

# ÉTAT DES LIEUX DE L'OUVERTURE ET DE L'UTILISATION DES DONNÉES DE MOBILITÉ



> Juin 2022

**Crédits photos couverture :**

Shopify, le point d'accès national aux données de transports (<https://transport.data.gouv.fr/>),  
Handimap, A. Gougoua.

# Sommaire

# 03

## SYNTHÈSE 05

### 01

#### Une ouverture des données de mobilité pour fournir des services d'information multimodale (SIM) aux voyageurs 09

La directive européenne relative aux systèmes de transports intelligents (STI) pose l'innovation dans les transports, notamment le développement des SIM, comme une solution pour limiter la congestion routière et la consommation d'énergie

La publication des données de mobilité requise par le règlement européen délégué 2017/1926 est nécessaire au déploiement de SIM

L'Autorité de régulation des transports a été désignée par le législateur pour contrôler le respect des exigences relatives à l'ouverture des données d'intérêt collectif

### 02

#### État des lieux de la publication des données de mobilité 17

Les données statiques relatives aux transports en commun représentent une forte majorité des données de mobilité ouvertes sur le point d'accès national

Deux formats sont utilisés pour publier les données statiques relatives aux transports en commun

La lisibilité et la qualité des données publiées sont essentielles à leur bonne utilisation, elles figurent donc parmi les priorités fixées par l'Autorité de régulation des transports

L'évaluation de la qualité des données par les producteurs et les utilisateurs de données nécessite de s'appuyer sur des outils informatiques adaptables aux différents formats

Les premiers tests automatiques réalisés témoignent d'un relatif respect des exigences de publication et de qualité de la part des producteurs

#### Les marchés aval de l'information sur les déplacements et la circulation : étude comparative des calculateurs d'itinéraires 41

L'utilisation des données de mobilité publiées sur le point d'accès national aux données (PAN) s'opère sur des marchés multiples et segmentés

La nature des exigences en matière de réutilisation de données et la diversité des acteurs concernés nécessitent d'examiner le fonctionnement des services fournis par ces derniers

L'étude des calculateurs révèle des fonctionnements hétérogènes, reposant sur des paramètres et des fonctionnalités variés et partiellement opaques

L'analyse du marché et du fonctionnement des calculateurs, au travers du prisme de la transparence et la non-discrimination, permettra d'orienter l'action de l'Autorité de régulation des transports dans l'exercice de ses missions

#### Glossaire 57

## ANNEXES

#### L'utilisation des formats de données 62

Exemple de lecture d'un jeu de données GTFS et NeTEx 62

Structuration d'un jeu de données GTFS 66

#### Précisions sur l'état des lieux de l'ouverture des données 68

Couverture géographique des données statiques des transports en commun selon le standard utilisé 68

Les jeux de données GTFS incluent fréquemment des fichiers « facultatifs » 68

#### Des instructions ex ante et/ou ex post : l'exemple de Vianavigo 70



## Synthèse

**En favorisant un usage mieux coordonné et plus sûr des différents modes de transport, le numérique constitue une des solutions permettant de réduire la congestion routière et la consommation d'énergie, et ainsi de limiter l'incidence de l'usage des transports sur l'environnement.**

Afin de permettre le déploiement de tels services numériques de transport (ou services de transports dits « intelligents »), la directive 2010/40/UE du Parlement européen et du Conseil du 7 juillet 2010<sup>1</sup> identifie plusieurs actions prioritaires, dont le développement de services d'information destinés aux utilisateurs des transports, appelés « services d'information multimodale » (SIM). Ces services visent à fournir des informations notamment sur les itinéraires permis par les différents modes de transport (ferroviaire, aérien, autocars...) et sur le trafic que les utilisateurs des transports peuvent être amenés à rencontrer sur leur trajet.

L'accès à des données de mobilité (telles que les arrêts desservis, les horaires de passage des transports collectifs ou les déplacements effectués) constituant une des conditions de réussite du déploiement des SIM, la Commission européenne a précisé les données concernées, ainsi que les modalités de mise à disposition et d'accès à ces données par son règlement délégué (UE) 2017/1926 du 31 mai 2017<sup>2</sup>. Celui-ci prévoit :

- pour les producteurs de données de mobilité, l'obligation de mettre à disposition les données dont ils disposent sur un point d'accès unique, suivant des formats déterminés et avec des conditions d'utilisation aussi peu restrictives que possible ;
- pour les utilisateurs de ces données, l'obligation de respecter différentes conditions d'utilisation.

Ces obligations visent d'une part à permettre aux fournisseurs de SIM un accès effectif aux données de mobilité qui constituent une **ressource essentielle pour le développement des SIM** et d'autre part à assurer la **fourniture aux voyageurs d'une information claire, complète et fiable**.

Dans ce cadre, l'Autorité de régulation des transports (« Autorité ») s'est vu confier, par la loi n° 2019-1528 du 24 décembre 2019 d'orientation des mobilités (LOM), la mission de :

- contrôler la conformité des publications de données effectuées par les acteurs soumis à cette obligation ;
- contrôler les modalités de réutilisation de ces données par les acteurs présents sur les marchés aval ;
- publier un rapport biennal portant sur ces contrôles.

L'Autorité devient ainsi le régulateur sectoriel de l'accès aux données de mobilité.

---

<sup>1</sup> Directive 2010/40/UE du Parlement européen et du Conseil du 7 juillet 2010 concernant le cadre pour le déploiement de systèmes de transport intelligents (STI) dans le domaine du transport routier et d'interfaces avec d'autres modes de transport (« directive 2010/40/UE » dans la suite du rapport).

<sup>2</sup> Règlement délégué (UE) 2017/1926 de la Commission du 31 mai 2017 complétant la directive 2010/40/UE du Parlement européen et du Conseil en ce qui concerne la mise à disposition, dans l'ensemble de l'Union, de services d'informations sur les déplacements multimodaux (« règlement délégué 2017/1926 » dans la suite du rapport).

**Le présent rapport, qui constitue la première publication de l'Autorité sur l'ouverture et l'utilisation des données de mobilité, établit un premier état des lieux de la situation, notamment une analyse des données disponibles au 1<sup>er</sup> décembre 2021 et de leur conformité sur la base de premiers tests de contrôle.**

Dans le cadre de ce premier état des lieux, l'Autorité a examiné les données de mobilité statiques, décrivant notamment les trajets disponibles par autocar ou par train (par opposition aux données dynamiques ou temps réel) qui constituent la priorité du règlement délégué 2017/1926<sup>3</sup>, et en particulier celles concernant les transports en commun (par opposition aux données de transport à la demande et de déplacement personnel).

**L'Autorité a développé une méthodologie de contrôle visant à vérifier d'une part que l'ouverture des données est effective et conforme aux standards de qualité requis (amont) et d'autre part que les utilisateurs respectent les conditions de réutilisation des données mises à disposition (aval).**

La méthodologie de contrôle de l'ouverture des données par les producteurs s'établit en trois niveaux et vise à vérifier le respect des obligations de qualité et d'interopérabilité, condition indispensable à une réutilisation effective des données :

- le premier niveau vise à évaluer l'accès effectif à la donnée au travers des critères d'accessibilité (présence de métadonnées), de normalité (respect des formats requis) et d'actualité (publication d'un jeu à jour) ;
- le second vise à évaluer la qualité du jeu de données, et notamment sa complétude (présence des informations requises pour chaque jeu de données disponible), sa pertinence (données correctes) et sa cohérence (absence de données en double ou inutiles) notamment ;
- enfin, un troisième niveau vise à s'assurer de l'exhaustivité (présence de l'ensemble des données de mobilité existantes) et de l'exactitude des données au regard de la réalité qu'elles décrivent.

Ce premier état des lieux montre plus particulièrement les travaux sur les deux premiers niveaux de contrôle qui peuvent faire l'objet de contrôles automatisés, répliquables à grande échelle. Ces contrôles s'appuient sur un outil informatique dénommé « validateur » développé en propre qui permet de lire de manière automatisée les données déposées sur le point d'accès national aux données (PAN).

Sur le marché aval très fragmenté des utilisateurs de données, l'Autorité a concentré ses travaux sur les calculateurs d'itinéraires et s'est attachée à vérifier le respect par ces acteurs des exigences relatives à la réutilisation des données. La méthodologie de contrôle comporte également trois niveaux :

- le premier niveau vise à s'assurer d'une réutilisation transparente des données (via l'indication de critères de classement, des sources, etc.) ;
- une fois les critères de classement identifiés, le second niveau vérifie la neutralité (absence d'offre commerciale favorisée) et la non-discrimination (absence d'utilisation de l'identité de l'utilisateur pour établir la réponse à la requête formulée) ;
- enfin, le troisième niveau permet d'évaluer la conformité d'une réutilisation « sans biais » (traitement différencié des situations ou des catégories d'individus par l'algorithme) des données du PAN.

---

<sup>3</sup> L'article 2 du règlement délégué 2017/1926 définit les métadonnées comme « une description structurée du contenu des données qui aide à les retrouver et à les utiliser ». Elles correspondent aux données d'informations telles que la date de publication et de dernière mise à jour des données, l'identité du producteur, etc.

**Les données statiques de mobilité publiées sur le PAN et soumises aux tests de conformité de l'Autorité montrent une relative conformité des données avec les exigences d'accessibilité et de qualité requises.**

L'analyse des jeux de données recensés sur le PAN fait tout d'abord apparaître quelques caractéristiques notables :

- les autorités organisatrices de la mobilité (AOM) régionales constituent les principaux contributeurs, avec près de la moitié des jeux de données disponibles, suivies des collectivités et AOM locales (28 %) ;
- les données mises à disposition couvrent une très large part du territoire français avec des données déposées par toutes les régions métropolitaines, trois régions et une collectivité d'outre-mer notamment ;
- les jeux statiques représentent 88 % des jeux de données déposés sur le PAN ;
- la grande majorité des jeux de données (83 %) a été déposée selon un unique format<sup>4</sup>.

Si le contrôle des métadonnées effectué par l'Autorité atteste de **la bonne description des jeux de données, l'accès effectif à la donnée n'est pas complètement assuré en raison d'une insuffisante mise à jour des données** (pour 32 % des producteurs). Le format des jeux de données apparaît en revanche conforme au minimum requis pour pouvoir être réutilisé.

**Les jeux de données évalués sont pour une large majorité complets, comportent très peu de valeurs incorrectes et peu de valeurs incohérentes. Une faible proportion présente néanmoins une qualité très nettement inférieure à la moyenne.** Ainsi, seules 8,4 % des cellules ne sont pas renseignées, et un nombre marginal de valeurs sont incorrectes (0,1 % du total des cellules remplies). Même si seuls 25 % des jeux de données ne présentent aucune non-conformité, la très grande majorité des problèmes constatés dans les autres jeux de données s'avèrent en pratique mineurs. Pour la grande majorité des données examinées, la cohérence est également très satisfaisante. Les incohérences apparaissent ainsi concentrées dans certains jeux de données qui tendent à cumuler plusieurs types d'incohérences et peuvent dans certains cas présenter près de 90 % de valeurs incohérentes.

**Sur le marché aval, très segmenté, des services d'information multimodale, l'Autorité a constaté que n'étaient pas systématiquement respectées les règles de réutilisation et en particulier l'obligation de classer de manière transparente les propositions d'itinéraires fournies aux voyageurs.**

L'observation de l'écosystème des utilisateurs des données disponibles sur le PAN<sup>5</sup> montre une **réutilisation des données de mobilité qui se fait sur plusieurs marchés, relativement segmentés**. À l'image des services de transport, eux-mêmes gérés par différents acteurs, les calculateurs d'itinéraires des SIM – catégorie sur laquelle les contrôles de l'Autorité se sont plus spécifiquement portés – présentent une forte fragmentation, ainsi que des fonctionnalités et paramétrages très divers. Les résultats fournis en réponse à une requête d'utilisateur varient en conséquence fortement, en fonction des périmètres respectifs des SIM (services et modes couverts, périmètres géographiques), des spécificités pouvant être renseignées par les

---

<sup>4</sup> Il s'agit du format General Transit Feed Specification (« GTFS »), largement utilisé au niveau mondial par les acteurs du secteur des données de mobilité. L'autre format utilisé pour déposer les données est le format Network Timetable Exchange (« NeTEx »), défini et maintenu au niveau européen. Voir glossaire.

<sup>5</sup> Le règlement délégué 2017/1926 dresse une liste non-exhaustive des utilisateurs en les définissant comme « une entité publique ou privée qui utilise le point d'accès national, telle que les autorités chargées des transports, les opérateurs de transport, les fournisseurs de services d'informations sur les déplacements, les producteurs de cartes numériques, les fournisseurs de services de transport à la demande et les gestionnaires d'infrastructure » (point 11 de l'article 2).

usagers (options, critères de recherche, formulation de la recherche notamment), mais aussi des instructions et règles suivies par les algorithmes au sein des calculateurs.

**L'étude des résultats de recherche d'itinéraires questionne tant sur le degré de transparence que sur les critères de classement des calculateurs.** Certains des critères visibles ne peuvent en effet être qualifiés de transparents du fait de leur faible intelligibilité (tels que l'indication d'itinéraires « recommandés »). L'Autorité a également pu identifier l'existence de critères invisibles : certains de ces critères peuvent être déduits par l'analyse statistique des résultats fournis par les calculateurs des SIM, tandis que d'autres restent inaccessibles pour l'utilisateur – contrevenant ainsi directement au principe de transparence, essentiel à la fourniture d'un service d'information fiable et de qualité.

\*\*\*

Le présent rapport rappelle d'abord le cadre réglementaire applicable et notamment les obligations relatives à la publication de données et à leur utilisation (section 1). Il présente ensuite un état des lieux de la publication des données, les méthodologies et outils qui seront mobilisés pour réaliser les contrôles relatifs à l'ouverture des données et les premiers résultats obtenus (section 2). Enfin, une première analyse de l'utilisation de données et du respect des obligations qui incombent aux utilisateurs, est faite à travers le cas des calculateurs d'itinéraires (section 3).

## 1 UNE OUVERTURE DES DONNEES DE MOBILITE POUR FOURNIR DES SERVICES D'INFORMATION MULTIMODALE (SIM) AUX VOYAGEURS

### 1.1 La directive européenne relative aux systèmes de transports intelligents (STI) pose l'innovation dans les transports, notamment le développement des SIM, comme une solution pour limiter la congestion routière et la consommation d'énergie

Afin de répondre à l'augmentation du transport routier dans l'Union européenne - pouvant provoquer la congestion des infrastructures routières et générer des externalités négatives d'ordre environnemental notamment - la directive 2010/40/UE du Parlement européen et du Conseil du 7 juillet 2010 concernant le cadre pour le déploiement de systèmes de transport intelligents dans le domaine du transport routier et d'interfaces avec d'autres modes de transport (ci-après « la directive 2010/40/UE »), prévoit la mise en œuvre d'un ensemble d'actions destinées à favoriser le déploiement de « Systèmes de Transport Intelligents » (STI).

Ces systèmes, définis comme des applications avancées<sup>6</sup> permettant la fourniture de services innovants, visent à permettre à leurs utilisateurs d'être mieux informés et de faire un usage plus sûr et plus coordonné des réseaux de transport. **Ils constituent ainsi un moyen de limiter la congestion routière, et plus largement de favoriser des déplacements moins carbonés.**

Les STI ont historiquement été déployés dans le secteur du transport routier de voyageurs et de marchandises en s'appuyant sur des spécifications techniques hétérogènes, limitant leur bonne interopérabilité et leur continuité géographique dans l'ensemble de l'Union européenne.

Afin de favoriser cette interopérabilité, la directive 2010/40/UE met en avant la nécessité de :

- Concevoir et mettre en œuvre des spécifications ou des normes techniques régissant le fonctionnement des STI ;
- Rendre ces spécifications et normes accessibles à tous les fournisseurs et utilisateurs de STI.

Ce besoin de spécifications et de normes concerne quatre domaines prioritaires<sup>7</sup> :

1. **L'utilisation optimale des données relatives à la route, à la circulation et aux déplacements (par exemple pour permettre aux usagers de la route de planifier des trajets) ;**
2. La continuité des services STI de gestion de la circulation et du fret (c'est-à-dire des services ininterrompus lorsque les camions traversent les frontières) ;
3. Les applications de STI à la sécurité et à la sûreté routières (par exemple, une alerte en cas de risque de visibilité réduite ou de présence de personnes, d'animaux ou de débris sur la route) ;
4. Le lien entre le véhicule et les infrastructures de transport (par exemple, l'équipement des véhicules pour permettre l'échange de données ou d'informations).

---

<sup>6</sup> Pour plus de précisions, voir glossaire à la fin du rapport.

<sup>7</sup> Domaines définis à l'article 2 de la directive 2010/40/UE.

