de trafic en cas d'incident ou de crise doivent être pris en compte. Pour répondre à ces préoccupations, la réutilisation peut être subordonnée au respect de règles de gestion de flux ; à cet effet, les gestionnaires des services de transport et les autorités responsables en cas de crise mettront à disposition du réutilisateur les plans de transport adaptés aux situations de crise ou de perturbations, par analogie avec les bonnes pratiques mises en place dans le domaine de l'information routière ;
- pour la « position du véhicule », on vérifiera auprès de la CNIL que cette ouverture ne soulève pas d'objections liées au respect de données personnelles propres aux conducteurs ;
- lorsque la mise à disposition des données entraîne des coûts significatifs, il apparaît justifié d'offrir la possibilité de rendre payant l'accès aux données afin de couvrir les coûts entraînés par cette mise à disposition.

Sur le principe, le Comité n'exclut pas que cette mise à disposition payante puisse intervenir pour d'autres données que les données en temps réel, en cas de coûts de mise à disposition significatifs. Toutefois, en dehors des données en temps réel, il n'a pas identifié de cas concrets où la mise à disposition entraînerait des coûts significatifs. La mise à disposition de données en temps réel n'entraîne pas nécessairement des coûts significatifs.

Pour les données « d'usage » et « billettiques », des attentes fortes se manifestent, de sorte que l'ouverture de certaines données paraît constituer une opportunité ; toutefois, le GART et l'UTP estiment que ces données pourraient être protégées au nom du secret commercial ; il est proposé que la mise à disposition des données puisse se faire selon un certain niveau d'agrégation, laquelle devrait toutefois rester limitée pour ne pas vider les données de l'essentiel de leur intérêt. La mise à disposition doit également respecter la réglementation concernant les données personnelles.

L'ouverture des classes « données d'usage » et « données billettique » contribuerait à la transparence de l'action publique, qui est l'un des objectifs forts de l'ouverture des données publiques.



Compte tenu de l'ensemble de ces éléments, le Comité adopte les recommandations suivantes, pour les données des services publics de transport (existantes ou à venir). Il s'agit successivement :

- des recommandations en matière d'ouverture des données ;
- des recommandations sur les conditions permettant de remédier aux risques.

Recommandations sur l'ouverture des données des services publics de transport [SPT]

SPT1 : Pour les classes « topographie », « arrêts », « structure du réseau », « données horaires théoriques » et « offre tarifaire » :

Les données de ces classes sont réutilisables.

Cas de l'offre tarifaire : à titre exceptionnel, l'ouverture des données de l'offre tarifaire peut ne pas être complète : les informations qui alourdiraient inutilement la présentation ou demanderaient une expertise additionnelle pour être correctement interprétées peuvent être omises.

SPT2 : Pour la classe « information en temps réel » :

Les données suivantes sont réutilisables :

- messages diffusés sur le réseau, y compris messages de perturbation ;
- position des véhicules ;
- prochain passage du véhicule à l'arrêt ;
- état des équipements (escalators, ascenseurs...).

Cette liste pourra faire l'objet de complément au fil du temps.

SPT3 : Pour les classes « usage » et « billettique » :

Les données suivantes sont réutilisables :

- origine/destination des déplacements ;
- usage des transports (nombre de voyageurs par segment et par période, serpent de charge) ;
- informations de sûreté (vols, agressions) ;
- données de validation.

Ces données peuvent être mises à disposition sous une forme agrégée, l'ampleur de cette agrégation devant toutefois rester limitée. La mise à disposition des données doit également respecter la réglementation concernant les données personnelles.

Recommandations sur les conditions permettant de remédier aux risques identifiés [SPT]

SPT4 : Pour toutes les classes de données :

La réutilisation peut être soumise aux conditions suivantes :

- la réutilisation ne doit pas induire en erreur des tiers quant au contenu de « *l'Information* » et à sa date de mise à jour ;
- en vue de préciser l'exigence précédente, le fournisseur de données peut demander que l'information réponde à des exigences de complétude et des critères de neutralité ; afin de faciliter les réutilisations, les fournisseurs de données concernés définiront ces exigences de complétude et ces critères de neutralité dans une rédaction commune ;
- le fournisseur des données ne peut garantir l'absence de défaut de la donnée et ne peut être tenu pour responsable des pertes, préjudices ou dommages de quelque sorte causé à des tiers du fait de la réutilisation ;
- prix de l'accès aux données : si, contrairement à ce qui est observé actuellement, les coûts de mise à disposition devaient se révéler significatifs pour des données autres que celles en temps réel, les dispositions prévues en **SPT5** pourraient être mises en œuvre pour ces données.

Remarque : la consultation en cours pour préparer la loi sur le numérique met en avant la notion de loyauté⁴³, qui pourra être utilisée pour préciser le critère de complétude.

⁴³ Sur cette question, voir notamment : <http://www.numerama.com/media/ConseilEtat-numerique-DroitsFondamentaux.pdf>

SPT5 : Classe « Informations en temps réel » :

- la réutilisation peut être soumise au respect de règles de gestion de flux, notamment en cas d'incident ou de crise. Ces règles sont fixées par l'autorité responsable en cas d'incident ou de crise ;
- prix de l'accès aux données : lorsque les coûts de mise à disposition sont significatifs, cette mise à disposition peut être payante ; le total des recettes ne peut excéder les coûts supplémentaires engendrés par la mise à disposition des données ; les éléments de calcul permettant de justifier le prix de l'accès aux données à partir des coûts supplémentaires sont publiés ; la mise à disposition reste gratuite pour les petits utilisateurs.

Remarque : le sous-groupe technique a été invité à approfondir les aspects relatifs aux coûts de mise à disposition des données. Pour certains membres du sous-groupe « *ces coûts sont très faibles (quel que soit la nature des données, temps réel ou pas), alors que pour d'autres ces coûts sont indiqués comme étant très élevés (en particulier pour le temps réel)* ». Cette question mérite une réflexion complémentaire.

Recommandations relatives aux services concurrentiels de transport longue distance :

Des opérateurs de transport ont décrit un risque d'abus de position dominante de certains intermédiaires de l'économie numérique. De l'analyse menée par le Comité, il ressort que ce risque de captation de la valeur des données est crédible.

Pour y remédier, une solution à approfondir serait que les rapports entre opérateurs de transport en situation de concurrence et fournisseurs d'information fassent l'objet d'une régulation, sur le modèle de ce qui a été fait pour les systèmes informatisés de réservation dans le transport aérien (cf. § II. 2) : mise en place d'un règlement européen, imposant la neutralité de l'affichage des informations, ainsi que des commissions d'intermédiation raisonnables.

L'échelon européen, déjà adopté pour le transport aérien, paraît le plus pertinent pour cette régulation. En cas de manquement, les sanctions prévues par le règlement n° 80/2009 s'appliquant aux services d'information et de réservation sont dissuasives : les amendes peuvent s'élever jusqu'à 1% du chiffre d'affaires.

La régulation pourrait être mise en place parallèlement aux actions menées par la commission en matière d'information multimodale (cf. feuille de route publiée en juin 2014⁴⁴), notamment l'Initiative sur l'accès aux données ou l'action prioritaire a) définie par la directive ITS précitée.

Cette recommandation devrait être portée au niveau européen, notamment dans le cadre des travaux de mise en œuvre de la directive ITS.

Sur la base de ces éléments, nous formulons la recommandation suivante :

Recommandations sur l'ouverture des données des services concurrentiels de transport [SCT]

SCT : L'ouverture des données des services de transport en situation de concurrence ne devrait être réalisée que si elle est accompagnée de mesures robustes permettant de prévenir l'abus de position dominante d'un fournisseur de services d'information

Recommandations relatives aux nouveaux services de mobilité et au stationnement :

Lors de la réunion consacrée aux nouveaux services de mobilité et au stationnement, les intervenants ayant présenté les initiatives prises en matière d'ouverture des données, qu'il s'agisse de Citiz pour le vélo, de JCDecaux pour le vélo en libre-service, de la Feduco pour le covoiturage, d'Autolib' pour l'autopartage et de la FNMS pour le stationnement, se sont prononcés en faveur d'une ouverture des données, le cas échéant sous conditions.

⁴⁴ <http://ec.europa.eu/transport/themes/its/doc/swd%282014%29194.pdf>

Au cours de cette réunion ont été mises en évidence les particularités propres à chaque service de mobilité, mais aussi des approches convergentes pour les orientations à adopter en matière d'ouverture des données :

- pour le vélo, les données relatives aux véloroutes et voies vertes sont d'ores et déjà collectées et mises à disposition au niveau national (cf. ON3V⁴⁵) et un standard décrivant ces données a été élaboré par le CEREMA (DTecTV) et validé par la COVADIS⁴⁶. Dans le secteur urbain, il convient de distinguer :
 - o les données d'infrastructure, telles que celles décrivant le réseau cyclable, la voirie apaisée ou les voies mixtes ;
 - o les données associées aux parkings vélos (nombre d'arceaux vélos, le type de mobilier de stationnement, la localisation, etc.) ou à des services (parkings vélos sécurisés, location de vélo, etc.).

L'élaboration d'un standard portant sur la description du caractère cyclable des voies est souhaitée.

- pour le vélo en libre service, il apparaît opportun de généraliser la mise à disposition des données théoriques (localisation des bornes de VLS, capacité...) et d'ouvrir certaines données dynamiques (disponibilité de vélos et d'emplacements aux stations); à propos de l'ouverture des données d'usage ou de fréquentation est soulevée l'objection du secret des affaires, sans que cet argument soit étayé de façon solide; le Comité observe que la transparence de l'action publique est l'un des objectifs forts de l'ouverture des données à caractère public ;
- pour le covoiturage, l'ouverture des données relatives aux aires de covoiturage s'impose ;
- pour l'autopartage, comme pour le vélo en libre-service, il paraît opportun de généraliser la mise à disposition des données théoriques (arrêts, capacité...) et de certaines données dynamiques (disponibilité de véhicules et d'emplacements aux stations) ; le réseau Citiz a d'ores et déjà ouvert ces données⁴⁷ ;
- pour le stationnement, la même approche se dégage : mise à disposition de l'ensemble des données théoriques (localisation des parkings, capacité...) et ouverture de certaines données dynamiques. Les données de l'offre tarifaire mises à disposition seraient seulement celles que l'opérateur est légalement tenu d'afficher en entrée de parc ; les données relatives aux abonnements et aux offres promotionnelles ne seraient pas ouvertes. Pour les données dynamiques, les opérateurs de stationnement soulignent que des enjeux de concurrence sont à prendre en compte, puisque plusieurs sociétés peuvent être en concurrence sur un même territoire ; il convient de veiller tout particulièrement à ce que les fournisseurs de service ne puissent en aucun cas fournir une information biaisée ; à cet effet, la FNMS recommande d'adopter un indicateur de disponibilité de places de stationnement à trois positions : places libres, parc saturé, parc en voie de saturation ; il s'agit donc ici non pas d'ouvrir les données brutes elles-mêmes, mais un indicateur élaboré à partir de ces données.

Une fois identifiées les données concernées par l'ouverture, se pose la question de savoir à quels services au juste, parmi les services de mobilité tels que vélo, vélo en libre service, autopartage, covoiturage et stationnement, devrait s'appliquer une obligation d'ouverture des données.

Dans le cas des nouveaux services de mobilité, le Comité propose d'étendre le principe d'ouverture à certaines données produites en dehors du cadre strict d'une mission de service public et émanant de sociétés privées.

Cette ouverture porterait sur les données produites dans le cadre de services gérés par un opérateur privé dans le cadre d'une convention avec une personne publique ; par exemple, seraient réutilisables de droit les données relatives aux :

⁴⁵ <http://www.departements-regions-cyclables.org/page/telecharger--p-73.html>

⁴⁶ <http://www.certu.fr/geostandard-veloroutes-et-voies-a1119.html>

⁴⁷ <http://production.citiz.fr/doc/>

- services de VLS mis en place par un opérateur privé dans le cadre d'une convention d'occupation du domaine public ;
- services gérés par des personnes morales percevant une subvention publique: par ex., prêt longue durée de vélos géré par l'association Vélocampus ou service d'autopartage mis en place par la SCIC Autocool ;
- aires de covoiturage implantées sur un terrain appartenant à une société privée (par ex., parking du Super U d'Arthon en Retz) sous convention avec une personne publique (par ex., le conseil général de Loire Atlantique) dans le cadre d'une convention de signalisation de zone de covoiturage⁴⁸.

Nous présentons successivement :

- les recommandations en matière d'ouverture des données ;
- les recommandations sur les conditions permettant de remédier aux risques.

Rappel de la classification des données pour les nouveaux services de mobilité

	Vélo	VLS	Autopartage	Covoiturage courte distance	Stationnement
Données temps réel et autres données	- Disponibilité des vélos et des stations - Fréquentation des voies	- Disponibilité des vélos et des stations - Données d'usage	- Disponibilité des véhicules et des places - Données d'usage		- Disponibilité des places de parking - Données d'usage et de fréquentation
Offre théorique	- Capacité des parkings, tarifs, contrat avec l'exploitant... - Types de vélos proposés à la location	- Capacité des stations, tarifs...	- Liste des stations et des véhicules - Prix prévisionnels... - Planning à 24H	- Capacité des aires...	- Capacité des parkings, horaires d'ouverture, tarifs...
Topographie	- Types de voie - Localisation des parkings	- Localisation des stations	- Localisation des stations	- Localisation des aires	- Localisation des parkings

Recommandations sur l'ouverture des données des nouveaux services de mobilité [NSM]

Les recommandations suivantes s'appliquent aux nouveaux services de mobilité et au stationnement, que ces services soient exploités dans le cadre d'une mission de service public ou qu'ils donnent simplement lieu à une convention avec une personne publique.

NSM1 : Vélo en milieu urbain et interurbain :

Les données relatives aux infrastructures ou aux équipements utilisés par les vélos sont réutilisables.

NSM2 : Vélo en libre-service :

- les données des classes « arrêts », « structure du réseau », « données théoriques », « informations en temps réel » et « offre tarifaire » sont réutilisables ;
- les données de la classe « usage » sont réutilisables. Elles peuvent être mises à disposition sous une forme agrégée, l'ampleur de cette agrégation devant toutefois rester limitée et être compatible avec la réglementation concernant les données personnelles.

NSM3 : Covoiturage (courte distance) :

Les données relatives aux aires de covoiturage sont réutilisables.

NSM4 : Autopartage :

Les données sont réutilisables. Il s'agit, notamment, des données suivantes :

- pour la classe « arrêts » : la liste des stations (par une recherche en x, y ou par opérateurs)
- pour la classe « données théoriques » :

⁴⁸ Cette convention précise bien que la personne publique n'organise pas le covoiturage, mais met simplement « à disposition des intéressés les outils ou informations permettant de faciliter cette pratique. »

pour l'autopartage en boucle, la liste des véhicules à une station, le planning à 24h, le taux d'occupation à 24h
pour l'autopartage en trace directe, la capacité de chaque station

- pour la classe « informations en temps réel », dans le cas de l'autopartage en trace directe, la disponibilité des stations et des véhicules.

NSM5 : Pour le stationnement,

- les données des classes topographie et « données théoriques », hors « offre tarifaire », sont réutilisables ;
- les données retraitées (places libres, parc saturé, parc en voie de saturation) de la classe « information en temps réel » sont réutilisables ;
- les données de la classe « usage » sont réutilisables ; elles peuvent être mises à disposition sous une forme agrégée, l'ampleur de cette agrégation devant toutefois rester limitée ; l'ouverture des données doit respecter la réglementation concernant les données personnelles.

NSM6 : Pour l'ensemble des nouveaux services de mobilité et le stationnement :

Offre tarifaire

- À titre exceptionnel, l'ouverture des données de l'offre tarifaire peut ne pas être complète : les informations qui alourdiraient inutilement la présentation peuvent être omises.
- Pour le stationnement, les données faisant l'objet d'une obligation légale d'affichage en entrée de parc sont réutilisables.

Recommandations sur les conditions liées à l'ouverture des données, pour remédier aux risques [NSM]

NSM7 : La réutilisation des données des infrastructures ou des équipements relatifs aux vélos, des services de vélos en libre-service, de plate-forme de covoiturage, d'autopartage et de parcs de stationnement peut être soumise aux conditions suivantes :

- la réutilisation ne doit pas induire en erreur des tiers quant au contenu de « *l'Information* » et sa mise à jour. En vue de préciser cette condition, le fournisseur de données peut demander que l'information réponde à des exigences de complétude et des critères de neutralité qu'il définit ;
- le fournisseur des données ne peut garantir l'absence de défaut de la donnée et ne peut être tenu pour responsable des pertes, préjudices ou dommages de quelque sorte causés à des tiers du fait de la réutilisation ;
- prix de l'accès aux données : voir **SPT 5**

Il va de soi qu'il ne pouvait être question dans le cadre de la mission dévolue au Comité de passer en revue l'ensemble des données transport qu'il serait opportun d'ouvrir. Le cas des données des bornes de recharge pour les véhicules électriques, par exemple, n'a pu être examiné attentivement. Or, « *l'utilisation massive de véhicules électriques rechargeables est un objectif politique motivé par les enjeux énergétiques et environnementaux. Il ne pourra se réaliser que si l'usage de ces véhicules est facile et le conducteur libéré du souci de leur autonomie réduite et du temps de recharge long* ». L'un des moyens d'y parvenir serait d'ouvrir les données relatives aux bornes de recharge⁴⁹.

NSM8 : Le Comité recommande que les parties concernées examinent ensemble l'opportunité et les modalités d'une ouverture des données relatives aux bornes de recharge des véhicules électriques.

Recommandations relatives à l'information routière :

Lors de la réunion du 4 novembre ont été présentées les démarches d'ouverture des données routières menées dans le cadre de la mise en œuvre des directives Inspire et ITS. Les actions prioritaires a), b) et c) de cette seconde directive traitent en particulier des services d'information

⁴⁹ Cf. Annexe 8.

pour l'usager (cf. § I.2). La lettre de mission suggérait au Comité de se référer aux travaux menés par la Commission pour l'élaboration de spécifications relatives à ces trois actions prioritaires.

Le Comité constate :

- que les actions prioritaires b) et c) de la directive ITS traitent de l'ouverture des données de l'information routière sur le réseau principal (réseau de transport européen, dit RTE) ;
- que l'action prioritaire a) de la directive ITS a vocation à traiter de l'information routière en général, en tant que composante de l'information multimodale ;
- que, d'une manière générale, les données liées à l'information routière, étant produites dans l'exercice d'une mission de service public administratif, sont réutilisables.

Recommandations relatives à l'ouverture des services liés à l'information du voyageur :

Enfin, la recommandation ci-après porte sur les services liés à l'information du voyageur, tels que les calculateurs d'itinéraire mis en place par les personnes publiques dans le cadre de leur mission de service public.

Nous traitons ce cas à part, car ces calculateurs peuvent porter sur plusieurs des ensembles de modes de transport : services publics de transport, nouveaux services de mobilité, information routière. Ces calculateurs constituent un service qui implique l'intervention d'algorithmes et d'un savoir-faire qui va au-delà de la simple mise à disposition des données.

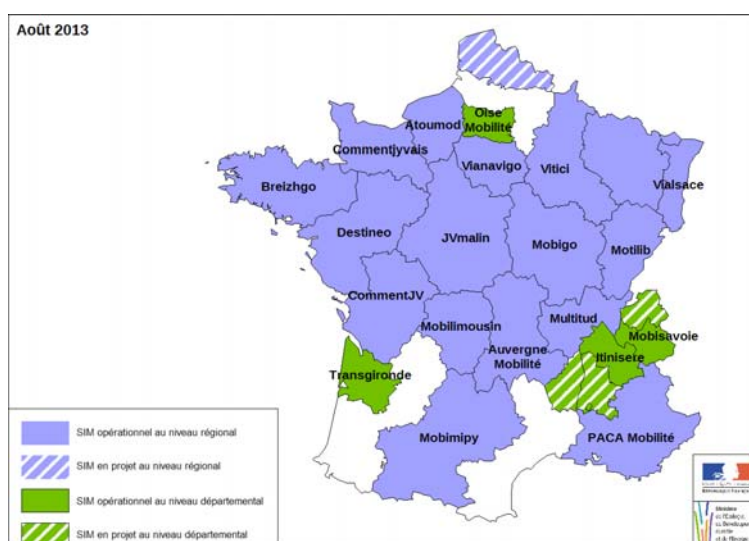
Recommandations sur les services liés à l'information du voyageur [IV]

IV : Les services liés à l'information du voyageur mis en place par les collectivités territoriales sont réutilisables, sous conditions. Il s'agit par exemple des services de calculateur d'itinéraires, de calcul d'isochrones ou du prix d'un trajet.

Prix de l'accès aux services : lorsque les coûts de mise à disposition du service sont significatifs, cette mise à disposition peut être payante ; le total des recettes ne peut excéder les coûts supplémentaires engendrés par la mise à disposition. Les éléments de calcul permettant de justifier le prix de l'accès aux services à partir des coûts supplémentaires sont publiés ; la mise à disposition reste gratuite pour les petits utilisateurs.

La SNCF a formulé une réponse réservée au sujet de l'ouverture d'un calculateur ferroviaire de gare à gare, invoquant un risque d'intermédiation.

Rappel : de nombreux services d'information multimodale (SIM) couvrent le territoire à différentes échelles, permettant localement le calcul d'un itinéraire multimodal. La carte ci-dessous montre les SIM existant au niveau des régions et départements qui seraient concernés par la mise en œuvre de la recommandation ci-dessus :



III.2. – Recommandations juridiques [RJ]

RJ1 : modification du cadre législatif

Insérer dans le Code des transports une disposition législative, concernant les services publics de transport, les nouveaux services de mobilité et le stationnement, et ayant pour objet de rendre réutilisables certaines données ou classes de données « *nécessaires à l'information du voyageur* ».

Aux termes de la loi CADA, seules les informations publiques, c'est-à-dire contenues dans des documents produits ou reçus dans l'exercice d'une mission de service public administratif, sont réutilisables de droit. C'est pourquoi le Comité recommande de soumettre les données nécessaires à l'information du voyageur à un régime distinct de celui des informations publiques.

Dans la partie législative du code des transports serait introduite une disposition créant un droit à réutilisation des informations nécessaires à l'information du voyageur produites dans le cadre d'une mission de service public. Pour les nouveaux services de mobilité et le stationnement, seraient concernés l'ensemble des services exécutés dans le cadre d'une mission de service public ou faisant l'objet d'une convention avec une personne publique.

La loi renverrait à un décret le soin de lister les ensembles de modes de transport concernés ainsi que les données ou classes de données que le Comité aura choisi de rendre réutilisables. Cette liste pourrait faire l'objet d'une mise à jour régulière.

La disposition créant un droit à réutilisation serait complétée par des dispositions relatives aux licences et au mode de calcul du prix éventuel destiné à couvrir le coût supplémentaire engendré par la mise à disposition des données, en investissement et fonctionnement.

Certaines dispositions s'appliquant aux informations publiques seraient reprises ; le prix éventuellement demandé en contrepartie de l'accès aux données ne devrait pas être discriminatoire (cf. art. 15 de la loi CADA), il devrait respecter le principe de tarification au coût marginal posé par la directive PSI, non encore transposée ; enfin, les conditions figurant dans la licence ne pourraient « *avoir pour objet ou pour effet de restreindre la concurrence* » (cf. art. 16 de la loi CADA).

Conformément à la lettre de mission du Comité, la rédaction circonscrirait le droit à réutilisation au champ des « *données nécessaires à l'information du voyageur* ». La référence à l'information du voyageur permet de rattacher l'ouverture des données au droit au transport créé par l'article L. 1111-1 du code des transports, lequel comprend, aux termes de l'article L. 1111-4, « *le droit pour l'usager d'être informé sur les moyens qui lui sont offerts et sur les modalités de leur utilisation* ». L'un des objectifs assignés à l'ouverture des données transport est, en effet, de faciliter l'émergence de services innovants susceptibles de faciliter la mobilité des personnes.

Il serait aussi envisageable de prendre appui sur l'article L. 1211-4 qui cite « *le développement de l'information sur le système des transports* » comme une mission de service public « *dont l'exécution est assurée par l'État, les collectivités territoriales et leurs établissements publics en liaison avec les entreprises privées ou publiques* ». Ce développement de l'information pourrait inclure l'ouverture des données.

RJ2 : élaboration d'un guide de bonne pratique « open data », proposant des clauses-types :

- pour les contrats en cours (délégations de service public, notamment), clauses-types à insérer dans des avenants au contrat, pour préciser les modalités de la mise à disposition des données, par la personne publique ou le délégataire ;
- pour les contrats en cours et les nouveaux contrats (notamment DSP et SIM), clauses-types relatives aux exigences techniques résultant du caractère réutilisable des données.

Dans le respect du principe de sécurité juridique, la loi ne devrait s'appliquer qu'aux données produites dans le cadre de nouveaux contrats (tels que délégation de service publics, marchés publics ou partenariats public-privé).

Pour modifier les contrats en cours, il semble plus prudent d'emprunter la voie de la négociation. Le Comité recommande l'élaboration d'un guide de bonnes pratiques à l'usage des collectivités et des autres acteurs du transport, proposant des clauses-types à insérer dans l'avenant au contrat.

Ces clauses pourraient être rédigées de la manière suivante, selon les cas :

a) les données sont mises à disposition par la personne publique (délégant de service public...) :

« Afin de permettre leur réutilisation dans les conditions de la loi n° 78-753 du 17 juillet 1978, le délégataire fournira [à titre gratuit ou onéreux] au délégant les jeux de données listés en annexe. Il autorise le délégant à réutiliser tout ou partie de ces données. Plus généralement le délégataire autorise la réutilisation des jeux de données [à titre gratuit ou onéreux] par des tiers dans les conditions du chapitre II de la loi précitée ».

b) les données sont directement mises à disposition par le co-contractant (délégataire...) :

« Afin de permettre leur réutilisation dans les conditions de la loi n° 78-753 du 17 juillet 1978, le délégataire s'engage à mettre à disposition [à titre gratuit ou onéreux] de tout réutilisateur, dans des standards ouverts (c'est-à-dire selon l'article 4 de la LCEN du 21 juin 2004 « tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérables et dont les spécificités techniques sont publiques et sans restriction d'accès ni de mise en œuvre ») tels que les formats de type .CSV, .ODS, .XML, .KML, .SHP, les jeux de données listés en annexe extraits d'une base constituée à l'occasion de sa mission de service public objet de la délégation. Ces jeux de données sont mis à disposition [à titre gracieux] par le délégataire qui autorise tout réutilisateur, sans accord préalable, à réutiliser tout ou partie de ces données, dans les conditions de la réutilisation des informations publiques au sens de la loi n° 78-753 du 17 juillet 1978 ».

Sur le plan technique, le guide pourrait proposer des clauses-types sur :

- Les formats attendus (cf. **RT1**) ;
- Les moyens de test et de qualification envisagés ;
- Les outils recommandés.

RJ3 : Licences

Les fournisseurs mettent leurs données à disposition sous Licence Ouverte (Etalab).

Lorsque l'accès aux données est payant, la Licence Ouverte est complétée d'une clause relative à la tarification.

La licence peut être complétée, le cas échéant, selon les recommandations de **SPT4** relatives aux conditions de réutilisation.

Toutefois, en attendant que soient adoptées des mesures de nature à garantir la loyauté des plates-formes, il est souhaitable, pour certaines données considérées comme "stratégiques" d'un point de vue économique, de laisser au fournisseur de données le choix du modèle économique de réutilisation qu'il entend promouvoir : modèle "ouvert", autorisant l'appropriation des données et la revente des enrichissements apportés aux bases, ou modèle coopératif, fondé sur le partage des enrichissements apportés à la base de données initiale.

Ces données peuvent être mises à disposition sous une licence alternative, de type ODbL.

Afin d'harmoniser les conditions de réutilisation, les parties concernées veilleront à poursuivre leur réflexion afin d'identifier, pour chaque ensemble cohérent de modes de transport, les données dont le caractère "stratégique" justifie une mise à disposition sous une licence coopérative.

La Licence Ouverte d'Etalab, incarnant le modèle ouvert, a paru, sous certaines réserves, adaptée au cas du transport.

Cette licence, compatible avec la loi CADA et les principales licences internationales, est de nature à répondre à la plupart des préoccupations exprimées par les collectivités territoriales et les opérateurs de services publics de transport, notamment celles relatives à l'exonération de leur

responsabilité en tant que producteur de données, ainsi qu'à la complétude et à la neutralité de l'information fournie à l'utilisateur.

La Licence Ouverte précise en outre que la mention de paternité « *ne doit ni conférer un caractère officiel à la réutilisation [...] ni suggérer une quelconque reconnaissance ou caution par le « Producteur», ou par toute autre entité publique, du « Réutilisateur» ou de sa réutilisation* », ce qui permet de protéger les éléments de marques.

Toutefois, le temps que soient adoptées des mesures garantissant la loyauté des plates-formes d'intermédiation, il est apparu opportun de laisser aux fournisseurs de données la possibilité d'ouvrir leurs données « stratégiques » sous une licence de type ODbL.

L'intérêt de l'ODbL est de comporter une clause de « partage à l'identique » (dite *share alike*) qui prévient toute appropriation des enrichissements apportés à la base de données initiale. Valorisant le modèle coopératif et permettant la constitution de biens communs, l'ODbL permet d'autres modèles économiques que ceux fondés sur la revente des données.

L'appréciation du caractère « stratégique » des données tient compte de différents éléments et notamment de la taille du réseau de transport. Ainsi l'ODbL pourrait-elle se justifier pour les données transport de grandes agglomérations.

Il convient de noter qu'à la différence de la Licence Ouverte, l'ODbL ne comporte pas de clause contraignant le licencié à ne pas dénaturer le sens de l'information ou à indiquer sa dernière date de mise à jour.

Des précisions pourraient compléter chacune de ces licences, selon une rédaction commune à l'ensemble des acteurs du transport. Pour les « informations temps réel », il pourra être précisé que les réutilisateurs doivent respecter des règles de gestion des flux de voyageurs. D'autres éléments de contrôle (CGU du portail, mentions légales, etc.) pourraient encore venir s'ajouter à la licence.

Toutefois, le Comité souligne qu'il convient de ne pas multiplier ces éléments, l'empilement normatif étant source d'inintelligibilité et d'imprévisibilité. Le Comité note également l'importance de veiller à la cohérence de ces différents éléments et recommande de rassembler les conditions de mise à disposition et de réutilisation des données dans un document unique, accessible via la page permettant au réutilisateur d'accéder aux données.

Dans le cas où le fournisseur souhaite rendre l'accès à ses données payant, la Licence Ouverte seule est proposée, complétée d'une clause relative à la tarification.

RJ4 : identification des réutilisateurs

Pour les informations temps réel, les données d'usage, de billettique et les services liés à l'information du voyageur, les producteurs de données peuvent demander l'identification du réutilisateur préalablement à la distribution de clef d'accès à l'API ; ils peuvent exiger, en complément, une déclaration d'usage.

Pour les autres données, les producteurs de données peuvent inviter les réutilisateurs à créer un compte, cette identification ne revêtant toutefois pas un caractère obligatoire.

Le Comité recommande de donner aux producteurs de données la possibilité d'inviter les réutilisateurs à créer un compte. En rendant possible un dialogue entre producteurs de données et réutilisateurs, la création d'un compte est de nature à stimuler le développement d'une communauté de développeurs. La création d'un compte sensibilise également le réutilisateur à l'importance de respecter les termes de la licence, l'anonymat étant de nature à créer un sentiment d'impunité.

Pour les informations temps réel et les services liés à l'information du voyageur, les producteurs de données peuvent faire de l'identification un préalable à la distribution de clef d'accès à l'API ; ils

peuvent exiger, en complément, une déclaration d'usage leur permettant de connaître les réutilisations qui sont faites de leurs données et d'agir en cas de mauvaises utilisations.

III.3. – Recommandations sur les dispositions techniques permettant de faciliter la réutilisation des données [RT]

Le sous-groupe technique avait pour mission de proposer des dispositions techniques destinées à faciliter la réutilisation des données. À cet effet, le sous-groupe a notamment :

- défini un état de l'art, en s'appuyant sur les retours d'expérience des membres du sous-groupe ;
- procédé à l'audition d'acteurs clés pour approfondir certaines thématiques : opérateurs de SIM, de SAE, des plates-formes de données multimodales, point de vue des réutilisateurs...