Six actions prioritaires contribuent, pour ces quatre domaines, à l'élaboration et à l'utilisation de spécifications et de normes<sup>8</sup>, trois d'entre elles visant explicitement la mise à disposition de services d'informations sur la circulation et les déplacements :

- Favoriser le développement, dans l'ensemble de l'Union, de services d'informations sur les déplacements multimodaux, c'est-à-dire concernant plusieurs types de transport (par exemple, le transport ferroviaire, routier, etc.), nommés « Services d'Information Multimodale » en France ou « SIM » ;
- Favoriser le développement, dans l'ensemble de l'Union, de services d'informations en temps réel sur la circulation ;
- Favoriser le développement de moyens permettant de fournir gratuitement aux usagers des informations minimales universelles liées à la sécurité routière.

Pour mettre en œuvre chacune de ces actions, la Commission européenne a adopté des règlements délégués complétant la directive 2010/40/UE<sup>9</sup>. Ainsi, pour la mise en œuvre de l'action relative aux SIM, le règlement délégué 2017/1926 (UE) de la Commission du 31 mai 2017<sup>10</sup> (ci-après le « règlement délégué 2017/1926 ») prévoit d'assurer **une ouverture généralisée des données existantes sur les déplacements et la circulation** (ci-après « données de mobilité ») comme les horaires des bus, les travaux ou accidents routiers, la disponibilité de vélos en libre-service.

## 1.2 La publication des données de mobilité requise par le règlement délégué 2017/1926 est nécessaire au déploiement de SIM

**Les Services d'Informations Multimodales (SIM) fournissent des informations sur les déplacements et la circulation, couvrant plusieurs modes de transport et permettant la comparaison entre ces modes<sup>11</sup>.** Ils s'appuient sur des cartes numériques<sup>12</sup> et, le plus souvent, sur une fonctionnalité de recherche d'itinéraires. **Leur fonctionnement repose sur le traitement de données statiques (ou théoriques) ou dynamiques (en temps réel) de mobilité.** Par exemple, la fourniture d'un itinéraire en transports en commun nécessite le traitement de données relatives à la topographie du réseau de transport (localisation des arrêts, lignes) et au plan de transport (horaires de passage, descriptif des trajets et dessertes). L'accès à de telles données est ainsi indispensable pour que puissent se développer les SIM et fournir une information multimodale à destination des usagers des transports.

---

<sup>8</sup> Actions définies à l'article 3 de la directive 2010/40/UE.

<sup>9</sup> Par exemple, pour la mise en œuvre de l'action relative aux circulations en temps réel, la Commission a adopté le règlement 2015/962 du 18 décembre 2014 complétant la directive 2010/40/UE en ce qui concerne la mise à disposition, dans l'ensemble de l'Union, de services d'informations en temps réel sur la circulation. Pour la mise en œuvre de l'action relative aux informations de sécurité routière, la Commission a adopté le règlement 886/2013 du 15 mai 2013 complétant la directive 2010/40/UE en ce qui concerne 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.

<sup>10</sup> Règlement délégué 2017/1926 (UE) de la Commission du 31 mai 2017 complétant la directive 2010/40/UE du Parlement européen et du Conseil en ce qui concerne la mise à disposition, dans l'ensemble de l'Union, de services d'informations sur les déplacements multimodaux.

<sup>11</sup> L'article 2 du règlement délégué 2017/1926 utilise le terme de service d'informations sur les déplacements pour faire référence à un SIM et le définit comme « un service STI, comprenant des cartes numériques, qui fournit aux utilisateurs et aux utilisateurs finaux des informations sur les déplacements et la circulation pour au moins un mode de transport ». De plus, la définition de la notion d'« informations sur les déplacements multimodaux » issue du même règlement, est la suivante : « des informations tirées de données statiques ou dynamiques sur les déplacements et la circulation, ou les deux, à l'intention des utilisateurs ou des utilisateurs finaux, par l'intermédiaire de tout mode de communication, couvrant au moins deux modes de transport et donnant la possibilité de comparer les modes de transport ». Voir glossaire également.

<sup>12</sup> Il s'agit notamment des cartes numériques utilisées dans les outils de navigation (« GPS »). Voir glossaire.

**Or, l'ensemble de ces données de mobilité demeure peu accessible aux utilisateurs<sup>13</sup> - dont les fournisseurs de SIM.** Ce déficit d'accessibilité est dû au fait notamment de l'importante fragmentation des producteurs de données (opérateurs de transport, autorités chargées des transports<sup>14</sup>, fournisseurs de services de transport à la demande et gestionnaires d'infrastructure). Il s'explique également par la non-diffusion (par choix ou non) des données disponibles. Ces entités ont pu déployer, sur la base de ces données, des systèmes numériques pour piloter et exploiter leurs services de transports ou infrastructures, et ont également pu mobiliser ces données pour alimenter une information communiquée aux usagers (par exemple pour de l'affichage et des annonces en stations et en gares), mais ne publient pas nécessairement ces données. En outre, l'absence de normalisation des formats et spécifications techniques, ainsi que de possibles restrictions de conditions d'utilisation, ont pu constituer des barrières à l'accès et à l'utilisation des données de mobilité. Dès lors est apparue la nécessité d'une réglementation permettant de lever ces barrières et ainsi favoriser l'émergence de SIM.

Le règlement délégué 2017/1926 exige des producteurs de données (1) de les publier, (2) selon des formats définis, (3) sur un point d'accès unique et (4) avec des conditions de licence et, le cas échéant, de compensations financières ne limitant pas inutilement leur réutilisation, et ouvre ainsi l'accès à un ensemble de données de mobilité (Figure 1). **En exigeant la publication de certaines données de mobilité afin de faciliter leur réutilisation, la Commission européenne entend favoriser l'émergence d'un ensemble plus riche de services d'information.** Ces services pourront, en délivrant une information plus complète et qualitative au voyageur, contribuer à une meilleure utilisation des réseaux de transports disponibles et favoriser les reports modaux, participant ainsi à l'atteinte des objectifs de réduction de la congestion routière de la directive 2010/40/UE.

Le règlement délégué 2017/1926 répond ainsi à plusieurs des freins précédemment évoqués : (1) lever les effets de verrouillage (intentionnels ou non) au travers d'une obligation de publication des données et de conditions d'utilisation peu limitatives ; (2) faciliter la réutilisation de ces données grâce à la standardisation des formats utilisés ; (3) réduire la fragmentation constatée en imposant une publication sur un point d'accès unique. Il met ainsi en place un dispositif d'ouverture des données de mobilité (« dispositif amont » ci-après) au bénéfice d'un marché aval sur lequel opèrent les utilisateurs de ces données.

**L'ouverture des données repose en premier lieu sur la création d'un Point d'Accès National** (ci-après « PAN »), dédié à l'hébergement et à la centralisation des données de mobilité, qui peut prendre plusieurs formes spécifiées par la directive (article 3)<sup>15</sup>. C'est sur cette infrastructure numérique unique que les producteurs doivent publier leurs données numériques<sup>16</sup>, y compris les métadonnées, c'est-à-dire les données descriptives des données (producteur de la donnée, date de mise à jour, etc.) qui permettent notamment aux utilisateurs de retrouver et d'utiliser aisément les séries de données disponibles. En France, le décret n° 2020-183 du 28 février 2020<sup>17</sup> a désigné comme PAN, la plateforme « [transport.data.gouv.fr](http://transport.data.gouv.fr) ».

<sup>13</sup> Cf. note 5 concernant la définition de la notion d'utilisateur au sens du règlement délégué 2017/1926.

<sup>14</sup> L'article L. 1115-1 du code des transports précise que ces derniers sont les Autorités Organisatrices de la Mobilité, l'État, les régions, les départements, les communes, les établissements publics de coopération intercommunale, les syndicats mixtes, Ile-de-France Mobilités, les autorités désignées à l'article L. 1811-2 et la métropole de Lyon.

<sup>15</sup> Le considérant 10 du règlement délégué 2017/1926 énumère les diverses formes que peut prendre un PAN, allant d'une base de données (forme centralisée) à un registre (forme décentralisée).

<sup>16</sup> Le considérant 14 du règlement délégué 2017/1926 précise que « (l)es exigences spécifiques relatives aux données statiques et dynamiques sur les déplacements et la circulation dans différents modes de transport ne devraient s'appliquer qu'aux données déjà collectées et disponibles dans un format lisible en machine. »

<sup>17</sup> Décret n° 2020-183 du 28 février 2020 portant désignation du point d'accès national aux données de l'information sur les déplacements multimodaux.

En outre, **les articles 4 et 5 du règlement délégué 2017/1926 encadrent les modalités d'échange de données d'une part en imposant l'utilisation de normes et de spécifications techniques permettant d'assurer l'interopérabilité des formats d'échanges de données**<sup>18</sup> et d'autre part en formulant des exigences sur la qualité des données et des métadonnées (article 8.1), le succès du déploiement des SIM reposant notamment sur l'exactitude et la fiabilité de l'information qu'ils délivrent. À cette fin, les producteurs sont tenus de mettre à jour leurs données et de corriger toute erreur leur étant signalée (article 6.2).

Enfin, **les modalités d'utilisation aux données publiées sur le PAN sont encadrées afin de faciliter leur réutilisation**<sup>19</sup>. D'une part, les producteurs de données peuvent définir des modalités d'utilisation des données de mobilité déterminées par des accords de licence. Dans ce cas, l'article 8.4 du règlement délégué 2017/1926 impose que ces conditions d'utilisation *« ne limitent pas inutilement les possibilités de réutilisation et ne soient pas utilisées pour restreindre la concurrence. Lorsqu'ils sont utilisés, les accords de licence restreignent dans tous les cas aussi peu que possible les possibilités de réutilisation »*. D'autre part, lorsqu'une compensation financière est prévue, elle doit être *« raisonnable et proportionnée aux coûts légitimes encourus pour la fourniture et la diffusion des données pertinentes sur les déplacements et la circulation »*.

**Le règlement délégué 2017/1926 soumet également les utilisateurs de données à des obligations visant notamment à assurer la fiabilité de l'information fournie aux voyageurs et la transparence des critères utilisés pour hiérarchiser les options de voyage**<sup>20</sup>. Ces obligations visent par ailleurs à permettre l'échange des résultats de recherches d'itinéraires afin d'assurer une couverture géographique qui aille au-delà du territoire de chaque État membre.

En premier lieu, tous les utilisateurs doivent ainsi réutiliser<sup>21</sup> les données du PAN de manière *« neutre, sans discrimination ni biais »* (article 8.2). Dans ce cadre, tout utilisateur est tenu d'indiquer la source des données et, pour les données statiques, la date et l'heure de leur dernière mise à jour (article 8.3). Par ailleurs, les SIM doivent être fondés sur des données mises à jour (article 6.1) et les critères utilisés pour le classement des options de voyage doivent être transparents, non discriminants et ne doivent pas induire les utilisateurs finaux (c'est-à-dire les utilisateurs des informations sur les déplacements produites à partir des données mises à disposition sur le PAN)<sup>22</sup> en erreur lors de la première présentation d'itinéraires de voyage (article 8.2).

---

<sup>18</sup> Le considérant 7 de la Directive 2010/40/UE prévoit que : *« (pour) que l'approche soit coordonnée, la Commission devrait assurer la cohérence des travaux du comité institué par la présente directive et ceux du comité institué par la directive 2004/52/CE du Parlement européen et du Conseil du 29 avril 2004 concernant l'interopérabilité des systèmes de télépéage routier dans la Communauté (4), du comité institué par le règlement (CEE) 3821/85 du Conseil du 20 décembre 1985 concernant l'appareil de contrôle dans le domaine des transports par route (5), du comité institué par la directive 2007/46/CE, ainsi que du comité institué par la directive 2007/2/CE du Parlement européen et du Conseil du 14 mars 2007 établissant une infrastructure d'information géographique dans la Communauté européenne (INSPIRE) »*.

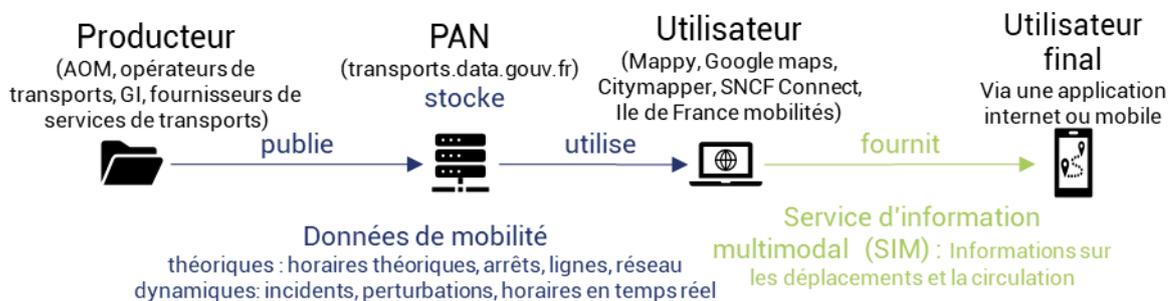
<sup>19</sup> Le règlement délégué 2017/1926 dresse une liste non-exhaustive des utilisateurs en les définissant comme *« une entité publique ou privée qui utilise le point d'accès national, telle que les autorités chargées des transports, les opérateurs de transport, les fournisseurs de services d'informations sur les déplacements, les producteurs de cartes numériques, les fournisseurs de services de transport à la demande et les gestionnaires d'infrastructure »* (point 11 de l'article 2).

<sup>20</sup> Considérant 20 du règlement délégué 2017/1926.

<sup>21</sup> Un utilisateur des données du PAN peut faire le choix de créer un SIM à destination d'autres utilisateurs ou usagers des transports. Dans ce cas-là il s'agit d'une « réutilisation de données ». Voir glossaire pour les définitions.

<sup>22</sup> Un utilisateur final est *« une personne physique ou morale, qui a accès aux informations sur les déplacements »* (article 2 du règlement délégué 2017/1926).

Figure 1 – schéma d'ouverture et d'utilisation des données



Source : ART.

**La loi n° 2019-1428 du 24 décembre 2019 d'orientation des mobilités (ci-après « LOM ») précise le règlement délégué 2017/1926<sup>23</sup>.** La loi LOM s'inscrit ainsi dans la continuité des objectifs définis par la directive 2010/40/UE et le règlement délégué 2017/1926, tout en précisant et complétant les exigences prévues par ce dernier. L'exposé des motifs de la LOM indique ainsi : « [la loi LOM] vise également à encourager le développement des innovations et des expérimentations, ainsi que le déploiement de services performants d'information multimodale et de billettique. Un tel déploiement doit favoriser le recours aux modes alternatifs à la voiture individuelle, lorsque cela est possible, et en particulier du transport en commun en intermodalité avec d'autres usages »<sup>24</sup>, tandis que les articles 25 et 27 de la LOM, d'où sont issus les articles L. 1115-1 à L. 1115-7 du code des transports, approfondissent et complètent les obligations prévues par le règlement délégué 2017/1926. Les compléments apportés par la LOM portent sur la désignation des acteurs responsables de la fourniture des données sur le PAN, l'extension du périmètre des données devant être publiées, le calendrier de publication des données, ainsi que sur les compensations financières qui peuvent être fournies aux producteurs de données.

La LOM précise ainsi que les autorités chargées des transports, tout en restant responsables de la fourniture des données sur le PAN, peuvent en confier la charge aux acteurs suivants<sup>25</sup> :

- l'opérateur de transport qui réalise le service de transport pour leur compte ou les opérateurs de système d'aide à l'exploitation et à l'information voyageur gérant le système d'information du service concerné ;
- le prestataire chargé de l'exécution d'un service de partage de véhicules, cycles et engins personnels qui réalise le service de transport pour leur compte ;
- le prestataire de la gestion du stationnement en ouvrage ou sur voirie qui réalise la gestion du stationnement pour leur compte.

<sup>23</sup> Voir Figure 2 pour la synthèse des exigences réglementaires selon les textes juridiques.

<sup>24</sup> Exposé des motifs : <http://www.senat.fr/leg/exposes-des-motifs/pjl18-157-expose.html>

<sup>25</sup> Aux 3°, 4° et 5° de l'article L. 1115-1 du code des transports.