À l'issue de cette démarche, le sous-groupe technique a élaboré des recommandations réparties selon trois axes :

1. La donnée transport elle-même ;
2. Les outils ;
3. D'autres éléments techniques.

Les paragraphes suivants présentent une synthèse des recommandations du sous-groupe technique. Des informations plus détaillées figurent en annexe 6 dans le rapport complet du sous-groupe technique. Par ailleurs, pour faciliter la lecture croisée des deux documents, est indiquée la correspondance entre la numérotation des recommandations RT du présent rapport et celle du sous-groupe technique.

Recommandations techniques sur la donnée transport :

Formats d'échange de données

Souhaitant concilier la facilité de réutilisation et l'interopérabilité des données d'une part, ainsi que les contraintes opérationnelles et financières des fournisseurs de données d'autre part, le sous-groupe technique a recommandé plusieurs solutions, par ordre de priorité, pour le choix du format d'échanges de données, dans le cas des services publics de transport.

La recommandation porte sur le choix entre les normes, les standards de fait, tels que GTFS, ou le format propre au producteur des données. Signalons que le format GTFS ne couvre qu'une partie des données transport.

RT1 : services publics de transport :
utilisation des normes et standards

Dans le domaine des services publics de transports, inciter les producteurs de données à mettre à disposition ces données dans les formats d'échange de données suivants, par ordre de priorité :

1. Dans toute la mesure du possible, à la fois aux formats normalisés (NeTEx, SIRI, ...) ET aux formats orientés réutilisation open-data (GTFS) ;
2. À défaut, aux formats orientés réutilisation open-data seulement ;
3. À défaut, au format propre au producteur des données, accompagné de toute documentation utile sur ce format.

(Correspondance avec le sous-groupe technique => recommandation R1)

Dans le champ des nouveaux services de mobilité (vélo, vélo en libre-service, covoiturage, autopartage, stationnement), il n'existe pas de normes équivalentes. Le sous-groupe technique recommande l'élaboration de telles normes permettant de faciliter les échanges de données.

Dans le cas du vélo, un format a été élaboré pour les véloroutes et voies vertes⁵⁰. Il serait très souhaitable d'étendre ce format aux données utilisées en milieu urbain.

⁵⁰ <http://www.certu.fr/geostandard-veloroutes-et-voies-a1119.html>

Les formats d'échange de données ne se conçoivent qu'en lien avec les modèles de données, dont ils dérivent. L'utilisation d'un modèle de données de référence est donc également recommandée.

RT4 : services publics de transports et nouveaux services de mobilité :
élaboration et utilisation de modèles de données en France

Inciter les acteurs transports (producteurs, réutilisateurs...) à utiliser les modèles de données existants, en particulier Transmodel pour les services publics de transport.

Inciter les instances de normalisation (CN03/GT7) à étendre ces modèles aux nouveaux services de mobilité.

(Correspondance avec le sous-groupe technique => recommandation R4)

Qualité des données

Le sous-groupe technique a ensuite formulé une série de recommandations destinées à encourager la mise à disposition de données de qualité et mises à jour. La publication des métadonnées (données sur les données) permet par ailleurs au réutilisateur de connaître les caractéristiques de la donnée.

RT5 : diffusion de données transports de qualité

Inciter les producteurs des données transports à tester la qualité des jeux de données avant leur diffusion et à joindre aux données ouvertes les résultats des tests de qualité qu'elles ont passés.

Inciter les acteurs du transport à utiliser les mêmes données pour la mise à disposition des tiers et pour leurs propres services d'information.

Inciter les producteurs des données transports à mettre en place une organisation pour assurer la mise à jour des données réutilisables. Toute démarche d'ouverture des données devrait inclure un volet « modalités de mise à jour ».

Inciter les producteurs des données transports à publier, pour chaque donnée mise à disposition, les métadonnées décrivant les données (ou l'API).

(Correspondance avec le sous-groupe technique => recommandations R5, R6 et R7)

Plate-forme de données multimodales à l'échelle territoriale

Afin de limiter le nombre de points d'accès et de fournir des données interopérables et de qualité, le sous-groupe technique recommande de généraliser la mise en place de plates-formes de données, sur le modèle de la base REFLEX en Île-de-France ou d'Optimod à Lyon.

Dans ce contexte, il recommande la constitution de bases de données d'arrêts à l'échelle territoriale, en assurant leur cohérence à l'échelle France entière. De telles bases de données permettent en effet :

- au niveau territorial, d'assurer l'interopérabilité des échanges de données sur l'ensemble du territoire considéré ; comportant l'attribution d'un identifiant unique à chaque arrêt, ces bases pourront être utilisées dans les différents métiers du transport (information voyageur, billettique, planification, exploitation), faciliter les échanges de données entre eux, ainsi que les évolutions vers le temps réel et la combinaison de l'information voyageur et la billettique ;
- au niveau France entière, une base de données d'arrêts facilite la mise en place d'un service performant de recherche d'itinéraires de porte-à-porte ; la Grande-Bretagne dispose de cette base de données et l'Allemagne a décidé de la réaliser.

Dans le cadre de l'acte III de la décentralisation, la loi portant Modernisation de l'Action Publique territoriale et d'Affirmation des Métropoles (MAPAM) a prévu la réalisation de Schémas Régionaux de l'Intermodalité (SRI). Adopté par la région en lien avec les autres collectivités, le SRI a pour objectif de coordonner les politiques de mobilité à l'échelle régionale en matière d'offre de services, d'information aux usagers, de tarification et de billettique.

Le Comité recommande que les dispositions législatives relatives au SRI soient complétées par une clause disposant que le schéma décrit l'organisation adoptée, au niveau régional, pour assurer la mise à disposition des données. Il s'agit d'inciter l'ensemble des partenaires, au niveau de la région, à mutualiser leurs moyens afin de faciliter la mise à disposition des données et à diminuer le nombre de points d'entrée pour les réutilisateurs.

Remarque : le SRI pourra notamment décrire s'il convient d'élaborer une ou plusieurs bases de données à l'échelle territoriale, et quel serait le rôle des différents acteurs.

RT6 : élaboration de plates-formes de données multimodales à l'échelle territoriale

Inciter les autorités organisatrices des transports et les autorités organisatrices de la mobilité, en lien avec les opérateurs de transport, à mettre en place des plates-formes de données multimodales. Ces plates-formes agrègent les bases de données des différents acteurs du territoire, les mettent en qualité et diffusent des données interopérables et de référence.

Promouvoir l'élaboration de bases de données arrêts à l'échelle territoriale, en assurant leur cohérence globale à l'échelle nationale.

Ajouter dans le Code des transports (cf. articles 1213-3-1 et 1213-3-2) une disposition disposant que le schéma régional de l'intermodalité décrit l'organisation adoptée sur le territoire régional, pour la mise à disposition des données liées à l'information du voyageur.

(Correspondance avec le sous-groupe technique => recommandations R15, R16 et R23)

Mode de diffusion des données transport

Comme indiqué au §I.3, les initiatives d'ouverture des données se limitent aujourd'hui, pour l'essentiel, à une mise à disposition statique des données (téléchargement sur un portail internet des fichiers informatiques à plat). Seuls quelques rares dispositifs d'accès dynamique, au travers de flux, commencent à apparaître : exemple de Rennes, de Tisséo ou de SNCF Proximité pour le temps réel.

RT7 : diffusion de données brutes ET de services d'information.

En complément à la mise à disposition de données brutes (sous forme de fichiers à plat), inciter les producteurs des données transports à généraliser l'ouverture de services, permettant au réutilisateur d'accéder à une information pré-traitée et simple d'utilisation. Il peut s'agir des services de recherche d'arrêt, de calcul d'itinéraire, d'affichage de résultat sur une carte, etc.

(Correspondance avec le sous-groupe technique => recommandation R8)

Recommandations sur les outils (logiciels pour manipuler la donnée transport) :

En complément à leur réflexion sur l'ouverture des données, les membres du sous-groupe technique ont exprimé le besoin d'utiliser des logiciels permettant de manipuler la donnée transport. Les fonctionnalités souhaitées par les producteurs de données transport concernent notamment le test de la qualité des données (cf. **RT5**), la conversion des données à un format normalisé ou bien l'automatisation de la publication des données sur un portail internet.

RT8 : logiciels facilitant la manipulation des données

Inciter les producteurs des données transports, dans leur démarche d'ouverture des données, à utiliser :

- des outils de vérification de la conformité aux profils normalisés. Un outil CHOUETTE open source existe déjà pour la norme NEPTUNE et pourrait être amélioré pour d'autres normes ;
- des outils de conversion d'un format à un autre. CHOUETTE propose également des fonctions de conversion ;
- des outils harmonisés de publication des données et métadonnées.

À des fins de mutualisation, ces outils pourraient être en open source et référencés.

(Correspondance avec le sous-groupe technique => recommandations R10, R11, R12, R13 et R14)

Autres recommandations techniques :

La mise en place de lieux d'échanges (tels qu'un forum de discussion) peut permettre aux réutilisateurs de confronter leurs expériences ou difficultés, d'exprimer le besoin d'enrichissements et de proposer la publication de nouveaux types de données.

RT9 : mise en place d'un écosystème « *developer-friendly* »

Inciter les producteurs des données transports à proposer aux réutilisateurs des lieux de discussion leur permettant d'échanger sur les services développés et les retours d'expériences, ainsi que de s'entraider dans leurs réalisations.

Encourager les opérateurs et les autorités publiques à développer un écosystème public-privé favorisant l'innovation, la qualité et le développement économique.

(Correspondance avec le sous-groupe technique => recommandations R18 et R20)

III. 4. – Recommandations destinées à favoriser le développement de l'économie numérique dans le domaine du transport [RE]

Le sous-groupe modèles économiques avait pour objectif de décrire les conditions permettant l'apparition de services ayant un modèle économique viable. Les recommandations formulées ci-après sont exclusivement d'ordre économique, les recommandations techniques ou juridiques ayant été présentées dans les deux premières sections.

RE1 : Afin de favoriser le décollage de marchés naissants, mettre les données publiques à disposition des tiers de manière gratuite et ce quel que soit le format de rediffusion.

La tarification des données est une barrière à l'entrée pour les acteurs de taille modeste, et donc un frein au décollage de marchés naissants. Les petits acteurs (PME, start-up) n'ont pas les moyens d'acheter les données, le risque est que seules les entreprises en position dominante le puissent.

Des conditions de service spécifiques peuvent cependant donner lieu à tarification (ex : garantir une bande passante ou une disponibilité 24/24 7/7 du flux de données..., nécessitant des investissements au-delà du service de base). Au-delà d'un certain seuil de sollicitations, un service peut nécessiter une infrastructure technique et une exploitation spécifiques pour le réutilisateur demandeur.

RE2 : Privilégier, pour le secteur public, la fourniture d'un service de base d'information voyageur afin de laisser des espaces au secteur privé.

Cet espace pourrait être, par exemple, le développement des services sur les applications mobiles et les services avancés sur internet.

Cela suppose que le secteur public accepte, par exemple, de ne plus développer et de retirer ses applications mobiles, qui, compte tenu des contraintes réglementaires, seront toujours en retard sur les développements technologiques et les usages.

RE3 : Donner de la visibilité aux entreprises innovantes en leur permettant de bénéficier des marques des territoires pour la diffusion de leurs produits.

Cela passe par une logique de labellisation afin de permettre à des produits numériques d'être mis en avant, par exemple via les sites web des acteurs publics urbains, ces sites internet étant parmi les plus consultés dans les métropoles notamment.

La labellisation devrait favoriser des modèles économiques autonomes des fonds publics.

L'attribution de labels ou de distinctions devra respecter les principes de transparence, d'égalité de traitement et de non discrimination.

RE4 : Donner gratuitement accès à des services clefs, qui ne sont maîtrisés aujourd'hui que par un faible nombre d'acteurs.

Il s'agit ici d'ouvrir le marché des services d'aide à la mobilité, pouvant combiner plusieurs modes de transport : transport collectif, vélo, covoiturage, etc.

Pour cela, le Comité recommande que les collectivités mettent à disposition des réutilisateurs leurs systèmes d'information transport, via des web services.

En cohérence avec les propositions précédentes, cette mise à disposition devrait se faire à titre gratuit, sauf cas particuliers.

III. 5. – Recommandations transversales [RTV]

Mise en place progressive des dispositions en matière d'ouverture des données

La mise à disposition des données ne représente pas des coûts élevés, dans l'ensemble du budget transport d'une autorité organisatrice. Toutefois, le Comité souhaite tenir compte du contexte économique difficile dans lequel les collectivités évoluent aujourd'hui. C'est pourquoi il formule la recommandation suivante :

RTV1 : Mettre en place de manière progressive les dispositions en matière d'ouverture des données : une première étape pourra concerner, outre les services régionaux et départementaux de voyageurs, les services de mobilité présents dans les agglomérations dont le nombre d'habitants est supérieur à un seuil à définir.

Répertoire des données réutilisables

Il semble opportun que certaines dispositions s'appliquant aux informations publiques détenues par l'État s'appliquent également aux données que le Comité aura rendues réutilisables.

Ainsi, le décret n° 2011-577 du 26 mai 2011 relatif à la réutilisation des informations publiques détenues par l'État et ses établissements publics administratifs, complété par la circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'État data.gouv.fr par la mission Etalab et à l'application des dispositions régissant le droit de réutilisation des informations publiques, dispose que :

« Chaque ministère désigne un interlocuteur unique pour « Etalab » [...] responsable [...] des points suivants :

- identifier les informations publiques produites ou reçues dans le cadre des missions de service public ;
- coordonner le recensement et la qualification des informations publiques ;
- mettre en place une méthode, avec la direction des systèmes d'information du ministère, pour transmettre régulièrement les informations publiques dans des formats exploitables et accompagnées de leurs informations descriptives (métadonnées) ;
- gérer l'attribution et le contrôle des droits d'accès à data.gouv.fr, et les réponses de son administration aux questions et aux demandes adressées par les réutilisateurs ».

Par analogie, le Comité recommande que les détenteurs d'informations réutilisables aient l'obligation d'établir et de mettre en ligne un répertoire des données réutilisables qu'ils détiennent.

Par conséquent, le Comité formule la recommandation suivante :

RTV2 : Inciter les fournisseurs de donnée à mettre en place un répertoire identifiant les données mises à disposition et les conditions d'accès (format, licences, services web...)

Mettre en place un annuaire national des données ouvertes.

Cas où plusieurs personnes détiennent les mêmes données réutilisables

RTV3 : Si plusieurs personnes détiennent les mêmes données réutilisables, elles peuvent décider entre elles laquelle assurera la mise à disposition des données et rendra accessible en ligne le répertoire des informations réutilisables ; elles peuvent également décider de confier ces tâches à un tiers.

Dispositif spécifique pour l'accès aux données à des fins d'études et de recherche

Dans le cas où l'accès aux données est conditionné au paiement d'une redevance, le Comité se dit favorable à l'application d'un principe de gratuité pour une réutilisation aux fins d'études et de recherche. Il est proposé de calquer le dispositif mis en place par l'IGN, qui met gratuitement ses données à disposition des organismes exerçant une mission de service public ni industrielle ni commerciale (CEREMA...) et des organismes de recherche et d'enseignement (ENTPE...) ⁵¹.

Le Comité convient de l'intérêt de mettre en place un dispositif permettant aux chercheurs d'avoir accès à des données publiques non ouvertes. Aucune obligation ne devrait toutefois peser sur les fournisseurs de données. La voie du partenariat et de la négociation est privilégiée.

Le Comité note que les données peuvent être communiquées sous une forme non anonymisée lorsque leur traitement se fait à des fins statistiques ou à des fins de recherche scientifique ou historique, conformément aux dispositions de l'article 6 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ⁵².

RTV4 : Lorsque le fournisseur fait payer l'accès à ses données, les organismes exerçant une mission de service public administratif ainsi que les organismes de recherche et d'enseignement devraient bénéficier de mesures d'exonération totale.

L'accès aux données non réutilisables pourrait se faire dans le cadre d'une convention pluri-annuelle liant l'organisme fournisseur et l'organisme demandeur et comprenant éventuellement une clause de confidentialité afin d'assurer la protection des données personnelles et/ou du secret commercial. Un modèle-type pourrait être proposé. La coopération entre les deux organismes s'inscrirait dans une logique de partenariat gagnant-gagnant, le demandeur devant communiquer au fournisseur le résultat de ses recherches. Après signature de cette convention-cadre, l'organisme demandeur devrait adresser, pour chaque projet de recherche, une demande d'accès à certaines données détenues par le fournisseur ; place serait alors laissée à la négociation.

Notion d'information d'intérêt général

Le débat sur l'ouverture des données porte sur le périmètre défini par le CIMAP, c'est-à-dire sur l'ensemble des données produites dans le cadre d'une mission de service public, alors qu'aux termes de la loi CADA, les seules données réutilisables sont les informations publiques, c'est-à-dire les informations produites dans l'exercice d'une mission de service public à caractère administratif (cf. schéma infra).

Le périmètre CIMAP représente donc un élargissement considérable du périmètre CADA.

Toutefois, il est apparu qu'il était difficile de définir le statut juridique de certains services de transport (cf. I.2) et, en particulier, de déterminer si certains nouveaux services de mobilité étaient bien réalisés dans le cadre d'une mission de service public ; par ex. un service d'autopartage géré par une société civile d'intérêt collectif subventionnée par une collectivité locale. Or les dispositions

⁵¹ <http://professionnels.ign.fr/gratuit>

⁵² « Un traitement ne peut porter que sur des données à caractère personnel qui satisfont aux conditions suivantes :

1° Les données sont collectées et traitées de manière loyale et licite ;

2° Elles sont collectées pour des finalités déterminées, explicites et légitimes et ne sont pas traitées ultérieurement de manière incompatible avec ces finalités. Toutefois, un traitement ultérieur de données à des fins statistiques ou à des fins de recherche scientifique ou historique est considéré comme compatible avec les finalités initiales de la collecte des données, s'il est réalisé dans le respect des principes et des procédures prévus au présent chapitre, au chapitre IV et à la section 1 du chapitre V ainsi qu'aux chapitres IX et X et s'il n'est pas utilisé pour prendre des décisions à l'égard des personnes concernées ; [...] »

à mettre en place devraient être d'application simple ; leur mise en œuvre ne devrait pas nécessiter une étude juridique pour déterminer le statut exact du service !

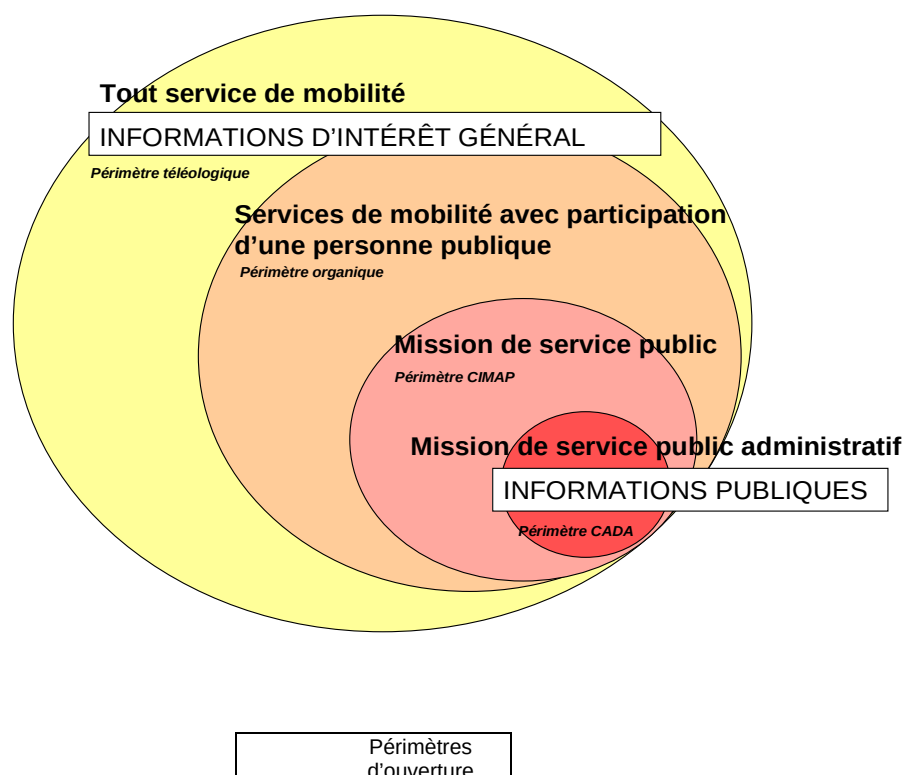
C'est pourquoi le Comité jugerait opportun d'élargir, au-delà des missions de service public au sens strict, le périmètre des services dont les données pourraient être rendues réutilisables.

Deux pistes d'élargissement du périmètre semblent pouvoir être explorées :

- une piste organique consistant à tenir compte de la participation d'une personne publique à la mise en place, au fonctionnement ou au financement du service considéré ; dans cette approche, le principe d'ouverture s'appliquant aux données publiques pourraient s'étendre à des données produites dans le cadre de services de mobilité donnant lieu à une convention avec une personne publique ;
- une piste téléologique consistant à considérer les données du point de vue de leur finalité et en introduisant une notion « d'information d'intérêt général » (IIG) ; une IIG est une information produite dans le cadre de services au public dont l'ouverture est considérée d'intérêt général, car elle rend possible la mise en place de nouveaux services à l'intention du public ; pour le domaine du transport, il s'agirait de services facilitant la mobilité.

Le Comité n'a pu procéder à l'analyse approfondie de ces pistes de travail. Toutefois, une comparaison sommaire met en évidence que :

- le périmètre organique, plus large que le périmètre CIMAP, permet d'englober des services ne relevant pas stricto sensu d'une mission de service public : par ex. une aire de covoiturage implantée sur le parking d'un supermarché sous convention avec un conseil général ; toutefois, il ne permet pas d'englober l'ensemble des services de mobilité ;
- le périmètre téléologique, très large, présente l'avantage de pouvoir englober toutes les données de l'information transport, en fonction des évolutions éventuelles.



Les données (ou classes de données) des services de mobilité dont l'ouverture serait jugée opportune pourraient donc être déclarées « information d'intérêt général ». D'autres données pourraient venir s'ajouter ultérieurement, si une évolution du périmètre était souhaitée.

À l'instar d'une information publique, une « information d'intérêt général » serait réutilisable de droit. Toutefois, ce droit s'exercerait selon une logique totalement distincte de celle de la loi CADA, puisqu'il :

- ne s'articulerait pas à un droit à communication des documents administratifs (une partie seulement de ces informations étant susceptible d'être contenue dans de tels documents) ;
- s'exercerait indifféremment de la distinction entre mission de service public administratif et mission de service public industriel et commercial ;
- pourrait, dans certains cas, s'étendre à des données produites par des personnes privées dans le cadre d'un service ne relevant pas d'un service public, mais faisant l'objet d'une convention avec une personne publique ;
- pourrait, in fine, s'étendre à des données purement privées. À titre d'exemple, des données relatives aux bornes de recharge des véhicules électriques (cf. annexe 8) ou au guidage indoor, de source privée, pourrait être déclarée IIG.

La notion « d'informations d'intérêt général » offre une plasticité appréciable et pourrait s'adapter, si nécessaire, aux évolutions du cadre européen.

Cette notion présente aussi l'intérêt de pouvoir être utilisée dans l'ensemble des domaines listés par le CIMAP : santé, éducation, logement etc. De ce fait, la piste de la création d'un statut « d'information d'intérêt général » pourrait être examinée dans le cadre des travaux préparatoires de la loi sur le numérique.

Grâce à cette notion d'IIG, une méthode générale serait donc disponible pour rendre les données réutilisables.

Compte tenu de l'ensemble de ces considérations, le Comité formule la recommandation générale suivante :

RTV5 : Dans le cadre des travaux préparatoires à la loi sur le numérique, examiner l'opportunité de créer une notion d'information d'intérêt général, applicable à de nombreux domaines (transports, logement, santé etc).

Les informations d'intérêt général seraient dotées d'un statut propre, distinct du statut des informations publiques défini par la loi CADA ; ce cadre offrirait toute souplesse pour accueillir différentes informations qui seraient rendues réutilisables au fur et à mesure des évolutions souhaitées ; il pourrait être plus large que le strict champ du service public.

CONCLUSION

Les données des services publics de transport nécessaires à l'information du voyageur ne sont pas réutilisables de droit. Faut-il les rendre réutilisables ? Telle était la principale question posée au Comité. Certes, des initiatives d'ouverture des données de plus en plus nombreuses sont prises dans le monde et en France, mais on savait les avis très partagés sur le sujet et la lettre de mission avertissait que « *les autorités organisatrices des transports et les opérateurs de transport ont des positions différentes sur leur rôle* ».

Pourtant, dès les premières réunions, au-delà de la variété des démarches et de la diversité des positionnements, le principe d'une ouverture des données des services publics de transport recueillait le consensus.

Les effets bénéfiques d'une politique globale d'open data ont été soulignés : amélioration de l'information délivrée à l'usager des transports, promotion des modes de transport alternatifs à la voiture individuelle, transparence de l'action publique, développement de l'économie numérique en France.

Dans ce contexte, pour les services publics de transport, les nouveaux services de mobilité (vélo, autopartage...) et le stationnement, le Comité recommande l'ouverture des données selon une règle simple et souple, s'appuyant sur une classification des données : selon la nature des données, il convient d'ouvrir l'ensemble de la classe de données ou uniquement certaines données de la liste, données dont la liste pourra évoluer.

Le Comité a souhaité entendre les participants qui ont fait valoir certains risques pouvant résulter de l'ouverture des données et proposé les conditions à imposer aux réutilisateurs des données pour remédier à ces risques. Il s'avère que plusieurs de ces conditions correspondent à des exigences déjà présentes dans la Licence Ouverte (Etalab) et il paraît justifié de les préciser dans le cas du transport. Pour les données en temps réel, une condition supplémentaire a paru légitime : celle de pouvoir imposer des règles de gestion des flux de voyageurs, comme cela se pratique déjà dans le domaine de l'information routière.

Le Comité a souhaité également être attentif aux préoccupations financières exprimées par les collectivités territoriales : c'est la raison pour laquelle il a prévu la possibilité de faire supporter par les réutilisateurs les surcoûts résultant de la mise à disposition des données, lorsque ces surcoûts sont significatifs : les données en temps réel peuvent être dans ce cas de figure. Ces dispositions respectent les règles de tarification au coût marginal fixées dans la directive PSI.

Dans le cas des services de transport en situation de concurrence, certains opérateurs ont décrit un risque d'abus de position dominante d'un fournisseur d'information. Les analyses effectuées ont confirmé l'existence de ce risque. En conséquence, l'ouverture de ces données ne devrait être effectuée que si elle est accompagnée de mesures robustes de nature à prévenir tout risque d'abus de position dominante.

Le Comité reconnaît qu'il ne pouvait pas dans les délais impartis analyser l'ensemble des données susceptibles d'être ouvertes. Son rapport propose néanmoins une méthode susceptible d'être appliquée à tout type de donnée et recommande que le travail d'examen des données cas par cas soit poursuivi : il suggère en particulier que l'ensemble des parties concernées puisse examiner prochainement le cas des bornes de rechargement des véhicules électriques.

Les travaux ont mis en évidence que le strict périmètre de la mission de service public n'était pas pleinement pertinent en matière d'ouverture des données nécessaires à l'information du voyageur : parmi les nouveaux services de mobilité, des services identiques peuvent avoir ou non le statut de service public.

Par conséquent, le Comité recommande d'examiner la possibilité de créer un statut « d'information d'intérêt général », qui permettrait d'élargir le périmètre des informations réutilisables au-delà de

celles produites dans le cadre d'une mission de service public. Le projet de loi sur le numérique en cours de préparation devrait en offrir l'opportunité.

Les réflexions du Comité ont également vocation à nourrir la position de la France dans les débats européens relatifs à la mise en œuvre de la directive ITS.

Le Comité s'est enfin attaché à formuler des recommandations sur les dispositions d'ordre technique ou organisationnel permettant de faciliter la réutilisation des données. En complément des mesures législatives proposées, la mise en œuvre de ces dispositions pourrait donner lieu à un plan d'action global.

Nous adressons nos vifs remerciements aux membres du Comité, aux personnes auditionnées, ainsi qu'aux participants aux groupes de travail et à leurs présidents, pour avoir activement contribué à l'élaboration de ce rapport.



PARTIE 4 : LES ANNEXES

Annexe 1 : Lettre de mission

Annexe 2 : Composition du Comité du débat

Annexe 3 : Tableau comparatif des licences de réutilisation

Annexe 4 : Rappels sur la loi CADA

Annexe 5 : Rapport du sous-groupe juridique

Annexe 6 : Rapport du sous-groupe technique

Annexe 7 : Rapport du sous-groupe modèles économiques

Annexe 8 : Le cas des données des bornes de recharge de véhicules électriques

Annexe 9 : Position de la SNCF

Annexe 10 : Position de la FNMS

Annexe 11 : Position de l'UTP

Annexe 1 : Lettre de mission adressée à Francis Jutand



MINISTÈRE DE L'ÉCOLOGIE, DU DÉVELOPPEMENT DURABLE ET DE L'ÉNERGIE

*Le secrétaire d'État chargé des
Transports,
de la Mer et de la Pêche*

Paris, le 7 mai 2014

Réf. : 14007566

Monsieur le Directeur,

Le gouvernement a souhaité donner un élan fort au mouvement d'ouverture des données publiques, dans le cadre de l'effort de Modernisation de l'action publique (MAP).

Le Comité interministériel pour la modernisation de l'action publique (CIMAP), présidé par le Premier ministre, a réaffirmé le principe de gratuité de la réutilisation des données publiques et décidé de mettre en œuvre des débats thématiques ouverts, afin de favoriser l'ouverture de données stratégiques ayant un fort impact sociétal et un fort potentiel d'innovation sociale et économique.

L'une des thématiques retenues pour ces débats est l'offre de transport. Une plus large ouverture des données dans ce domaine permettrait de favoriser le développement de services facilitant l'usage des transports publics et les nouvelles formes de mobilité. Cela permettrait également d'offrir au public des services innovants combinant des informations sur les transports et des informations relatives à d'autres domaines (tourisme, environnement, santé...).

Si les attentes en matière d'information sur l'offre de transport sont fortes, les préoccupations et les objectifs des acteurs sont variés : tandis que les usagers attendent de pouvoir préparer plus facilement leurs déplacements et d'être informés en temps réel, les autorités organisatrices des transports et les opérateurs de transport ont des positions différentes sur leur rôle. Certains prennent des initiatives en matière d'ouverture des données (*open data*) et d'autres sont désireux de pouvoir maîtriser la qualité de l'information fournie à leurs clients. Enfin, les acteurs de l'économie numérique souhaitent pouvoir développer de nouveaux services.

Comme je l'ai annoncé lors de la journée « Mobilité 2.0 » le 11 février, je vous confie la responsabilité de présider les travaux du Comité du débat. Ce Comité a pour mission d'élaborer des recommandations concernant l'ouverture des données publiques relatives à l'offre de transport, en précisant les bénéfices attendus.

Ce Comité, dont vous mettrez au point la composition sur la base des propositions que vous avez préparées, se réunira dès le mois de mai 2014. L'Agence française pour l'information multimodale et la billettique (AFIMB) se tient à votre disposition pour en assurer le secrétariat.

Monsieur Francis Jutand
Directeur scientifique
Institut Mines-Telecom
46 rue Barrault
75634 Paris cedex 13

Les données sur lesquelles porteront les travaux sont celles utiles à l'information du voyageur (arrêts, horaires de transport...). L'ensemble des modes de transport et des services de mobilité, tels que, par exemple, le covoiturage ou le vélo en libre-service, méritera d'être considéré. Toutefois, dans les différents modes de transport, les données à prendre en compte, selon le cadre défini par le CIMAP, sont celles des collectivités territoriales, des délégataires de service public, des missions de service public à caractère industriel et commercial et des entreprises publiques.

Parallèlement à l'ouverture des données proprement dite, le Comité pourra également examiner d'autres modalités de mise à disposition de l'information, par exemple sous forme de services de calcul d'itinéraire.

Votre réflexion ne se limitera pas à examiner quelles données rendre réutilisables, mais portera également sur les conditions de cette réutilisation. À cet égard, j'attends que le Comité formule notamment des recommandations sur les trois questions suivantes :

- convient-il, et dans quelle mesure, de subordonner la réutilisation de certaines données à des conditions liées à l'intérêt général ou au respect des politiques publiques ?
- quelles dispositions devraient être prises en vue d'éviter le risque de position dominante d'un fournisseur de services ?
- quels sont les grands principes d'un modèle de licence permettant, si cela s'avérait nécessaire, de prendre en compte les spécificités du transport et de la mobilité ?

Sur le plan technique, le Comité pourra examiner si des dispositions particulières (modèles de données, formats, outils de qualification des données...) devraient être adoptées en vue de faciliter la réutilisation.

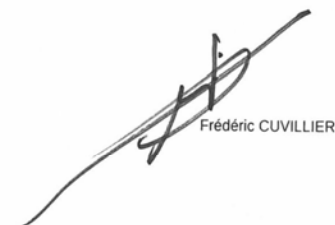
Je tiens à ce que le Comité puisse entendre les différents points de vue et, sur cette base, formuler des propositions qui prennent en compte l'objectif d'amélioration de l'information des voyageurs et les préoccupations légitimes exprimées par les acteurs du transport.

À cet égard, les travaux du Comité prendront en compte le risque de captation de la valeur pour les services de transport en situation de concurrence, en cas d'émergence d'opérateurs de service d'information acquérant une position dominante.

L'information routière est d'ores et déjà élaborée très largement à partir de données publiques et privées. Dans ce domaine, le Comité évitera l'inconvénient de mener une réflexion qui serait limitée aux seules données publiques. Il pourra, sur ce point, s'en remettre aux travaux menés à l'initiative de la Commission européenne pour la mise en œuvre de la directive sur les systèmes de transport intelligents.

Il me serait agréable que le Comité remette ses recommandations en novembre 2014.

Je vous prie d'agréer, Monsieur le Directeur, l'expression de ma meilleure considération.



Frédéric CUVILLIER

Annexe 2 : Composition du Comité du Débat

Président

JUTAND Francis, directeur scientifique de l'institut Mines-Télécom et membre du Conseil national du numérique

Réunions plénières du Comité

Membres

AYOUN Philippe, sous-directeur des Études, des statistiques et de la prospective, représentant la Direction générale de l'aviation civile (DGAC), Ministère de l'écologie, du développement durable et de l'énergie

BISSON Frédéric, vice-président de la Fédération nationale du covoiturage (Féduco)

BLANCHON Charles-Guillaume, chef du Bureau du pilotage des systèmes d'information, représentant le secrétariat général des ministères de l'écologie, du développement durable et de l'énergie et du logement.

BONDUELLE Ithier, chargé de la mission de transports intelligents (GART)

CAMBILLAU Gérard, délégué Transport et Mobilité à l'Institut Mines-Télécom

CHIGNARD Simon, expert des questions liées à l'ouverture des données

COLDEFY Jean, adjoint responsable service Mobilité Urbaine du Grand Lyon, en charge du projet Optimod-Opticities

CORRADI Olivier, *data scientist* (SNIPS)

DE TERNAY Dominique, responsable de l'Unité marketing au sein du département commercial de la RATP

DICIANNI Bruno, adjoint au sous-directeur des transports ferroviaires et collectifs et des déplacements (DGITM / DST)

DUQUESNE Christophe (Aurige), directeur du groupe de normalisation AFNOR CN03/GT7 en charge de l'information voyageur

FRANCHINEAU Jean-Laurent, directeur du programme Eco-Mobilité (Institut Védécom)

FREDERIC Anne-Marie, chargée de mission Aménagement du territoire et mobilités à l'Assemblée des départements de France (ADF)

GILLAIZEAU Aymeric, secrétaire de la FNAUT Pays de Loire

HESLOUIN Benoît, chef du projet ZenBus (JOUL)

KESSLER Bernadette, responsable du service Innovation numérique (Rennes Métropole)

LALANNE Romain, responsable Open data à la SNCF

LAZARD Jean-Marc, fondateur d'Open data soft

LEBORGNE Guillaume, en charge du projet SNCF Transparence, direction de la communication, SNCF

LEVALLOIS-BARTH Claire, maître de conférences en droit à Télécom ParisTech, coordinatrice de la chaire « valeur et politiques des informations personnelles » de l'Institut Mines-Télécom

LOMBARD Amaury, conseiller Infrastructures et Transports (ARF)

MEYER Anne, directrice du département des affaires économiques et techniques (UTP)

MICHAUD Véronique, secrétaire générale du Club des villes et des territoires cyclables (CVTC)

MICHEL Fabrice, représentant la fédération nationale des associations d'usagers des transports (FNAUT)

MIGNEREY Pascal, directeur de l'Autorité de la qualité du service dans les transports, au sein du conseil général de l'Environnement et du Développement durable (CGEDD)

PIRODON Patrick, président de la Fédération nationale des métiers du stationnement (FNMS)

PLACE Dominique, chef de la division mobilité des personnes (CGDD)

RIEG Julie, responsable des études et du développement du groupe Chronos, un cabinet d'études et de prospective animant le programme Datact

ROCHE Yvon, directeur des services d'information du réseau CITIZ (ex-France Autopartage)

SAINTILLAN Christophe, directeur des infrastructures de transport (DGITM / DIT)

SCEMAMA Gérard, directeur de recherche à l'IFFSTAR, co-responsable de l'équipe RESINI (REseau SErvice INterface Intermodalité)

SCHAUSS Samuel, chargé de mission « Ouverture des données » chez SNCF Voyages

SCHWOB Bernard, directeur de l'AFIMB

TALES Romain, responsable du recensement des données publiques au sein de la mission Etalab

VACHERET Olivier, responsable du pôle information transports (STIF)

Acteurs auditionnés

AGUILERA Vincent, chef de la Division Optimisation de l'Usage des Réseaux (CEREMA/DTecITM)
BARBIER David, regional Manager Content Operations France (Tom-Tom)
BERJOAN Sybille, directrice Monde des Études Transports Accenture Research (Paris)
BOUTEILLIER Catherine, doctorante au Laboratoire d'Économie des Transports (Lyon)
BREUIL Éric, responsable information voyageurs digitale et relation clients (SNCF)
CARIOU Alix, Club des villes et des territoires cyclables (CVTC)
CHEVEREAU Laurent, en charge de l'information multimodale (CEREMA)
DELORT Thierry, chef du bureau de l'observation des marchés (DGAC)
FIERLING Matthieu, chef de projet, syndicat mixte Autolib' Métropole
JAMET Anne-Sophie, chargée de mission Partenariat international, veille et expérimentations, Agence de la Mobilité, Direction de la Voirie et des Déplacements (Ville de Paris)
LANGE Thierry, chef du service des déplacements, direction de la voirie et des déplacements (Ville de Paris)
MARI Isabelle, directrice de la prospective et des études stratégiques (JC Decaux)
MORETTI Marc, directeur développement (APRR)
MARLIERE-CORFDIR Michèle, adjointe au chef de bureau des systèmes d'information (DGITM / SAGS)
NAVES Jacques, responsable des partenariats de contenus et d'affiliation pour les univers information routière, voyage et restaurant (Michelin Travel Partners)
ROCHE Emmanuel (Chambéry métropole)
SCHAUSS Samuel, chargé de mission « Ouverture des données » chez SNCF Voyages
VALENTINO Julien, chargé de mission auprès du sous-directeur des études, des statistiques et de la prospective à la direction du Transport aérien (DGAC)
WAGNER Laure, responsable de la communication chez Blablacar

Participants

AVERSENG Antoine, adjoint au chef du bureau des opérateurs de transport ferroviaire (DGITM / DST)
BURTZ Frédéric, Directeur général adjoint, en charge des Produits et Partenariats Stratégiques, VSC
GOZLAN Julie, chargée d'études développement et coordination de la politique intermodale du transport de voyageurs au bureau des politiques de déplacement (DGITM / DST)
LAGUNE Marina, adjointe au chef de bureau des politiques de déplacement (DGITM / DST)
LECA Philippe, chargé de Mission Service, Relation Clients et Espaces à la Direction Générale, RATP
MONTEIL Julien, ingénieur données (AQST)
PITTAUINO Alain, directeur métiers chez Transdev
RIQUELME Valérie, responsable des Systèmes d'information pour les vélos en libre-service (JC Decaux)
VERPEAUX Cédric, responsable du pôle « Ville numérique et durable » de la Caisse des Dépôts
ZACHEE Priscilla, cartographe (Tom-Tom)

Sous-groupe technique

Animateur

DUQUESNE Christophe (Aurige), directeur du groupe de normalisation AFNOR CN03/GT7 en charge de l'information voyageur

Membres

BYRD Andrew (Conveyal)
CAMBILLAU Gérard, délégué Transport et Mobilité à l'Institut Mines-Télécom
CARON Philippe, chef IT Architect (Keolis)
CHEVEREAU Laurent, en charge de l'information multimodale (CEREMA)

CHIGNARD Simon, expert des questions liées à l'ouverture des données
CHOLLET-MEIRIEU Gaël, expert Normalisation Systèmes d'Information Voyageurs (SNCF)
GENDRE Patrick, responsable du service d'études « Trafic et Information Multimodale » au CETE Méditerranée (CEREMA)
HESLOUIN Benoît, chef du projet ZenBus chez JOUL
LECARDEUX Benjamin, responsable de la division Urbanisme, Innovation et Expertise (SNCF)
MICHEL Jean-Pierre (Transdev)
OULD HAMICHE M'hammed, responsable de projet (RATP)
PAULAIS Julien, chef de projet SI et référentiel transports (STIF)
PRAT Jean-Luc, chargé de la diffusion de l'Information Multimodale (STIF)
SCEMAMA Gérard, directeur de recherche à l'IFFSTAR, co-responsable de l'équipe RESINI (REseau SErvice INterface Intermodalité)

Acteurs auditionnés

BAUDET Xavier (Navocap)
COLPAERT Pieter (Open Knowledge Foundation)
CROUIGNEAU Guillaume, CEO (Canal TP)
DESCOLAS Dominique (Digimobee)
FLORISSON Marc, chef de projets (Cityway)
FONBOUZE Hervé, responsable des SAEIV (INEO)
GOURAUD Stéphane, architecte SI (Keolis)

Sous-groupe juridique

Animatrice

LEVALLOIS-BARTH Claire, maître de conférences en droit à Télécom ParisTech, coordinatrice de la chaire « valeur et politiques des informations personnelles » de l'Institut Mines-Télécom

Membres

AUGER Delphine, adjointe au chef de bureau du droit pénal, du droit privé et de la déontologie (DAJ)
CORNEVIN Blandine, chef de projet pôle juridique (APIE)
FAUCONNIER Anne, chef de projet (APIE)
FROUARD Marina, chargée de projets juridiques, aux affaires juridiques, marchés publics et patrimoine (STIF)
GOZLAN Julie, chargée de mission au bureau des politiques de déplacement (DGITM / DST)
GUILBOT Michèle, directrice de recherche (IFFSTAR)
LAGUNE Marina, adjointe au chef de bureau des politiques de déplacement (DGITM / DST)
MONDIN-SIMON Sandrine, juriste nouvelles technologies et PI (SNCF)
TALES Romain, responsable du recensement des données publiques au sein de la mission Etalab
TROUBAT Axelle, chef de projet pôle juridique (APIE)

Personnes auditionnées

BONDUELLE Ithier, chargé de la mission de transports intelligents (GART)
CHIGNARD Simon, expert des questions liées à l'ouverture des données
DUQUESNE Christophe (Aurige), directeur du groupe de normalisation AFNOR CN03/GT7 en charge de l'information voyageur, animateur du sous-groupe technique
JOSSE Anne, secrétaire générale de la CADA

Sous-groupe modèles économiques

Animateur

COLDEFY Jean, adjoint responsable service Mobilité Urbaine du Grand Lyon, en charge du projet Optimod-Opticities

Membres

HESLOUIN Benoît, chef du projet ZenBus (JOUL)
LALANNE Romain, responsable Open data à la SNCF
MARZLOFF Bruno, sociologue, prospectiviste et directeur du Groupe Chronos
SCHAUSS Samuel, chargé de mission « Ouverture des données » chez SNCF Voyages
SEGRESTIN Antoine, manager chez Greenwich Consulting
VERPEAUX Cédric, responsable du pôle « Ville numérique et durable » de la Caisse des Dépôts
WAELEBROECK Patrick, spécialiste de la valeur de la donnée, maître de conférences en économie à Télécom ParisTech

Participants

ARANDEL Maxime, chargé de mission au pôle « Ville numérique et durable » de la Caisse des Dépôts
LAZARD Jean-Marc, fondateur d'Open data soft

Secrétariat du Comité

DALENS-MAREKOVIC Delphine, chargée de mission information multimodale et billettique (AFIMB)
SCHWOB Bernard, directeur de l'agence française pour l'information multimodale et la billettique (AFIMB)
SENG Jean, chargé de mission information multimodale (AFIMB)

Annexe 3 : Tableau comparatif des licences de réutilisation

POINTS DE COMPARAISON		ODbL v1.0 (V0 ou VF)	Licence Ouverte (Etalab)	Licences engagée / associée du Grand Lyon
champ d'application		bases de données	informations publiques	données publiques « sensibles » (et à fort potentiel commercial pour la « licence associée ») mises à disposition par le GL, à l'exclusion des marques, logos et signes distinctifs
droits conférés	actions autorisées	toutes	toutes, sous réserve de ne pas « induire en erreur des tiers quant au contenu de l'information »	toutes, sous réserve de ne pas « altérer » les données, notamment leur teneur et leur portée, ou de « dénaturer » leur sens
	finalité	utilisation commerciale, sans toutefois exclure tout autre champ d'activité, indépendamment des supports et des formats	exploitation à titre commercial	utilisation à des fins professionnelles ou privées compatibles avec l'intérêt général, notamment les politiques publiques mises en œuvre par la CU de Lyon ou le producteur
obligations	indication paternité / source	oui	oui, en veillant à ne pas entretenir de confusion sur le fait que le producteur ne reconnaît ni ne cautionne l'exploitation qui est faite de ses données	non
	date de dernière MàJ	non	oui	oui
	partage à l'identique	oui	non	non
	autres obligations	aucune	aucune	- ne pas porter atteinte à l'ordre public - s'abstenir de tout usage contraire aux lois et règlements - informer le Producteur des dysfonctionnements dans les modalités de mise à disposition des données
contrôle		aucun	aucun	- « déclaration de services » du réutili-sateur permettant un contrôle <i>a priori</i> de l'utilisation des données - pouvoir d'audit des SI du licencié
sanctions		résiliation en cas de manquement du licencié	résiliation en cas de manquement du licencié	- suspension à titre conservatoire de la mise à disposition des données en cas de manquement grave du licencié voire résiliation - résiliation pour des motifs d'intérêt général - résiliation en cas de manquement du licencié
redevance		non, sauf dans les pays où la renonciation aux droits de collecter des redevances n'est pas autorisée	non	non, sauf en cas de réutilisation à des fins commerciales par un réutilisateur dont le volume d'activité est supérieur à un certain seuil pour les données mises à disposition sous la « licence associée »

Annexe 4 : Rappels sur la loi CADA

Documents	Droit à communication	Droit à réutilisation
Article 1^{er} : Documents produits ou reçus, dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission, quels que soient leur date, leur lieu de conservation, leur forme et leur support (à l'exception des actes et documents produits ou reçus par les assemblées parlementaires).	Oui , sauf : • s'il s'agit de documents préparatoires à une décision administrative en cours d'élaboration (art. 2) ; • si ces documents ont fait l'objet d'une diffusion publique (art. 2).	Oui , sauf si les documents : • sont produits ou reçus dans le cadre d'une mission de SPIC (art. 10) ; • sont produits ou reçus par des établissements ou institutions d'enseignement et de recherche ou des établissements, organismes ou services culturels (art. 11) ; • sont grevés d'un droit de PI détenue par un tiers aux administrations mentionnées à l'article 1^{er} (art. 10) ; • font l'objet d'un droit d'exclusivité (art. 14) ;
Article 6, I : • Avis du conseil d'État • Documents dont la consultation ou la communication porterait atteinte : – au secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif ; – au secret de la défense nationale ; – à la conduite de la politique extérieure de la France ; – à la sûreté de l'État, à la sécurité publique ou à la sécurité des personnes ; – à la monnaie et au crédit public ; – au déroulement des procédures, ou d'opérations préliminaires, engagées devant les juridictions ; – à la recherche, par les services compétents, des infractions fiscales et douanières ; – aux autres secrets protégés par la loi.	Non	Non , sauf si les documents ont fait l'objet d'une diffusion publique.
Article 6, II : • Documents dont la communication porterait atteinte à la protection de la vie privée, au secret médical et au secret en matière commerciale et industrielle ; • Documents portant une appréciation ou un jugement de valeur sur une personne physique, nommément désignée ou facilement identifiable ; • Documents faisant apparaître le comportement d'une personne, dès lors que la divulgation de ce comportement pourrait lui porter préjudice.	Non , sauf à l'intéressé	Non , sauf : • si les documents ont fait l'objet d'une diffusion publique, sauf si les documents sont produits ou reçus dans le cadre d'une mission de SPIC (art. 10) ; • [pour les informations comportant des données à caractère personnel] si la personne intéressée a consenti à leur réutilisation, si l'autorité détentrice les a anonymisées ou si une disposition (législative ou réglementaire) le permet.

Annexe 5 : Rapport du sous-groupe juridique

Objectifs du sous-groupe :

Formuler des recommandations sur des dispositions permettant de lever les obstacles juridiques à la réutilisation des données et répondre aux trois questions figurant dans la lettre de mission du Comité :

- convient-il, et dans quelle mesure, de subordonner la réutilisation de certaines données à des conditions liées à l'intérêt général ou au respect des politiques publiques ?
- convient-il de prévenir un abus de position dominante d'un fournisseur de services ? le cas échéant, par quels moyens ?
- quels sont les grands principes d'un modèle de licence permettant, si cela s'avérait nécessaire, de prendre en compte les spécificités du transport et de la mobilité ?

I. – Sur les dispositions juridiques facilitant la réutilisation des données publiques de transport Méthode de travail du Comité :

L'article 10 de la loi CADA exclut du droit à réutilisation « *les informations contenues dans des documents : [...] b) [...] produits ou reçus par les administrations mentionnées à l'article 1er dans l'exercice d'une mission de service public à caractère industriel ou commercial [SPIC] ; c) [...] sur lesquels des tiers détiennent des droits de propriété intellectuelle* ».

Or les données transport sont majoritairement produites dans le cadre d'une mission de SPIC par des opérateurs de transport susceptibles de se prévaloir d'un droit de propriété intellectuelle, notamment au titre du droit *sui generis* des bases de données.

Pour permettre la réutilisation de ces données, le sous-groupe juridique propose l'adoption, à court ou moyen terme, de mesures législatives, lesquelles ne devront toutefois pas interférer avec le calendrier législatif, et notamment avec les travaux de :

- transposition en droit interne, vraisemblablement par le véhicule d'une loi sur le numérique modifiant la loi CADA, de la directive n° 2013/37/UE du 26 juin 2013 modifiant la directive 2003/98/CE⁵³ relative à la réutilisation des informations du secteur public ou Public Service Information (PSI) ;
- élaboration de spécifications pour l'action prioritaire a) (mise en œuvre d'un système d'information multimodale) de la directive n° 2010/40/EU du 17 juillet 2010⁵⁴, relative aux systèmes de transport intelligents ou Information Transport Services (ITS).

R1 : Le sous-groupe recommande d'insérer une disposition dans le code des transports rendant réutilisables les données « nécessaires à l'information du voyageur ».

Les données produites dans l'exercice d'un Service Public Industriel et Commerciale (SPIC) ne sont actuellement pas réutilisables de droit.

Le sous-groupe propose de soumettre les données publiques de transport à un régime dérogatoire à la loi CADA en introduisant dans la partie législative du code des transports une disposition créant un droit à réutilisation des données nécessaires à l'information du voyageur produites dans l'exercice d'une mission de service public (sans distinguer entre service public administratif et service industriel ou commercial).

La loi renverrait à un décret le soin de lister les données ou les classes de données que le Comité aura choisi de rendre réutilisables. Cette liste pourrait faire l'objet d'une mise à jour régulière.

Conformément à la lettre de mission du Comité, la rédaction proposée circonscrit le droit à réutilisation au champ des données nécessaires à l'information du voyageur. La référence à l'information du voyageur permet de rattacher l'ouverture des données au droit au transport créé par l'article L. 1111-1 du code des transports, lequel comprend, aux termes de l'article L. 1111-4 du code des transports, « *le droit pour l'utilisateur d'être informé sur les moyens qui lui sont offerts et sur les modalités de leur utilisation* ». L'un des objectifs assignés à l'ouverture des données transport est, en effet, de faciliter l'émergence de services innovants susceptibles de faciliter la mobilité des personnes.

Il serait aussi envisageable de prendre appui sur l'article L. 1211-4 du code des transports qui cite « *le développement de l'information sur le système des transports* » comme une mission de service public « *dont l'exécution est assurée par l'État, les collectivités territoriales et leurs établissements publics en liaison avec les entreprises privées ou publiques* ». Ce développement de l'information pourrait inclure l'ouverture des données.

⁵³ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:FULL:FR:PDF>

⁵⁴ <http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32010L0040&from=EN>

Le sous-groupe attire toutefois l'attention du Comité sur le fait qu'étendre le principe d'ouverture à l'ensemble des données transport, y compris celles non nécessaires à l'information du voyageur, rendrait délicate la justification d'un régime dérogatoire pour les données SPIC relatives aux transports. Cette extension ferait ainsi courir le risque d'une contamination à d'autres SPIC (collecte et traitement des déchets, distribution d'eau etc.)

En complément de cette disposition, si l'idée de généraliser les plates-formes territoriales de données est approuvée par le Comité, il sera possible de prendre appui sur l'article L. 1213-3-1 du code des transports relatif au schéma régional de l'intermodalité ; le 4e alinéa dispose que ce schéma « *prévoit les mesures de nature à assurer une information des usagers sur l'ensemble de l'offre de transports [...]* » ; il pourrait être inséré la phrase suivante : « *il décrit l'organisation globale de l'ouverture des données de mobilité à l'échelle régionale* ».

R1bis : Le sous-groupe juge intéressant l'extension du principe d'ouverture des données au-delà du périmètre fixé par le CIMAP

Deux pistes ont été envisagées :

1. une modification de la loi CADA afin d'abolir la différence de traitement entre données SPA et SPIC ; toute donnée produite dans le cadre d'une mission de service public serait réutilisable de droit ;
2. la création d'un statut « d'information d'intérêt général », susceptible de s'appliquer à des données de statut juridique divers (public ou privé) dont l'ouverture serait jugée d'intérêt général.

Le sous-groupe juridique note que la directive PSI n'interdit pas de telles extensions.

Le sous-groupe attire toutefois l'attention du Comité sur le fait que ces pistes :

- vont bien au-delà de la mission dévolue au sous-groupe, limitée aux données transport ;
- sont proposées tardivement, rendant leur examen hâtif ;
- élargissent considérablement le périmètre initial alors même que le Comité a été bien plus circonscrit dans ces recommandations d'ouverture.

C'est pourquoi le sous-groupe recommande qu'il soit procédé à une étude d'impact de ces possibles extensions, sous un angle à la fois juridique, technique et économique.

Le sous-groupe observe que :

- dans l'hypothèse 1, il semblerait très difficile de faire un cas à part des données TGV, à moins de considérer que ces données sont protégées par le secret industriel et commercial mentionné à l'article 6 de la loi CADA ; toutefois, cet article ne concerne que le droit à communication ; or les données TGV sont publiées...
- dans l'hypothèse 2, il serait nécessaire de faire une analyse de la notion d'intérêt général à travers ses applications sectorielles en consultant, notamment, la jurisprudence du Conseil d'État et de la CADA.

R2 : Le sous-groupe recommande l'adoption d'une mesure législative rendant obligatoire l'insertion d'une clause open data dans les nouveaux contrats, et notamment dans les contrats de délégation de service public de transport.

Le droit *sui generis* des producteurs de bases de données (cf. articles L. 341-1 et suivants du code de la propriété intellectuelle) protège le contenu des bases de données, puisqu'il permet au producteur d'interdire toute extraction ou réutilisation d'une partie qualitativement ou quantitativement substantielle des bases.

Ainsi, les opérateurs de transport délégataires d'une mission de service public peuvent se prévaloir de ce droit *sui generis* sur les bases de données produites dans le cadre de l'exploitation de ce service dès lors qu'ils prennent l'initiative et le risque d'un investissement substantiel (financier, humain ou matériel) pour constituer la base de données.

Or, aux termes de l'article 10 de la loi CADA, les informations contenues dans des documents sur lesquels des tiers aux administrations mentionnées à l'article 1er détiennent des droits de propriété intellectuelle ne sont pas réutilisables de droit.

Il semblerait qu'un opérateur de transport délégataire de service public doive bien être regardé comme un tiers qui détient des droits de propriété intellectuelle au sens de l'article 10 ; « *le délégataire n'est délégataire que du SPIC, et non du SPA. S'il ne saurait être regardé comme un tiers s'agissant du premier, il semble qu'il peut en revanche l'être s'agissant du second, auquel il ne participe pas directement dans le cadre de sa délégation [...]. Les informations étant indissociables de la base de données détenue par le délégataire et à laquelle l'administration a seulement accès, la base et son contenu ne pourront être regardés comme « détenus » par l'administration* ».

Pour empêcher les délégataires de faire obstacle, au nom de leurs droits de propriété intellectuelle fondés sur les articles L.341-1 et suivants du CPI, à la réutilisation de leurs bases de données, le sous-groupe juridique propose d'insérer une disposition dans la partie législative du code des transports rédigée ainsi :

« Lorsque le délégataire (ou co-contractant) de service public de transport détient des droits de propriété intellectuelle sur les bases de données produites dans l'exercice de la mission qui lui a été confiée, les modalités par lesquelles ces bases de données seront ouvertes à la réutilisation par des tiers doivent être explicitement prévues par un contrat conclu entre l'administration et le délégataire. Si, dans l'exercice de la mission qui lui a été confiée, le délégataire (ou co-contractant) fait appel à un prestataire, la cession des droits de propriété intellectuelle susceptibles de grever les bases de données produites par le prestataire doit être organisée dans le contrat conclu entre le délégataire et son prestataire ».

Les entreprises délégataires ne seraient ainsi plus en mesure de négocier l'existence même de cette clause, mais seulement les conditions – notamment financières – de la mise à la disposition de tiers de données ou de jeux de données à des fins de réutilisation.

En droit, la non-rétroactivité est un principe. Aussi, dans un souci de sécurité juridique, la loi ne s'appliquerait qu'aux nouveaux contrats. Pour modifier les contrats en cours, le sous-groupe recommande d'emprunter la voie de la négociation (signature d'un avenant).

En parallèle, le sous-groupe recommande l'élaboration d'un guide de bonnes pratiques à l'usage des collectivités et des autres acteurs du transport, proposant des clauses-types à insérer dans les contrats.

Les clauses-types pourraient être rédigées ainsi :

Cas où les données de transport sont mises à disposition par le délégant dans le cadre de sa politique OPEN DATA

« Afin de permettre leur réutilisation dans les conditions de la loi n°78-753 du 17 juillet 1978, le délégataire fournira [à titre gratuit] au délégant les jeux de données listés en annexe. Il autorise le délégant à réutiliser tout ou partie de ces données. Plus généralement le délégataire autorise la réutilisation des jeux de données [à titre gratuit ou onéreux] par des tiers dans les conditions du chapitre II de la loi n°78-753 du 17 juillet 1978 ».

Alternative ou complément dans le cas où tout ou partie des données de transport concernées par cette politique d'OPEN DATA seront directement mises à disposition par le délégataire

« Afin de permettre leur réutilisation dans les conditions de la loi n°78-753 du 17 juillet 1978, le délégataire s'engage à mettre à disposition [à titre gratuit] de tout réutilisateur, dans des standards ouverts (c'est-à-dire selon l'article 4 de la LCEN du 21 juin 2004 « tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérables et dont les spécificités techniques sont publiques et sans restriction d'accès ni de mise en œuvre ») tels que les formats de type .CSV, .ODS, .XML, .KML, .SHP, les jeux de données listés en annexe extraits d'une base constituée à l'occasion de sa mission de service public objet de la délégation. Ces jeux de données sont mis à disposition [à titre gracieux] par le délégataire qui autorise tout réutilisateur, sans accord préalable, à réutiliser tout ou partie de ces données, dans les conditions de la réutilisation des informations publiques au sens de la loi n°78-753 du 17 juillet 1978 »

II. – Sur les grands principes d'un modèle de licence et la question de la subordination de la réutilisation de certaines données à des conditions liées à l'intérêt général ou au respect des politiques publiques :

R3 : Le sous-groupe recommande de ne pas apporter de restrictions disproportionnées à la liberté de réutilisation prévue par la loi CADA.

En contraignant le réutilisateur à mettre gratuitement à disposition de la communauté les enrichissements qu'il apportés aux bases de données, la clause de partage à l'identique (*share alike*) des données contenue dans l'Open Database License (ODbL) ne permet pas un modèle économique fondé sur la revente de bases de données enrichies.

C'est pourquoi le sous-groupe se prononce en défaveur d'une mise à disposition des données exclusivement sous ODbL. Toutefois, le sous-groupe n'exclut pas la possibilité d'une mise à disposition des données sous plusieurs licences, dont l'ODbL, liberté étant laissée au réutilisateur de choisir la licence convenant à ses besoins.

R4 : Le sous-groupe considère qu'il n'y a pas lieu de créer une licence spécifique pour les données transports, la Licence Ouverte pouvant constituer une base commune

La Licence Ouverte, déclinaison contractuelle de la loi CADA, semble de nature à répondre aux craintes exprimées par les collectivités territoriales et les opérateurs de services publics de transport.

La Licence Ouverte stipule, en effet, que le réutilisateur s'engage à « *ne pas induire en erreur des tiers quant au contenu de « l'Information », sa source et sa date de mise à jour* ». Cette clause, qui ne figure pas dans l'ODbL, garantit la complétude et la neutralité de l'information délivrée et permet au réutilisateur d'apprécier son niveau de fraîcheur.

La Licence Ouverte précise également que la mention de paternité « *ne doit ni conférer un caractère officiel à la réutilisation [...] ni suggérer une quelconque reconnaissance ou caution par le « Producteur », ou par tout autre entité publique, du « Réutilisateur » ou de sa réutilisation* ».

La Licence Ouverte intègre enfin une clause contractuelle permettant de ne pas engager la responsabilité du producteur de données au regard de la qualité de l'information fournie. Cette clause précise, en effet, qu'« *il ne peut garantir l'absence de défauts ou d'irrégularités éventuellement contenues dans « l'Information »* ».

En complément de la base commune constituée par la Licence Ouverture, le sous-groupe n'exclut pas la mise en place, au cas par cas (pour des données sensibles notamment) d'autres éléments de contrôle, précisant, par exemple, que la réutilisation ne doit pas porter atteinte à la sécurité des personnes. Ces éléments comprennent, notamment, les CGU du portail, les mentions légales, la protection des éléments de marque, les conditions spécifiques liées aux API et au web service, les labels.

R5 : Le sous-groupe recommande de ne pas multiplier les éléments de contrôle et d'être attentif au principe de sécurité juridique

Le sous-groupe préconise de ne pas multiplier les éléments de contrôle (licence, CGU, mentions légales etc.), cet empilement normatif étant source d'inintelligibilité et d'imprévisibilité.

Le sous-groupe souligne également l'importance de veiller à la cohérence de ces différents éléments et recommande de rassembler toutes les conditions de mise à disposition et de réutilisation des données dans un document unique, directement accessible via la page qui permet au réutilisateur d'accéder aux données.

R6 : Le sous-groupe souligne son attachement au principe de gratuité, mais attire l'attention du Comité sur le fait qu'il conviendra de résoudre la question de la répartition des coûts de mise à disposition de certaines données.

Le sous-groupe juridique rappelle le principe de gratuité de la réutilisation des données publiques promu par le Gouvernement⁵⁵ et affirmé dans la directive PSI (2013).

Toutefois, il observe que les modalités de mise à disposition des données (notamment via une API) peuvent représenter un coût significatif.