**Elle impose par ailleurs la fourniture de données complémentaires qui sont nécessaires aux SIM pour pouvoir proposer aux usagers des modes alternatifs à la voiture individuelle.** Ainsi, les articles L. 1115-1, L. 1115-6 et L. 1115-7 du code des transports étendent la liste des données concernées par l'obligation d'ouverture<sup>26</sup>, celle-ci couvrant en conséquence :

- les données dynamiques sur les déplacements et la circulation, qui n'étaient qu'une option du règlement délégué 2017/1926 ;
- les données statiques, historiques et dynamiques sur les déplacements de services de partage étendues à tous les engins de déplacement personnel au-delà des seuls véhicules et cycles prévus par le règlement délégué 2017/1926 ;
- les données relatives aux points de recharge publics pour véhicules électriques ou hybrides rechargeables plus précises que les données prévues dans l'annexe du règlement délégué 2017/1926 notamment leur puissance, leur accessibilité aux personnes handicapées ou à mobilité réduite, leur disponibilité et les éventuelles restrictions d'accès liées au gabarit du véhicule ;
- les données sur l'accessibilité des services réguliers de transport public aux personnes handicapées ou à mobilité réduite dont la collecte devient obligatoire (alors que le règlement délégué 2017/1926 n'impose que la publication de données existantes) et les données relatives à l'accessibilité aux personnes handicapées ou à mobilité réduite des principaux itinéraires pédestres situés dans un rayon de deux cents mètres autour des points d'arrêt prioritaires au sens de l'article L. 1112-1 du code des transports<sup>27</sup> qui ne sont pas mentionnées par le règlement délégué 2017/1926.

En outre, **la loi a modifié le calendrier de fourniture des données statiques de mobilité fixé par le règlement délégué 2017/1926, en avançant de deux ans l'échéance finale de publication de toutes les données relatives aux différents réseaux de transport, soit au 1er décembre 2021.**

Enfin, la LOM précise les modalités selon lesquelles une compensation financière peut être versée aux producteurs de données (article L. 1115-3 du code des transports), en étendant, sous réserve du respect de certaines conditions, la gratuité à certaines données<sup>28</sup>, dont les données dynamiques (article R. 1115-4 du code des transports<sup>29</sup>). Elle prévoit que la compensation financière, quand elle est prévue, doit être raisonnable et proportionnée aux coûts légitimes encourus pour la fourniture et la diffusion des données de mobilité. Elle précise également que « *le produit total de la compensation financière ne peut excéder le montant des coûts d'investissement et de fonctionnement résultant directement de la mise en œuvre des obligations de publication de données* ». Les producteurs de données doivent fixer le montant de la compensation selon des critères objectifs, transparents, vérifiables et non discriminatoires, ce qui passe notamment par l'obligation de publier les modalités de calcul de la compensation financière sous format électronique<sup>30</sup>.

---

<sup>26</sup> L'article L. 1115-6 prévoit la collecte des données sur l'accessibilité des services réguliers de transport public aux personnes handicapées ou à mobilité réduite. Les données mentionnées à cet article mais aussi à l'article L. 1115-7 sont soumises aux mêmes conditions de publication et de réutilisation que les données de mobilité.

<sup>27</sup> Avant-dernier alinéa de l'article L. 141-13 du code de la voirie routière mentionné dans l'article L. 1115-5 du code des transports. Les trois derniers points sont des catégories de données ouvertes dans le cadre de l'article 27 de la LOM.

<sup>28</sup> Il s'agit des données sur la localisation des passages à niveau qui sont rendues gratuitement accessibles et librement réutilisables par le gestionnaire de l'infrastructure ferroviaire (article L. 1115-4 du code des transports). La gratuité s'impose aussi aux données dynamiques qui ne dépassent pas un seuil de nombre de requêtes par jour et heure déterminé à l'article R. 1115-3 du code des transports.

<sup>29</sup> L'article R. 1115-4 du code des transports a été créé par le décret n° 2020-1753 du 28 décembre 2020 relatif à certaines conditions de mise à disposition des données numériques destinées à faciliter les déplacements.

<sup>30</sup> Article R. 1115-4 du code des transports.

### 1.3 L'Autorité de régulation des transports a été désignée par le législateur pour contrôler le respect des exigences relatives à l'ouverture des données d'intérêt collectif

**Pour garantir la mise en œuvre de ses dispositions, notamment en matière d'ouverture des données, et permettre ainsi le déploiement de SIM fournissant une information multimodale aux voyageurs, le règlement délégué 2017/1926 impose aux États membres d'évaluer la conformité de la publication et de la réutilisation des données du PAN avec les exigences de ses articles 3 à 8 (article 9.1).** Aux fins de cette évaluation, les États membres peuvent notamment demander aux autorités chargées des transports, aux opérateurs de transport, aux gestionnaires d'infrastructure, aux fournisseurs de services de transport à la demande et aux fournisseurs de services d'informations sur les déplacements de fournir une déclaration de leur conformité fondée sur des données concrètes (article 9.2), dont ils contrôlent l'exactitude de manière aléatoire (article 9.3). À cet égard, l'article L. 1115-5 du code des transports, introduit par la LOM, impose à ces entités de transmettre annuellement une déclaration de conformité<sup>31</sup> au ministre chargé des transports, lequel met cette déclaration à la disposition de l'Autorité.

**Le législateur a désigné l'Autorité pour effectuer les contrôles du respect des exigences réglementaires susmentionnées.** En contribuant à une ouverture effective des données de mobilité ainsi qu'à un suivi de leur bon usage, cette mission de contrôle s'inscrit à la fois dans l'intérêt des voyageurs qui pourront disposer d'une information plus riche et multiple et dans les missions de régulation de l'Autorité consistant à veiller au développement et au maintien d'un cadre concurrentiel équitable au service des acteurs et des usagers des transports, dans une approche multimodale<sup>32</sup>. Les objectifs premiers de la directive 2010/40/UE consistant à favoriser un usage plus coordonné des réseaux de transports et ainsi contribuer à la lutte contre le changement climatique par une meilleure utilisation des transports peu carbonés trouvent également un écho dans les orientations stratégiques de l'Autorité présentées dans son document de référence 2021-2022.

L'Autorité est chargée de contrôler de manière aléatoire l'exactitude des déclarations de conformité (comme prévu à l'article 9.3 du règlement délégué 2017/1926). Elle peut pour ce faire effectuer des contrôles d'office et des contrôles à la demande des AOM ou des associations de consommateurs agréées au titre de l'article L. 811-1 du code de la consommation. Elle dispose aussi de pouvoirs d'enquête afin de constater d'éventuels manquements aux articles L. 1115-1 à L. 1115-7 du code des transports.

Aux fins de ces contrôles et enquêtes, sur le fondement des articles L. 1115-5 et L. 1264-2 du code des transports, l'Autorité dispose d'un droit d'accès aux documents et informations nécessaires auprès des gestionnaires d'infrastructure, des exploitants et fournisseurs de service de transport, des entreprises de transport public, des fournisseurs de services numériques multimodaux et des autorités organisatrices de la mobilité. Elle peut enfin entendre toute personne dont l'audition lui paraît susceptible de contribuer à son information.

---

<sup>31</sup> Le décret précisant le contenu et les modalités de publication de la déclaration de conformité n'était pas encore paru à la date de publication du présent rapport. Le décret n° 2020-1753 du 28 décembre 2020 relatif à certaines conditions de mise à disposition des données numériques destinées à faciliter les déplacements précise que la déclaration de conformité doit être transmise tous les ans et qu'une déclaration modificative doit être transmise en cas de changement de circonstances dans un délai de trois mois.

<sup>32</sup> Comme précisé dans le [Document de Référence 2021-2022 de l'Autorité](#), page 12 notamment.

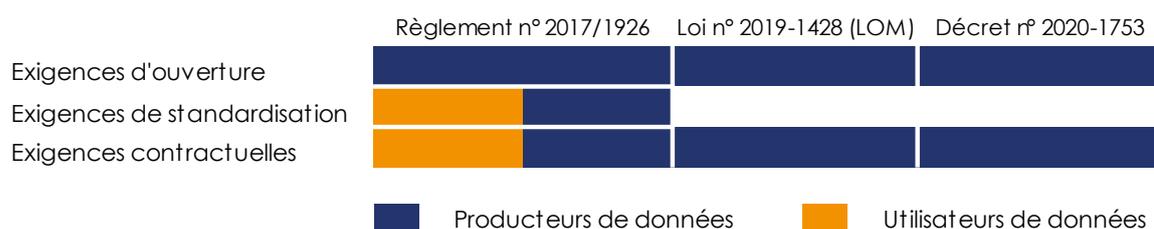
**Pour que l'action de l'Autorité soit performative et puisse effectivement bénéficier aux producteurs de données comme aux utilisateurs, ce pouvoir de contrôle s'accompagne d'une d'un pouvoir de sanction.** L'Autorité peut ainsi sanctionner les manquements aux exigences prévues par le règlement délégué 2017/1926 et par le code des transports (article L. 1264-7 du code des transports). A cet égard, l'article 15 de la loi n° 2021-1308 du 8 octobre 2021, dite loi « DDADUE »<sup>33</sup>, a modifié l'article L. 1264-1 du code des transports afin de prévoir que l'Autorité peut, soit d'office, soit à la demande d'une entité énumérée à cet article, procéder à la recherche et à la constatation de ces manquements et, à cette fin, utiliser les pouvoirs d'enquête prévus par cet article.

**L'Autorité peut également être saisie de règlements de différend relatifs à la mise à disposition des données sur les déplacements et la circulation par les producteurs de données** (article L. 1263-4 du code des transports).

L'article L. 1115-5 confie aussi à l'Autorité le soin d'établir un rapport sur les contrôles qu'elle est tenue d'effectuer. L'Autorité se propose d'établir le rapport requis par le code des transports à un rythme au moins biennal. Ce rapport permettra notamment d'effectuer un bilan sur le niveau de conformité des acteurs concernés à la réglementation et sur les non-conformités identifiées lors des contrôles réalisés par l'Autorité. La publication de ce rapport favorisera la mise en conformité - ex-ante et ex-post - de l'ensemble des acteurs. Cette publication permettra ainsi de mettre en application l'un des piliers socles de l'Autorité : améliorer la connaissance des marchés régulés et favoriser la mise en œuvre d'une régulation par la transparence, éclairée par les données.

En guise de rapport introductif, en amont des contrôles effectifs qui peuvent dès à présent être menés par l'Autorité et dont il sera rendu compte dans les prochaines éditions, l'Autorité présente dans le présent rapport un état des lieux de l'ouverture des données, en complément des informations déjà disponibles sur le PAN ainsi qu'une première analyse des marchés de l'information multimodale et, plus particulièrement, des fournisseurs de service d'informations multimodales.

**Figure 2 – Matrice synthétique de la répartition des exigences selon les textes réglementaires et les entités concernées**



Source : ART.

<sup>33</sup> Loi n° 2021-1308 du 8 octobre 2021 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière économique et financière.

## 2 ETAT DES LIEUX DE LA PUBLICATION DES DONNEES DE MOBILITE

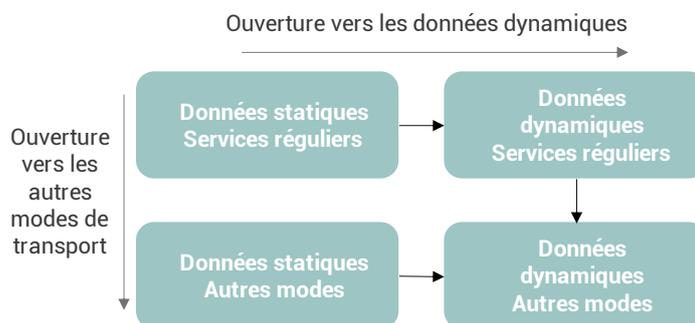
### 2.1 Les données statiques relatives aux transports en commun représentent une forte majorité des données de mobilité ouvertes sur le point d'accès national

Les premières analyses conduites montre que la majeure partie des données de mobilité mises à disposition sur le PAN consiste en des données statiques concernant les transports en commun. L'analyse des métadonnées du catalogue du PAN recense en effet sur un total de 689 jeux de données :

- 560 jeux de données statiques, portant dans leur totalité sur des données de transport en commun, et représentant 81 % des jeux de données du PAN et 88 % des jeux de données de transport en commun ;
- 129 jeux de données dynamiques (c'est-à-dire en temps réel) qui comportent, par nature, des enjeux techniques de mise à disposition supérieurs aux données statiques et font l'objet d'une publication plus complexe à mettre en place. Ceux-ci se répartissent en 75 jeux de données de transports en commun et 54 jeux de données de services de transport à la demande tels le partage de véhicules, cycles et engins de déplacements personnels (trottinettes en libre-service, etc.) fonctionnant avec ou sans stations.

L'état des lieux, réalisé dans le cadre du présent rapport, présente ainsi une évaluation de la publication des données statiques des transports en commun jusqu'au 1<sup>er</sup> décembre 2021, hormis quelques informations directement issues du PAN en date du 1<sup>er</sup> mars 2022. En effet, les données statiques sont la première priorité du règlement délégué 2017/1926, les données dynamiques ou en temps réel n'étant qu'une option de celui-ci. La publication des données de transports en commun (par opposition au transport à la demande et aux déplacements individuels) est aussi largement majoritaire sur le PAN à la date de publication de ce rapport du fait des priorités d'ouverture de données faites par ce dernier<sup>34</sup>. **Les données statiques de mobilité, essentielles pour l'information et la planification avant le voyage ont ainsi été ouvertes en priorité, en commençant par les données relatives aux différents modes de transport en commun**<sup>35</sup> (Figure 3).

Figure 3 – Dynamique de mise à disposition des catégories de données du PAN



Source : ART.

<sup>34</sup> <https://doc.transport.data.gouv.fr/documentation/liste-des-rencontres-publiques/tour-de-france>

<sup>35</sup> Equivalent aux services réguliers au sens du règlement délégué 2017/1926.

## 2.2 Deux formats sont utilisés pour publier les données statiques relatives aux transports en commun

**Un standard (*General Transit Feed Specification, GTFS*) et une norme (*Network Timetable Exchange, NeTEX*) coexistent et sont mobilisés dans le cadre de la publication des jeux de données statiques des transports en commun sur le PAN :**

- **GTFS<sup>36</sup>** est une spécification technique, développée initialement par l'entreprise Google afin d'intégrer les services de transports publics dans Google Maps ; il a ensuite été repris par un groupe d'autorités chargées des transports et d'opérateurs de transports nord-américains. Depuis 2009, il s'agit d'un format libre de droit maintenu par une association de producteurs de données MobilityData<sup>37</sup>, le format GTFS s'est largement diffusé dans le monde et apparaît aujourd'hui comme le standard le plus utilisé par les producteurs de données. Lorsqu'elles sont fournies en format GTFS, les données d'un même service de transport sont stockées dans différents fichiers : chaque fichier correspondant à une information spécifique sur le service concerné. Par exemple, le fichier `calendar.txt` contient le calendrier de circulation et le fichier `route.txt` présente le nom et la direction des lignes desservies.
- **NeTEX<sup>38</sup>** est une norme définie au niveau européen<sup>39</sup> (CEN<sup>40</sup>/TS<sup>41</sup> 16614) en vue de faciliter l'interopérabilité entre services de transports et avec les services numériques de mobilité. La norme NeTEX, se base sur le langage XML (eXtended Markup Language) et suit le modèle de données Transmodel<sup>42</sup>. La norme est divisée en plusieurs parties et permet la description détaillée de chaque élément du service de transport (topologie du réseau, tarifs, etc.)<sup>43</sup>. Le règlement délégué 2017/1926 prévoit que les producteurs peuvent publier leurs données au format NeTEX en suivant un « profil minimal » défini par chaque État membre (2° de l'article 4). Le profil minimal définit, à partir du schéma général de la norme NeTEX, la façon dont les fichiers sont échangés, leur structure ainsi qu'un contenu constituant une partie seulement de la norme NeTEX (profils correspondant aux arrêts, aux horaires ou aux tarifs par exemple). Actuellement, les profils validés par la commission de normalisation des transports publics au sein du bureau de normalisation des transports, des routes et de leurs aménagements<sup>44</sup> (BNTRA) n'ont pas tous été officiellement publiés. Ainsi, la publication des jeux de données statiques mobilise essentiellement (à hauteur de 83 %) le format GTFS. Le format NeTEX est mobilisé par seulement 17 % des jeux publiés et concerne principalement les régions Nouvelle-Aquitaine et Grand-Est qui publient leurs jeux de données en utilisant les deux formats GTFS et NeTEX<sup>45</sup>.

En matière de publication des jeux de données dynamiques, le standard GTFS-RT (General Transit Feed Specifications – Real Time) s'impose face à la norme SIRI (Service Interface for

<sup>36</sup> Source : <https://developers.google.com/transit/gtfs?hl=fr>

<sup>37</sup> <https://gtfs.org/>

<sup>38</sup> Source : <https://netex-cen.eu/>

<sup>39</sup> Source : <http://www.normes-donnees-tc.org/format-dechange/donnees-theoriques/netex/>

<sup>40</sup> Le Comité Européen de Normalisation (CEN) est une association qui regroupe les organismes nationaux de normalisation de 34 pays européens. Source : <https://www.cencenelec.eu/about-cen/>

<sup>41</sup> Technical Standard (TS) : standard technique en français.