C'est pourquoi le sous-groupe recommande au Comité d'être attentif aux modalités de la répartition des coûts entre les acteurs impliqués dans la mise à disposition des données.

Si la piste d'une redevance devait être retenue, il conviendrait de la justifier dans un contexte général de gratuité des données. La création de nouvelles redevances est, en effet, strictement encadrée. En outre, les données transport entrant dans le champ d'application de la directive PSI, c'est-à-dire détenues par des organismes publics, ne peuvent être tarifées qu'au coût marginal de mise à disposition.

III. – Sur l'opportunité de prévenir l'apparition d'une position dominante d'un fournisseur de services et les moyens permettant d'y parvenir :

Le sous-groupe juridique estime que les discussions ont fait ressortir qu'il existe un risque non négligeable de captation de la valeur des données et d'abus de position dominante d'un fournisseur d'information en position de monopole. Par conséquent, il a étudié la validité juridique des mesures susceptibles de remédier à ces risques.

1. Pour les raisons exposées en **R3**, le sous-groupe écarte la solution consistant à ouvrir les données sous ODbL, qui comporte une clause de *share alike* dont l'effet répulsif sur les intermédiaires de l'économie numérique a été constaté. Le sous-groupe observe, en outre, que pour un opérateur de transport en position de monopole, imposer une obligation de partage à l'identique des données qu'il met à disposition peut être assimilé à un abus de position dominante.

⁵⁵ http://www.france-universite-numerique.fr/IMG/pdf/feuille_de_route_du_gouvernement_sur_le_numerique.pdf ; cf. p. 29.

2. Le sous-groupe écarte la solution consistant à instaurer une redevance anti-monopole sur le modèle de celle du Grand Lyon, c'est-à-dire modulée en fonction du volume d'activité du produit ou du service créé, censée traduire « *la position occupée par le réutilisateur sur son marché* ». D'un montant faible, voire nul, pour les petits acteurs, la redevance s'élève à un montant dissuasif pour les entreprises en position dominante. Le sous-groupe observe que la redevance, qui doit être la contrepartie d'un service rendu, n'apparaît pas comme un outil juridique pertinent pour lutter contre les monopoles. Par son mode de calcul, cette redevance est susceptible d'apparaître comme discriminatoire. Une exonération totale pour la majorité des réutilisateurs pose problème au regard du principe d'égalité, la différence de traitement n'apparaissant pas fondée sur une différence objective de situation au regard du service rendu. La mise en œuvre du critère part de marché pourrait prêter à contestation compte tenu des difficultés méthodologiques auxquelles se heurte la quantification d'une part de marché ou d'activité, surtout au niveau local. Sur un plan technique, ce type de critère paraît peu adapté pour définir un seuil de déclenchement d'une obligation de payer : le montant d'une redevance ne saurait reposer sur des données potentiellement variables dans le temps, dépendantes des autres acteurs sur le marché.
3. Le sous-groupe juridique a également écarté les solutions élaborées par le sous-groupe modèles économiques, qui propose d'interdire l'accès aux données aux entreprises :
 - a. ayant été condamnées par la CNIL (en s'inspirant de l'article 43 du code des marchés publics) ;
 - b. ne s'étant pas acquittées de leurs obligations fiscales et sociales (sur le modèle du même article) ;
 - c. en position de monopole (en s'inspirant d'une décision du conseil de la concurrence de 2001 interdisant à Vinci Park de répondre à des appels d'offre pendant trois ans dans les villes du sud de la France où il était en position ultra-dominante).

Ces restrictions, et tout particulièrement les deux premières, visent à empêcher Google d'accaparer la valeur des données.

La proposition a) suscite des réserves en raison de l'absence de rapport entre l'objet de la licence et la nature de l'infraction qui ferait obstacle à sa souscription. En outre, cette clause n'empêcherait pas une entreprise comme Google de signer la licence, puisqu'elle a formé un recours devant le Conseil d'État (sa condamnation n'est donc pas définitive).

Le sous-groupe juridique observe que la proposition b) ne permet pas d'écarter les intermédiaires de l'économie numérique pratiquant l'optimisation fiscale (le redressement fiscal récemment subi par Google ne constitue pas une condamnation).

Concernant la proposition c), le sous-groupe juridique observe que la situation de monopole de Google sur le marché pertinent n'est pas prouvée et qu'en outre l'on ne s'inscrit pas, en l'espèce, dans le cadre d'un litige portant sur l'attribution de marchés publics.

Par conséquent, le sous-groupe a retenu les recommandations suivantes :

R7 : Le sous-groupe recommande de donner de la visibilité à des entreprises proposant des services innovants.

Le sous-groupe approuve les initiatives de labellisation des services à l'échelle locale, qui lui semble de nature à rééquilibrer le jeu de la concurrence en faveur des petits acteurs, et notamment des start-up.

Le sous-groupe attire toutefois l'attention du Comité sur le fait que l'attribution de label ou de distinctions diverses devra respecter les principes de transparence, d'égalité de traitement et de non discrimination.

R8 : Le sous-groupe promeut l'instauration d'un règlement européen encadrant les rapports entre opérateurs de transport en situation de concurrence et fournisseurs d'information.

Pour prévenir les risques d'abus de position dominante d'une entreprise telle que Google et répondre aux préoccupations exprimées par les opérateurs de transport en situation de concurrence, le sous-groupe recommande, à l'instar du rapport *All Ways travelling* publiée par la DG Move⁵⁶ et sur le modèle de ce qui a été fait dès 1989 pour les systèmes informatisés de réservation dans le transport aérien, de mettre en place un code de conduite régulant les rapports entre les opérateurs de transport et les fournisseurs de moteurs de voyage.

Ce code encadrerait le montant des commissions perçues par l'intermédiaire et imposerait aux moteurs de voyage de fournir une information neutre. Ces dispositions préserveraient les marges des opérateurs de transport et empêcheraient qu'une meilleure visibilité ne soit accordée au plus offrant.

⁵⁶ <http://ec.europa.eu/transport/themes/its/studies/doc/20140812-july9thversion-awtfinalreport.pdf>

Il est probable que l'échelon européen soit le plus pertinent pour cette régulation. Ce dernier a déjà été adopté pour le transport aérien. En cas de manquement, les sanctions prévues par le règlement n° 80/2009 (qui a abrogé le règlement n° 2299/1989) sont dissuasives : les amendes sont calculées sur le chiffre d'affaires.

La régulation pourrait se faire dans le cadre de la mise en œuvre de l'action prioritaire a) définie par la directive ITS précitée.

Annexe 6 : Rapport du sous-groupe technique

Objectifs du sous-groupe :

Formuler des recommandations sur des dispositions techniques permettant de faciliter la réutilisation des données transport.

Méthode de travail du sous-groupe :

Pour mener sa réflexion, le sous-groupe technique a :

- répondu aux sollicitations de la plénière (notamment concernant la création d'une classification des données transport) ;
- procédé à un état de l'art grâce à des retours d'expérience de ses membres ;
- procédé à des auditions de parties prenantes dans les données transports (fournisseurs de SIM, SAE, plateformes de données, acteurs de l'open data).

Plusieurs axes de recommandations ressortent des travaux du sous-groupe :

1. Un axe portant sur la donnée elle-même ;
2. Un axe portant sur les outils permettant de manipuler ces données (produire, lire, contrôler, modifier) ;
3. Un axe portant sur les conditions de mise à disposition des données.

Les recommandations sont formulées selon ces trois axes. Elles sont, de plus, complétées par des attentes non techniques, qui ne sont donc pas du ressort du sous-groupe, mais pour lesquelles des recommandations pourraient être formulées (de la part du groupe juridique ou de la plénière elle-même) : ces attentes sont présentées en ANNEXE-A.

Recommandations du sous-groupe technique :

Synthèse des recommandations

Le tableau ci-dessous présente une synthèse des recommandations, la priorité qui leur a été attribuée (de 1:« très prioritaire », à 3: « moins prioritaire ») et le niveau de consensus (une croix par avis favorable/défavorable). Même si certaines recommandations, comme R.9, sont plus controversées, le niveau moyen de consensus est très bon (95% de favorable).

Note : certains acteurs du sous-groupe n'ont pas rempli ce tableau. Seul les avis exprimés ont été pris en compte.

		Synthèse			
		Priorité 1/2/3 1=très prioritaire	Favorable	Défavorable	
R1	Formats de mise à disposition	1	xxxxxxx	x	Données
R2	Traiter les formats non disponibles	3	xxxxxxxxx		
R3	Définir des profils d'échanges	2	xxxxxxxxx		
R4	Harmonisation des concepts utilisés	2	xxxxxxxxx		
R5	Fournir des indicateurs de qualité des données ouvertes	2	xxxxxxx	x	
R6	Fournir des données à jour	2	xxxxxxx	xx	
R7	Fournir les métadonnées	1	xxxxxxxxx		
R8	Proposer des données et des API	2	xxxxxxxxx		
R9	Données théoriques et temps réel	2	xxxxx	xxxx	
R10	Outils pour vérifier la qualité des données	2	xxxxxxxxx		Outils
R11	Outils de conversion et production aux formats	2	xxxxxxxxx		
R12	Outils de publication	2	xxxxxxxxx		
R13	Les outils doivent être open source	2	xxxxxxxxx		
R14	Mettre en place un référencement des outils	2	xxxxxxxxx		
R15	Mise en place de données de référence (référentiels)	2	xxxxxxxxx		
R16	Mise en mise en cohérence et fusion des données d'un territoire	2	xxxxxxxxx		Conditions de mise à disposition
R17	Éléments de contrôles et CGU harmonisés	2	xxxxxxxxx		
R18	Information sur les réutilisations	3	xxxxxxxxx		
R19	Agir sur les cahiers des charges	2	xxxxxxxxx		
R20	Sites et lieux d'échange et partage	2	xxxxxxxxx		
R21	Formations	2	xxxxxxxxx		
R22	Annuaire des données disponibles en France	2	xxxxxxxxx		Réflexions non techniques
NT1	Les données doivent être disponibles	NA	xxxxx	xx	
NT2	La possibilité de réutilisation doit pouvoir être identifiée de façon claire	NA	xxxxx	xx	
NT3	La documentation associée aux données ouvertes doit être libre d'accès et gratuite	NA	xxxxxxxxx	x	
NT4	Les données d'un maximum de modes (TC et non-TC) sont utiles pour les services à la mobilité	NA	xxxxxxxxx		

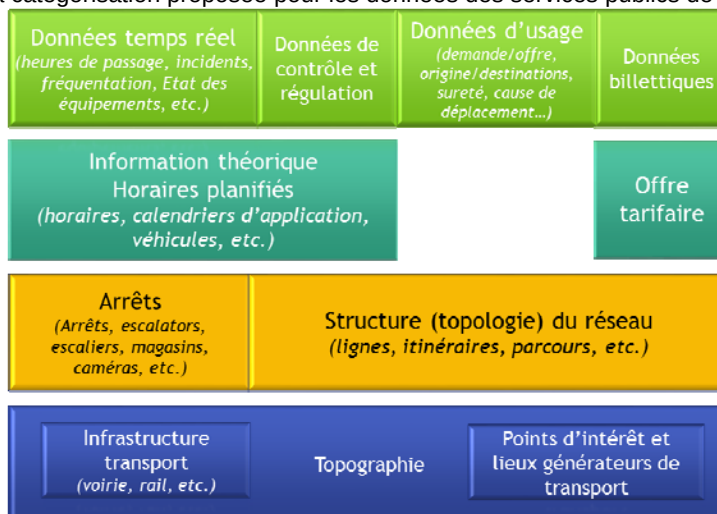
Données et classification des données

À la demande du Comité, le sous-groupe a élaboré une classification des données « transport » et une liste des principales données par classe. Cet exercice est destiné à aider le Comité dans sa réflexion sur les données susceptibles d'être ouvertes.

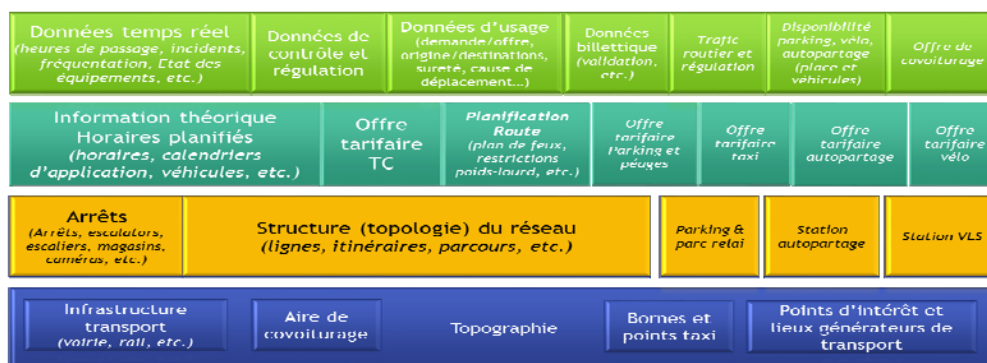
L'exercice est apparu d'autant plus déterminant que le Comité a adopté, en septembre, comme première orientation, le principe que :

- pour certaines classes, les données seraient ouvertes ;
- pour d'autres classes, il y aurait lieu de préciser quelles données de ces classes seraient ouvertes.

La figure ci-dessous présente la catégorisation proposée pour les données des services publics de transport.



La seconde figure, ci-dessous, complète la première en y ajoutant les données des modes de transport connexes.



Les catégories proposées sont décrites plus en détail dans le tableau ci-dessous.

Catégorie de données	Description	Normes, textes réglementaires et standards
Topographie et Infrastructure		
Topographie	La topographie concerne la description du terrain en général, incluant le relief et toutes les caractéristiques naturelles et artificielles. Elle détermine la position et la forme de chacun de ses constituants.	INSPIRE GDF GML
Infrastructures de transport	L'infrastructure de transport est un sous-ensemble spécifique de la topographie, décrivant les routes, rails, piste cyclable, marche, chemin, etc.	GDF Thème transport INSPIRE
Point d'intérêt	Les points d'intérêt sont un sous-ensemble spécifique de la topographie, décrivant les lieux connus étant souvent utilisés comme départ ou destination d'un voyage.	IFOPT NeTex GDF INSPIRE
Station de taxis	Endroit où les taxis peuvent attendre leurs clients	GDF IFOPT
Aire de covoiturage	Emplacement, généralement en latéral de voirie ou sur des parkings où des personnes peuvent sans risque emprunter ou descendre de véhicules en covoiturage.	Travaux en cours à la FEDUCO (basé sur NeTex/Transmode I)
Description du réseau - Infrastructure exploitée		
Parking, P+R et autopartage	Les parcs de stationnement ne sont pas seulement un sous-ensemble spécifique de la topographie : ils sont aussi les nœuds pour les connexions entre plusieurs modes : par exemple voiture<->transports en commun, marche<->vélo, etc. Ils nécessitent donc un ensemble d'attributs spécifiques : est-il relié à un arrêt de transport en commun, est-il inclus dans une offre de P+R, est-il relié à un service de vélo-partage, etc. Ils disposent aussi d'un ensemble d'information comme : nombre de places, horaires, entrées et sorties, etc. Les parcs de stationnement comprennent également des parcs de vélos.	IFOPT NeTex GDF INSPIRE

Catégorie de données	Description	Normes, textes réglementaires et standards
<i>Station d'autopartage</i>	<i>Les stations d'autopartage sont des endroits spécifiques (habituellement des aires de stationnement ou une partie d'un parking) où un service d'autopartage est disponible. Les informations associées sont par exemple : nombre de places/véhicules, lien vers le service de réservation, tarifs, connexion avec d'autres services (transport en commun), etc.</i>	<i>IFOPT NeTEx</i>
<i>Station de partage de vélos</i>	<i>Les stations de vélos partagés sont des endroits spécifiques (habituellement des aires de stationnement ou une partie d'un parking) où un service vélos partagés est disponible. Les informations associées sont par exemple : nombre de places/vélo, lien vers le service de réservation, tarifs, connexion avec d'autres services (transport en commun), etc.</i>	<i>IFOPT NeTEx</i>
<i>Point d'arrêt</i>	<i>Les arrêts des transports publics sont les endroits où les gens peuvent embarquer et débarquer des véhicules. Ils comprennent l'ensemble des caractéristiques des arrêts (localisation, entrées/sorties, équipements, cheminements, etc.)</i>	<i>TRANSMODEL IFOPT NeTEx</i>
<i>Description des réseaux de transport en commun</i>	<i>La description des réseaux de transport en commun est la partie structurale des offres: les lignes, itinéraires, correspondances, etc.</i>	<i>TRANSMODEL NeTEx</i>
Information planifiée sur l'offre de transport		
<i>Informations horaires de transport en commun</i>	<i>L'information horaire de transport en commun comprend principalement les heures de passage aux arrêts, calendriers d'application, types de jour, etc.</i>	<i>TRANSMODEL NeTEx</i>
<i>Offre tarifaire de transport en commun</i>	<i>L'offre tarifaire de transport en commun décrit est à distinguer de la billettique (titres, contrôle, validation, etc.) et décrit des concepts tels que les produits tarifaires, les droits d'accès qu'ils octroient, les conditions d'utilisation, les prix, etc.</i>	<i>TRANSMODEL NeTEx</i>
<i>Tarifs de stationnement et de péage</i>	<i>Description des tarifs pour les péages et le parking (prix avec informations horaires associées, abonnements, etc.).</i>	<i>NeTEx (en partie)</i>
<i>Tarifs de covoiturage</i>	<i>Tarifcation associée au service de covoiturage.</i>	<i>Pourrait être adapté de NeTEx</i>
<i>Course en taxi</i>	<i>Il s'agit de la description de la structure tarifaire des taxis</i>	<i>Pourrait être adapté de NeTEx</i>
<i>Tarifs de vélos partagés</i>	<i>Tarifcation associée au service de vélos partagés.</i>	<i>Pourrait être adapté de NeTEx</i>
Informations produites en cours d'exploitation		
<i>Voir 0-Précisions sur les données produites en cours d'exploitation</i>		

Précisions sur les données produites en cours d'exploitation

À la demande du Comité, et du sous-groupe juridique, le sous-groupe a affiné la classification des données temps réel de façon à permettre un positionnement différencié pour chaque sous-catégorie et ainsi mieux prendre en compte les spécificités de production et les enjeux concurrentiels liés à ces données.

Données temps réel pour le transport en commun:

Heures de passage estimées et réalisée et informations associées (quai, etc.)	Position des véhicules	Messages diffusés sur le réseau	État des équipements (aux arrêt et dans les véhicules)	Évènements, incidents et leurs conséquences	Taux de remplissage de véhicules et fréquentation
---	------------------------	---------------------------------	--	---	---

Données de contrôle TC:

Ajout de services	Suppression de services	Modification de services (modification de trajet, interruption, modification des horaires, etc.)	Modification de véhicule	Changement de conducteur
-------------------	-------------------------	--	--------------------------	--------------------------

Données temps réel d'usage:

Heures de passage estimées et réalisée et informations associées (quai, etc.)	Position des véhicules	Messages diffusés sur le réseau	État des équipements (aux arrêt et dans les véhicules)	Évènements, incidents et leurs conséquences	Taux de remplissage de véhicules et fréquentation
---	------------------------	---------------------------------	--	---	---

Si ces données sont souvent utilisées en différé (et non en temps réel), elles sont bien créées dans le contexte de l'usage réel des transports.

Données temps réel billettique:

Données de validation à l'entrée (et à la sortie si disponible)	Type de titre utilisé	Mode de paiement (si un paiement est induit)	Information sur les contrôles (nombre de contrôles, verbalisations et causes de verbalisation)
---	-----------------------	--	--

Données temps réel de trafic routier:

Vitesses et flux	Incident et accidents	Messages diffusés (PMV)	Fermetures de voies, limitation de vitesse, etc.
------------------	-----------------------	-------------------------	--

Données temps réel de parking (et auto/vélo-partage):

Nombre de places libres	Nombre de véhicules disponibles	Indicateurs synthétiques (saturé, disponible, ouvert, fermé)
-------------------------	---------------------------------	--

Données temps réel covoiturage:

Listes des offres pour une Origine/ Destination	Liste globale des offres	Liste des demandes de covoiturage	Évolution des offres (disparition/ apparition)
---	--------------------------	-----------------------------------	--

Note : les données de covoiturage longue distance s'apparentent souvent plus à des données planifiées (définies à l'avance) alors que le covoiturage courte distance/urbain s'apparente plus au temps réel. Par souci de simplicité, les deux sont conservés dans la même catégorie.

La donnée

La partie précédente a permis d'identifier quelles étaient les données transports potentiellement concernées par l'ouverture (le choix d'ouverture ou non restant, pour chaque catégorie du ressort du Comité lui-même). Les lignes qui suivent portent des recommandations sur les conditions techniques de mise à disposition de ces données, et visent à répondre aux questions suivantes :

- comment remédier à l'hétérogénéité des caractéristiques techniques des données transports ?
- quels formats les données transports doivent-elles respecter ? standards, normes, format natif ?
- comment garantir un certain niveau de qualité des données ?
- comment assurer la cohérence de plusieurs jeux de données complémentaires ?

L'objectif de ces recommandations est de répondre aux constats trop fréquents :

- les données sont incomplètes ou de mauvaise qualité, ou non mises à jour ;
- les données sont incohérentes entre les territoires et d'un producteur à l'autre (format différent, interprétations différentes d'un même format, incompatibilité des éventuels éléments communs, etc.) ;
- il n'y a, dans la grande majorité des cas, pas de métadonnées disponibles.

Ces recommandations ont de plus été comparées aux « *Best Practice* » proposées par le W3C pour la mise à disposition des données (http://www.w3.org/2013/dwbp/wiki/Best_practices_guidelines). Le sous-groupe adhère globalement à ces « *Best Practice* », même s'il n'a pas émis de recommandation pour chacune d'elles. On peut globalement remarquer que les recommandations ci-dessous s'inscrivent en complément et précision de celles du W3C.

Note : les recommandations font régulièrement apparaître la notion de « producteurs des données » dans la rubrique « Acteurs concernés ». Cette terminologie générique est utilisée à dessein pour indiquer que la donnée peut être produite par une AOT, un de ses prestataires ou délégataires, un transporteur, une SEM, un EPIC, une startup, etc.

R.1- Formats de mise à disposition

Il est recommandé de mettre à disposition, de manière complémentaire, les données des services publics de transport selon les scénarios suivants :

1. Aux formats normalisés (NEPTUNE, NeTEx, SIRI, etc.) en respectant les profils disponibles ET aux formats orientés « réutilisation open-data » (GTFS pour le moment, mais il ne couvre qu'une partie des données transport)

2. A défaut, si cela n'est pas possible : aux formats orientés « réutilisation open-data » (GTFS pour le moment, mais il ne couvre qu'une partie des données transport)

3. A défaut, si cela n'est pas possible : au format natif, accompagné de toute documentation utile sur ce format (l'un des objectifs est alors que des réutilisateurs assurent la conversion vers des formats réutilisables)

<i>Priorité</i>	1
<i>Actions induites</i>	Développement d'un outil de mise au format et/ou de conversion (cf R.11)
<i>Acteurs concernés</i>	Les producteurs des données Les réutilisateurs quand seule l'option 1 est retenue, pour la conversion en format réutilisable ou normalisé
<i>Niveau de consensus</i>	89%

Précision sur la notion de « **format natif** » : format de la donnée, au sein d'un système, autre que standardisé ou normalisé. Même quand aucun format de mise à disposition n'a été précisé, si une donnée existe au sein d'un système elle y est au moins dans un format natif, elle est donc ainsi potentiellement publiable. Si un mécanisme d'échange de donnée est disponible, alors un format de fichier, ou un web service ou encore un API ont été développés (même s'ils sont spécifiques ou propriétaires), et s'il n'y a pas de mécanisme d'échange il y a souvent une base de données qui dispose d'un mécanisme d'export standard de ses tables (format SQL, CSV, DBF, etc.). Ce qu'on appelle format natif est le format existant de ces exports fichier, web service, API ou base de données spécifique au système.

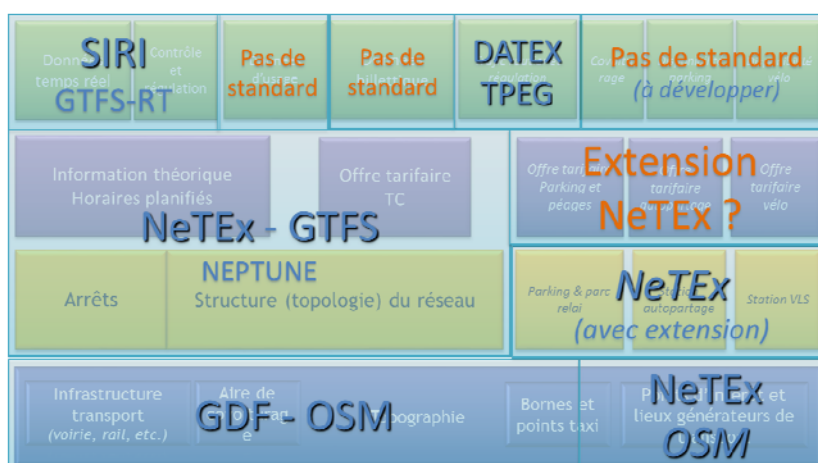
Même si elle évoque certains formats, à titre d'exemple, c'est volontairement que la recommandation ne spécifie pas de format d'échange : en effet les formats évoluent relativement rapidement (on peut citer les dates d'apparition des formats : TRIDENT-2002, GTFS-2005, SIRI-2006, NEPTUNE-2009, NETEX-2014, sachant que ces formats évoluent pour répondre aux attentes et aux marchés), cette recommandation restera en effet valable, quelle que soit l'évolution des formats normalisés.

Le sous-groupe insiste sur la nécessité de prendre en compte les travaux et usages internationaux pour spécifier les formats à utiliser.

La notion de « *formats orientés réutilisation open-data* » qu'évoque la recommandation, fait simplement référence aux formats dont l'utilisation s'est généralisée dans le domaine open-data : dans le domaine transport, seul GTFS est concerné (et pas GTFS-RT).

En l'état actuel des choses, le sous-groupe technique a identifié les formats disponibles et recommandés pour les différentes catégories de données présentées en **Erreur ! Source du renvoi introuvable.** La figure ci-dessous et le tableau qui la suit présentent ces formats. Dans la figure, la typographie a été utilisée pour mettre en valeur les formats les plus recommandés

On peut toutefois noter qu'il a été mentionné à plusieurs reprises, par les réutilisateurs et les producteurs que l'important est d'avoir un format partagé par tous, outillé et capable de véhiculer les informations nécessaires : si ces objectifs sont atteints, l'origine du format est alors secondaire.



Nom	Statut	Référence	Domaine	Commentaire
Transmodel	Norme européenne	EN 12896	Couvre tous les aspects du transport public	Transmodel n'apparaît pas dans la figure ci-dessus, car il ne s'agit pas d'un format d'échange, mais d'un modèle de données. Il est toutefois important de la mentionner, car c'est sur Transmodel que reposent tous les formats d'échange Européen liés au Transport Public (en particulier NeTeX et SIRI). La version 6 de Transmodel est en cours de finalisation et inclut IFOPT.
IFOPT	Norme européenne	EN 28701	Description des arrêts du transport public	Comme Transmodel, IFOPT est modèle de données sur lesquels reposent les formats d'échange. IFOPT signifie : Identification of Fixed Objects in Public Transport. IFOPT fusionne avec Transmodel dans Transmodel v6.
NeTeX	Norme européenne	CEN/TS 16614-1 Description du réseau CEN/TS 16614-2 Information horaire CEN/TS 16614-3 Description de l'offre tarifaire	Transport public : information planifiée (réseau, horaire et offre tarifaire)	Échange des données planifiées du transport public. Basé sur Transmodel 6 (intégrant IFOPT)
SIRI	Norme européenne	EN 15531-1 - Cas d'usage case EN 15531-2 - Communication EN 15531-3 - Services TS 15531-4 – Service de surveillance des équipements TS 15531-5 – Service d'échange de situation	Transport public : information temps réel	Échange des données temps réel du transport public : heures de passage, événements, position des véhicules, etc.
INSPIRE	Directive européenne	Directive 2007/2/EC http://inspire.ec.europa.eu/	Données géographiques nécessaires à la cartographie et métadonnées associées.	INSPIRE couvre un large éventail d'informations. Il contient un ensemble de couches dédiées au transport (route, rail, eau, câble), qui se concentrent principalement sur la description de l'infrastructure et des informations géographiques associées.

Nom	Statut	Référence	Domaine	Commentaire
DATEX	Norme européenne	CEN/TS 16157 partie 1 à 5 (6 en cours)	Données de trafic routier	DATEX est scindé en 5 parties : Partie 1: Contexte et cadre général Partie 2: Localisation Partie 3: Publication de situations Partie 4 : Publication de VMS Partie 5 : Publication de données mesurées et de données calculées Partie 6 : Publication parking La version courante est Datex 2
TPEG	Norme internationale et européenne (ISO-CEN, via TISA)	ISO/TS 18234 parties 1 à 11	Données de trafic routier et de déplacement à destination des usagers	TPEG est scindé en un grand nombre de parties, en particulier : RTM – Messages trafics routiers (Road Traffic Message) PTI – Information de transport public (Public Transport Information) Loc – Localisation utilisé en lien avec des applications (Location referencing, used in conjunction with applications) PKI – Information parking (Parking Information) CTT – Congestion et temps de trajet (Congestion and Travel-Time) TEC – Événement trafic compact (Traffic Event Compact) WEA – Information météorologique pour les voyageurs (Weather information for travellers)
GDF	Norme internationale et européenne (ISO-CEN)	ISO 14825:2011	Description du réseau routier pour le calcul d'itinéraire	GDF (Geographic Data Files) vise à fournir des données de référence pour les systèmes de navigation embarqués ou mobiles, les centres de gestion de trafic ou les services liés aux systèmes de gestion routière, y compris les systèmes de transports publics. La version actuelle: GDF 5
NEPTUNE	Norme française	PR NF P99-506	Offre planifiée de transport public, pour l'information voyageur	Norme française pour l'échange de données d'information des voyageurs dans les transports en commun (réseau, horaire et calendrier) Basé sur Transmodel 4
GTFS	Standard de fait	https://developers.google.com/transit/gtfs/	Offre planifiée de transport public, pour l'information voyageur	Format le plus utilisé en Open data. C'est le format reconnu comme étant « celui de Google », mais initialement défini par la ville de Portland (http://sf.streetsblog.org/2010/01/05/how-google-and-portlands-trimet-set-the-standard-for-open-transit-data)
GTFS-RT	Standard de fait	https://developers.google.com/transit/gtfs-realtime/	Transport public : information temps réel	Complément de GTFS pour le temps réel. Son adoption est toutefois moins large que celle de GTFS. SIRI est probablement plus utilisée.
Open Street Map (OSM)	Standard de fait	http://www.openstreetmap.org/about	Données géographiques nécessaires à la cartographie et métadonnées associées.	OpenStreetMap (OSM) est un projet collaboratif pour créer une carte gratuite et modifiable du monde. OSM couvre un large éventail d'objets, y compris dans le domaine du transport en commun (http://wiki.openstreetmap.org/wiki/Public_transport), et du réseau routier (http://wiki.openstreetmap.org/wiki/Highways)

R.2- Traiter les formats non disponibles

Certaines catégories de données, pour lesquelles l'ouverture pourrait être envisagée, ne disposent pas de format d'échange standardisé ou normalisé.

Il est recommandé de rapidement développer des normes d'échange pour cette catégorie.

En corollaire, compte tenu du mandat « Standard request on Urban ITS » de la commission vers le CEN, il convient d'impliquer les groupes de normalisation Français et de soutenir les travaux qui vont être engagés au CEN sur ces domaines.

Ces travaux devront naturellement être menés en cohérence avec les normes déjà élaborées (NeTEx et SIRI en particulier).

Les principales catégories aujourd'hui identifiées sont :

- **Le covoiturage**
- **L'autopartage**
- **L'information temps réel sur les parkings (tous modes : vélo et 2 roues, voitures ; en partage ou non)**
- **L'information dédiée au vélo**

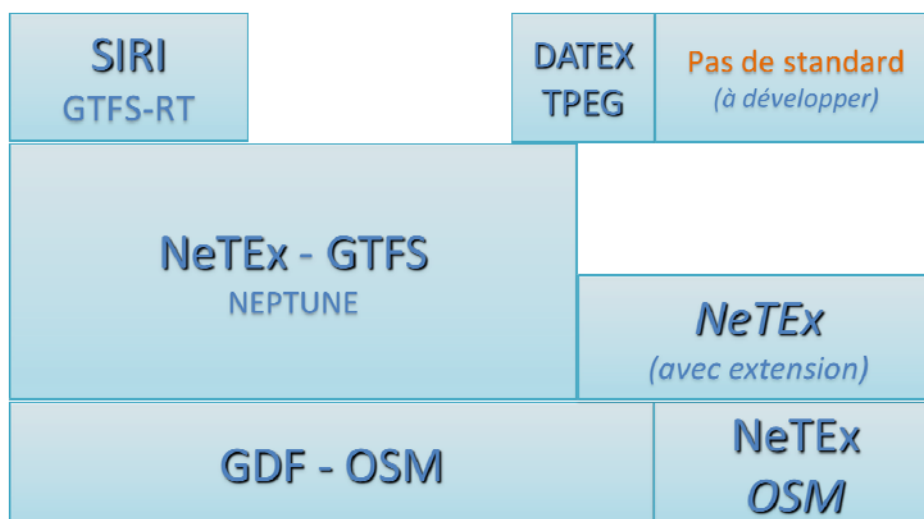
Priorité	3
Actions induites	Engager des actions de normalisation et d'échange avec les acteurs du domaine
Acteurs concernés	AFIMB et AFNOR pour l'élaboration des normes Acteurs des domaines concernés, pour l'expression des besoins et l'expertise fonctionnelle et technique
Niveau de consensus	100%

Si cette recommandation est simple et bien partagé, le sous-groupe souhaite toutefois alerter sur le fait que certains des domaines concernés (le covoiturage en particulier) relèvent en bonne partie de startup et le « temps normatif », traditionnellement assez long, est souvent difficilement compatible avec le temps très court des cycles d'évolution des startups (avec en conséquence un risque d'aboutir à une norme ne prenant pas en compte les dernières évolutions du domaine). Si l'implémentation ne fait pas partie du cycle normatif, il semble toutefois particulièrement important de l'intégrer dans ce contexte, de façon à raccourcir les cycles de production.

Il est aussi noté que l'absence de norme ou standard ne signifie pas que des travaux n'ont pas déjà été anticipés :

- Covoiturage : échange d'offre avec RDEX (<http://www.feduco.org/articles/actus/rdex>)
- Covoiturage : les travaux en cours de la FEDUCO sur les aires de covoiturage
- Autopartage et vélopartage : travaux Allemand IXSI (Interface for X-Sharing Information: <http://ixsi-schnittstelle.de>)

La figure ci-dessous reprend la juste les normes et standards retenue sur les catégories qui semblent les plus pertinentes pour l'ouverture (en plus gros caractères celle à favoriser dans un contexte d'ouverture des données) et fait apparaître plus explicitement les domaines sur lesquels des travaux restent nécessaires.



R.3- Définir des profils d'échanges

Les formats sont souvent vastes et sont tous soumis à interprétation et il est important d'en harmoniser l'usage: pour tout format recommandé ou imposé, il est nécessaire de définir un profil d'utilisation (et ceci est aussi vrai pour GTFS).

Les profils doivent être harmonisés à un niveau national.

Ils peuvent être élaborés en contexte normatif.

Priorité	2
Actions induites	Production et diffusion nationale des profils
Acteurs concernés	AFIMB, AFNOR pour l'élaboration des profils (dans l'hypothèse où l'on souhaite en faire des normes françaises).
Niveau de consensus	100%

Les normes, dans leur définition même, sont des « documents établis par **consensus** ». Elles sont de plus établies à un niveau européen, en prenant donc en compte des exigences qui dépassent souvent le périmètre national. Elles ont vocation à avoir une durée de vie (et une stabilité) la plus longue possible (ce qui n'empêche pas leur mise à jour quand nécessaire) et doivent, à ce titre, être définies en prenant une posture prospective, de façon à bien prendre en compte l'ensemble des besoins sur une fenêtre temporelle d'une dizaine d'années (pour les normes ici évoquées).

Il en résulte des normes qui sont relativement volumineuses et dont le périmètre dépasse souvent largement les besoins d'une utilisation donnée. Ainsi, à titre d'exemple, SIRI propose toute une série d'options ou de mécanismes dont la vocation est d'assurer la compatibilité avec les systèmes développés en Allemagne dans le contexte des VDV453/454. De même, SIRI propose des services dédiés à la gestion des correspondances garanties, services qui, s'ils sont dès aujourd'hui pertinents en Suisse ou en Allemagne, sont pratiquement inexistantes en France.

De plus, un certain nombre de spécificités locales ou nationales peuvent amener à préciser l'usage ou la codification qui sera utilisée pour certaines informations. Par exemple, les Anglais disposant d'un référentiel national d'identification des points d'arrêts (NaPTAN), ils imposeront naturellement que cette codification soit utilisée dans les échanges SIRI, ce que ne feront pas les autres pays européens.

Enfin, certains éléments proposés par les normes sont facultatifs et il convient, lors d'une implémentation, de décider si ces éléments seront ou non implémentés.

L'utilisation des normes liées à l'implémentation de l'interopérabilité pour le transport en commun passe donc systématiquement par la définition d'un profil (*local agreement*, en anglais). Concrètement, le profil est un document complémentaire à la norme et qui en précise les règles de mise en œuvre dans un contexte donné. Le profil contient donc des informations comme :

- détail des services utilisés ;
- détails des objets utilisés dans un échange ;
- précisions sur les options proposées par la norme ;
- précision sur les éléments facultatifs ;
- précision sur les codifications à utiliser ;
- etc.

Il est aussi recommandé qu'un profil soit accompagné de la définition des tests permettant d'établir si une donnée (ou une API) est conforme.

L'absence de profil ne doit toutefois être un frein à l'ouverture des données (voir recommandation R.1).

Les principaux profils que l'on peut actuellement évoquer en France sont (mais cela pourrait être étendu à tous les formats des recommandations précédentes) :

Profil	Avancement
NEPTUNE	Disponible
Profil SIRI (défini par le STIF et remis au GT7)	Disponible pour la partie IDF, à finaliser pour la généralisation France
Profil GTFS	En cours d'élaboration (premier travail réalisé dans le cadre de CHOUETTE)
Profil NeTEx pour les arrêts	En cours de finalisation (première version proposée)
Profil NeTEx pour les réseaux	En cours d'élaboration
Profil NeTEx pour les horaires	En cours d'élaboration
Profil NeTEx pour les tarifs	À faire

Il est recommandé, à chaque fois que cela sera possible, de travailler sur les profils en coordination avec nos voisins Européens.

R.4- Harmonisation des concepts utilisés

Les concepts manipulés par les différents fournisseurs et réutilisateurs doivent être harmonisés et banalisés autant que possible.

A minima ils doivent être documentés.

Si cette recommandation a d'évidentes conséquences techniques, elle dépasse aussi ce cadre technique et touche au champ global de la communication: de très nombreuses incompréhensions tiennent au simple fait que les acteurs n'utilisent pas les mêmes termes, ou cachent des définitions très différentes derrière des termes identiques.

Le domaine du transport public a le gros avantage de disposer d'un excellent référentiel de terminologie en la norme Transmodel.

Priorité	2
Actions induites	Communication adaptée à destination des différents acteurs (réutilisateurs, AOT, transporteurs, industriels et bureau d'étude) en s'appuyant sur Transmodel. Élaboration (et diffusion) des termes et définition pour les nouveaux modes de Transport et les nouveaux usages.
Acteurs concernés	AFIMB, bureaux d'étude et experts AFNOR (et CEN) pour la cohérence des normes
Niveau de consensus	100%

Cette recommandation sur le besoin d'harmoniser, vulgariser et communiquer la terminologie est très largement partagée et ne nécessite que peu d'informations complémentaires.

Les normes à utiliser comme base pour cette recommandation sont rappelées ci-dessous. Il est de plus rappelé qu'il est indispensable de faire vivre et d'enrichir ce dictionnaire de données au fur et à mesure de l'évolution des pratiques du transport public.

Nom	Statut	Référence	Domaine	Commentaire
Transmodel	Norme Européenne	EN 12896	Couvre tous les aspects du transport public	Transmodel est un modèle de données. C'est sur Transmodel que reposent tous les formats d'échange Européen liés au Transport Public (en particulier NeTEx et SIRI). La version 6 de Transmodel est en cours de finalisation et inclu IFOPT.
IFOPT	Norme Européenne	EN 28701	Description des arrêts du transport public	Comme Transmodel, IFOPT est modèle de données sur lesquels reposent les formats d'échange. IFOPT signifie : Identification of Fixed Objects in Public Transport IFOPT fusionne avec Transmodel dans Transmodel v6.

On peut aussi noter que la communication devra être adaptée aux différentes cibles : il est inutile et même contre-productif de communiquer de la même façon vers les réutilisateurs, les transporteurs et les AOT !

*Note: Cet aspect de communication pourra être intégré à la recommandation **R-21** si elle est retenue.*

R.5- Fournir des indicateurs de qualité des données ouvertes

Il est recommandé de tester la qualité des jeux de données avant leur diffusion et d'accompagner les données ouvertes des résultats des tests de qualité qu'elles ont passés (contrôles syntaxiques de respect des formats ainsi que des contrôles sémantiques).

Il est de plus fondamental que les tests de qualité réalisés soient connus et harmonisés au niveau national.

Priorité	2
Actions induites	Définir et documenter les tests de qualité pour tous les formats et profils retenus (et outiller ces tests dans un second temps).
Acteurs concernés	AFIMB et AFNOR pour la définition Producteur de données pour la mise en œuvre
Niveau de consensus	89%

La qualité des données est une thématique récurrente de l'interopérabilité : une donnée ne respectant pas les formats imposés, une donnée incohérente ou trop incomplète est en effet souvent une donnée inutilisable.

Les tests de qualité vérifient les règles d'écriture et d'organisation des données d'un jeu de données, notamment :

- règles syntaxiques: il s'agit de vérifier les principalement règles de forme (présentation) *du format retenu* ainsi que la présence dans le jeu de données de certaines données obligatoires imposées par la spécification (à comparer à la vérification de l'existence d'un verbe en présence d'un sujet dans une proposition en langage naturel) ;
- les règles relatives à la structure et la cohérence *des données*, imposées par la spécification (profil et norme): il s'agit (principalement) de vérifier les liens entre certains éléments (par exemple le fait qu'un arrêt référencé sur un parcours soit bien défini quelque part dans le jeu de données) ;
- les règles relatives aux valeurs des données et à leur signification permettent de vérifier que les valeurs ont un sens dans un contexte donné (par exemple vérification de la valeur des coordonnées dans un contexte géographique: les arrêts d'une ligne francilienne doivent être localisés en Ile-de-France et non sur l'équateur !).

R.6- S'assurer de la mise à jour des données à jour

Il est recommandé de prêter une attention particulière à la mise à jour des données : une donnée dépassée n'est en effet souvent plus d'aucune utilité et risque d'induire les utilisateurs en erreur.

Une organisation est à mettre en place pour que les données rendues réutilisables soient à jour : une démarche d'ouverture des données devrait inclure un volet « modalités de mise à jour ».

Priorité	2
Actions induites	Mettre en place un processus de mise à jour des données. Communiquer la période de validité des données.
Acteurs concernés	Producteurs de données Consommateurs de données pour garantir l'utilisation du jeu de donnée le plus à jour
Niveau de consensus	67%

La notion de mise à jour des données peut sembler être une évidence, mais est trop souvent négligée voir oubliée. Il est indispensable qu'un réutilisateur dispose de données valide et non dépassée et puisse savoir quand le jeu de donnée qu'il utilise doit être mis à jour.

On peut aussi lier cette recommandation à la bonne pratique 17 du W3C: *BP17. Versioning information should be available*. C'est en effet lors des mises à jour que les versions des données sont modifiées. Si la gestion des versions des données ne peut être imposée aux systèmes existants, elle doit toutefois être recommandée pour tous les nouveaux systèmes. On pourra distinguer différents niveaux de version :

- la version globale du jeu de données (à rapprocher du concept de *FRAME* de Transmodel) ;
- la version de chacun des objets contenus dans le jeu de données.

Dans un certain nombre de cas (le temps réel en particulier), il sera aussi important d'être en mesure de préciser si l'absence de mise à jour est due au fait que l'information n'évolue pas, ou bien à une défaillance dans la chaîne de production de la donnée.

Cette recommandation est en lien avec la recommandation R.7, les métadonnées pouvant préciser la période de validité des données ou encore le délai entre les mises à jour.

R.7- Fournir les métadonnées

Publier de manière harmonisée les métadonnées décrivant les données ou l'API (on pourra notamment s'inspirer de la structuration de métadonnées INSPIRE) décrivant notamment :

- **modes de transport pris en compte, et non prise en compte**
- **nature des données (infrastructure, arrêts, réseau, horaires, temps réel, etc.)**
- **couverture géographique,**
- **période de validité temporelle,**
- **fréquence de mise à jour,**
- **formats de données ou type d'API,**
- **précision des informations (géographique notamment)**
- **producteur(s),**
- **les différents acteurs impliqués et leurs rôles (propriétaire, producteur, source, exploitant...)**
- **etc.**

Il conviendra de préciser si toutes les informations sont obligatoires ou si certaines peuvent être facultatives (cela peut dépendre des producteurs et de leur contexte).

Priorité	1
Actions induites	Définition de métadonnées transport harmonisées.
Acteurs concernés	AFIMB et AFNOR pour la définition des métadonnées Producteurs de données pour la production des métadonnées
Niveau de consensus	100%

Une métadonnée est une donnée servant à définir ou décrire une autre donnée. Elle permet de savoir ce que contient un jeu de données et d'avoir une idée de ce que l'on peut en faire sans avoir à lire les données elles même. C'est, d'une certaine façon, la « quatrième de couverture » d'un jeu de données.

Le «Dublin Core» (http://fr.wikipedia.org/wiki/Dublin_Core), soutenu par la Commission Européenne, est la principale initiative visant à la convergence des éléments de métadonnées générique à utiliser.

La directive INSPIRE a établi ses métadonnées⁵⁷ en respectant le «Dublin Core», et ces travaux pourraient être utilisés comme base pour définir les métadonnées transport.

De plus NeTEx offre la possibilité de véhiculer la plupart des métadonnées envisageable pour les données transport planifiées et pourra être utilisé à cet effet.

Il est recommandé d'envisager la définition de métadonnées transport à un niveau International plutôt que national (à minima initier un travail national pour très vite le proposer à un niveau Européen).

Le sous-groupe considère que les métadonnées doivent rester relativement simples dans leur structure et leur technicité.

⁵⁷<http://cnig.gouv.fr/wp-content/uploads/2014/01/Guide-de-saisie-des-éléments-de-métadonnées-INSPIRE-v1.1-final-light.pdf>

R.8- Proposer des données et des API

Si les données sont indispensables, il est aussi très utile de les compléter par des services (API ou Web Service) qui permettent au réutilisateur d'éviter d'avoir à apprendre et comprendre le domaine transport, qui est complexe et peut ne le concerner que de façon très marginale (par exemple : intégration d'information transport sur un site de services immobiliers...).

On pourra donc lui proposer des services de recherche d'arrêt, de calcul d'itinéraire, d'affichage de résultat sur une carte, etc.

Les API de service aujourd'hui disponibles sont généralement spécifiques et un travail d'harmonisation s'impose (utiliser les normes quand elles existent, et définir des standards dans les autres cas).

Priorité	2
Actions induites	Définir des API standardisées quand il n'en existe pas. Mettre en place les API en complémentarité des données
Acteurs concernés	AFIMB et AFNOR pour la définition des API manquantes Producteurs de données pour la mise en place
Niveau de consensus	100%

Note : il faut noter que l'API peut être un simple choix technique de mise à disposition des données. Ce qui est traité ici est l'accès à un service élaboré (recherche d'itinéraire, recherche de l'offre à proximité à une heure donnée, etc.) et non la technique de mise à disposition des données (voir R.1 pour cet aspect-là).

Il faut noter que la mise en place d'une API peut, dans certain contexte et pour certains acteurs, s'avérer d'une complexité, et donc d'un coût, supérieur à la simple mise à disposition des données (impact sur le développement et l'exploitation). À ce titre la mise à disposition d'une API peut être recommandée, mais il semble difficile de l'imposer.

Les normes récentes comme NeTeX et SIRI intègrent maintenant systématiquement une possibilité d'accès via un Web Service SOAP. Plusieurs choix d'implémentation sont disponibles pour ces deux normes, et c'est la version la plus simple qu'il est recommandé de retenir pour l'open data (implémentation dite SIRI Lite, basée sur une interface technique REST/JSON : http://fr.wikipedia.org/wiki/JavaScript_Object_Notation).

Mais SIRI et NeTeX ne proposent des services que pour l'accès à la donnée elle-même et non des services élaborés comme la recherche d'itinéraire qui fait l'objet d'une attente particulière. S'il n'existe pas de norme, ni de standard de fait, pour les API de recherche d'itinéraire (et globalement tous les services élaborés pour le transport), on peut noter des travaux menés par l'Allemagne dans la cadre du projet IP-KOM (travaux menés par le VDV : <http://www.ip-kom.net>) dont l'API TRIAS connaît un certain succès actuellement et dont s'inspirent certains travaux en cours au CEN (TC278/WG3/SG8 pour le calcul d'itinéraire distribué) ou encore sont étudiés par des projets Européens (comme OPTICITIES <http://www.opticities.com>). TRIAS respecte les choix établis par la norme SIRI et propose un certain nombre de services présentés dans le tableau ci-dessous et il semble opportun de l'utiliser comme base pour la standardisation des échanges :

Nom du service	Description
BookingInfo	Information sur les réservations
ConnectionDemand	Information sur les correspondances
FacilityRequest	Information sur les services et les équipements
FaresRequest	Information sur les tarifs
GeoCoordinatesRequest	Information sur bas de critère géographique (autour d'un point, etc.)
ImageCoordinatesRequest	Demande d'image et pictogramme (cartographie)
IndividualRouteRequest	Recherche d'itinéraires (TC uniquement)
LocationInformationRequest	Information sur un lieu
MapServiceRequest	Service de cartographie
PersonalisationRequest	Personnalisation (liste de paramètres et choix personnels: vitesse de marche, modes préférés, etc.)
PositioningRequest	Localisation sur le réseau de transport public
ServiceRegisterRequest	Service de notification (alerte, modification d'itinéraire, etc.)
StopEventRequest	Heures de passage à l'arrêt
StopRequestRequest	Demande d'arrêt au prochain arrêt possible du véhicule
TripInfoRequest	Information sur les courses
TripRequest	Recherche d'itinéraires (multimodale)
VehicleDataRequest	Information sur les véhicules
Authentication	Authentification

R.9- Données théoriques et temps réel

Il est recommandé d'envisager la possibilité d'une mise à disposition progressive (par étapes) en commençant par les données planifiées puis, dans un second temps, des données temps réel.

En effet, du point de vue des réutilisateurs, le besoin est aussi fort pour les données planifiées (théorique) que pour les données temps réel (la demande est même souvent plus forte sur le temps réel), il semblerait donc cohérent de ne pas faire de distinctions planifié-temps réel pour l'ouverture des données.

Toutefois suivant les contextes de productions et l'historique des systèmes existant (tels qu'ils ont été exposés lors des différentes auditions), la mise à disposition de l'information temps réel s'avère pour certains acteurs plus délicate que la mise à disposition des données planifiées.

Priorité	2
Actions induites	Définir des paliers de mise à disposition des différentes catégories de données.
Acteurs concernés	Producteur de données.
Niveau de consensus	56%

La mise à disposition des données temps réel passe obligatoirement par la mise en place d'une API et d'une plateforme de diffusion. Il faut noter que la mise en place de cette structure (voir recommandation précédente) peut parfois s'avérer d'une complexité, et donc d'un coût, supérieur à la simple mise à disposition des données (impact sur le développement et l'exploitation).

Si l'opération peut être complexe pour un système existant elle est beaucoup plus accessible dans le cadre d'un renouvellement (et pourrait alors être exigée par l'AOT ou le transporteur dans leurs cahier des charges).

Il faut noter que la donnée temps réel est une donnée complémentaire à la donnée planifiée : sans un minimum d'information sur les données planifiées il n'est pas possible d'obtenir une information temps réel (il est impossible de demander les horaires de passage à un arrêt si on ne connaît pas l'arrêt !).

Note : la formulation « envisager la possibilité de » utilisée par la recommandation reflète la très grande disparité d'avis des acteurs (disposant d'un retour d'expérience) quant au niveau de difficulté de la mise à disposition de l'information temps réel (de « très simple » à « très compliqué »).

Les outils

Le besoin d'outils est régulièrement ressenti comme aussi, voire plus important que celui du choix du format.

Beaucoup des recommandations concernant les outils sont naturellement une prolongation directe des recommandations portant sur les données.

Les normes et standards considérés étant majoritairement internationaux, une collaboration, à minima européenne, pourra être envisagée sur ces outils.

R.10- Outils pour vérifier la qualité des données

Il est recommandé de mettre en place des outils de vérification de la qualité des données (syntaxique et sémantiques) pour les formats et profils recommandés ou exigés.

Note : cette recommandation est naturellement un prolongement de la recommandation R5

Priorité	2
Actions induites	Développement des outils de qualification/vérification.
Acteurs concernés	AFIMB Communauté open source
Niveau de consensus	100%

La nécessité de tels outils tient à plusieurs points :

- si l'on souhaite un bon niveau d'interopérabilité, il n'est pas efficace qu'un développeur valide ses interfaces avec ses propres outils : un outil externe et neutre apporte une bien meilleure garantie ;
- les AOT ont besoin d'outil pour réceptionner les systèmes qu'elles commandent et vérifier les caractéristiques d'interopérabilité qu'elles ont exigées, et le développement d'outils de ce type n'est pas à la portée de toutes les AOTs ;
- le développement d'un tel outil implique un niveau de technicité assez important et un investissement en conséquence. De plus un tel outil ayant vocation à être utilisé par plusieurs acteurs, un développement mutualisé prend ici tout son sens ;
- le partage d'un même outil de validation améliore l'interopérabilité par le fait qu'il favorise une même interprétation de la norme ou du standard sur les points sur lesquels une certaine latitude d'interprétation est permise.

On peut noter que l'outil CHOUETTE (www.chouette.mobi) a été développé dans ce sens et permet une très bonne validation du format NEPTUNE.

Cette recommandation s'applique à toutes les catégories de données.

R.11- Outils de conversion et production aux formats

Il est recommandé de mettre à disposition des outils permettant de simplifier la mise en place de la chaîne de production (outils de conversion d'un format connu à un autre, etc.).

Priorité	2
Actions induites	Développement des outils
Acteurs concernés	AFIMB Communauté open source Les producteurs de données (qui pourront intégrer ces outils dans leurs systèmes)
Niveau de consensus	100%

La nécessité de tels outils tient à plusieurs points :

- des outils (open source) permettant la mise au format, ou la lecture d'un format données sont propres à faciliter l'adoption du dit format, car ils simplifient, accélèrent et rendent moins onéreux la mise en conformité des systèmes existants et des nouveaux systèmes ;
- les AOT peuvent accompagner leurs exigences vis-à-vis d'un format d'outil en facilitant l'usage, limitant ainsi l'impact sur les coûts et délais, et rendant l'accès aux marchés publics plus concurrentiel (les outils sont alors, généralement, disponibles mais pas d'usage obligatoire) ;
- le développement d'un tel outil peut impliquer un niveau de technicité assez important (variable suivant le format) et un investissement en conséquence. Un développement mutualisé prend donc tout son sens.

On peut noter que les outils CHOUETTE et IRYS (www.chouette.mobi) ont été développés dans cet objectif (mais ne couvrent pas encore tous les formats).

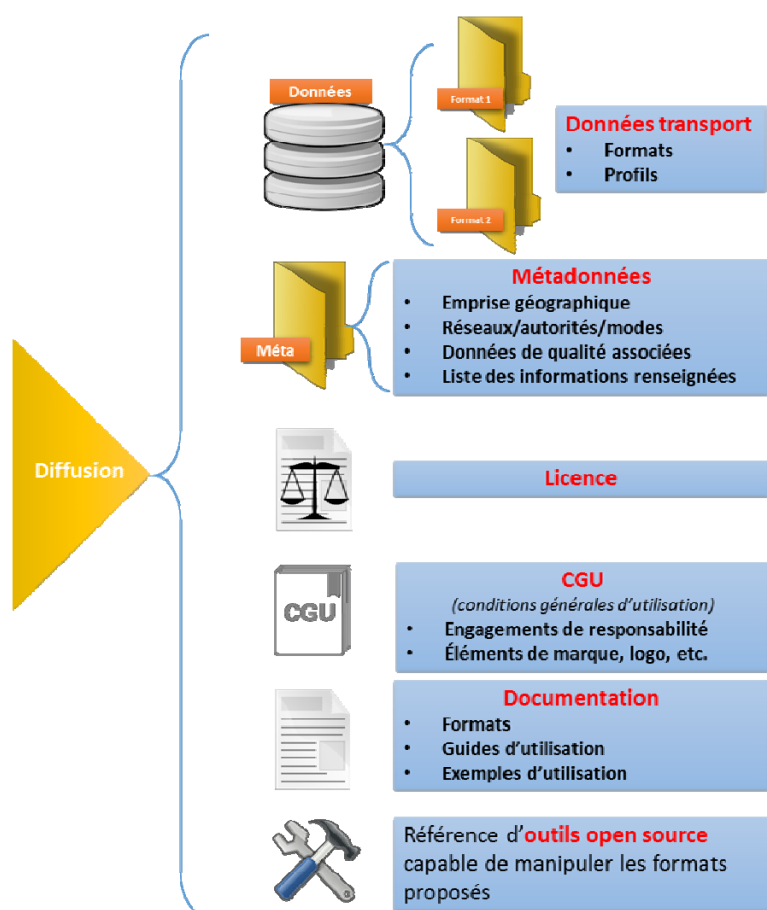
On pourra aussi envisager la mise à disposition d'outils supportant les formats recommandés et facilitant la mise en place de services basés sur ces données comme le calcul d'itinéraire (comme le font déjà OTP <http://www.opentripplanner.org>, SYNTHÈSE <https://extranet.rcsmobility.com/projects/synthese> ou navitia.io <http://www.navitia.io>), la production de plan de réseau, de fiches horaires, l'analyse de l'offre, etc.

R.12- Outils de publication

Il est recommandé de mettre à disposition des outils harmonisés de publication des données (et métadonnées, CGU, licence, etc.), de façon à garantir que la donnée et tous les éléments d'accompagnement nécessaire sont bien systématiquement disponibles.

Priorité	2
Actions induites	Définition d'un « référentiel » de publication de données transport et développement de l'outil correspondant.
Acteurs concernés	AFIMB Communauté open source AOT pour la mise en œuvre
Niveau de consensus	100%

Ainsi que cela a été évoqué à plusieurs reprises, la donnée seule n'est généralement pas suffisante (et métadonnées, CGU, licence, etc.). La mise à disposition d'outils de publication prenant en charge l'ensemble des éléments présentés par la figure ci-dessous serait de nature à garantir la mise à disposition de l'ensemble des informations requises.



R.13- Les outils doivent être open source

L'objectif des recommandations sur les outils est la mutualisation de l'effort de développement : il est donc indispensable que ces outils soient réalisés dans un contexte open source.

Priorité	2
Actions induites	Choix de cadre de réalisation des outils. Animation d'une communauté open source
Acteurs concernés	AFIMB, Adullact (https://www.adullact.org) Communauté open source
Niveau de consensus	100%

- des outils utilisables et intégrables par les industriels du domaine ;
- des outils utilisables et intégrables par les réutilisateurs ;
- la possibilité de mutualiser l'effort de développement et de maintenance ;
- une possibilité d'implication d'expertise venant des différents domaines transports ;
- la possibilité d'une meilleure réactivité aux évolutions du domaine ;
- la possibilité de s'ouvrir vers l'international.

On peut noter que les outils CHOUETTE (pour NEPTUNE et GTFS, et NeTex à terme) et IRYS (pour SIRI) sont déjà open-source.

R.14- Mettre en place un référencement des outils

Il est recommandé de bien référencer et suivre les outils existant pour que les producteurs et réutilisateurs les trouvent facilement.

Priorité	2
Actions induites	Mise en place d'un service de référencement. Action de communication.
Acteurs concernés	AFIMB, Adullact (https://www.adullact.org)
Niveau de consensus	100%

Pour que les outils soient pleinement efficaces, il est nécessaire que tous les utilisateurs potentiels en aient connaissance : cela passe par des actions de communication et un bon référencement des outils.

R.15- Mise en place de données de référence (référentiels)

De façon à faciliter l'utilisation conjointe de plusieurs jeux de données (plusieurs sources sur un même territoire, territoires en chevauchement, etc.), il est recommandé de mettre en place des bases de données partagées par tous, pour des données de référence, en commençant par les arrêts.

Priorité	2
Actions induites	Mise en place d'un référentiel d'arrêt à une échelle nationale : développement de l'outil, mise en œuvre et exploitation
Acteurs concernés	AFIMB (pour l'organisation et le pilotage) Producteurs de données (alimenteront ce référentiel et l'utiliseront pour les mises à disposition de données: les données ouvertes s'appuieront sur les données de référence).
Niveau de consensus	100%

Il est important d'encourager la mise en place de bases de données partagées par tous les acteurs d'un territoire (régional ou national) pour certaines données « clé » comme les arrêts (comme l'ont fait les Anglais avec NaPTAN, ou comme sont en train de le faire les Allemands). Ces bases de données partagées sont parfois appelées « référentiels des données partagés ».

Les données ainsi mutualisées facilitent la réutilisation par divers acteurs (applications, organismes), et par conséquent l'interopérabilité, en particulier dans les situations où il est nécessaire de fusionner/rapprocher des données issues de plusieurs fournisseurs sur un même territoire.

Cette recommandation est à mettre en lien avec les « best practice » du W3C : « BP12. Each data resource should be associated with a unique identifier ». L'un des intérêts d'un tel référentiel est en effet de disposer, pour chaque objet, d'un identifiant unique et pérenne (et partagé).

On pourra dans ce contexte s'appuyer sur les premiers travaux engagés par le CEREMA et surtout sur le Modèle d'Arrêt Partagé défini par le groupe de travail Qualité de Données de l'AFIMB déjà utilisé pour le profil Netex français pour les arrêts de TC.

R.16- Mise en cohérence et fusion des données d'un territoire

Les données d'un territoire sont souvent issues de sources multiples (plusieurs exploitants par exemple). Il n'est pas toujours simple de mettre en cohérence et fusionner de telles données, et quand ce travail est fait, il est utile qu'il soit mis à la disposition des réutilisateurs.

Priorité	2
Actions induites	Mettre en place une organisation (et des outils) de fusion des données territoriales.
Acteurs concernés	AOT (mise en place de bases de données territoriales fusionnées) AFIMB (pour d'éventuels outils mutualisés)
Niveau de consensus	100%

Une des problématiques lors de l'utilisation des données multisources est la cohérence d'informations fournies par différents fournisseurs (Nom d'arrêt, de ligne, mais aussi cohérence des correspondances ...). Le travail de mise en cohérence et fusion des données peut s'avérer relativement complexe, et la mutualisation de ce travail sera un gain pour l'ensemble de l'écosystème.

Il faut noter que si la recommandation R.15 est retenue, cette recommandation R.16 perd un peu de son importance car l'intégration et la fusion des données en seront grandement facilités.

Conditions de mise à disposition des données

R.17- Éléments de contrôles et CGU harmonisés

L'ouverture des données est aussi un moyen de les contrôler : quand les données ne sont pas ouvertes, elles sont souvent produites par une communauté d'utilisateurs qui en a besoin, et ce sans aucun contrôle. Cette problématique de contrôle vaut aussi pour toute donnée ouverte sans élément d'accompagnement.

Il est donc recommandé que les données ouvertes soient accompagnée de mécanismes de contrôle (licence, CGU: conditions générales d'utilisation, etc.) de façon à en garantir un usage conforme aux attentes des propriétaires et producteurs.