<sup>42</sup> <https://www.transmodel-cen.eu/>

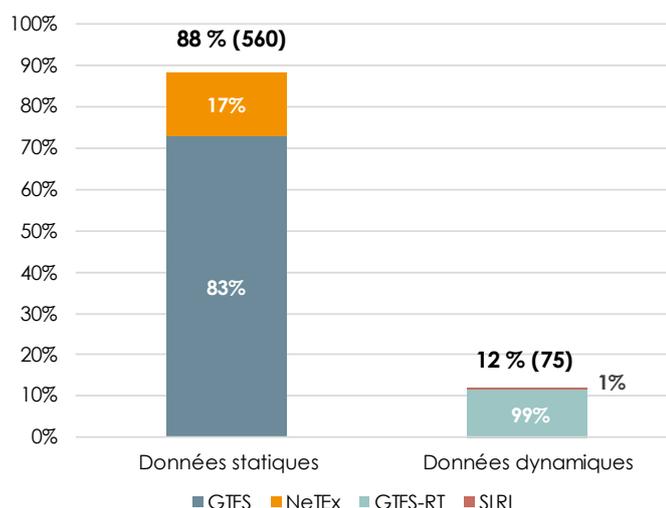
<sup>43</sup> La topologie du réseau – arrêts et lignes – (partie 1), les horaires et calendriers (partie 2) et les informations tarifaires (partie 3). Il intègre en plus d'autres informations dont celles du profil européen d'information sur les voyageurs (partie 4), du format d'échange des modes alternatifs et du profil européen d'accessibilité aux informations sur les voyageurs (partie 5).

<sup>44</sup> Le Bureau de normalisation pour les transports, les routes et leurs aménagements (BNTRA) est l'un des bureaux de normalisation sectoriels du Système Français de Normalisation. Il couvre, par délégation de l'AFNOR, les secteurs de la géotechnique, de la route et de ses équipements, des ouvrages d'arts ainsi que l'ingénierie du trafic. <https://www.cerema.fr/fr/cerema/directions/cerema-infrastructures-transport-materiaux/bureau-normalisation-transports-routes-leurs-amenagements>

<sup>45</sup> Voir Annexe 1 - Éléments complémentaires sur l'utilisation des formats.

Real time Information), ceci notamment du fait de l'utilisation massive de données statiques en GTFS sur lesquelles les données dynamiques au format GTFS-RT se fondent<sup>46</sup> (Figure 4).

Figure 4 – Répartition des jeux de données des transports en commun selon la catégorie de données et le format utilisé (nombre de jeux de données entre parenthèses)



Note : le pourcentage global correspond à la proportion de jeux de données selon une distinction données statiques/données dynamiques. Le pourcentage dans les histogrammes correspond à la proportion de jeux de données de chaque catégorie selon le standard utilisé.

Source : ART, à partir des métadonnées du PAN extraites le 01/03/22.

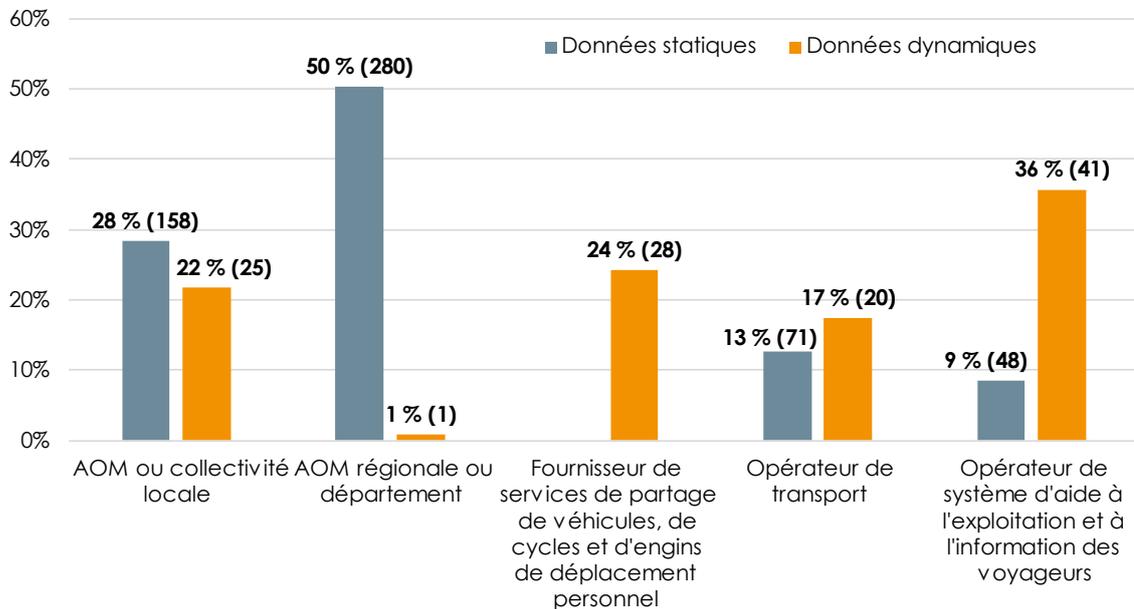
**Les AOM régionales sont les contributeurs majoritaires de l'offre de données statiques des transports en commun puisqu'elles fournissent 50 % des 557 jeux de données pour lesquels le producteur est identifié<sup>47</sup>.** Elles publient à la fois les données des réseaux de cars interurbains et celles des réseaux de transport urbains à la demande des AOM et collectivités locales (Figure 5). Ces dernières privilégient la fourniture des données de leur propre réseau de transport (28 % des jeux de données) plutôt que de déléguer cette charge à leurs opérateurs de transport (13 %) ou à leurs opérateurs de système d'aide à l'exploitation et à l'information des voyageurs (9 %).

En dépit du fait que, sur les 129 jeux de données dynamiques recensés, seuls 115 soient associés à des fournisseurs identifiables, ceux-ci témoignent d'une variété plus importante de ce type de producteurs de données, du fait notamment de la contribution des fournisseurs de services de partage. Ces derniers fournissent 24 % des jeux de données dynamiques, soit un peu plus que les AOM ou collectivités locales (22 %). Les opérateurs de système d'aide à l'exploitation et à l'information des voyageurs, souvent en charge de mettre en place le système d'information dynamique pour les transports en commun, sont les premiers producteurs avec 41 jeux de données dynamiques publiés, soit 36 % de la totalité des jeux de cette catégorie.

<sup>46</sup> A noter que les jeux de données de services de transport à la demande sans stations sont publiés sur le PAN au format GBFS (General Bikeshare Feed Specification) uniquement.

<sup>47</sup> Sur 560 jeux de données statiques de transport en commun, le producteur de données n'a pas pu être correctement identifié pour 3 d'entre eux.

Figure 5 – Part de jeux de données publiés par catégorie de données et catégories de producteurs (nombre de jeux de données entre parenthèses)



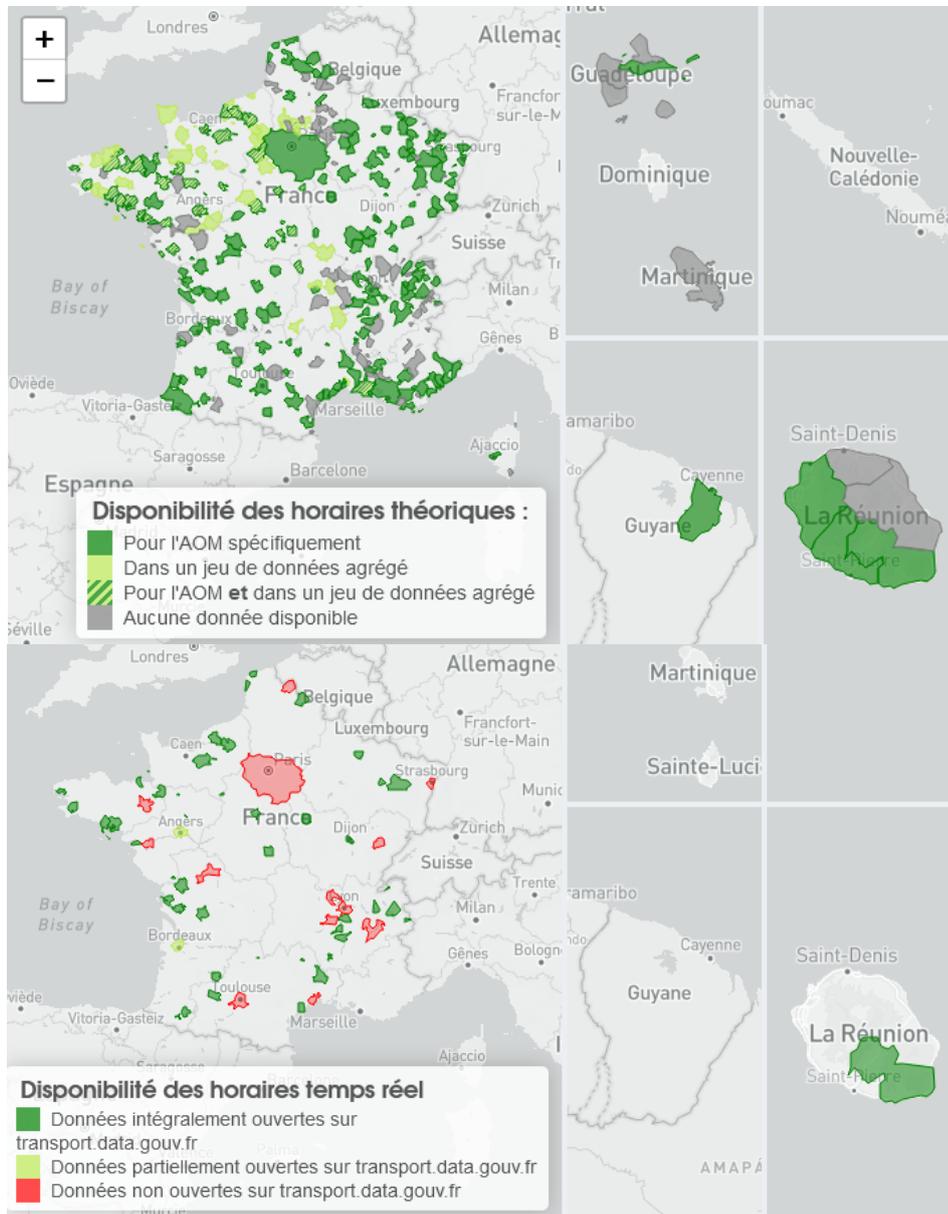
Note : la part de jeux de données statiques dont le type de producteur est identifié est de 99,4 % (557 jeux sur 560), celle des jeux de données dynamiques s'élève à 89,1 % (115 jeux sur 129). Le pourcentage indiqué se base sur le total de jeux de données selon la catégorie de données (statiques ou dynamiques).

Source : ART, à partir des métadonnées du PAN extraites le 01/03/22.

À la date de réalisation de ce rapport, toutes les régions métropolitaines (ainsi que trois régions et une collectivité d'outre-mer) ont publié leurs données statiques sur le PAN. En revanche, **sur les 335 autres AOM, 236 ont mis à disposition des données statiques, et seules 67 des données dynamiques**<sup>48</sup> (Figure 6). Ces 236 AOM représentent néanmoins une large proportion (44 millions des 49 millions) des Français vivant dans une AOM.

<sup>48</sup> Le PAN recense 18 jeux de données supplémentaires non comptabilisés car ils ne respectent pas l'ensemble des exigences techniques formulées par le PAN.

Figure 6 – Cartographies des données statiques (en haut) et dynamiques (en bas) des transports en commun publiées par le PAN

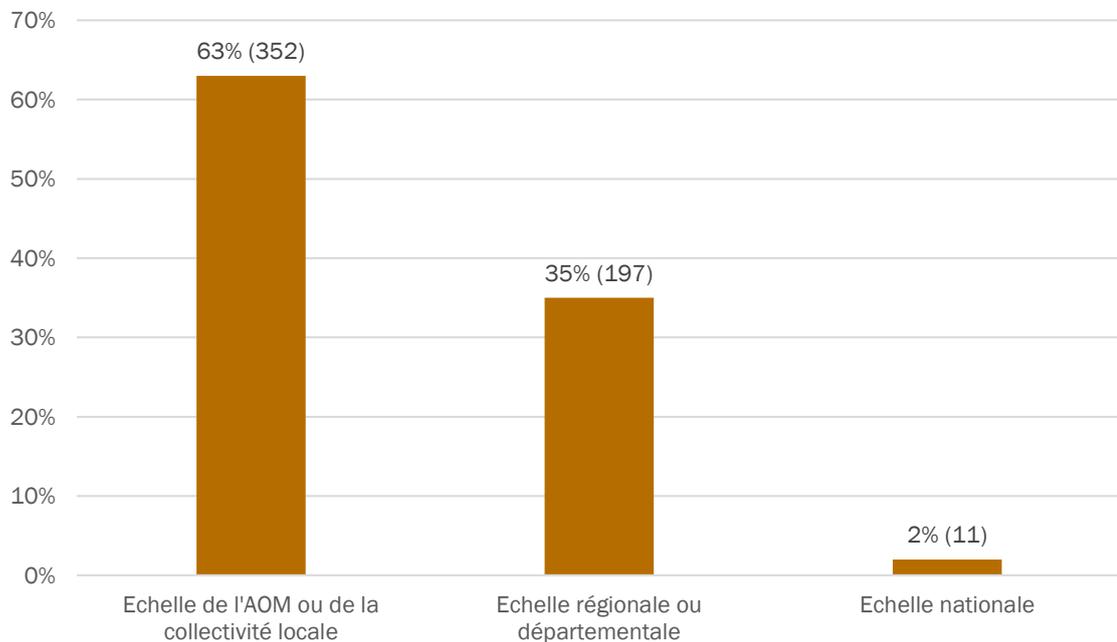


Source : transport.data.gouv.fr, captures d'écran du 01/03/2022.

Enfin, les jeux de données publiés diffèrent au regard de la couverture géographique des données qu'ils intègrent ainsi que les modes de transports qu'ils concernent<sup>49</sup>. Logiquement, 63 % des jeux de données recouvrent le territoire d'une collectivité locale du fait du nombre important de collectivités locales, en tant qu'AOM, parmi les producteurs de données (Figure 7). Cela permet d'approcher la part des réseaux urbains dont les données sont publiées sur le PAN, mais donne peu d'indications sur les modes de transports inclus dans les jeux concernés. Les jeux de données régionaux ou nationaux sont logiquement moins nombreux.

<sup>49</sup> Voir la Figure 44 en Annexe 2 pour plus de précisions sur la couverture géographique des données selon le format utilisé.

Figure 7 – Nombre de jeux de données statiques publiés selon l'échelle géographique



Source : ART, à partir des métadonnées du PAN extraites le 01/03/22.

### 2.3 La lisibilité et la qualité des données publiées sont essentielles à leur bonne utilisation, elles figurent donc parmi les priorités fixées par l'Autorité de régulation des transports

L'Autorité a établi un processus de contrôle comportant trois niveaux successifs. Chaque niveau se décline en « critères de qualité » définis de sorte à évaluer (1) la structure et l'organisation des jeux de données publiés, (2) leur adéquation avec les formats exigés, (3) la présence des informations obligatoires, (4) la lisibilité des jeux et la qualité de données qu'ils comprennent<sup>50</sup>.

Les trois niveaux qui structurent le contrôle de l'Autorité sont les suivants :

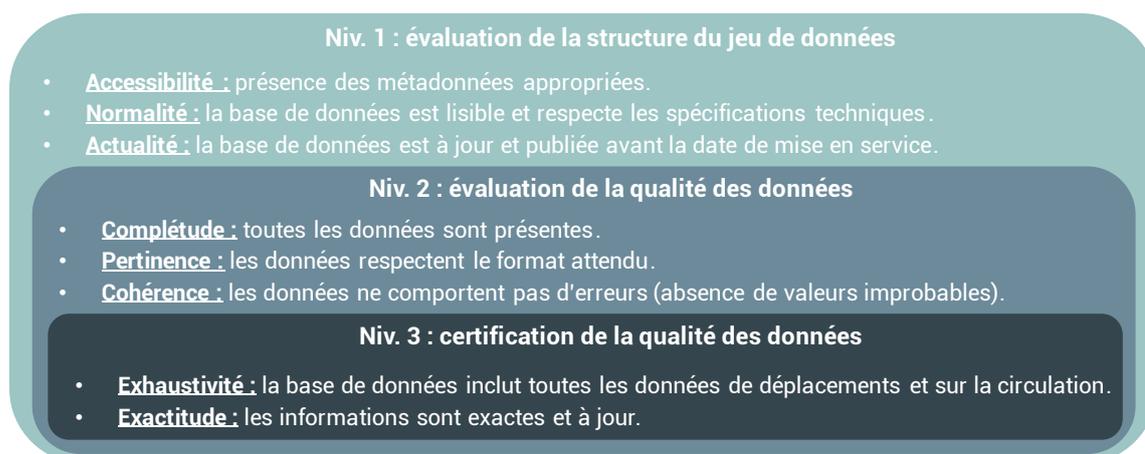
- les contrôles du format des données et des métadonnées associées, permettant de vérifier le respect des exigences de standardisation et par conséquent de lisibilité et de description des jeux de données (niveau 1 de contrôle) ;
- les contrôles du contenu des données, permettant d'en vérifier la bonne mise en qualité en matière de complétude et de cohérence (niveau 2 de contrôle) ;
- un troisième niveau de contrôle visant à confronter les données publiées avec la réalité du service de transport décrit, et permettant donc d'en vérifier la fiabilité.