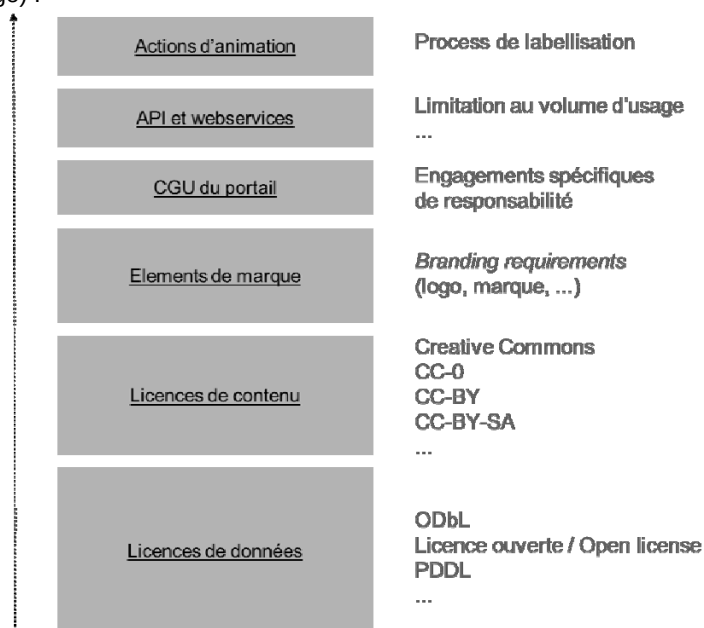
Il est de plus recommandé d'harmoniser ces éléments de contrôle (plus simple à mettre en place pour le producteur et garantie de reproductibilité des actions pour le réutilisateur).

Priorité	2
Actions induites	Définir les CGU harmonisées.
Acteurs concernés	AFIMB et producteurs (GART, etc.) pour la définition des CGU harmonisées. Producteur pour leur mise en œuvre.
Niveau de consensus	100%

Note : cette recommandation n'est pas technique mais a des inductions techniques.

Une bonne communication et une harmonisation des éléments de contrôle (licence, CGU, etc.) est un point aussi important pour le réutilisateur que pour le producteur des données.

La figure ci-dessous (source Simon Chignard) présente les principaux mécanismes de contrôle dont la plupart devraient être harmonisés (on peut toutefois laisser une certaine liberté locale, par exemple sur la labellisation ou encore les éléments de marque et de logo) :



R.18- Information sur les réutilisations

Il est recommandé de mettre en place les éléments nécessaires à une communication sur les réutilisations observées.

Cela est aussi important pour les producteurs qui pourront ainsi savoir comment leurs données sont utilisées que pour les réutilisateurs qui disposeront d'un outil de communication, pourront établir des collaborations ou encore s'inspirer de travaux réalisés.

Priorité	3
Actions induites	Mettre en place un environnement permettant de décrire les réutilisations.
Acteurs concernés	Les réutilisateurs pour la communication sur leurs réalisations, Les producteurs pour la mise en place du système de communications sur les réalisations ETALAB
Niveau de consensus	100%

L'environnement permettant de décrire les réutilisations pourra être réalisé dans le cadre d'un outil mutualisé.

Cette action nécessite que les réutilisateurs s'enregistrent auprès de l'outil réalisé à cet effet et décrivent leur réalisation.

Le producteur pourra, s'il le souhaite, utiliser cette information pour « labelliser » certaines applications et mettre en valeur celles qui sont les plus en adéquation avec la politique du territoire.

R.19- Agir sur les cahiers des charges

Note : si cette recommandation peut sembler non technique au premier abord, elle a été conservée, car les cahiers de charges contiennent de nombreuses exigences techniques.

Il est recommandé que les donneurs d'ordre puissent disposer de modèles de clauses réutilisables et insérables dans les marchés publics (SIM, DSP, etc.) en vue d'intégrer la réutilisation, et toutes les exigences techniques induites, comme objectif fonctionnel.

Ces clauses (en plus de leurs aspects juridiques hors des compétences de ce sous-groupe) incluront en particulier :

- **Les formats attendus (voir R.1)**
 - **Les moyens de test et de qualification envisagés**
 - **Les outils recommandés**
 - **Les attentes en termes de qualité des données**
- Les clauses devront être partagées à un niveau national.**

Priorité	2
Actions induites	Produire les clauses réutilisables, et les diffuser auprès des donneurs d'ordre.
Acteurs concernés	AFIMB, GART, AOT, Open Data France
Niveau de consensus	100%

Il faut noter qu'une recommandation juridique est attendue en complément de celle-ci, notamment pour traiter de la propriété des données (et de tous les éléments propres aux marchés publics, aux possibilités d'allotir le marché pour distinguer données et service, etc.).

On pourra distinguer trois cas pour ces clauses réutilisables :

1. contrat de transport (entre 2 acteurs, AO et exploitant) ;
2. partenariat (convention SIM, etc.) ;
3. fourniture de système (marché public ou PPP).

Ces trois cas sont aujourd'hui clairement des obstacles à la réutilisation qui se cumulent (les données « passant » d'un contrat à l'autre) et qu'il faut lever.

Il faut souligner que même si l'on n'ouvre pas les données, les besoins d'échange et de partage de données existent depuis des (dizaines d'années et subissent des difficultés récurrentes, bien connues de tous ceux qui travaillent dans le domaine: contrats de transport, convention d'échanges entre collectivités / partenaires, systèmes techniques, etc.

On peut, dans ce cadre, noter des travaux en cours sur le sujet au niveau de l'AFIMB ou encore d'Open Data France. Il conviendra donc de travailler avec ces organismes.

R.20- Sites et lieux d'échange et partage

Il est recommandé de proposer aux réutilisateurs des systèmes et lieux d'échange leur permettant d'échanger sur leurs retours d'expériences et de s'entre-aider dans leurs réalisations.

Priorité	2
Actions induites	Mise en place des systèmes et lieux d'échange (forum, wiki, Stack Exchange, etc.)
Acteurs concernés	ETALAB, AFIMB (pour éviter l'inutile multiplication des sites) Communauté des développeurs
Niveau de consensus	100%

Le principe est d'établir un forum de discussion (en complément de contacts courriel du ou des responsables de la plateforme) permettant aux utilisateurs de confronter leurs expériences ou difficultés, d'exprimer le besoin d'enrichissement, de proposer de publication de nouveaux types de données, etc.

Un tel espace de discussion peut être mutualisé (à un niveau national) en particulier si les pratiques et formats le sont aussi.

R.21- Formations

Le domaine du transport est relativement technique, et l'interopérabilité passe aussi par une bonne compréhension (du domaine et entre acteur).

À ce titre il est recommandé de mettre en place un support et des formations en direction des différents acteurs :

- ***Producteurs de données***
- ***Réutilisateurs***
- ***Industriels et exploitants***

Priorité	2
Actions induites	Mettre en place les formations. Communication sur ces formations.
Acteurs concernés	Organismes de formation et écoles Experts du domaine DGITM, AFIMB, ETALAB
Niveau de consensus	100%

Les formations devront naturellement avoir des formes adaptées aux différents publics ciblés : si elles peuvent rester relativement classiques pour les producteurs de données, industriels et exploitants (mais avec des niveaux techniques adaptés), il conviendra de se concentrer sur les présentations, formation en ligne, ainsi que sur des outils comme les MOOC pour les réutilisateurs (on pourra notamment s'inspirer du travail réalisé autour du GTFS, par Google et les réutilisateurs).

Dans ce cadre on peut d'ores et déjà citer les informations mises à disposition par <http://www.normes-donnees-tc.org/> ou <https://www.data.gouv.fr/fr/> .

Dans ce contexte, la mise en place d'un "centre de ressource national" sur la donnée transport devrait probablement être envisagée.

R.22- Annuaire des données disponibles en France

Il est essentiel pour connaître l'ouverture des données transport en France , et pour piloter et suivre dans la durée la mise en œuvre des présentes recommandations, de disposer d'un annuaire identifiant, pour chaque AOT, quelles sont les données mises à disposition et les conditions d'accès (format, licences, services web, etc.)

Il est important pour le réutilisateur de pouvoir identifier les sources de données dont il a besoin.

Il est donc recommandé de mettre en place un annuaire national des données transport ouvertes.

Priorité	2
Actions induites	Mise en place d'un annuaire des données transport ouvertes au niveau national.
Acteurs concernés	CEREMA, DGITM, ETALAB Producteur de données pour la mise à jour régulière concernant les données qu'ils produisent.
Niveau de consensus	100%

Trouver la donnée dont on a besoin est souvent beaucoup plus délicat que ce que l'on pourrait imaginer (quel est l'AOT de l'agglomération ? Quel est son exploitant ? Comment s'appelle son site web ? Y a-t-il un site open data dédié ? La donnée open data transport est-elle disponible sur le site open data de l'agglomération ou sur un autre site ? Quelles sont les données transport ouvertes ? Etc.). La mise en place d'un annuaire a pour vocation de faciliter cette recherche. Cela permettra aussi de suivre la mise en œuvre des recommandations du Comité (qui publie quoi à quel format, etc.). L'annuaire peut être utile pour piloter l'ouverture des données transport, pour toutes les instances concernées. Dans une logique similaire à celle de la directive Inspire (et que la directive ITS imposera peut être en ce qui concerne les « traveller info data »), la publication les conditions d'accès et de réutilisation des données peut être considéré comme un préalable à la publication des données elle-même.

Il semble que le plus pertinent soit de s'appuyer sur l'existant : PASSIM (<http://petitpois.passim.info/poi/search?ack=&k=&q=open+data&w=&s=active>) et les travaux d'ETALAB.

Les métadonnées évoquées plus haut sont un des éléments dont l'utilisation pourrait grandement simplifier la mise en place d'un tel annuaire.

R.23- Plateformes

Le sous-groupe a été interrogé sur l'opportunité de la mise en place de plateformes de données territoriales : au final il apparaît que la question est plus du ressort du sous-groupe modèle économique qui aborde cette thématique dans son rapport.

Points non techniques :

Les points ci-dessous ne sont pas de nature technique et, à ce titre, ne peuvent pas faire partie des recommandations du sous-groupe. Ils ont cependant été abordés dans le cadre des discussions du groupe et il semble important qu'ils soient analysés par le sous-groupe juridique ou par la plénière.

NT.1- Les données doivent être disponibles

Avant d'évoquer les aspects techniques de la donnée transport, le tout premier prérequis est que la donnée existe.

Si ce n'est pas l'avis de tous les participants du sous-groupe, de nombreux autres acteurs, en particulier les réutilisateurs, considèrent généralement que l'ouverture des données transport doit être la règle et la non-ouverture l'exception. Le principe serait alors que la loi, sous une forme ou sous une autre, établisse cette ouverture par défaut sans interdire le fait de ne pas ouvrir, mais qui devra être justifié au cas par cas.

Le périmètre exact des données concernées pourra, par exemple, être précisé sur la base du travail réalisé sur la classification des données transport.

D'un point de vue technique, ce point se justifie entre autre par le fait que les technologies actuelles permettent aux réutilisateurs de s'organiser pour produire de façon non contrôlée les données non mise à disposition.

On notera que cette recommandation est portée à un niveau international par l'Open Knowledge Foundation (OKFN : <http://fr.okfn.org/> pour son antenne en France et <http://transport.okfn.org/> pour son groupe de travail transport).

Niveau de consensus	78%
---------------------	-----

NT.2- La possibilité de réutilisation doit pouvoir être identifiée de façon claire

Si la possibilité de réutilisation est liée à la nature publique des données, cela implique qu'un potentiel réutilisateur soit en mesure de juger de cette nature publique, ce qui n'est en rien évident (y compris pour un juriste d'ailleurs). Il serait souhaitable d'utiliser des moyens non ambigus pour qualifier la possibilité de réutilisation: par exemple la nature de données en utilisant des définitions normatives ou encore les données relatives à un format d'échange et profil technique

par exemple, comme cela est fait dans d'autres pays (on peut, entre autre, citer l'exemple de l'Angleterre avec le format TransXChange).

Niveau de consensus	78%
---------------------	-----

NT.3- La documentation associée aux données ouvertes (exclusivement) doit être libre d'accès et gratuite

Il est indispensable, si l'on souhaite que les formats normalisés puissent être utilisés dans le contexte de l'open data, de faire en sorte que les documents normatifs et éléments documentaires associés aux données ouvertes (et uniquement celles-là) soient librement, facilement et gratuitement accessibles et à tous (aujourd'hui ils sont difficiles à trouver, introuvable pour certains, et payants).

Ce point vaut essentiellement pour les normes dont l'usage serait recommandé, voir imposé, pour l'open data.

Niveau de consensus	89%
---------------------	-----

NT.4- Les données d'un maximum de modes (TC et non-TC) sont utiles pour les services à la mobilité

Pour les services à la mobilité, il ne faut pas se concentrer uniquement sur la donnée de transport en commun, mais sur les données relatives à l'ensemble des modes de transport (routier planifié et temps réel, modes doux, nouveaux services à la mobilité, etc.).

La plupart des réutilisations pour le transport en commun sont en effet généralisable à tous les modes, et la disponibilité de tous les modes combinés ouvre de nouvelles perspectives de réutilisation.

Niveau de consensus	100%
---------------------	------

Éléments sur les inductions financières des recommandations techniques :

Les demandes et échanges autour de l'induction financière de l'ouverture des données ont été nombreux.

Il est en final très difficile d'apporter une réponse consensuelle à cette question. En effet suivant les acteurs, certains indiquent que ces coûts sont très faibles (quel que soit la nature des données, temps réel ou pas), alors que pour d'autres ces coûts sont indiqués comme étant très élevés (en particulier pour le temps réel).

On notera aussi que les acteurs dont il est question ont tous des retours d'expérience opérationnels de mise à disposition de données (pas toujours en open data toutefois) pour des données planifiées et temps réel. C'est donc sur ces retours d'expérience qu'ils appuient ces avis très divergents.

Rappels de « best practice » du W3C pour la mise à disposition de données. :

Source : http://www.w3.org/2013/dwbp/wiki/Best_practices_guidelines

☐ BP1. Datasets should be available in an open format
☐ BP2. Datasets should be available in a machine-readable format
☐ BP3. Datasets should be available in multiple formats
☐ BP4. Datasets should be accessible in different ways
☐ BP5. Datasets should be available in standard data formats
☐ BP6. Datasets should be described by metadata
☐ BP7. Metadata should be available in a machine-readable format
☐ BP8. Standard vocabularies should be used to define metadata
☐ BP9. Datasets should be available at different levels of granularity
☐ BP10. Datasets selected for publication should be of high-value
☐ BP11. Datasets should be available in an up-to-date manner
☐ BP12. Each data resource should be associated with a unique identifier
☐ BP13. Datasets should be suitable for industry reuse
☐ BP14. Provenance information should be available
☐ BP15. Quality information should be available
☐ BP16. Usage information should be available
☐ BP17. Versioning information should be available
☐ BP18. Licensing information should be available
☐ BP19. Vocabularies should be well documented
☐ BP20. Existing reference vocabularies should be reused where possible
☐ BP21. Vocabularies should be shared in an open way

Cas d'utilisation associés : <http://www.w3.org/TR/2014/WD-dwbp-ucr-20140605/>

Annexe 7 : Rapport du sous-groupe modèles économiques

1 - Objet de la note :

Le présent document constitue la formalisation des réflexions et propositions du sous groupe modèles économiques sur l'ouverture des données transports.

La réflexion s'est structurée selon le plan de cette note à savoir :

- convient-il, et dans quelle mesure, de subordonner la réutilisation de certaines données à des conditions liées à l'intérêt général ou au respect des politiques publiques ?
- ouvrir les données, pour servir quels objectifs ? Les acteurs en présence et leurs attentes.
- quelles sont les difficultés à adresser en matière de modèles économiques pour les services de mobilité, ou d'autres services numériques ?
- quelles sont les propositions du sous-groupe permettant d'adresser les difficultés, avec des réponses aux questions clés que pose l'ouverture des données transport ?

2 - Introduction :

Nous pensons que l'ouverture des données transports représente un potentiel relai de croissance pour l'économie et les entreprises françaises, en particulier dans le numérique.

Quatre facteurs confortent cette affirmation :

- **une demande / un intérêt très fort des entreprises** : les données transports lorsqu'elles sont disponibles sont aujourd'hui parmi les plus utilisées et génèrent le plus d'applications et de services (cf résultats des villes ayant ouverts leurs données, avec ~60% des applications étant dédiées à la mobilité), en particulier dans le milieu urbain, qui concentre l'essentiel de la démographie et du potentiel économique du pays ;
- **des données existantes de qualité et nombreuses : la réflexion intervient également après une tendance de fond visant à optimiser l'usage des réseaux de transports grâce au numérique.** En effet, après un demi-siècle d'investissements massifs depuis les années 1960 dans les infrastructures de transports, pour arriver sur le plan national et local à construire des réseaux routiers, ferroviaires et de transports collectifs urbains performants, les années 1990 ont vu l'éclosion des premiers systèmes de gestion temps réel de trafic (routier, puis ensuite des transports collectifs), qui ont permis d'optimiser l'exploitation des réseaux pour les gestionnaires, et améliorer l'information aux usagers nées. Ce sont elles qui alimentent aujourd'hui les services d'information des usagers des réseaux de transports.
- **une maturité technologique** : Sur le plan technologique et sociétal **deux développements majeurs vont influencer durablement sur le développement du numérique et le positionnement des acteurs publics et privés** :
 - o la libéralisation de la géolocalisation satellitaire GPS et les développements cartographiques numériques associés ;
 - o l'explosion de la diffusion des smartphones, téléphones intelligents connectés à internet et embarquant une puce GPS/Galiléo, accessibles aujourd'hui au plus grand nombre.De nombreux services sont aujourd'hui techniquement et économiquement possibles grâce à la combinaison de ces technologies, des données des gestionnaires de réseaux, et certainement des données produites par les usagers.
- **Contexte économique** : la situation de crise des finances publiques pourrait servir paradoxalement de catalyseur pour faire de cette contrainte budgétaire un levier pour adapter notre organisation et mode de travail entre acteurs publics et privés : **le public ne peut plus tout financer, et le secteur privé tout comme le particulier doit s'adapter à cette nouvelle donne.**

Ce contexte favorable est une chance pour le développement du numérique. Les sociétés françaises qui bénéficieraient d'accès aux données, de cofinancement publics (investissements d'avenir) et un travail partenarial avec le secteur public peuvent largement être en tête dans la course mondiale sur le numérique (cf exemple de CityWay, partenaire d'Optimod'Lyon, qui co-financé par le programme des investissements d'avenir, et après un travail avec les équipes du Grand Lyon, vient de remporter un contrat de 8 millions de dollars au Canada, pour l'information voyageur de l'agglomération de Toronto agglomération de 5.5 millions d'habitants).

Pour identifier les modèles économiques potentiels, **nous pensons que l'approche ne doit pas se limiter aux seules données des transports publics, mais intégrer toutes les données de mobilité en général.** En effet, si les données des transports publics sont essentielles dans l'éco-système de mobilité urbaine, elles sont loin d'être les seules : les données des services de véhicules partagés (vélos, voiture en autopartage, en covoiturage), de trafic routier, de parkings, etc. reflètent toutes les possibilités de déplacement et donc de services numériques. Il serait préjudiciable à la réflexion et à contre courant des stratégies multimodales actuelles, de limiter le champ de la réflexion aux seuls transports publics.

En revanche, le sujet est plus complexe qui n'y paraît. Des interrogations, débats et parfois controverses se font jour : N'y a-t-il pas un risque de perte de contrôle par le public et a contrario de prise de pouvoir par le privé ? L'open data est-il la solution pour le développement des innovations et des nouveaux services ? La ville intelligente va t'elle être un relais de croissance fort pour les entreprises françaises ? Existe-t-il un modèle de valorisation autre que le gratuit ? et beaucoup d'autres questions liées à la ville intelligente que nous n'aborderons pas ici.

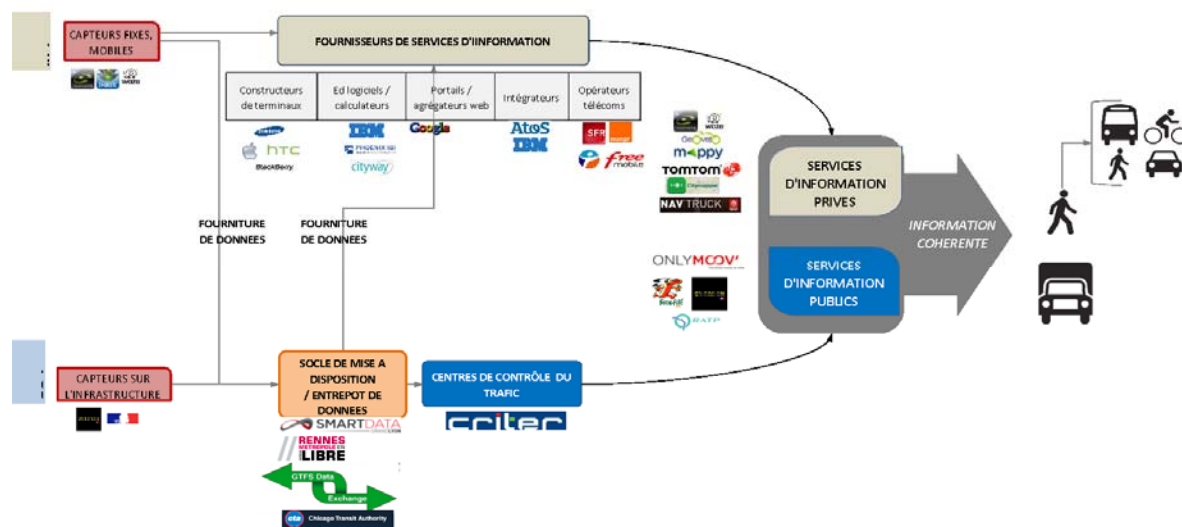
3 - Les acteurs en présence et leurs attentes :

Avant d'étudier les stratégies possibles pour développer de nouveaux services numériques utilisant les données mobilité, il convient de resituer les acteurs en présence et le cadre dans lequel ils agissent.

Globalement les acteurs en présence sont de qu

- **les acteurs publics**, qui agissent aujourd'hui en maître d'ouvrage, construisant pour les usagers (et électeurs) des infrastructures de transports, leur offrant des services d'exploitations de ces infrastructures, et également des services d'information sur les réseaux de transports. **Ils déploient une politique de mobilité pour garantir l'accessibilité de leur territoire, limiter les impacts environnementaux de la mobilité** et en particulier l'usage excessif de la voiture, et **renforcer le développement économique de leur territoire**. Le cadre réglementaire (code des marchés publics) ne leur permet pas l'agilité et l'adaptativité que requiert le numérique et l'évolution des technologies ;
- **les usagers (personnes et fret urbain)**, qui aujourd'hui consomment ces offres de transports et services associés, en n'en payant qu'une fraction du cout d'exploitation (ex en matière de TC : en moyenne 35% du cout d'exploitation est financé par l'utilisateur en France, Lyon faisant figure d'exception avec 50%). **Les usagers (personnes et fret urbain)**, demandent de pouvoir **se déplacer de la manière la plus efficiente** (c'est-à-dire avec le meilleur compromis temps de parcours/ coût) **tout en assurant un moindre impact environnemental**, compte tenu des problèmes récurrents de pollution de l'air en ville ;
- **les ré-utilisateurs des données mobilité** : éditeurs de logiciel, fournisseurs de services et/ou contenus, médias. Ils développent de nouveaux services, auprès de leurs clients, publics et privés, délivrant souvent ces services gratuitement ou sans modèle économique pérenne (cf expériences des concours suite aux démarches d'open data). Les modèles économiques reposent soit sur des fonds publics, soit sur des revenus publicitaires en particulier à Paris qui est un cas très particulier dans le paysage français. **Les acteurs privés de la mobilité**, qui, au-delà des opérateurs de réseaux de transport et leurs filiales, ont pour clients essentiellement les acteurs publics dans le secteur de la mobilité. **Ils ont besoin de revenus leur permettant de poursuivre et développer leurs activités et leur R&D. Ils souhaitent développer de nouvelles offres et des modèles économiques stables ;**
- **les acteurs privés de la mobilité**, agissant essentiellement dans le cadre de contrats publics (essentiellement des marchés publics, DSP et concessions) qui exploitent les réseaux de transports, et les services d'informations sur ces réseaux. Ils ont in fine des objectifs similaires aux ré-utilisateurs de données mobilité, avec un champ d'application plus restreint (celui de la mobilité), et ne souhaitent pas forcément couvrir tous les champs numérique de la mobilité (en fonction de la stratégie de chacun des acteurs).

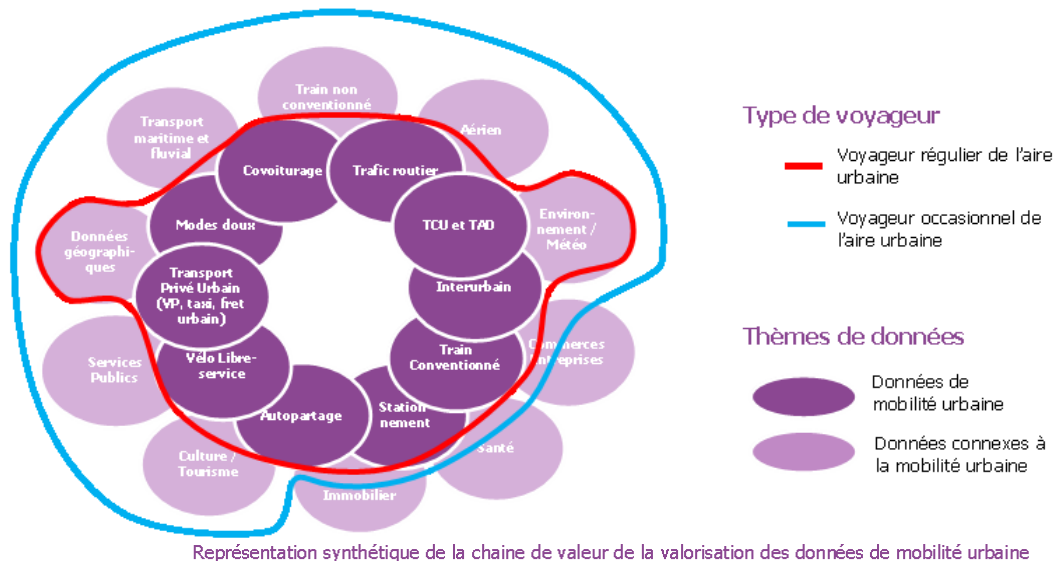
Schéma cible de services d'information voyageurs publics et privés :



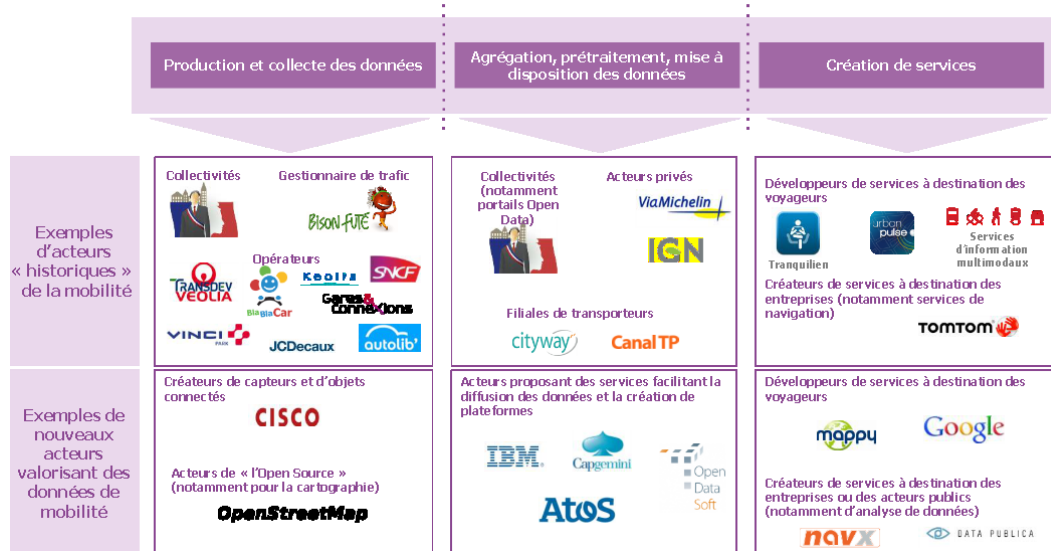
4 - Les services possibles et les principes des chaînes de valeur :

Il n'est pas du ressort du groupe d'imaginer l'ensemble des services possibles construits en utilisant les données mobilité. Néanmoins, il est apparu intéressant de resituer, au travers de deux schémas issus d'une étude menée par la Caisse des dépôts, d'une part, l'environnement de données d'une aire urbaine, et d'autre part, le principe des chaînes de valeur.

Représentation schématique à titre illustratif des thèmes de données nécessaires pour proposer des services aux différents types de voyageurs



Représentation synthétique de la chaîne de valeur de la valorisation des données de mobilité urbaine



Les modèles économiques peuvent être nombreux. Citons par exemple - en dehors du marché public, dont il n'est pas question dans le cadre de ce travail :

- la publicité ce qui suppose une assiette assez conséquente de populations, et donc des aires urbaines larges (>1 à 2 M habitants vraisemblablement ?), ou une portabilité aisée des services d'une aire géographique à une autre, ce qui suppose une mise aux normes de la plupart des référentiels de données existants ;
- l'achat par les usagers : de même, il faut une assiette de population assez large ou une portabilité aisée des services d'une aire géographique à une autre, ce qui suppose une mise aux normes de la plupart des référentiels de données existants ;
- un mix des deux modèles précédents en jouant sur les publics occasionnels (tourisme, business) ;
- le rachat par un gros acteur du secteur, en jouant sur l'aspect start-up. Ce modèle perturbe fortement les modèles économiques classiques et se situent parfois en dehors de toute rationalité économique (Waze a ainsi été racheté par Google ~37 fois son CA cumulé ...)

5 - Nécessité de faire émerger de nouveaux modèles économiques, plus indépendants des fonds publics :

Il est clair qu'aujourd'hui perdurer sur les seuls fonds publics est un vrai facteur de risque dans la période actuelle de tension budgétaire. Si l'on souhaite que l'ouverture des données produise de la valeur ajoutée économique et des relais de croissance pérennes, ceux-ci doivent reposer sur des acteurs privés et des modèles économiques solides.

C'est dans ce contexte particulier - *a priori* très favorable - qu'il faut penser la construction de nouveaux modèles économiques pour les services liés aux données de mobilité.

C'est pourquoi le sous groupe s'est accordé sur le fait que ses réflexions doivent avoir pour objet de créer un environnement favorable au développement de modèles économiques de services numériques liés aux données mobilité, dans le respect des politiques publiques et de la libre concurrence.

Pourtant, si l'on reste sur le champ de l'information sur la mobilité, jugé le plus prometteur en matière d'économie numérique, il n'y a pas à ce jour de modèle économique autonome des fonds publics. Mis à part l'information routière, financée par l'utilisateur – la plupart du temps sans qu'il en ait conscience via un coût attaché à l'achat du véhicule et redistribué aux opérateurs d'information routière -, il n'y a pas d'exemple de services d'information de mobilité financés autrement que par des fonds publics.

L'une des raisons invoquées - à juste titre- est l'absence de disponibilité des données. Avec l'*open data* il s'agit de libérer les énergies et la créativité de nombreux acteurs et les services aux citoyens vont se diffuser largement et en nombre.

Force est de constater que si les acteurs – PME et start-up – sont nombreux et créatifs, le bilan économique de l'*open data* n'est pas au rendez-vous de la promesse : il y a bien eu la création d'applications sur la mobilité issue des démarches d'*open data*, mais ces applications n'ont généré quasiment aucun *business* significatif et leur pérennité s'est avérée très fragile une fois les fonds publics d'aide au lancement épuisés.

La raison en est en fait assez simple, **l'ouverture des données est certainement une condition *nécessaire* à l'émergence de nouveaux services, mais ce n'est pas une condition *suffisante*.** Il y a bien d'autres facteurs empêchant la diffusion de services numériques et pour les adresser il faut aller au-delà de la démarche d'*Open Data*, en se rappelant que c'est un moyen et non une fin :

- **des services incomplets et donc à faible valeur ajoutée :** les services d'information sur la mobilité sont aujourd'hui organisés par silo de données et sont le reflet du millefeuille institutionnel : vous avez autant d'applications et sites internet que d'organisations : trafic routier par zone, TER, vélo libre service, parkings, autopartage, covoiturage, transports en commun urbains, aéroport, taxis, ...etc. Pris individuellement ces services ont peu de valeur marchande leur permettant de se développer. Ils n'ont pas tous une couverture géographique suffisante, ne disposent pas toujours d'information temps réel et proposent dans le meilleur des cas une juxtaposition des modes de transports, mais sans combinaison des modes ;
- **un accès aux données parfois limité :** la situation a particulièrement évolué sur le sujet, grâce au mouvement open data. **Les acteurs publics sont aujourd'hui engagés dans une mise à disposition des données et la question n'est plus de savoir s'il faut le faire mais comment.** Ces acteurs demandent en particulier qu'il y ait une cohérence entre leurs politiques publiques et les services produits avec leurs données. Par ailleurs **de nombreuses données échappent au mouvement open data et dans le domaine de la mobilité certaines sont particulièrement critiques. Ainsi les données des opérateurs privés, agissant souvent, mais pas systématiquement, dans le cadre de contrats ou statuts publics, ne sont pas toutes accessibles :** données ferroviaires, parkings et autoroutes concédés etc. Si la négociation échoue, comment accéder à ces données ? Il en va de même des données issues des usagers, parfois sans leur consentement formel, qui peuvent pourtant très utilement compléter les bases de données des réseaux et services de transports, pour peu que l'on assure la protection de la vie privée et que l'on reste dans un cadre légal sur les données personnelles.
- **des acteurs dominants qui perturbent une logique économique rationnelle :** il y a d'abord - on l'oublie - les acteurs publics qui ont tendance à diffuser une large gamme de services bien au-delà de leurs obligations réglementaires, mais aussi bien sur de grands acteurs en particulier (mais pas seulement) de l'internet – notamment Google – qui diffusent aujourd'hui des services très performants et ce gratuitement, leur modèle économique étant fondé sur la publicité et des commissions sur la mise en relation entre les particuliers et de multiples services. Ces opérateurs ont des stratégies et pour certains déjà des positions monopolistiques sur l'information (ex 70% de taux de pénétration pour Google sur l'agglomération lyonnaise, à quasi égalité avec le service public d'information sur les transports en communs). Ils perturbent le marché avec des phénomènes qui ne sont pas sans rappeler la bulle internet du début des années 2000 : Waze, start-up fournissant de l'information routière, a été ainsi été racheté plus de 1 milliard de dollars, soit 37 fois son chiffre d'affaire cumulé ! La culture du tout gratuit qu'ils diffusent est également un non sens économique et un outil au service de leur stratégie de monopole.

- **une notoriété insuffisante et des circuits de distribution difficiles à monter** : Corollaire du point précédent, quand bien même une entreprise arriverait à élaborer un service performant, la difficulté pour elle va être de le faire connaître du plus grand nombre et de le diffuser sur les territoires urbains de manière à créer son modèle économique. Tout le monde ne s'appelle ni Google, ni SNCF ni Lille, Rennes Métropole etc.

L'objectif de l'ouverture des données c'est donc aussi la création de nouveaux services et de modèles économiques indépendants des fonds publics. Dans des marchés non matures, il faut travailler en partenariat avec le secteur privé pour créer les conditions d'émergence de nouveaux modèles économiques.

Ainsi, la situation de crise des finances publiques pourrait servir de catalyseur pour faire de cette contrainte budgétaire un levier pour adapter notre organisation et mode de travail entre acteurs publics et privés.

En synthèse :

- **des services d'information voyageurs qui n'ont pas de modèles économiques parce que trop restreints et sans valeur ajoutée réelle ;**
- **des données encore insuffisamment disponibles, en particulier les données opérées sous contrats publics (marchés, DSP, concession...) par des opérateurs privés ;**
- **un besoin de s'assurer que le secteur privé ne développe pas des offres allant à l'encontre des politiques publiques avec les données publiques ;**
- **des monopoles privés et publics, toujours destructeurs à terme de valeur et d'innovation, ne laissant que très peu de chance au secteur privé PME, ETI... de se développer ;**
- **une posture de soutien au marché au-delà des aspects financiers, à adopter par le secteur public pour faire décoller des offres qui rencontrent leur clientèle.**

Il n'est donc pas surprenant qu'il soit particulièrement complexe pour une PME de pouvoir lancer un service viable économiquement sans fonds publics. *A contrario*, l'ouverture des données en *open data* est pour les grands monopoles privés une véritable aubaine et constitue une menace certaine pour les *pure players* de l'information mobilité.

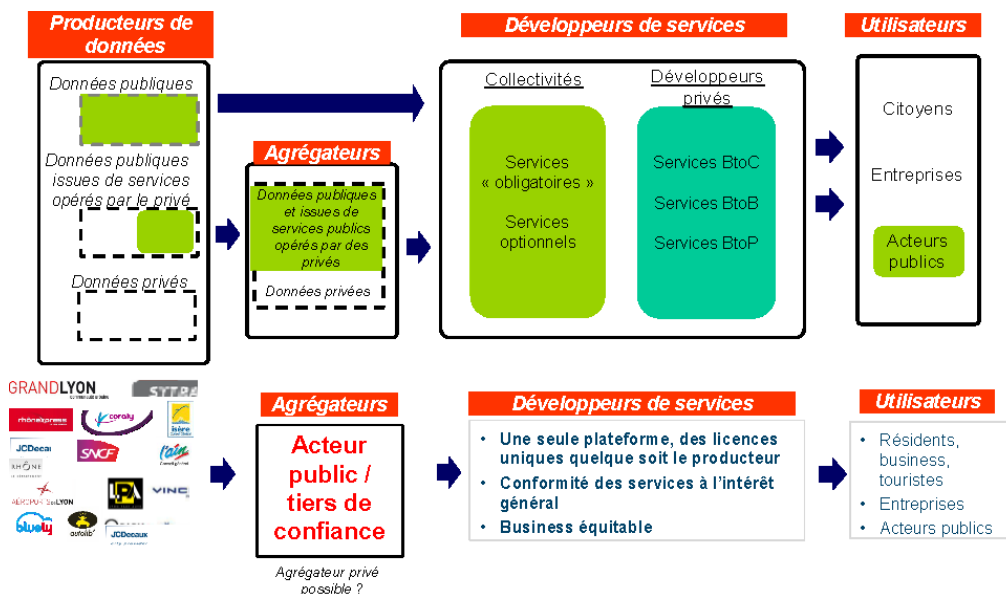
Ce contexte a figé la situation côté fournisseurs, avec des entreprises liées aux opérateurs de transports publics, donc très peu nombreuses, financées exclusivement sur fonds publics et qui travaillent sur des niches non encore accessibles à des Google faute d'accès aux données. Cet environnement a institué *de facto* des oligopoles sur le plan national, ce qui n'est pas favorable à l'innovation.

Comment en sortir ?

6 - Les propositions du groupe de travail :

En premier lieu il est apparu au groupe particulièrement complexe de chercher à décrire les modèles économiques possibles. Ceux-ci peuvent être très variés et il apparaît illusoire de vouloir les imaginer tous et les décrire précisément (financement direct par les usagers, par la publicité, financement par des professionnels, logique de levée de fonds et de rachat par des grands groupe à l'image de Waze et sans doute demain Moovit etc).

Le schéma suivant, issu d'une étude de la Caisse des Dépôts nous paraît assez bien résumer les chaînes de valeurs et les multiples possibilités, dans des marchés qui sont en train de se construire :



Nous avons cherché à adresser les points de blocage identifiés au chapitre précédent et avons identifié une série de questions auxquelles nous avons cherché à répondre collectivement. Lorsqu'un consensus a été trouvé celui-ci est clairement indiqué, et lorsque cela n'a pas été possible également nous avons précisé les différentes positions.

Nous classons nos propositions en deux types :

- **les évolutions du cadre réglementaire** (à étudier par le sous-groupe juridique) ;
- **les bonnes pratiques qui ne relèvent pas d'une modification du cadre réglementaire.**

ce à partir d'une position schématique suivante :

- **le secteur public est en charge de construire des entrepôts de données publiques et privées sur son territoire et le rendre accessible.** Il doit assurer ses obligations de services publics en veillant cependant à laisser des espaces au secteur privé en particulier sur le numérique ;
- **Le secteur privé est en charge de développer des services**, en évitant l'écueil de construire des modèles économiques basés sur des fonds publics ;
- **Les licences de mise à disposition des données doivent permettre de gérer les risques** que sont les monopoles et la non prise en compte par les opérateurs des contraintes induites par les politiques publiques.

6.1. - Construire des services à valeur ajoutée qui trouvent leur modèle économique :

Il n'appartient pas au groupe de travail de faire des propositions en matière de modèles économiques possibles. A noter cependant que des études et expérimentations (en particulier Optimod'Lyon) ont validé le principe que **des services complets (tous modes, couverture géographique cohérente) et de haut niveau (temps réel, historique, prédictif), ont une valeur ajoutée perçue comme forte avec une propension à payer très importante** : 80% des usagers sondés (cf étude Optimod'Lyon) !

Ceci confirme le potentiel des services numériques construits avec des données mobilité. Il convient de laisser l'initiative privée construire les offres, le secteur public pouvant grandement créer des conditions favorables pour aider des marchés encore non matures.

6.1. - Rendre accessible les données :

A l'évidence, le carburant des services ce sont les données, qu'elles soient publiques ou privées.

Sur les données privées non produites dans le cadre de contrats publics, il ne peut y avoir qu'accord entre les parties pour y avoir accès. Le sous-groupe s'est donc focalisé sur les données publiques, et celles produites dans le cadre de contrats publics.

1^{ère} proposition : Donner un caractère public à certaines données issues de contrats publics - privés

Sur les données publiques et les données aujourd'hui à statut privé, le groupe de travail propose de traiter de manière particulière les données d'information voyageurs élaborées dans le cadre de contrats publics avec rémunération financière : il s'agit donc des marchés publics, des DSP, des concessions et PPP.

Cette disposition ne concerne pas les données SNCF voyage, qui ne font pas l'objet d'un contrat public.

P1 : La proposition à caractère réglementaire, serait que tout jeu de données relatives à l'information voyageur et aux flux de mobilité (données statiques, théoriques, temps réel, historiques...) issu de contrats publics rémunérés reviennent de droit au cocontractant public et qu'il en ait toute liberté d'usage. Ceci reviendrait à dire que ces données aient un statut public.

Il faut cependant étudier finement les choses pour éviter des effets indésirables et respecter la propriété intellectuelle.

Cette disposition permettrait également d'éviter les pertes de temps liées au millefeuille territorial et aux problèmes locaux, tout acteur public pouvant avoir accès aux données d'un autre acteur public, de droit.

Le sous-groupe tient à faire remarquer que dans le cas des DSP ou concession, l'ouverture des données est une condition favorisant une mise sur pied d'égalité pour les procédures d'appels d'offres de DSP et concession. Cette proposition permettrait à la collectivité d'avoir accès à des données lui permettant d'assurer l'équité entre les candidats.

2^{ème} proposition : Fournir les données gratuitement

La tarification des données est économiquement un frein au décollage de marchés naissants et par ailleurs ne permettra jamais d'assurer des rentrées financières significatives aux acteurs publics. Les rares expériences en la matière l'ont prouvé. Enfin, la valeur d'une donnée ne provient pas de son coût mais de sa vente : dans des marchés naissants, les petits acteurs (PME, start-up) n'ont pas les moyens de les acheter, le risque est que ce soit les situations monopolistiques qui seules en aient les moyens.

P2 : Afin de renforcer les possibilités de décollage de marché naissant, les données publiques sont mise à disposition des tiers de manière gratuite sauf exceptions (voir §6.3) et ce quel que soit le format de rediffusion (voir P6). La mise à disposition des données doit se faire de telle sorte qu'elle permette d'opérer un service sans tarification.

Des conditions de service spécifiques peuvent cependant donner lieu à tarification (ex garantir une bande passante... nécessitant des investissements spécifiques au-delà du service de base, ou une disponibilité particulière du flux de données).

C'est le cas par exemple d'un calculateur d'itinéraire (qui est un service et non une donnée), qui au-delà d'un seuil de sollicitations, peut nécessiter une infrastructure technique et une exploitation spécifiques pour le ré-utilisateur demandeur, et donc amener à tarifier cette infrastructure et le service associé (surveillance du flux 7/7 24/24 etc) pour des ré-utilisateurs demandant des conditions techniques particulières.

3^e proposition : Des licences favorisant la réutilisation commerciale

En cohérence avec l'objectif de créer un environnement favorable au développement de services numériques indépendants des fonds publics, les licences de mise à disposition doivent favoriser la réutilisation commerciale et en aucun cas l'entraver.

P3 : Des études conduites ont montré que la licence ODBL est une gêne et que la licence ouverte dite « Etalab » semble plus appropriée et permet plus de modèles économiques que la licence ODBL. Ce point est confirmé par le sous-groupe juridique.

4^e proposition : Des réutilisations cohérentes avec les politiques publiques

P4 : L'accès aux données peut être conditionné à la prise en compte des politiques publiques par les services développés à l'aide des données publiques. Cette disposition peut aujourd'hui être gérée dans le cadre du droit actuel, qui ouvre une limitation à la réutilisation dans la limite de l'intérêt général, notion large. Le sous-groupe juridique peut évaluer la nécessité de préciser ou pas cette notion pour éviter d'être trop soumis à des interprétations de la jurisprudence.

Comment en effet comprendre qu'un service utilisant des données publiques envoie le trafic routier devant les écoles parce que la rue parallèle est encombrée ? Ou qu'un autre dirige de manière préférentielle les usagers vers son parking, son réseau de transport ?

Cette disposition doit permettre de rassurer les craintes, justifiées, d'acteurs publics sur la réutilisation de leurs données, sans pour autant que cela soit un prétexte à ne pas les ouvrir.

5^e proposition : Des données facilement ré-exploitable

Le sous-groupe considère par ailleurs que c'est à l'acteur public (ou privé sous contrat public) de fournir des données permettant une interprétation facilitée pour le ré utilisateur.

Par ailleurs les données brutes sont souvent en termes de mobilité porteuses d'erreur d'analyse et peuvent conduire à délivrer des informations erronées. Un filtrage, nettoyage des données est nécessaire avant une exploitation fiable et facilitée par ce filtrage.

P5 : En conséquence, les données mises à disposition doivent a minima être les mêmes que celles utilisées par les opérateurs de mobilité ou leurs AOT pour leur propre services(d'information, ...).

Le groupe préfère laisser ouvert l'obligation ou pas (selon les possibilités juridiques) de fournir les données issues directement des systèmes d'exploitation. Nous renvoyons aux travaux du groupe technique sur le sujet.

1^{ère} Bonne pratique : Construire des plateformes de données rassemblant l'ensemble des données d'un territoire, publiques et privées

Pour faciliter les réutilisations et le développement de services à haute valeur ajoutée, l'accès à un ensemble intégré de données est nécessaire : c'est une condition nécessaire au développement de services de haut niveau pouvant en conséquence trouver leur modèle économique. Cet accès doit se faire avec des licences ouvertes et permettant de gérer les risques de compatibilité avec les politiques publiques et le risque de monopoles.

BP1 : A l'image de ce qui se construit sur certains territoires et du résultat de projets et d'études, il est recommandé, que les territoires construisent à une échelle cohérente d'un point de vue utilisateurs (donc non forcément de manière bijective avec les frontières administratives), un entrepôt rassemblant l'ensemble des données mobilité, quel que soit leur statut, public ou privé, et en assurer l'accès à des tiers avec un seul type de licence par donnée.

Ce set de données mobilité doit être **contrôlé par les acteurs publics**, avec un chef de file clairement identifié, d'une part parce que les acteurs publics sont les principaux fournisseurs de données, et d'autre part parce qu'ils ont en charge la gestion des réseaux urbains de transports, et qu'à ce titre ils ont une fonction de régulation publique qui n'est pas déléguable. Enfin, il est infiniment plus facile aux autorités publiques d'un territoire d'assurer le rôle de tiers de confiance pour trouver les accords permettant d'enrichir le set de données avec des données privées. L'échelle la plus pertinente en terme d'usagers semble clairement les agglomérations. Les données

interurbaines doivent permettre de connecter les villes entre elles. Leur niveau de gestion (national / régional) le plus pertinent reste à déterminer.

Ainsi, un réutilisateur avec les dispositions proposées aurait accès à un nombre important de données (il est fréquent d'avoir plus de 15 sources de données mobilité sur une métropole) en un seul lieu, avec des licences en nombre très restreint et cohérente entre elles. L'APIE et Etalab pourraient travailler en sus de la licence ouverte à un modèle de licence gérant la cohérence avec l'intérêt général et un écosystème concurrentiel équitable (voir §4.3). Un dispositif d'homogénéisation des licences et des formats de données au niveau national serait certainement souhaitable afin de favoriser la portabilité des services.

On pourrait ainsi avoir deux niveaux territoriaux pour les données :

- **des plateformes à l'échelle d'une aire urbaine** (donc au-delà des frontières administratives d'une métropole ou communauté de communes) afin d'assurer une gestion cohérente des données du point de vue de la mobilité des citoyens du bassin de vie visé ;
- **des plateformes de données permettant de connecter les villes entre elles**, au niveau régional ou national.

Le sous-groupe recommande évidemment d'éviter les doublons sur un territoire donné. Il s'interroge sur le caractère obligatoire à donner à cette bonne pratique.

6.3 Mettre en place les conditions d'un écosystème concurrentiel équitable

2^{de} Bonne pratique : Laisser des espaces au secteur privé

Pour laisser des espaces, le public doit réinterroger ses services et en particulier les médias de diffusion : par exemple, est-ce à lui de diffuser et maintenir des applications mobiles, alors que les procédures publiques ne lui permettront jamais de suivre le rythme des évolutions technologiques et des usages ?

Des services à haute valeur ajoutée peuvent rencontrer leur public et permettre de sortir du tout gratuit actuel du numérique, mortifère économiquement. Comment cela serait-il possible si le secteur public fournit gratuitement (c'est-à-dire financé par le contribuable) ces services ?

Le sous-groupe propose de recentrer les services numériques publics afin de permettre l'émergence de modèles économiques privés autonomes des fonds publics. Si le secteur public a des obligations en matière d'information multimodale, cela ne signifie pas pour autant qu'il doive tout faire sur ce champ. Par ailleurs le carcan réglementaire public empêche l'agilité nécessaire au monde du numérique (évolution rapide des technologies versus processus de commande lents et rigides).

La puissance publique porterait ainsi sous maîtrise d'ouvrage publique le développement et l'exploitation des services de base, ce qui en termes de données signifie : **la constitution de données fiables et leur mise à disposition, l'information sur tous les modes de manière individuelle et combinée sur site web.**

Les services visés ici concernent les trafics routiers, les services de vélo en libre service, vélo, les transports en commun sur des médias de masse (internet fixe et radio) etc.

BP2 : La recommandation est de laisser au secteur privé le développement des services sur les applications mobiles et les services avancés sur internet et de focaliser le secteur public sur une offre de base sur ses sites internet. Ceci suppose que le secteur public accepte par exemple de ne plus développer et retirer ses applications mobiles, qui compte tenu du carcan réglementaire, seront toujours en retard sur les développements technologiques et les usages.

Le public peut aussi mettre en open service des applicatifs dont il a l'utilité et dont la mise à disposition permet la construction de nombreux autres services. Nous visons ici en particulier les calculateurs d'itinéraires multimodaux maîtrisés par peu d'acteurs en France.

6ème proposition : Des licences permettant d'assurer un écosystème concurrentiel équitable

P6 : La proposition est d'intégrer dans les licences d'accès aux données des clauses permettant de gérer les situations monopolistiques actuelles et futures.

Cette disposition nécessite une évolution réglementaire (loi CADA, voir impact sur d'autres code article de lois).

Néanmoins, les cas où la DGCCRF est sollicitée pour éviter des situations monopolistiques ou des perturbations d'une libre concurrence sont, au final, assez nombreuses.

Le Grand Lyon a mis en place une licence avec une clause prévoyant une tarification dissuasive se déclenchant à partir d'un seuil élevé de part de taux de pénétration du marché, que le Grand Lyon évalue lui-même par enquête représentative de sa population. Pour une PME ou une ETI se développant l'accès est totalement gratuit. Pour une entreprise avec des visées monopolistiques le principe de la tarification et son niveau est suffisamment dissuasif pour éviter qu'il ne s'empare d'un marché ou y perdure avec une position dominante excessive. Ce dispositif est suffisamment dissuasif pour que les monopoles ne signent de telles licences. Une autre idée pourrait être de donner un statut de taxe au lieu de redevance avec la faculté de mise en œuvre par les autorités

publiques locales, de manière simple et encadrée. Cela permettrait de gérer l'écueil juridique potentiel pointé par la DGCCRF sur la licence associée du Grand Lyon (légitimité à agir sur le champ de la concurrence).

On pourrait envisager également une licence « anti monopole » sans redevance avec des clauses limitant / interdisant l'accès aux données à des entreprises en situation de monopole sur des bases objectives (ex Vinci Park a dans le passé était interdit de concourir à des AO de parkings compte tenu de sa position ultra dominante, cf passage par les USA du monopole d'IE en interdisant à MS de le vendre installé avec son OS) ou en non-conformité avec certaines dispositions légales (ex : art 50 du CMP). D'autres limitations comme le respect de la vie privée, le fait d'être en règle de ses obligations fiscales et sociales comme pour un marché public, ... (le fait de ne pas être condamné pour ces motifs) peuvent aussi être envisagées. Ces dernières dispositions ne visent cependant pas spécifiquement les situations de monopoles et n'ont pas la faveur du groupe de travail.

Le fait que les entreprises soient en situation de monopoles pourrait se mesurer selon des méthodes propres à chaque acteur public, mais encadré par la loi.

Il peut y avoir d'autres dispositifs, la logique et la finalité étant de favoriser une concurrence équitable si l'on veut éviter que l'ouverture des données ne produise le résultat inverse de celui recherché, c'est-à-dire la destruction d'un éco système naissant et la délocalisation des emplois et des revenus du numérique.

Le sous-groupe juridique doit se pencher sur cette question, avec la seule finalité de permettre aux acteurs publics de développer des licences permettant de gérer les risques de situations monopolistiques, qui empêcheront le décollage des marchés du numérique et sont toujours à terme destructeur de valeur et freinent l'innovation.

3^{ème} Bonne Pratique : Donner accès à des canaux de distribution à des entreprises peu connues

Il s'agit de mettre le pied à l'étrier à des entreprises n'ayant pas la notoriété de grands groupes pour que leurs services soient connus et diffusés au grand public.

BP3 : La recommandation est de **permettre - de manière non discriminatoire - à des entreprises de bénéficier des marques des territoires pour la diffusion de leurs produits**. Cela passe sans doute par une logique de labellisation afin de permettre à des produits numériques d'être diffusés via les sites web des acteurs publics urbains, sites internet étant parmi les plus consultés dans les métropoles notamment. La labellisation doit favoriser des modèles économiques autonomes des fonds publics. Elle ne doit donc pas interdire les services payants, ni restreindre les modèles économiques.

4^e Bonne Pratique : Donner accès à des services clefs, mais maîtrisés aujourd'hui par un faible nombre d'acteurs

Il s'agit ici d'ouvrir le marché des services mobilités, en mettant en open service (et donc de manière différente de données) des services nécessitant un savoir faire ou des technologies maîtrisés par peu d'acteurs et donc empêchant le développement de services les intégrant.

BP4 : La recommandation est de **mettre à disposition de tiers, dans les mêmes conditions que les données** (gratuitement sauf cas particuliers), **des services verrouillant ou pénalisant le décollage de marchés**, services étant par ailleurs acquis par la collectivité.

C'est l'exemple type des calculateurs multimodaux (en dehors du strict périmètre des TC).

Dans ce cas précis, la collectivité si elle achète un tel service, dont la propriété reste aux acteurs privés, doit pouvoir le mettre à disposition sur la plateforme d'ouverture de données, via des web services.

En cohérence avec les propositions précédentes, cette mise à disposition doit se faire gratuitement sauf cas particuliers.

7^e proposition : utiliser les normes existantes ou en cours de développement pour diminuer les coûts d'entrée et étendre la surface des marchés.

Si l'on veut permettre à des applications et services de fonctionner dans différentes villes, l'utilisation de normes européennes et mondiales est nécessaire. Cette disposition permettra par ailleurs de diminuer le coût de développement des services (parfois l'adaptation à des standards spécifiques coutent jusqu'à 50% du coût du développement du service). Cette condition est essentielle pour qu'au-delà des grandes métropoles, les services puissent se déployer dans des marchés de surfaces plus faibles et donc moins attractifs pour les opérateurs.

P7 : La proposition est de **rendre obligatoire l'usage des normes existantes pour les données publiques transports**, qu'elles émanent d'entités publiques ou d'entités privées agissant dans le cadre de contrats publics (marchés, DSP, concessions, PPP). Cette proposition ne fait pas obstacle à ce que d'autres formats soient également disponibles sur les plateformes de mise à disposition (GTFS par exemple), format plus simple que certaines normes et donc favorisant le développement d'applicatif à moindre coût, mais moins riches fonctionnellement.

5^e Bonne pratique : Si l'utilisation des normes existantes devenait obligatoire, il est nécessaire que l'État mobilise ses moyens pour développer et maintenir les outils permettant de certifier les bases de données mobilité.

Les membres du groupe de travail font état d'un problème de l'usage de l'outil Chouette pour qualifier si les données TC sont conformes à la norme TRIDENT.

Par ailleurs, il est aussi nécessaire que les services de l'État se mobilisent pour simplifier l'usage des normes existantes ou en construction qui sont jugées parfois trop complexes et génèrent un coût d'usage (et donc un frein) important.

BP5 : L'État devrait mobiliser certains de ses moyens (CEREMA) afin de développer, maintenir et diffuser les outils de certification de conformité des bases de données aux normes ainsi que pour étudier la simplification de l'usage de normes jugées parfois trop complexes.

7 - Les questions diverses abordées en groupe de travail

Q1 : En quoi une situation de monopole est-elle un problème pour le développement du numérique ?

En quoi Google ou tout autre monopole est-il une menace pour l'innovation et le développement économique ?

Les monopoles entravent par nature l'innovation, car ils ne permettent pas, ou très difficilement, à des concurrents de proposer des services alternatifs.

Dans le cas du numérique, le monopole n'est pas un monopole de droit, mais de fait : certains acteurs comme Google concentrent l'audience sur internet et/ou des capacités d'investissement leur permettant de préempter de nouvelles technologies. Ces capacités financières leur permettent en outre de financer à perte des activités dans le temps, excluant les autres acteurs ayant, eux, besoin d'un retour sur investissement.

Au-delà de Google il faut se rappeler comment Microsoft a verrouillé le secteur informatique pendant 20 ans, une fois sa situation de monopole établie. Pourquoi l'innovation ne devrait-elle passer que par un acteur ?

Face à ce constat, l'État doit étudier comment éviter que la valeur potentielle du numérique soit captée par ces grands groupes, avec au final peu d'emploi et de recettes fiscales localisées en France. Les acteurs locaux doivent étudier comment construire des offres, en partenariat avec le secteur privé répondant plus finement et de manière plus adaptée aux besoins locaux.

Q2 : Comment rendre disponible des données de crowdsourcing, gérées par des services privés (Coyote sur le trafic, Google, ...) ?

Les données collectées par des acteurs privés auprès ou à travers leurs abonnés ou utilisateurs de leurs services constituent une propriété inaliénable. Cela ne peut donc passer que par des accords spécifiques.

En revanche, si des données sont collectées par des acteurs privés dans le cadre d'un marché public (DSP, PPP...), la proposition est de donner un statut public à ses données.

Q3 : Est-il réaliste d'imaginer des situations non monopolistiques sur des villes en dehors de Paris ?

L'exploitation des données locales est rendue complexe et coûteuse en raison de leur hétérogénéité. La normalisation des données et leur mise à disposition au sein d'entrepôts eux aussi normalisés favoriserait le développement des services pour les agglomérations de taille moyenne. Plus l'agglomération est petite, plus le degré de normalisation doit être grand pour permettre un retour sur investissement.

Par ailleurs, les petites agglomérations représentent des marchés de niche adaptés à de petits acteurs, qui n'intéressent pas forcément les grands acteurs. Si monopole il y a (car le marché est trop petit), il n'est plus forcément le fait d'un acteur national ou international, et est donc très local. Est-ce vraiment un monopole ?

Q4 : Quelle place pour les SIM régionaux existants dans le contexte évoqué ?

Les SIM régionales existantes constituent un effort de mutualisation sur des territoires où l'offre et la demande de mobilité sont limitées, en dehors des grandes agglomérations qui disposent de leur propre SIM. Il est peu probable que le privé aille sur fonds propres construire des services dans ces territoires. Une mutualisation à l'échelle régionale avec intégration des fonctions d'achat est une piste prometteuse, en veillant au rapport coût/service rendu, compte tenu des fréquentations faibles de ces SIM à l'heure actuelle.

À noter que le gouvernement du Royaume-Uni arrête fin septembre 2014 l'exploitation de « transport direct », la plateforme nationale sur les transports.

Annexe 8 : Le cas des données des bornes de recharge de véhicules électriques

Gilles BERNARD

le 17/12/2014



Ouverture des données pour le développement des offres de transport, cas de l'itinérance de la recharge des véhicules électriques.

Concernant l'ouverture des données en matière d'offres de transports, il convient d'évoquer le développement des voitures électriques. L'utilisation massive de véhicules électriques rechargeables est un objectif politique motivé par les enjeux énergétiques et environnementaux. Il ne pourra se réaliser que si l'usage de ces véhicules est facile et le conducteur libéré du souci de leur autonomie réduite et du temps de recharge long. La réussite des enjeux et de ce développement sera possible par la fourniture de services d'accompagnement des utilisateurs reposant sur un accès aux données des bornes de recharge.

En premier lieu, ils permettront en effet de localiser les points de charge disponibles depuis les véhicules, de les réserver, et d'être autorisé à y accéder au titre du contrat de l'utilisateur avec son fournisseur de services. Ultérieurement, la gestion énergétique des recharges, et l'accès à d'autres services de transports (stationnement, déplacements multi-modes etc.) reposeront aussi sur des échanges de données entre le fournisseur de services à l'utilisateur et les opérateurs des bornes de recharge et d'autres ressources de transport.

Pour que ces services soient efficaces et sans frontière, l'accès aux données doit être ouvert et non discriminatoire entre tous les opérateurs, dans une organisation de marché concurrentielle à maille européenne.

D'ores et déjà, plusieurs textes réglementaires renvoient à cette perspective :

- l'article 10 de la petite loi sur la transition énergétique pour favoriser l'utilisation mutualisée des points de charge ;
- le décret d'application de la loi n° 2014-877 du 4 août 2014 facilitant le déploiement d'un réseau d'infrastructures de recharge de véhicules électriques sur l'espace public, pour prescrire la publication des données ;
- L'article 7.7 de la Directive Européenne 2014/94 du 22/10 2014 sur le déploiement d'une infrastructure pour carburants alternatifs, en attente de déclinaison française, pour un accès ouvert et non discriminatoire aux données sur les bornes de recharge.

Des initiatives industrielles sont lancées en France et en Europe à cet effet : GIREVE, créée par la Caisse des Dépôts, la Compagnie Nationale du Rhône, EDF, ERDF et Renault, se propose comme « hub » d'échanges de données facilitant l'accès aux données statiques et dynamiques des bornes de recharge, évitant la multiplicité de connexions bilatérales. Deux autres plateformes équivalentes existent en Europe : e-clearing.net et Hubject.

Néanmoins, ces textes et ces développements industriels sont encore très préliminaires ; un nombre insuffisant d'acteurs politiques ou industriels impliqués dans le développement des VE sont conscients des enjeux de cette ouverture des données, ou tout simplement favorables eu égard à la recherche de rémunération par captation de marché. Des dispositions précises et opératoires doivent être prises.

L'intérêt collectif est de créer les conditions d'un accès universel aux données, avec un marché concurrentiel de services sans frontière, ouvert aux nouveaux entrants, et d'éviter la captation par un petit nombre d'acteurs dominants. Il faut sensibiliser les pouvoirs publics et les collectivités sur les dispositions qui doivent être prises à cet effet, à l'aube du déploiement massif de l'infrastructure de recharge.