La validation des niveaux 1 et 2 permet d'évaluer la conformité aux exigences réglementaires de normalisation et vérifie la « cohérence interne » du jeu de données. La validation du niveau 3 permet de contrôler la conformité du jeu de données au service de transport qu'il décrit, soit sa « cohérence externe ». Cela revient notamment à vérifier que les données et les

<sup>50</sup> Le non-respect par un producteur de la structure définie par un standard peut réduire voire compromettre la « lisibilité en machine » de son jeu de données, alors même que les données seraient exactes et à jour, du fait de l'incapacité de programmes informatiques à correctement interpréter le jeu ou à y retrouver les informations nécessaires à sa lecture.

métadonnées sont « *exactes et à jour* » (1° de l'article 8 du règlement délégué 2017/1926). Au regard du caractère imbriqué des niveaux et critères de contrôle, il apparaît peu utile de mener un contrôle de niveau 3, si celui-ci ne valide pas au préalable les critères de normalité du niveau 1 par exemple. Afin de préciser les niveaux de contrôle et cibler les points précis sur lesquels chaque jeu doit être évalué, chaque niveau de contrôle a été subdivisé en critères de qualité<sup>51</sup> (Figure 8).

Figure 8 – Schéma des différents niveaux de contrôle et critères de qualité



Source : ART.

Pour le premier niveau (validité de la structure des jeux de données), ces critères sont :

- l'« **accessibilité** » : présence et exactitude des métadonnées en les comparant avec les informations contenues dans les jeux de données. Ces métadonnées sont essentielles à tout utilisateur afin qu'il puisse identifier et comprendre le contenu d'un jeu de données sur le PAN ;
- l'« **actualité** » : conformité des métadonnées associées à la « validité » du jeu de données. Un jeu de données est considéré comme valide si le service de transport qu'il concerne est toujours d'actualité à la date de consultation du jeu de données ;
- la « **normalité** » : adéquation de la structure des données par rapport au format requis, notamment la lisibilité des jeux en format compressé.

Pour le niveau 2 (évaluation de la qualité du contenu des jeux de données), les critères sont les suivants :

- la « **complétude** » : présence des données obligatoires selon le format utilisé pour leur publication ;
- la « **pertinence** » : conformité de la donnée au format prévu (exemple : les informations relatives au calendrier d'exploitation du service de transport sont au format « date ») ;
- la « **cohérence** » : analyse à plat et croisée d'indicateurs issus d'un jeu de données afin de vérifier la fourniture d'une information cohérente (exemple : absence d'arrêts en double, de temps de trajet nul entre deux arrêts, de vitesses excessives, etc.).

<sup>51</sup> Issus des travaux Open Data Support (2013) "La qualité des données et métadonnées ouvertes", <https://joinup.ec.europa.eu/collection/open-government/document/tm22-la-qualite-des-donnees-et-metadonnees-ouvertes-fr>

Enfin, le niveau 3 (certification du jeu de données) complète l'évaluation en confrontant les données publiées à la réalité du service décrit. Il repose sur deux critères :

- l'« **exhaustivité** » : intégration effective dans le jeu de données de l'ensemble des données décrivant le service de transport telles que précisées par la réglementation ;
- l'« **exactitude** » : fiabilité des données pour décrire le service de transport (exemple : les horaires d'un jeu de données dynamiques correspondent bien aux horaires affichés en temps réel à chaque arrêt).

## **2.4 L'évaluation de la qualité des données par les producteurs et les utilisateurs de données nécessite de s'appuyer sur des outils informatiques adaptables aux différents formats**

**L'emploi de formats différents complexifie tant l'utilisation des données que leur contrôle du fait des nombreuses spécifications techniques - qui leur sont propres - à considérer.** Le volume de données à contrôler et en conséquence la variété des questions particulières qui se posent pour chaque jeu de données contribuent encore à accentuer cette complexité. Plus de 500 jeux de données statiques étaient ainsi publiés au 1er mars 2022 sur les transports en commun, chaque jeu contenant en outre une quantité importante de données. Par exemple, le jeu de données GTFS décrivant le service de navettes de Tignes, une des plus petites AOM, contient près de 99 000 valeurs sur l'ensemble de ses tables. Les fichiers stop\_time.txt et offre.xml du service de bus « Buséolien » de Puteaux présenté en annexe contiennent respectivement 41 450 cellules et plus de 32 000 lignes de codes. D'autre part, la structure des jeux de données et les formats de fichiers, conçus pour une lecture par des outils numériques de programmation informatique, rendent très difficile voire impossible une lecture directe (humaine).

**Afin de faciliter la vérification de la conformité des jeux de données nombreux, volumineux et lisibles uniquement par un programme informatique, il est ainsi nécessaire de s'appuyer sur un validateur,** un programme informatique permettant de lire les jeux et valider l'exactitude syntaxique des fichiers constituant ces jeux<sup>52</sup>. Ainsi, l'utilisation d'un validateur paramétré selon les critères de contrôle mentionnés dans la partie précédente permet d'évaluer au moins en partie la conformité des jeux de données.

**La mobilisation de validateurs publiquement disponibles peut appuyer la publication des données par les producteurs de données ainsi que les contrôles conduits par l'Autorité.** Depuis l'officialisation du format GTFS en 2009, plusieurs validateurs de ce format ont vu le jour<sup>53</sup>. Afin de juger de leur adéquation pour un appui au contrôle des données, l'Autorité a parangonné quatre validateurs GTFS existants choisis pour leur compatibilité avec différents systèmes d'exploitation, l'accès à leur code en open-source, l'intelligibilité de ce code et la possibilité d'extraire les résultats.

Quatre validateurs ont été testés par l'Autorité : le validateur officiel du PAN, deux autres issus de communautés de producteurs de données et un dernier provenant d'une entreprise privée spécialisée dans la fourniture de solutions techniques dans le domaine de la mobilité (Figure 9).

---

<sup>52</sup> Les validateurs sont d'usage fréquent en informatique. À titre d'exemple on peut citer les validateurs des logiciels de traitement de texte permettant de contrôler l'orthographe d'un mot par rapport au dictionnaire auquel il fait référence.

<sup>53</sup> On en recense au moins 33 sur la plateforme GitHub dédiée aux applications et codes librement accessibles ou ouverts.

Figure 9 – Présentation des validateurs GTFS choisis pour l'étude

Validateur	Origine	Lien GitHub
GTFSvtor 	Publié par Mecatran, société privée de développement de logiciel fournissant des produits et services pour le secteur de la mobilité, et plus particulièrement pour les transports publics.	<a href="https://github.com/mecatran/gtfsvtor">https://github.com/mecatran/gtfsvtor</a>
Transitfeed 	Open mobility data, est une communauté d'utilisateurs et producteurs engagée dans l'amélioration de l'accessibilité de l'information des transports publics. Il s'agit maintenant d'un projet soutenu par MobilityData.	<a href="https://github.com/google/transitfeed">https://github.com/google/transitfeed</a>
Mobility Data 	MobilityData est une association canadienne rassemblant des producteurs et utilisateurs de données à une échelle internationale, engagés dans la standardisation de la spécification GTFS.	<a href="https://github.com/MobilityData/gtfs-validator">https://github.com/MobilityData/gtfs-validator</a>
Transport.data.gouv 	Point d'Accès National géré par le Ministère de la transition écologique. Le validateur a été développé par Etalab.	<a href="https://github.com/etalab/transport-validator/">https://github.com/etalab/transport-validator/</a>

Source : ART.

Quoiqu'aucun validateur relatif à la norme NeTeX n'ait été identifié lors de ce recensement, il existe des programmes informatiques permettant de valider un fichier au format xml (utilisé par la norme NeTeX) par rapport à son schéma de données XSD<sup>54</sup> (« xmllint »<sup>55</sup> en est un exemple).

**Une comparaison a été réalisée entre les quatre validateurs sélectionnés afin d'identifier leurs fonctionnalités et comprendre les paramétrages de chacun d'entre eux.** Ce travail de parangonnage, qui s'appuie sur le code source ouvert par chaque validateur et sur la comparaison de leurs résultats, révèle des fonctionnalités et des paramétrages différents d'un validateur à un autre.

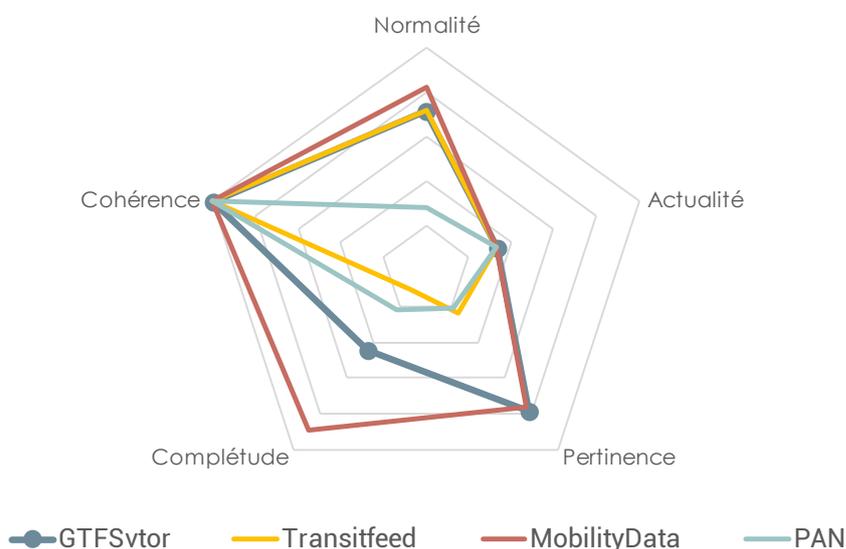
D'une part, les validateurs étudiés prennent en charge de multiples fonctionnalités et évaluent de manière variable les critères de contrôle mentionnés précédemment (Figure 10). La plupart

<sup>54</sup> Un schéma XSD est un fichier qui décrit la forme et la structure des informations attendues par un fichier au format xml.

<sup>55</sup> Source : <http://xmlsoft.org/xmllint.html>

de ces validateurs ont développé des fonctionnalités exhaustives pour évaluer la qualité de la structure des données d'un jeu GTFS, excepté celui du PAN qui ne vérifie que la présence des champs obligatoires selon le standard (situés dans les fichiers obligatoires et certains fichiers facultatifs). Par ailleurs, les validateurs analysent peu le critère *d'actualité* et ne prennent pas en compte *l'accessibilité* puisque ces deux critères tiennent compte de l'analyse des métadonnées des fichiers de données et que ces validateurs n'ont pas été conçus pour assurer cette analyse. Enfin, bien que ces validateurs évaluent la *cohérence* de manière complète, le niveau d'évaluation de la qualité des données diffère sur les critères de *complétude* et de *pertinence*. Notamment, les fonctionnalités ne prennent pas toujours en compte l'absence des champs-clés et de leurs liaisons (*complétude*) et la vérification de tous les formats des champs (*pertinence*).

Figure 10 – Comparaison des validateurs GTFS selon leur niveau de couverture des critères de qualité définis dans les niveaux de contrôle



Note : le critère d'accessibilité n'est pas présenté sur ce graphe car aucun validateur ne le prend en compte.

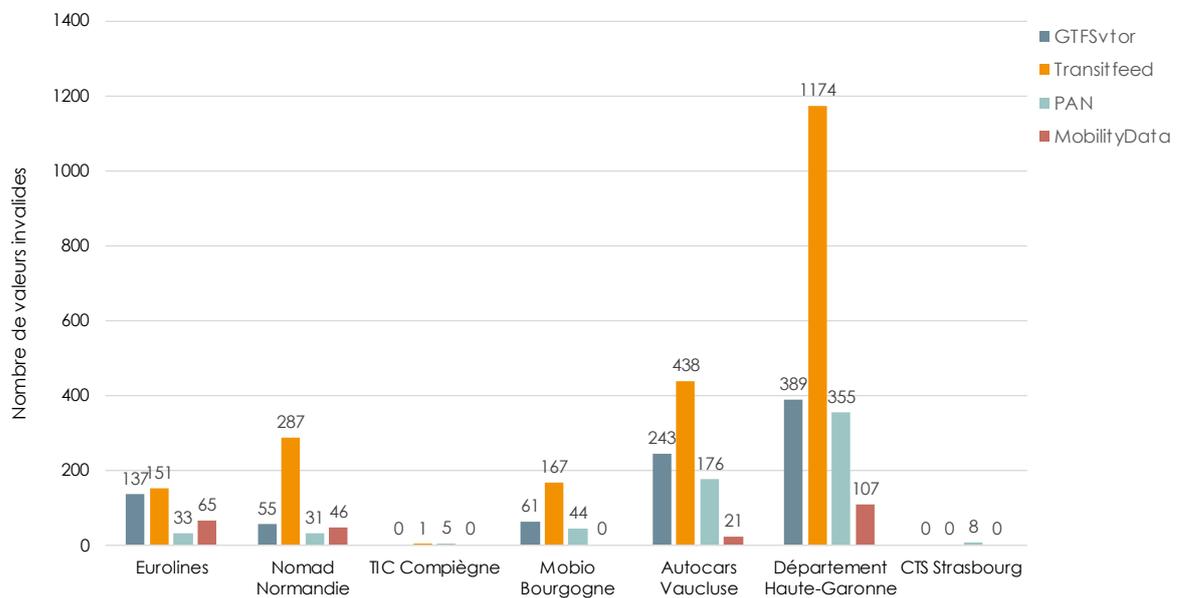
Source : ART.

D'autre part, les différentes fonctionnalités semblent paramétrées différemment d'un validateur à un autre. Ainsi, les fonctionnalités liées à la *cohérence* des données montrent, au travers de l'analyse de 51 jeux de données GTFS téléchargés sur le PAN, un nombre différent de types de valeurs incohérentes détectées par chaque validateur :

- le validateur Transitfeed détecte le plus d'incohérences pour les critères liés au temps de trajet nul et de vitesses excessives<sup>56</sup> (Figure 11) ;
- concernant l'analyse des arrêts (arrêts et zones d'arrêts en double ou proches), c'est GTFSvtor qui détecte le plus de valeurs incohérentes sur tous les jeux analysés ; Transitfeed est le validateur du PAN le moins sensible à la détection de valeurs invalides sur ces paramètres (Figure 12).

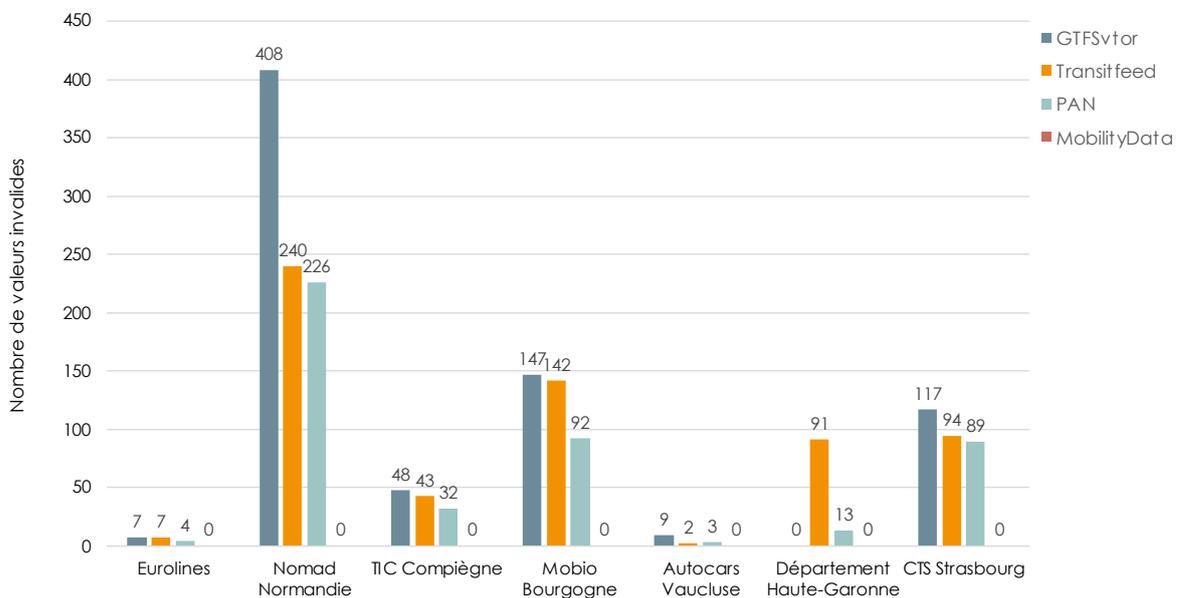
<sup>56</sup> Le premier est calculé en faisant la différence entre les horaires de deux arrêts consécutifs, tandis que le second prend en compte la distance entre ces deux mêmes arrêts pour calculer la vitesse.

Figure 11 – Nombre de valeurs invalides (vitesse excessive / temps de trajet nul) selon les validateurs



Source : ART, à partir de jeux de données GTFS publiés sur le PAN.

Figure 12 – Nombre de valeurs invalides (arrêts / zones d'arrêts en double ou proches) selon les validateurs



Source : ART, à partir de jeux de données GTFS publiés sur le PAN.

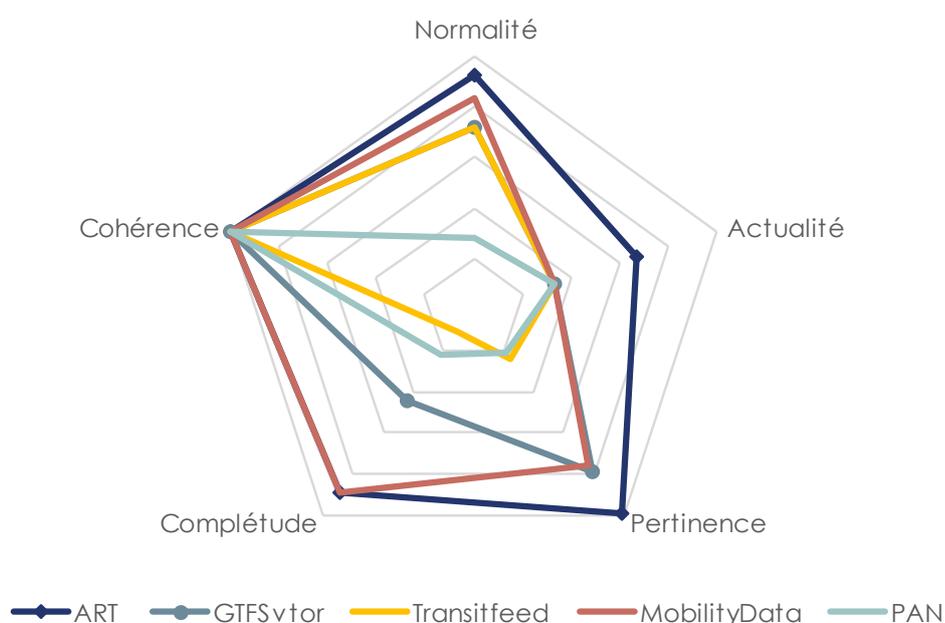
Les résultats de l'analyse comparative des validateurs montrent donc des paramétrages variables pour une même catégorie d'erreur. Afin de rendre compte des différences en matière de définition retenue d'un validateur à un autre, est examiné l'exemple de la détection de la vitesse excessive (vitesse entre deux arrêts consécutifs durant un trajet) qui montre que la valeur d'une vitesse peut être considérée comme invalide par un validateur et valide par un autre.

Ainsi, une vitesse est considérée comme excessive à partir de niveaux différents :

- GTFSvtor considère un avertissement pour une vitesse excessive entre 100 km/h et 300 km/h et comme une erreur une vitesse excessive strictement supérieure à 300 km/h ;
- Transitfeed détecte une erreur dès que la vitesse est supérieure à 100 km/h, d'où le nombre très élevé de valeurs invalides pour l'ensemble des fichiers analysés ;
- Le validateur du PAN définit une vitesse maximale en fonction du mode utilisé. Par exemple, le validateur considère une erreur lorsque la vitesse est supérieure à 120 km/h pour le bus et supérieure à 140 km/h pour le métro.
- MobilityData, tout comme le validateur du PAN, définit une vitesse excessive en fonction du mode utilisé mais les seuils considérés sont différents. Par exemple, la vitesse sera qualifiée d'excessive à partir de 150 km/h aussi bien pour un bus que pour un métro.

**L'étude comparative réalisée montre ainsi les limites des validateurs étudiés, limites dont les producteurs de données doivent être conscients s'ils souhaitent utiliser ces seuls outils pour valider leurs jeux de données**, ces outils ne peuvent constituer qu'un appui à la publication de jeux de données de qualité. En effet, la validation du niveau 1 de contrôle (accessibilité, actualité et normalité) n'est permise que par quelques validateurs et celle du niveau 2 (complétude, pertinence et cohérence) bien que plus largement développée reste partielle. En outre aucun de ces outils ne permet, par conception même, la validation du niveau 3 (exhaustivité et exactitude) qui ne peut reposer que sur un contrôle humain.

Figure 13 – Comparaison du validateur GTFS développé par l'ART par rapport aux validateurs GTFS étudiés selon les critères de qualité



Source : ART.

**En conséquence, l'Autorité a fait le choix de développer en complément ses propres outils de test GTFS et NeTEx**, ainsi qu'une plateforme de contrôle utilisant plusieurs validateurs. Ces outils offrent pour l'Autorité, l'opportunité d'en maîtriser intégralement le paramétrage, et de pouvoir maintenir et faire évoluer aisément le programme à la base de leur fonctionnement. Les outils de test de l'Autorité (nommés « validateurs ART »<sup>57</sup>) couvrent un périmètre fonctionnel permettant une évaluation plus large des niveaux 1 et 2 de contrôle et ainsi un respect plus étendu des exigences définies par le règlement délégué 2017/1926 (Figure 13).

Par ailleurs, l'Autorité a mis en place une plateforme de contrôle permettant l'analyse de multiples jeux de données au prisme des validateurs ART (GTFS et NeTEx) ainsi que des quatre validateurs issus du parangonnage afin de comparer les résultats obtenus. Cette plateforme permet de vérifier aléatoirement et automatiquement la qualité des jeux de données (évaluation de la conformité des jeux de données définie dans les niveaux 1 et 2, seule permise par ces outils).

## **2.5 Les premiers tests automatiques réalisés témoignent d'un relatif respect des exigences de publication et de qualité de la part des producteurs**

Le premier test de contrôle effectué par l'Autorité, qui a pris fin le 1<sup>er</sup> décembre 2021, visait tant à expérimenter en grandeur réelle la plateforme développée qu'à procéder à une première évaluation de la conformité des jeux de données publiés sur le PAN. Cette évaluation s'est focalisée sur la conformité aux exigences de format, point essentiel du dispositif réglementaire, et renvoie aux deux premiers niveaux de contrôle décrits précédemment :

- (niveau 1) conformité de la structure des jeux de données, notamment vis-à-vis du respect de la structure imposée par les standards GTFS et NeTEx ;
- (niveau 2) analyse de la qualité des jeux de données, selon les critères de complétude, de pertinence et de cohérence du deuxième niveau de contrôle.

Le panel de jeux de données contrôlés a été sélectionné sur le PAN de manière aléatoire, au sein d'un échantillon filtré :

- d'abord sur les jeux de données statiques qui constituent le périmètre identifié comme le plus prioritaire dans le règlement délégué 2017/1926, et en particulier les données de transports en commun qui ont fait l'objet d'une des priorités du PAN et qui constituent ainsi la très grande majorité des jeux publiés à la date du premier test de contrôle ;
- puis sur les jeux de données publiés par les AOM ou collectivités locales uniquement.

Le choix de ce second critère repose sur le fait que le PAN met à disposition plus de métadonnées exploitables pour cette catégorie de fournisseurs ce qui permet d'effectuer un test de contrôle plus complet, notamment pour le niveau 1 de contrôle appliqué à ces métadonnées (accessibilité et actualité). A l'inverse, les jeux de données dits « agrégés » (par exemple, les données de plusieurs AOM d'une région sont réunies ou agrégés au sein d'un seul jeu de données par certaines régions) sont exclus de la sélection, car les métadonnées relatives à la validité de ces jeux de données ne sont pas présentes sur le PAN. Au sein de ce panel représentant 100 jeux de données (collectés à la fois en format GTFS et en équivalent NeTEx permettant de juger de la qualité des jeux dans les deux formats<sup>58</sup>), un tirage aléatoire a retenu 67 jeux de données GTFS (et 67 fichiers NeTEx associés), produits par 48 entités locales.

<sup>57</sup> Les validateurs sont disponibles en open-source sur GitHub.

<sup>58</sup> Cependant, la grande majorité des jeux de données NeTEx sont les versions traduites par le PAN de jeux de données GTFS.

2.5.1 L'analyse des métadonnées témoigne d'une description correcte des données de mobilité et d'un bon respect des formats requis, mais d'une actualisation des jeux de données qui n'est pas systématique

**Accessibilité (niveau 1)** - Le processus de publication des données sur le PAN oblige les producteurs à renseigner les métadonnées de chaque jeu de données, dont la plupart doivent être sélectionnées dans une liste fermée de propositions. **Les jeux de données sont donc ainsi aisément accessibles au travers du moteur de recherche du PAN et relativement bien décrits.**

L'Autorité note cependant que certaines pratiques de publication – qui restent toutefois relativement marginales – peuvent réduire l'accessibilité de certains jeux de données. Bien que chaque producteur publie généralement l'ensemble de ses données statiques pour un service de transport en commun au sein d'une page (ou onglet) qui lui est dédiée, l'analyse de l'échantillon de jeux GTFS montre que certaines AOM décomposent leurs jeux de données en fonction de la saison ou de la période du calendrier de service ou par sous-réseau de transport par exemple (Figure 14). Quand bien même les métadonnées de ces jeux de données sont conformes et permettent de les retrouver au sein du PAN, une publication dans des onglets différents rend plus difficile un accès direct au jeu de données valide et à ses différentes mises à jour.

Figure 14 – Capture d'image de la recherche des données d'Alès sur le PAN

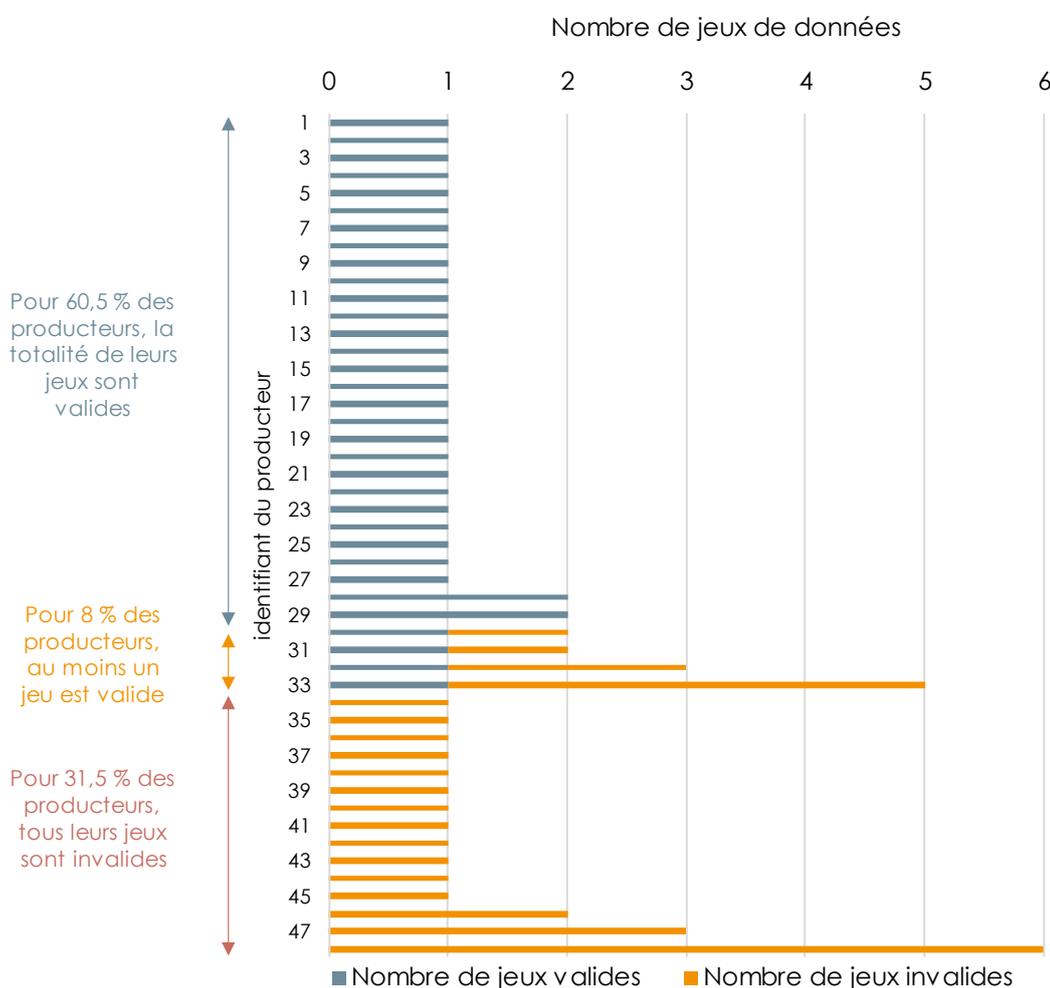
	<b>Réseau interurbain Arc-en-Ciel</b> Occitanie	
ajouté le 2019-08-24	Format <b>GTFS</b>	Transport public collectif - horaires théoriques
	<b>Réseau urbain et interurbain NtecC - Période Eté</b> SMTP du Bassin d'Alès	
ajouté le 2020-07-08	Format <b>GTFS</b>	Transport public collectif - horaires théoriques
	<b>Agrégat de réseaux interurbains de la région Occitanie</b> Occitanie	
ajouté le 2020-02-26		Transport public collectif - horaires théoriques
	<b>Réseau urbain et interurbain NtecC - Période Scolaire</b> SMTP du Bassin d'Alès	
ajouté le 2019-05-15	Format <b>GTFS</b>	Transport public collectif - horaires théoriques
	<b>Réseau interurbain liO Arc-en-Ciel - Haute-Garonne (31)</b> Occitanie	
ajouté le 2019-02-03	Format <b>GTFS</b>	Transport public collectif - horaires théoriques

Source : PAN, capture d'écran le 01/03/2022.

**Actualité (niveau 1)** - Par ailleurs, le critère d'actualité des jeux de données n'est pas toujours respecté (Figure 15). Or, du point de vue de l'utilisateur des données du PAN, la validité du calendrier d'exploitation, ainsi que la fiabilité des métadonnées indiquant les dates de début et de fin de ce calendrier, sont des critères primordiaux pour la réutilisation d'un jeu de données :

- **31,5 % des 48 producteurs contrôlés ne publient aucun jeu de données qui soit valide lors de la réalisation du test de contrôle au 1er décembre 2021 ;**
- 60,5 % d'entre eux ont produit et maintenu un, voire deux jeux valide(s) (i.e. « à jour ») ;
- et les 8 % restants possèdent à la fois un ou plusieurs jeu(x) valide(s) et non valide(s).

Figure 15 – Répartition des 48 producteurs de données sélectionnés selon la validité du calendrier d'exploitation du service de transport

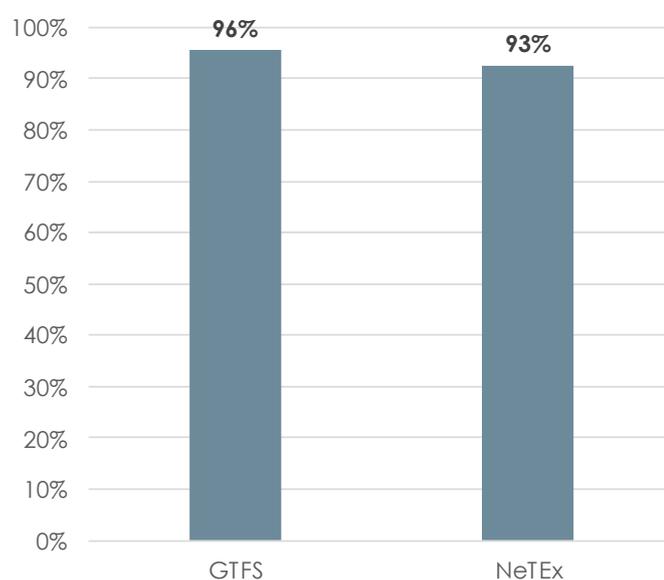


Source : ART, à partir des jeux de données contrôlés.

**Normalité (niveau 1)** - A l'aune du critère de normalité qui vérifie le respect des formats, en analysant à la fois la bonne lisibilité des jeux de données et le degré de conformité des fichiers et des champs inclus dans les jeux de données, l'Autorité a constaté que **certains jeux de données étaient illisibles**.