—

Annexe 9 : Position de la SNCF



Monsieur Francis Jutand
Institut Mines-Télécom
46, rue Barrault
75634 Paris Cedex

Paris, le 12 janvier 2015

Monsieur le Président,

Vous avez souhaité, par votre lettre du 29 avril 2014, que SNCF participe au Comité du débat sur l'ouverture des données de transport. Lors des différentes réunions qui se sont tenues, nous avons partagé nos convictions et notre expérience de l'Open Data.

Ce débat offre l'occasion de poser les conditions pour dynamiser l'ouverture des données de transport, un sujet qui cristallise beaucoup d'attentes et de préoccupations chez les autorités organisatrices et les opérateurs de transports. Nous nous réjouissons de la tenue de ce débat et de la volonté publique d'ouverture, dont nous partageons l'intention. Au terme de cette consultation, il nous semble important de partager avec vous les éléments suivants.

SNCF est engagé dans l'Open Data depuis 2010 avec la diffusion de données et la mise à disposition d'API. Ces ressources – horaires théoriques Transilien, TER et Intercités, régularité des trains, liste des incidents de sécurité, description du réseau ferroviaire ou données théoriques et temps réel de plusieurs réseaux Keolis – sont accessibles à tous dans une posture globale d'Open Innovation et de transparence. SNCF a identifié une centaine de réutilisations de données dans des domaines variés (information voyageurs, cartographie, API agrégeant plusieurs sources de données, objets connectés, datajournalisme). Les API SNCF totalisent une moyenne de 4,5 millions d'appels par mois.

Au-delà des données, SNCF via sa filiale Keolis a ouvert en avril 2014 le code source de son moteur de calcul d'itinéraire Navitia afin de lever la principale barrière à l'entrée dans la réalisation de service d'itinéraire par des start-ups. L'Open Source représente pour nous un levier d'Open Innovation puissant.

De manière générale, nous apprécions que le débat se fasse au cas par cas avec les producteurs de données afin de définir, pour chaque ressource, s'il y a exigibilité d'ouverture, pour quelle finalité et sous quelles modalités. A ce titre, le concept d'« information d'intérêt général » nous semble intéressant puisqu'il

replaces le débat de l'ouverture dans une problématique économique qui transcende la nature publique ou non du producteur de données, en intégrant l'ensemble des acteurs présents dans le champ concurrentiel. Il est donc primordial que l'application éventuelle du principe d'« information d'intérêt général » aille de pair avec un traitement équitable des acteurs publics et des acteurs privés.

Si la loi devait introduire un principe d'ouverture des données de transport, il nous semble important que cette ouverture se limite strictement aux transports conventionnés et de veiller à ce que les producteurs de données conservent la relation directement avec les développeurs pour mieux répondre à leurs besoins et mesurer l'impact d'une démarche d'ouverture. Il est donc nécessaire que les producteurs de données conservent sur leurs plateformes la source des données et API – ce qui n'empêche évidemment pas que d'autres plateformes moissonnent les données comme le fait déjà Etalab.

Par ailleurs, cette limitation aux transports conventionnés ne nous interdit en rien à envisager une ouverture des données des services de transport concurrentiel. L'autonomie dans le choix et les modalités d'ouverture nous apparaît être un élément nécessaire pour créer les conditions d'un débat interne. Nous y travaillons activement.

La sécurité est une préoccupation constante pour SNCF. Nous partageons avec les autorités organisatrices l'objectif que l'ouverture des données ne se traduise pas, en cas de situation perturbée, par un report modal qui pourrait mettre en tension nos capacités à gérer l'affluence des usagers sur un réseau de transport, dans les gares comme dans les trains. A ce titre, il ne faudrait pas que l'information en temps réel vienne aggraver une situation perturbée par le report massif de voyageurs sur un autre mode de transport, ou par la fourniture d'une information inadaptée susceptible d'engager la responsabilité juridique de l'opérateur. D'autre part, certaines données comme la localisation et le taux de remplissage des trains peuvent faciliter des actes de malveillances. L'ouverture des données doit donc, dans chaque cas, s'évaluer au regard du risque potentiel que cette ouverture représente pour la sécurité publique.

Si l'ouverture des données nous semble souhaitable, il convient aussi de rappeler qu'elle n'est pas neutre financièrement. L'Open Data engendre des coûts de collecte, de mise en qualité et de mise à disposition sur un portail, incluant les serveurs qui fournissent les données et API ainsi que le fonctionnement et la maintenance des scripts permettant de convertir certaines données d'un format métier très technique vers un format compréhensible par un développeur.

Nous tenons par ailleurs à rappeler que plusieurs classes de données ne sont pas simples à collecter et à maintenir. C'est le cas, par exemple, de l'état en temps réel des équipements en gare, dont la donnée n'existe souvent pas.

Au cours du débat, il nous est apparu que la sous-estimation des coûts de l'ouverture est forte. L'Open Data ne doit pas se traduire par un simple report financier sur les producteurs de données, auquel cas nous irions très probablement vers des ouvertures déceptives, tant en termes de qualité des données que de fréquence de mise à jour. Poser la question du coût, c'est aussi se poser pour nous-même l'impératif d'avoir une démarche d'ouverture pérenne afin de fournir des données à un niveau de qualité satisfaisant.

SNCF se dirige ainsi vers un modèle d'ouverture des données par plateforme d'API sous modèle Freemium. Outre qu'un modèle Freemium nous permet de générer une valeur partagée avec les plus gros réutilisateurs, il lève les barrières à l'entrée pour les jeunes startups. Il s'agit d'un élément central pour dynamiser l'économie numérique.

Cette importance accordée aux modèles économiques témoigne aussi d'une nouvelle orientation que nous souhaitons donner à l'Open Data – vous l'avez également rappelé en Comité du débat. Notre ambition est de passer à l'étape suivante, celle d'une industrialisation qui puisse reprendre les bonnes pratiques observées dans l'économie numérique.

Cela suppose donc de bien spécifier les différents modèles d'affaires pour les réutilisateurs mais aussi pour les producteurs des données, tout en prenant activement en compte l'un des déterminants de l'économie numérique : la valeur des données, que cette valeur soit directe ou indirecte. La donnée est une ressource qui a une valeur d'usage. Il faut donc veiller à préserver pour les producteurs la possibilité de développer de la valeur avec les données au-delà du coût marginal.

Nous restons à votre disposition pour vous apporter toute précision complémentaire.



Yves Tyrode
Directeur général Digital et Communication

Copie :

- Monsieur le Président de SNCF
- Madame la Directrice générale Voyageurs
- Cabinet de Madame la Secrétaire d'État chargée du Numérique
- Cabinet de Monsieur le Secrétaire d'État chargé des Transports, de la Mer et de la Pêche

SOCIÉTÉ NATIONALE DES CHEMINS DE FER FRANÇAIS - R.C.S. BOBIGNY 5 522 019 447

—

Annexe 10 : Position de la FNMS



Ouverture des données Stationnement Comité du débat 23 septembre 2014

Présentation FNMS

La Fédération Nationale des Métiers du Stationnement réunit les exploitants privés et semi-privés de stationnement et les fabricants et distributeurs de matériel de péage et d'horodateurs, éditeurs de solutions logicielles de contrôle d'accès ou de paiement, bureaux d'études, cabinets de conseils spécialisés en ingénierie de la mobilité et du stationnement, architectes.

La FNMS rassemble plus d'une centaine d'entreprises des métiers du stationnement présentes en France et pour certaines dans le mode entier. Elle représente la profession et défend les intérêts professionnels de ses adhérents auprès des pouvoirs publics.

Le contexte

Le Gouvernement a souhaité donner un élan fort au mouvement des données publiques dans le cadre de la politique de modernisation de l'action publique.

Le Comité interministériel pour la modernisation de l'action publique (CIMAP) a réaffirmé le principe de la gratuité de la réutilisation des données publiques et décidé, pour favoriser l'ouverture des données publiques ayant un fort impact sociétal et un fort potentiel d'innovation sociale et économique, d'organiser une large concertation autour de débats thématiques au sein du Comité du Débat.

Le stationnement a été rattaché dans cette démarche aux nouveaux services de mobilité.

Terminologie

Il convient de préciser qu'une donnée publique peut, sous conditions, être mise à disposition pour une nouvelle utilisation par un tiers tandis qu'une donnée ouverte (open data) est une donnée numérique proposée librement et souvent sans contrepartie financière par celui qui la produit. En conséquence, il convient de bien distinguer ces deux types d'informations et leurs conditions de mise à disposition conditionnelles.

Le cadre réglementaire

L'ouverture des données s'inscrit dans un cadre législatif et réglementaire qui comprend notamment :

- Loi n°78-753 du 17 juillet 1978
- Directive 2003/98/CE du 17 novembre 2003
- Ordonnance n°2005-650 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques
- Décret n°2005-1755 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques

Depuis l'entrée en vigueur de l'ordonnance de 2005, modifiant la loi de 1978, les informations publiques sont en principe librement réutilisables. Cependant, l'article 10 de la loi de 1978 précise que ne constituent pas des informations publiques :

- les informations présentes dans les documents administratifs élaborés par l'Etat, les collectivités locales et les organismes publics ou privés chargés d'une mission de service public qui ne sont pas accessibles selon les termes de la loi du 10 juillet 1978. Cependant, si ces mêmes documents ont été diffusés publiquement, les informations qu'ils contiennent sont alors réputées réutilisables.
- Les informations qui participent à une mission de service public industriel et commercial
- Les informations présentes dans ces documents sur lesquels des tiers détiennent des droits de propriété intellectuelle, sous conditions d'accords de réutilisation exprimés par les tiers

Les professionnels du stationnement et l'ouverture des données

La production et la mise à disposition de données représentent des investissements et des charges d'exploitation qui doivent nécessairement s'intégrer dans l'équilibre économique de contrats d'exploitation principalement concessifs, de longue durée.

L'appartenance du stationnement au secteur concurrentiel constitue une contrainte majeure à l'ouverture des données au public dans la mesure où l'ouverture de données notamment dynamiques engendrerait une situation de distorsion de concurrence.

Les professionnels du stationnement, par leurs activités, produisent des informations créatrices de valeur. Leur ouverture pose le problème du modèle économique applicable à des informations à valeur ajoutée dans le cas d'un désintermédiation ou dans celui de leur réutilisation par des tiers les proposant à titre onéreux.

Les professionnels du stationnement peuvent garantir la fiabilité des informations qu'ils produisent mais ne peuvent garantir leur fiabilité lorsque celles-ci sont réutilisées par des tiers.

Le panorama des données envisagées couvre à la fois des données statiques et des données dynamiques notamment relative à la disponibilité des places dans les parcs de stationnement.

Les professionnels du stationnement sont favorables à l'ouverture des données statiques relatives à :

- la localisation des parcs de stationnement
- le nombre total de places de stationnement dans un parc
- les horaires d'ouverture du parc
- les services complémentaires proposés dans les parcs
- les modes de paiement

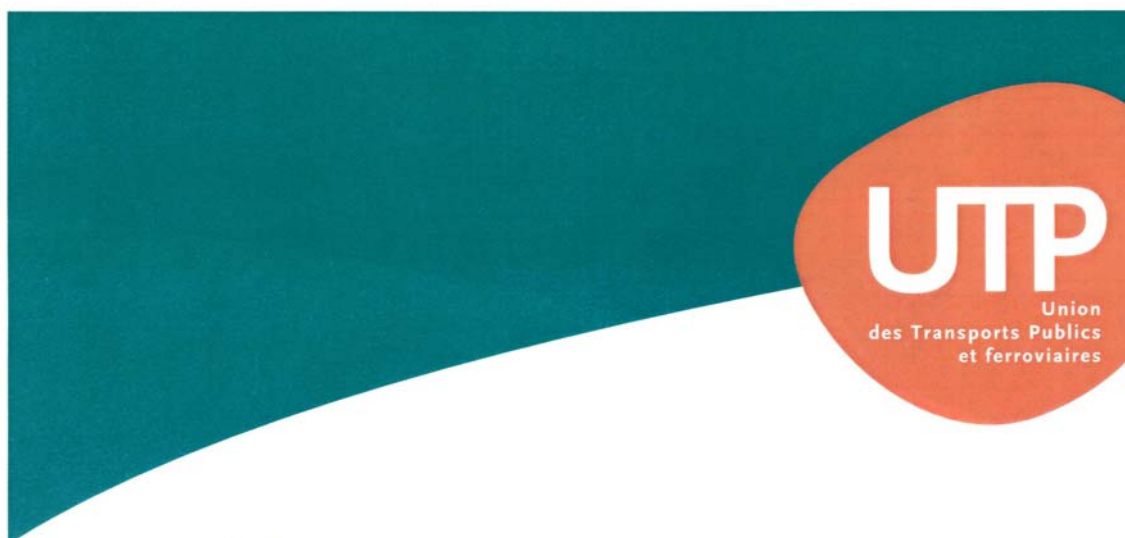
Chacun relève que ces informations sont d'ores et déjà rendues disponibles soit par les exploitants de stationnement eux-mêmes soit par des tiers sur des systèmes d'information en mobilité.

Les professionnels du stationnement seraient favorables à l'ouverture des données dynamiques dès lors que cette ouverture :

- n'engendre pas de situation de distorsion de concurrence
- ne fragilise pas l'équilibre économique des contrats si des investissements étaient rendus nécessaires et des charges d'exploitations venaient à être générées
- définit contractuellement les conditions d'ouverture et de réutilisation des données en termes de :
 - o modèle économique de la réutilisation
 - o fiabilité des données

En effet, les professionnels du stationnement estiment que la mise à disposition des informations relatives au stationnement permettrait une visibilité accrue sur les services proposés par les opérateurs du stationnement à la politique de la mobilité des collectivités locales.

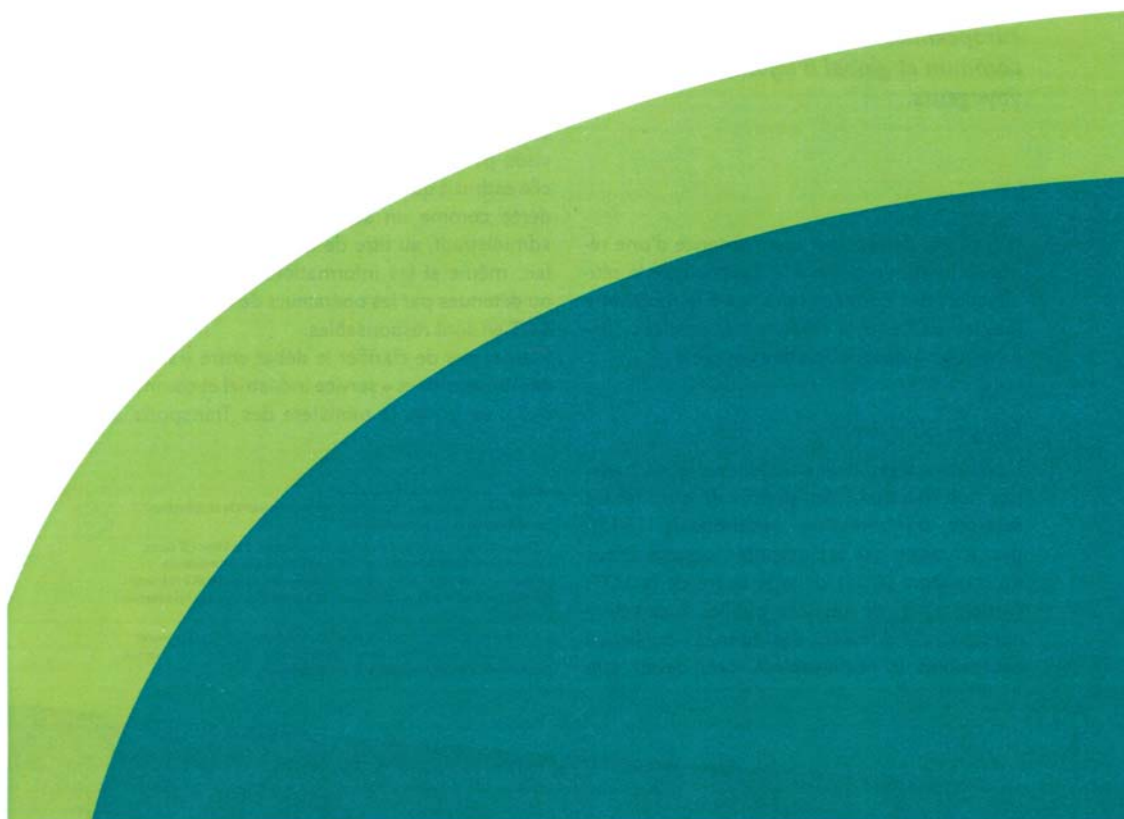
—



Position

Enjeux de l'ouverture des données Transport
pour le secteur du transport public

adoptée le 3 juillet 2014





La question de la donnée n'est pas nouvelle et figure dans l'agenda de l'UTP depuis le milieu des années 2000. Aujourd'hui, elle revient plus particulièrement sur le devant de la scène avec la décision du Gouvernement de favoriser l'ouverture des données publiques dans le domaine des transports et, tout récemment, avec les discussions devant les institutions européennes en faveur d'un système commun et global d'informations voyageurs.

Il convient de préciser que l'absence d'une réponse juridique claire à la question de la réutilisation des données relatives à la mobilité a longtemps freiné la création des services d'information à destination des voyageurs.

Rappel historique

L'évolution législative européenne¹ et sa transposition nationale² ont permis de qualifier les services d'information multimodale (SIM), mis en place par les autorités organisatrices de transport (AOT) dans le cadre de la LOTI (Article 27-1), de services publics à caractère administratif diffusant des données publiques accessibles et réutilisables³. Ceci devait être

de nature à lever les obstacles juridiques alors même que le transport public est un service industriel et commercial (SPIC).

Toutefois, en juillet 2008, la plate-forme de recherche et d'expérimentation pour le développement de l'information multimodale (PREDIM), émanation du ministère des Transports, a rédigé une première analyse juridique qui a conduit à s'interroger sur la qualification de «données publiques» pour les informations produites ou détenues par les AOT et leurs opérateurs. La PREDIM en a déduit que les informations concernant les transports publics entraient bien dans la catégorie des documents administratifs visés par l'ordonnance de 2005. Tout comme elle estimait que l'information devait être considérée comme un service distinct, par nature administratif, au titre de la LOTI, et que, de ce fait, même si les informations sont produites ou détenues par les opérateurs de transport, les AOT en sont responsables.

Pour tenter de clarifier le débat entre «service administratif» et «service industriel et commercial», en 2009, le ministère des Transports a

1. Directive 2003-98 du 17 novembre 2003 concernant la réutilisation des informations du secteur public.

2. Ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques et Décret n° 2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques.

3. Au sens de la loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal.

saisi la Commission d'Accès aux Documents Administratifs (CADA), qui s'est déclarée non compétente.

Dans son «Manifeste de la Mobilité durable», publié en juin 2010, l'UTP a fait de l'information voyageurs un des axes d'interpellation de la profession vers les autorités organisatrices de transport urbain (AOTU). Ainsi, l'UTP préconise *«que les AOTU intègrent dans leur cahier des charges une composante pédagogique de valorisation de l'offre locale. Il conviendrait en particulier qu'elles exigent la mise à disposition des données théoriques (horaires, localisation des arrêts), voire en temps réel, pour permettre de construire des services d'information plus accessibles, simples, homogènes.»*

L'une des premières actions de l'Agence Française pour l'Information Multimodale et la Billettique (AFIMB), créée en juillet 2010 dans le cadre du Grenelle de l'Environnement, fut de ressaisir la CADA sur le statut des données publiques émises notamment par des SPIC. Dans sa réponse du 21 février 2011, la CADA considère que les données collectées par les AOT dans le cadre de la mission de développement de l'information revêtent le caractère d'informations publiques, sauf celles dûment protégées par le secret des affaires ou propriété d'un tiers.

De son côté, au cours de l'année 2011, l'UTP a organisé une série de conférences techniques sur l'information voyageurs, en partenariat avec les Ingénieurs et Scientifiques de France (IESF), qui a abouti à une publication commune présentée au Salon européen de la mobilité de juin 2012. Le document «Informer le voyageur: agir ensemble» vise la mise en

place de systèmes d'informations dynamiques multi-modes et multi-réseaux à l'échelle des bassins de mobilité et décline dix pistes d'actions concrètes pour améliorer la mobilité quotidienne. Une des pistes énoncées par l'UTP et les IESF interpelle très clairement l'AFIMB sur la question de la propriété des données et de leur mise à disposition.

Les opérateurs, propriétaires des données qu'ils produisent

Afin de clarifier les rôles de chacun, l'AFIMB a constitué, fin 2011, un groupe de travail sur la réutilisation des données, réunissant des représentants des AOT, des opérateurs et des sociétés de service ainsi que l'Agence du Patrimoine Immatériel de l'État (APIE). Il s'agissait notamment de trouver une réponse à l'une des préconisations de l'UTP et des IESF qui énonce que *«sans attendre les décisions de Bruxelles, [que] l'AFIMB clarifie le statut des données publiques et les conditions de leur réutilisation»*. C'est ainsi que ce groupe de travail a tenté de traiter des problématiques de réutilisation des données, du caractère public des données transport et des questions juridiques autour du «service public d'information».

Face aux divergences d'interprétation des membres du groupe de travail, les services du ministère des Transports ont été sollicités pour un examen juridique. Ils ont conclu que les données transport n'étaient pas réutilisables de droit.

Cette analyse vient valider la position constante de l'UTP, à savoir que l'ouverture des données passe par le duo «AOT/opérateur de transport».

Les opérateurs, prêts à jouer le jeu de l'ouverture des données

En parallèle, depuis 2010, des initiatives de libération de données dans le domaine de l'information transport sont prises par certaines AOTU et opérateurs de transport pour permettre à des tiers de les réutiliser. Citons quelques précurseurs comme Rennes et son opérateur de transport urbain, Bordeaux, Montpellier, Nancy, Nantes, Paris, Toulouse, SNCF Transilien et TER, la RATP...

Cette libre décision d'ouverture des données vise à faciliter l'émergence de nouveaux services à disposition du grand public pour favoriser le report modal vers les modes alternatifs à la voiture particulière, mais aussi une meilleure intégration des transports publics dans l'univers quotidien des citadins et dans l'accès aux aménités urbaines.

Le double mouvement actuellement observable avec, d'une part, des démarches Open Data qui se développent localement et, d'autre part, le souhait du Gouvernement d'ouvrir et de partager les données publiques⁴, semble tendre vers un objectif partagé : une large ouverture des données pour développer des services innovants. Toutefois, il ne facilite pas la compréhension des acteurs en présence et pose un certain

nombre d'interrogations qu'il convient de lever. Le lancement, le 27 mai 2014, du Comité du débat national sur l'ouverture des données publiques dans le domaine des transports⁵ est une bonne opportunité pour clarifier les enjeux pour notre profession.

Nos préconisations

Tendre vers une définition des données transport

En préalable, il convient de définir précisément quelles sont les données à ouvrir, au titre des données publiques, et selon quelles priorités.

Si, par exemple, pour les opérateurs et les AOT, les horaires théoriques des services entrent dans ce champ, il n'en va pas forcément de même pour des données enrichies, comme les données en temps réel ou celles issues des SIM...

Par ailleurs, certaines données paraissent relever de la stricte exploitation, comme le taux de remplissage ou les constituants de l'offre de transport, ou d'autres de stratégies commerciales, comme la tarification et ses composants, et à ce titre, elles ne peuvent être diffusées.

4. Dans sa feuille de route en matière d'ouverture et de partage des données publiques, publiée le 28 février 2013, le Gouvernement a confirmé «souhaiter élargir cette stratégie à tous les établissements publics administratifs, aux délégataires de service public, aux collectivités territoriales et aux autorités administratives indépendantes mais également aux missions de service public à caractère industriel et commercial, aux entreprises publiques et à la sphère associative.»

5. L'UTP y est conviée, tout comme le GART et la FNMS. Les recommandations devront être remises au Premier ministre en novembre 2014.

Élaborer un cadre clair entre AOT et opérateur

Il est indispensable que les AOT et les opérateurs s'organisent pour préciser le cadre de la mise à disposition et de la réutilisation des données, comme cela est déjà le cas dans un certain nombre de réseaux, notamment pour « alimenter » des systèmes partenaires comme les SIM développés par les Régions. Par voie contractuelle ou accord formalisé, il convient de fixer les conditions de fourniture et de redistribution des données produites ou détenues par l'opérateur : type des données ouvrables, mise au format, qualité, mise à jour, restrictions d'usage, compensations financières...

La sous-estimation des coûts de l'ouverture des données est forte, notamment de la part des tiers intéressés. Ainsi, la question de la mise au format ne doit pas être éludée car il peut s'agir d'une opération coûteuse, tout comme les frais des serveurs d'hébergement et de sécurité.

De plus, si l'enjeu de la mutualisation des données entre acteurs publics est important pour le développement de la mobilité durable, il convient de rester vigilant dans les relations entre les opérateurs et les différents niveaux d'AOT. Le droit d'usage des données transmises par un exploitant est limité à son AOT. Le principe de réutilisation valant pour un usage premier, les données des opérateurs restent soumises à une autorisation préalable de leur part, en cas de projet d'ouverture à des tiers, y compris à d'autres AOT.

En parallèle, il importe également d'encourager les opérateurs de transport, propriétaires des données, à poursuivre leurs opérations de libération des données et, à ce titre, à définir des politiques de diffusion ou de partenariat avec des développeurs.

Conserver la maîtrise des données

Un des enjeux de la profession est la qualité et la fiabilité des données. Aussi, l'ouverture des données doit se faire de façon maîtrisée. Actuellement, toutes les parties prenantes (AOT, opérateurs et représentants des usagers) s'accordent sur la nécessité d'ouvrir les données théoriques dans une première étape, même si le risque peut exister d'une prise de contrôle de l'information voyageurs par les acteurs de l'Internet. La question des données en temps réel soulève plus de débats. Il ne faudrait pas que l'information en temps réel vienne, par exemple, aggraver une situation perturbée ou que le voyageur se voit délivrer une information inadaptée susceptible d'engager la responsabilité juridique de l'opérateur.

Il convient également de maintenir les exigences de sécurité tant sur l'accès aux données que sur les systèmes d'exploitation.

Il est impératif que les AOT et leurs opérateurs restent les référents des clients-voyageurs. À ce titre, sachant qu'il est difficile d'éviter l'édition de services défaillants, la mise en place d'une politique de label pour les meilleures réutilisations offrirait une garantie de qualité.

Jouer la complémentarité entre « données brutes » et « Open Services »

En matière d'utilisation des données, deux démarches sont adoptées :

- soit la mise à disposition de jeux de données brutes telles que les horaires théoriques, pour générer l'émulation et la créativité des développeurs en faveur de nouveaux services innovants,

- soit l'accès à des systèmes d'informations structurés (services web ou Open Services), élaborés par l'AOT à partir des données collectées auprès des opérateurs, tels qu'un calculateur d'itinéraires. Ce choix d'un mode de diffusion élargie de services est, par exemple, celui du Syndicat des Transports d'Île-de-France (STIF), qui met à disposition des sociétés commerciales son module de recherche d'itinéraires.

L'accès aux données via des interfaces de programmation d'applications (API)⁶ permet aux producteurs d'en conserver la maîtrise, de contrôler leur actualisation et leur usage de bout en bout. Cette approche favorise également l'enrichissement des services.

Promouvoir une licence partagée entre les AOT et les opérateurs

Il est impératif de travailler conjointement sur les licences de réutilisation, pour garantir la protection des intérêts des parties.

Ces licences doivent :

- conditionner l'utilisation des données aux exigences du secteur des transports et dans le respect des missions de service public,
- clarifier les questions de mise à jour des données,
- définir les principes tarifaires qui permettent de limiter les risques de captation des données par les tiers non autorisés, comme par exemple les groupes internationaux omniprésents dans la sphère numérique,

- protéger les opérateurs et les AOT, ainsi que les voyageurs, de toute mauvaise redistribution des données.

Limiter le principe de gratuité

Un des arguments de la gratuité des données publiques tient au fait que ces données sont déjà payées par les citoyens via les tarifs des services ou leurs impôts nationaux ou locaux. Cette considération est à prendre en compte mais il ne faut pas faire preuve de naïveté en la matière, compte tenu des retombées économiques de certaines applications ou services élaborés à partir de ces données et que les utilisateurs paieront, même indirectement (publicité, démarchage...).

L'UTP s'inscrit dans les préconisations formulées dans le rapport sur l'ouverture des données publiques⁷. C'est ainsi que la profession souhaite se placer dans une logique d'ouverture des données tendant vers la gratuité, notamment pour certains publics (étudiants, chercheurs, associations sans but lucratif, etc.)

6. En février 2012, l'UTP, en association avec trois sociétés de services et avec le soutien financier de la DGTIM, a édité des « Spécifications communes d'une interface (API) de recherche d'itinéraire en transports collectifs ». Cette étude, en amont de la normalisation, devait tendre vers la définition d'un protocole d'échanges de données entre SIM et favoriser le développement d'applications et de services, dotés de fonctions de calcul d'itinéraires de transports publics.

7. Mohammed Adnène Trojette, Ouverture des données publiques. Les exceptions au principe de gratuité sont-elles toutes légitimes ?, Rapport au Premier ministre, juillet 2013.

et avec rémunération pour service rendu pour les sociétés qui retireront une valeur ajoutée dans la réutilisation de ces données.

La question du prix de mise à disposition rejoint celle de la licence. Aussi, compte tenu de la diversité des jeux de données et de leurs usages, différentes modalités de redistribution et de rémunération devront être recherchées. Il est, en effet, important de trouver un modèle économique sur la réutilisation des données qui ne vienne pas accroître les charges du budget des transports publics.

Forts de ces préconisations, les opérateurs et leurs représentants sont prêts à s'engager dans une démarche proactive, en partenariat avec les AOT, les associations d'usagers et les autres parties prenantes afin de favoriser le développement de la valorisation numérique des informations qu'ils détiennent au service de la mobilité durable et citoyenne.

Petites précisions sémantiques sur l'ouverture des données

- Une donnée qualifiée de « publique » peut être mise à disposition de tiers pour une réutilisation sous conditions d'usage (ne pas l'altérer et en dénaturer le sens, ne pas porter atteinte à l'intérêt du fournisseur), avec possibilité de redevance.
- Une donnée largement diffusée n'est pas, de fait, une donnée « publique ».
- Une « open data » ou « donnée ouverte » est une donnée numérique, « libérée » par un producteur (c'est lui qui décide et choisit ce qu'il donne) sans sollicitation de tiers, souvent gratuitement et réutilisable librement par tous, sous réserve d'en respecter la paternité.

Par facilité d'usage, aujourd'hui, on qualifie d'open data le principe de rendre accessible au citoyen certaines données publiques, avec ou sans licence d'utilisation.

L'UTP

L'Union des Transports Publics et ferroviaires (UTP) est l'**organisation professionnelle regroupant les entreprises de transport public, les entreprises ferroviaires (fret et voyageurs) et les gestionnaires d'infrastructure en France.**

Elle représente la profession et défend les intérêts collectifs de ses adhérents auprès des institutions françaises et européennes.

L'UTP représente plus de 160 entreprises de transport urbain réparties sur le territoire français.

Certaines sont liées à des groupes de transport comme Car Postal, Keolis, Groupe RATP, SNCF Proximités, Transdev, Vectalia France. D'autres sont indépendantes et peuvent, à ce titre, adhérer à l'association AGIR.

Depuis 2006, l'UTP fédère les entreprises ferroviaires et les accompagne vers l'ouverture du marché du transport de voyageurs. Il s'agit notamment d'Euro Cargo Rail, Europorte, Eurostar International, Keolis, Groupe RATP, SNCF, Thello, Transdev, VFLI.

Depuis janvier 2013, l'UTP accueille également des gestionnaires d'infrastructure (Eurotunnel, Liséa, RFF) et incarne l'unité de la branche ferroviaire.

Contacts

Anne Meyer - Département affaires économiques et techniques
ameyer@utp.fr Tél. : +33 (0)1 48 74 73 28

Dominique Fèvre - Département Valorisation et Communication
dfevre@utp.fr Tél. : +33 (0)1 48 74 73 46

UTP
17, rue d'Anjou – 75008 Paris

Tél. : +33 (0)1 48 74 63 51
Fax : +33 (0)1 40 16 11 72

www.utp.fr



Impression: Imprimé sur papier Symbol Freefile certifié FSC - Imprimerie LFT 93100 Montreuil - Juin 2014