La lisibilité des jeux de données s'évalue d'abord par la possibilité de lire informatiquement l'ensemble des fichiers les composant. Ces fichiers sont, pour les formats GTFS comme NeTEx, rassemblés dans un fichier global compressé en format « .zip ». Ensuite, la lisibilité s'évalue selon l'exécution plus ou moins complète du programme du validateur. Si la totalité des jeux de données contrôlés ont pu être décompressés, le validateur ART a pu exécuter son programme sans interruption sur 96 % des fichiers GTFS et 93 % des fichiers NeTEx (Figure 16). Ces chiffres peuvent témoigner d'une lisibilité partielle des fichiers évalués<sup>59</sup>.

Figure 16 – Répartition des jeux de données décompressés selon l'aboutissement du processus de contrôle par le validateur ART



Source : ART, à partir des jeux de données contrôlés.

**Toutefois, le degré de conformité des jeux de données demeure élevé. En effet, l'ensemble des fichiers obligatoires selon les spécifications du standard GTFS sont renseignés. Par ailleurs, les champs obligatoires sont inclus de manière quasi-exhaustive au sein de chaque fichier (niveau 1, « normalité »).** Le degré de conformité des jeux de données aux standards GTFS et NeTEx s'évalue quant à lui au travers de la présence complète ou partielle des fichiers et champs obligatoires au sein de chaque jeu. Dans ce cadre, les fichiers et champs des jeux de données examinés peuvent être regroupés en trois catégories :

- fichiers obligatoires et champs clés/obligatoires<sup>60</sup> ;
- fichiers et champs obligatoires sous conditions ;
- fichiers et champs facultatifs

L'analyse des jeux de données publiés au format NeTEx témoigne de leur conformité en ce que l'ensemble des fichiers et balises requis sont renseignés et leur organisation respecte le

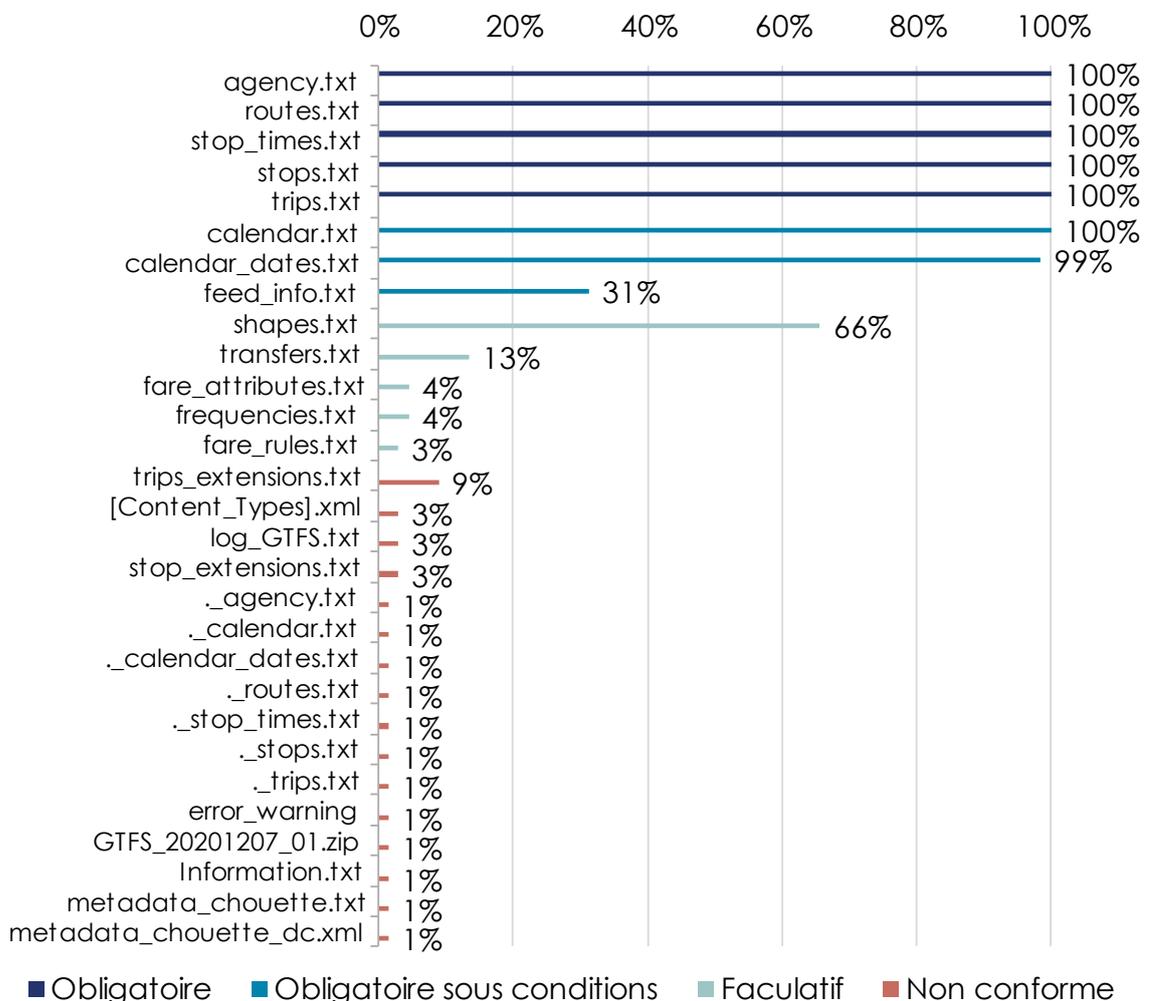
<sup>59</sup> Une partie de ces interruptions pourraient également être dus à des erreurs d'exécution, dont les validateurs ne sont jamais totalement exempts.

<sup>60</sup> Un champ obligatoire est un champ requis pour que le fichier soit utilisable (comme l'heure de passage à l'arrêt dans le fichier horaire), un champ clé est un champ qui permet de faire le lien entre plusieurs fichiers (numéro de ligne pour le faire le lien entre le fichier horaire et le fichier arrêt).

schéma de données requis (schéma XSD). De la même manière, les jeux de données GTFS contiennent l'ensemble des fichiers essentiels à leur réutilisation (Figure 17).

**Précision méthodologique :** la première analyse de normalité présentée ici porte sur le caractère facultatif ou obligatoire des fichiers et des champs selon les spécifications du standard GTFS (voir Annexe 1 2 - Structuration d'un jeu de données GTFS). Cette évaluation ne présage donc pas du caractère obligatoire ou facultatif de la publication de ces champs et de ces fichiers selon les dispositions réglementaires liées à l'ouverture des données. L'analyse de normalité au regard de ces dispositions nécessite ainsi une analyse complémentaire, qui doit être effectuée par un agent, à celle détaillée ci-après. L'analyse de normalité du format NeTEx repose intégralement sur le respect du schéma de données XSD et ne permet pas d'entrer dans le caractère obligatoire ou pas de certaines données.

Figure 17 – Pourcentage de fichiers présents dans les dossiers GTFS

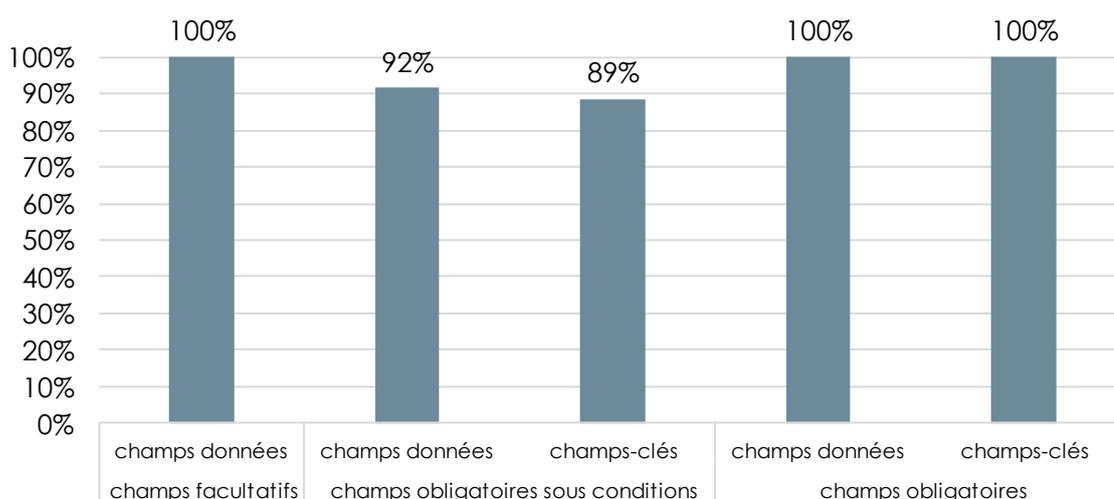


Source : ART, à partir des jeux de données contrôlés.

Au-delà de la présence des fichiers obligatoires, cette évaluation de la normalité permet de noter l'ajout par certains producteurs de fichiers additionnels non exigés ou facultatifs<sup>61</sup>. C'est le cas notamment des fichiers utilisés pour la projection cartographique des informations<sup>62</sup> (66 % des jeux de données contiennent ce type de fichier). Par ailleurs, entre 3 et 4 % des jeux de données intègrent des informations liées à la structure tarifaire. La qualité des jeux de données apparaît en revanche altérée par la présence de fichiers complémentaires mais non conformes aux spécifications GTFS. Cela peut s'avérer problématique à plusieurs égards, notamment dans le cas où ces fichiers contiendraient les données devant être présentes dans un autre fichier, ou si la seule présence de ces fichiers nuit à leur interopérabilité et à leur lecture par des outils informatiques.

Outre l'évaluation de la présence plus ou moins exhaustive des fichiers au sein des jeux de données, le contrôle de la normalité nécessite également d'évaluer la présence des champs composant ces fichiers. **L'essentiel des jeux de données contrôlés semblent conformes à cette exigence en ce que 100 % des champs obligatoires et facultatifs des jeux de données sous format GTFS sont renseignés, dont les champs-clés nécessaires à la liaison entre les fichiers (Figure 18). De surcroît, 98 % des champs de ces fichiers ont un libellé conforme aux spécifications requises par ce format.** Néanmoins, certains champs-clés et champs qui ne sont obligatoires que si certaines conditions sont respectées, sont vides. Ceci est principalement observé pour le fichier contenant les tarifs « fare\_attributes », dans lequel le champ indiquant l'identifiant du transporteur, « agency\_id », peut rester vide si un seul transporteur organise le service de transport (Figure 19). A l'inverse, des champs conformes sont renseignés bien qu'ils ne soient pas nécessaires. C'est le cas par exemple du champ « agency\_id » qui indique l'identité de la compagnie de transport et le nom du réseau. Celui-ci est parfois renseigné alors que le jeu de données ne décrit qu'un seul réseau de transport. L'absence de ces champs obligatoires sous conditions, ou leur présence lorsque cela n'est pas requis, n'entrave cependant pas la qualité de l'information délivrée (cf. statistiques portant sur les contrôles de niveau 2 publiées ci-après)<sup>63</sup>.

Figure 18 – Pourcentage de champs renseignés selon leurs modalités de publication



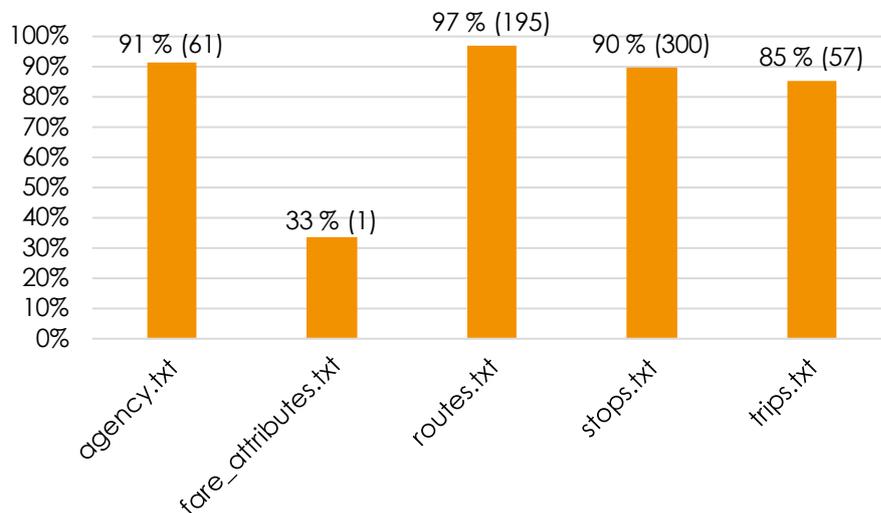
Source : ART, à partir des jeux de données contrôlés.

<sup>61</sup> 55 jeux de données du panel analysé, soit 82 %, comprennent des fichiers facultatifs - voir Figure 45 en Annexe 2.

<sup>62</sup> Fichiers au format « shapes.txt ».

<sup>63</sup> Un contrôle de niveau 3 permet de déterminer si les champs manquants devraient effectivement être présents.

Figure 19 – Pourcentage de champs obligatoires sous conditions renseignés par type de fichier (nombre de champs entre parenthèses)



Source : ART à partir des jeux de données contrôlés.

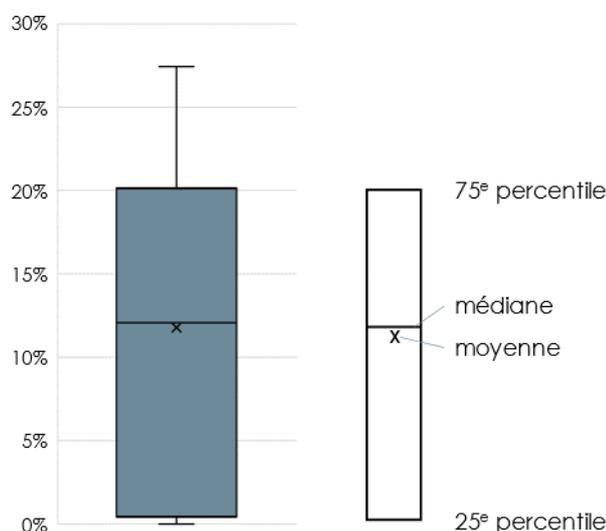
2.5.2 Les jeux de données évalués sont en moyenne très complets, comportent très peu de valeurs incorrectes et peu de valeurs incohérentes, même si des différences de qualité parfois substantielles apparaissent

**Complétude et pertinence (niveau 2)** - Le validateur ART détecte peu de cellules sans valeurs ou dont les valeurs sont incorrectes vis-à-vis du format demandé. Ce constat varie selon les jeux de données, les fichiers et les champs considérés. En effet, certains jeux de données comportent une proportion importante de valeurs manquantes ou incorrectes. **L'évaluation des jeux de données atteste toutefois que sur l'ensemble des jeux de données contrôlés, seulement 8,4 % des cellules ne sont pas renseignées et parmi les cellules remplies, 0,1 % d'entre elles possèdent une valeur incorrecte** (Figure 20). Ainsi les critères de complétude et de pertinence des données semblent dans l'ensemble bien respectés.

**Précision méthodologique** : le critère de complétude est ici considéré de manière binaire au regard de l'absence ou de la présence de valeurs dans les cellules des fichiers. Selon le format GTFS cependant, certains champs peuvent ne pas être renseignés si une ou plusieurs conditions sont remplies ou si l'absence de valeurs peut être considérée comme une donnée. Concernant le format NeTEx, l'analyse de complétude et de pertinence n'est pas effectuée par les outils de validation du schéma de données XSD, et fera l'objet de développements ultérieurs dans les outils de validation.

**Les taux de valeurs manquantes varient fortement selon les jeux de données.** Ainsi, les jeux de données évalués se caractérisent par un taux de données manquantes moyen de 12 %. 75 % des jeux ont un taux de données manquantes inférieur à 20 % et 25 % ont un taux proche de 0 %. Par jeu de données, le taux moyen de données incorrectes de 0,4 % est très faible, 75 % des jeux de données possédant moins de 0,3 % de valeurs incorrectes. Aucun jeu de données ne se démarque par des taux très élevés : le maximum atteignant 27 % de valeurs manquantes et 12 % de valeurs renseignées au format incorrect.

Figure 20 – Distribution des taux de données manquantes des jeux de données



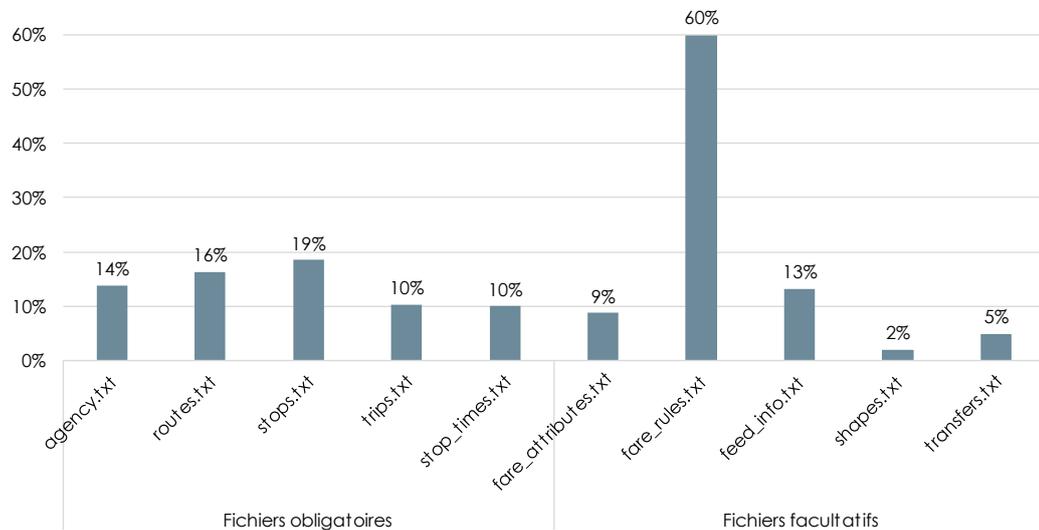
Note de lecture : 50 % des jeux de données ont des taux de valeurs manquantes comprises entre 0,5 % et 20 %, en moyenne les jeux de données ont 12 % de valeur manquante et 50 % des jeux de données ont des taux d'erreur inférieurs à 12,5 % (médiane).

Source : ART, à partir des jeux de données contrôlés.

**La variation observée, selon les jeux de données, en matière de complétude et de pertinence, existe aussi selon les fichiers considérés.** En particulier, les proportions de valeurs manquantes (Figure 21) apparaissent plus faibles au sein des fichiers facultatifs qu'au sein des fichiers obligatoires (excepté pour le fichier « fare\_rules.txt »). Ceci rend compte du souci qu'ont les producteurs d'intégrer un maximum de données dans les fichiers obligatoires. Par ailleurs, si les résultats des contrôles de niveau 1 témoignent d'une publication de données excédant parfois les exigences minimales *via* l'intégration de fichiers facultatifs et un taux de complétude relativement élevé, certains d'entre eux contiennent de nombreuses valeurs manquantes et peuvent ainsi présenter une valeur ajoutée limitée. C'est le cas du fichier « fare\_rules.txt » dont 60 % des cellules ne sont pas renseignées, l'information sur les tarifs pouvant ainsi s'avérer incomplète. Cependant les données manquantes peuvent s'expliquer à la fois par un remplissage incomplet mais aussi par des valeurs nulles (qui peuvent se coder par un champ vide) ou par des données inexistantes pour le producteur, il convient donc d'analyser les taux de données manquantes avec précaution.

De façon similaire, les proportions de valeurs incorrectes (Figure 22), parmi les données complétées, apparaissent plus faibles au sein des fichiers facultatifs qu'au sein des fichiers obligatoires, le respect des spécifications techniques liées aux formats de données restant tout de même très élevé avec peu de valeurs incorrectes.

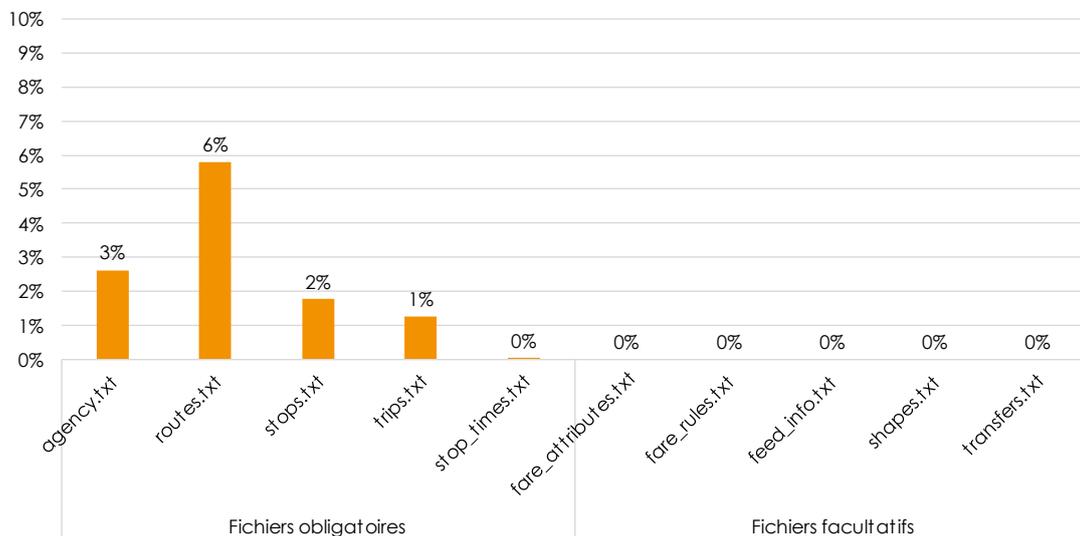
Figure 21 – Pourcentage de valeurs manquantes par type de fichier



Note : les pourcentages sont calculés à partir du total des cellules de l'ensemble des jeux de données.

Source : ART, à partir des jeux de données contrôlés.

Figure 22 – Pourcentage de valeurs incorrectes par type de fichier



Note : les pourcentages sont calculés à partir du total des cellules non vides de l'ensemble des jeux de données.

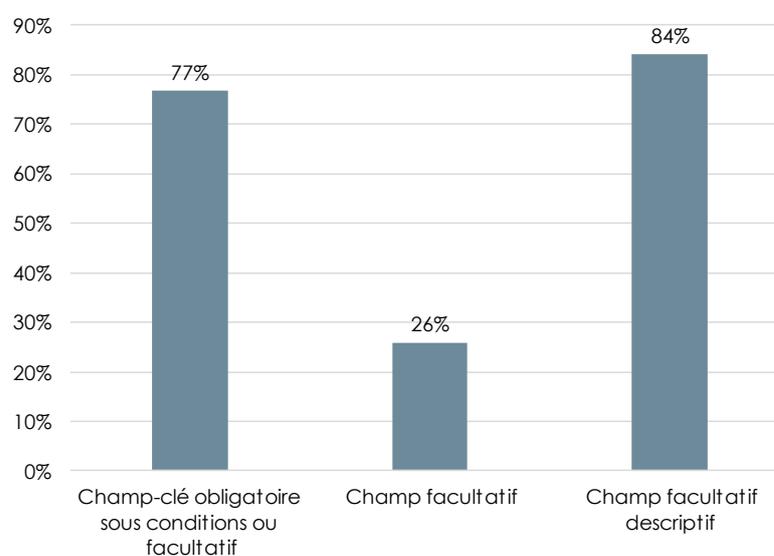
Source : ART, à partir des jeux de données contrôlés.

**Le degré de complétude et de pertinence varie également selon les champs considérés.** Les champs facultatifs sont particulièrement concernés par la présence de valeurs manquantes (Figure 23). C'est le cas par exemple des deux champs facultatifs descriptifs faisant référence à des données en format texte (84 % de valeurs manquantes), mais aussi pour les champs-clés obligatoires sous conditions ou facultatifs (77 % de valeurs

manquantes). Cette dernière catégorie de champs présents principalement dans les fichiers obligatoires peuvent rester vides si les fichiers facultatifs avec lesquels ils font le lien, sont absents du jeu de données, il convient donc d'analyser les taux de données manquantes avec précaution. Les autres champs (champ-clé obligatoire, champ obligatoire et champ obligatoire sous conditions), contenant des données principalement numériques, sont relativement complets.

Enfin, la proportion de données non-pertinentes est quasiment nulle au regard des différentes catégories de champs (le maximum atteint 0,2 % pour les champs obligatoires).

Figure 23 – Pourcentage de valeurs manquantes selon la catégorie de champs



Note : Les catégories de champ « champ-clé obligatoire », « champ obligatoire » et « champ obligatoire sous conditions » n'apparaissent pas sur le graphique car le pourcentage de valeurs manquantes est proche de 0.

Source : ART, à partir des jeux de données contrôlés.

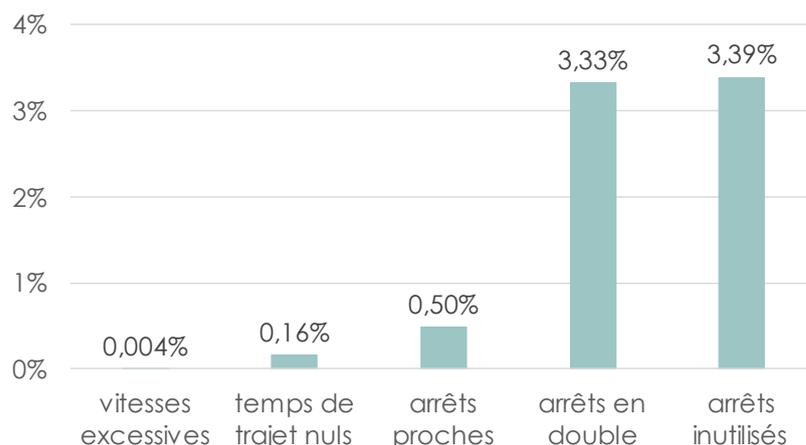
**Cohérence (niveau 2) - Les données de mobilité contrôlées montrent une cohérence élevée. Ce constat varie toutefois selon les jeux de données considérés, quelques rares jeux présentant même des incohérences très élevées.**

**Précision méthodologique** : les catégories d'erreurs présentées ici sont celles considérées par le validateur de l'Autorité. Il convient cependant d'analyser de manière prudente les résultats techniques et les erreurs apparentes associées à certains critères (notamment les critères d'arrêts proches et en double), qui peuvent traduire de réelles incohérences dans les jeux de données, mais également une potentielle inadéquation du format GTFS pour permettre une description complète d'une information de desserte de transport en commun. Par ailleurs d'autres catégories d'erreurs devront être intégrées dans les évolutions du validateur de l'ART afin de compléter l'analyse de cohérence au regard des exigences réglementaires.

Concernant le format NeTEx, l'analyse de cohérence est effectuée par les outils de validation de l'Autorité, mais n'est pas présentée ici dans la mesure où les fichiers NeTEx analysés sont des traductions directes des fichiers GTFS testés et ne présentent donc pas d'écart autre que de formats.

Parmi les données renseignées (i.e. les données issues des cellules non vides), le nombre de valeurs jugées incohérentes du fait notamment de la présence d'arrêts de transport en double ou inutilisés (catégories d'incohérence les plus fréquentes) atteint un peu plus de 3 %. Les autres catégories d'incohérence restent inférieures à 0,5 % (Figure 24).

Figure 24 – Pourcentage de valeurs incohérentes détectées sur l'ensemble des jeux de données (données non vides) selon la catégorie d'incohérence considérée



Note : les pourcentages sont calculés à partir des cellules non vides.

Source : ART, à partir des jeux de données contrôlés.

Tout comme les critères de **complétude** et **pertinence**, les résultats issus de l'évaluation du critère de « cohérence » varient d'un jeu de données à un autre. Certains jeux cumulent par ailleurs plusieurs catégories d'incohérence. En particulier, les incohérences liées à la présence d'arrêts en double ou d'arrêts inutilisés varient fortement selon le jeu considéré. Bien que 50 % des jeux de données ne comportent quasiment aucune valeur invalide toute catégorie confondue, certains jeux de données présentent un taux de valeurs incohérentes de, respectivement, 63 % et 92 % au maximum (Figure 25).

Figure 25 – Statistiques descriptives des taux de valeurs incohérentes par jeu de données selon la catégorie d'incohérence considérée

Catégorie d'incohérence	Minimum	Premier quartile	Médiane	Troisième quartile	Maximum	Moyenne
Vitesses excessives	0,00%	0,00%	0,00%	0,00%	2,64%	0,06%
Temps de trajet nuls	0,00%	0,03%	0,16%	0,34%	2,33%	0,25%
Arrêts proches	0,00%	0,00%	0,00%	0,43%	10,23%	0,73%
Arrêts en double	0,00%	0,00%	0,41%	3,95%	92,82%	4,88%
Arrêts inutilisés	0,00%	0,00%	0,00%	0,17%	63,24%	3,22%

Note : les statistiques sont calculées à partir des cellules non vides.

Source : ART, à partir des jeux de données contrôlés.

Par ailleurs, la corrélation entre les incohérences est forte entre les arrêts en double et les arrêts proches (69 %) (Figure 26). Cela signifie que les jeux de données avec le plus fort taux d'arrêts en double affichent souvent aussi un taux d'arrêts proches élevé.

Figure 26 – Matrice de corrélation par catégorie de validité

	Arrêts proches	Arrêts en double	Arrêts inutilisés
Arrêts proches		0,686	0,411
Arrêts en double	0,686		0,361
Arrêts inutilisés	0,411	0,361	

Note : les statistiques sont calculées à partir des cellules non vides.

Note de lecture : les coefficients de corrélation indiquent le degré de dépendance entre deux variables et sont compris entre 0 (variables indépendantes) et 1 (variables dépendantes). Un coefficient de corrélation supérieur à 0,6 indique une forte corrélation, un coefficient compris entre 0,4 et 0,6 décrit une corrélation moyenne tandis qu'un coefficient inférieur à 0,4 révèle une corrélation faible. Ici, les variables « arrêts proches » et « arrêts en double » sont fortement corrélées (0,686). Elles sont ainsi relativement dépendantes : lorsque les valeurs de l'une augmentent, les valeurs de l'autre suivent la même tendance.

Source : ART, à partir des jeux de données contrôlés.

### 3 LES MARCHES AVANT DE L'INFORMATION SUR LES DÉPLACEMENTS ET LA CIRCULATION : ÉTUDE COMPARATIVE DES CALCULATEURS D'ITINÉRAIRES

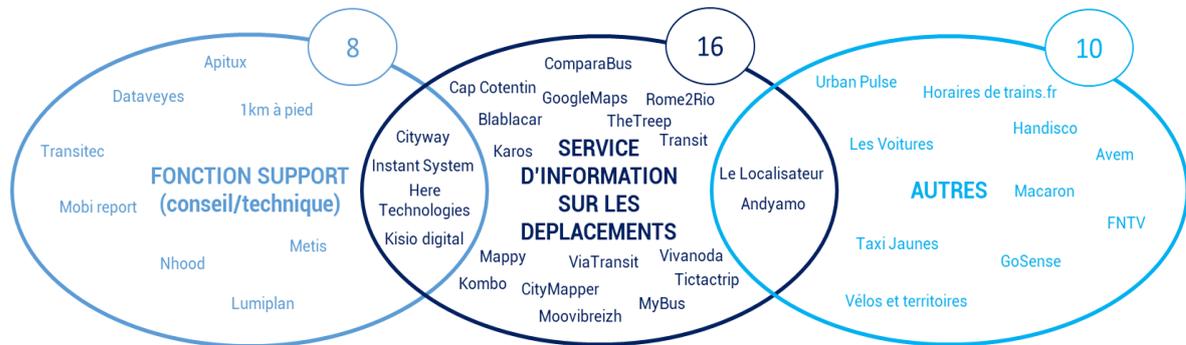
#### 3.1 L'utilisation des données de mobilité publiées sur le PAN s'opère sur des marchés multiples et segmentés

**La mise à disposition des données de mobilité donne lieu à des réutilisations variées, couvrant de multiples marchés.** L'identification exhaustive des utilisateurs des données de mobilité du PAN est complexe, contrairement à celle des producteurs de données. Dans une logique de données ouvertes librement accessibles, le PAN a fait le choix, pour les jeux de données mis à disposition, d'un accès sans identification préalable de l'utilisateur. Dès lors, le recensement des utilisateurs se limite aux seules déclarations de réutilisation fournies sur la base du volontariat et listées sur le PAN.

En 2022, 40 entités se sont identifiées comme utilisatrices des données du PAN. Si la liste n'est pas exhaustive, son étude apporte toutefois un éclairage sur la nature de l'utilisation des données. La variété des usages témoigne de la diversité des marchés sur lesquels évoluent les acteurs recensés. La Figure 27 propose une classification des utilisateurs de données de mobilité du PAN en trois catégories : les « *fonctions support* », les « *services d'informations sur les déplacements* » et une troisième catégorie « *autres* » :

- la catégorie « *fonction support* » regroupe des entités dont l'activité principale consiste en l'accompagnement de leurs clients (entreprises et administrations) dans la mise en œuvre de solutions digitales dans le domaine des transports ou en l'aide à l'élaboration de projets « *Mobility as a Service* ». Il peut s'agir d'éditeurs de logiciels de planifications d'itinéraires (dont Cityway, Instant System, Here Technologies), de sociétés de conseil (l'entreprise 1km à pied propose notamment de diagnostiquer et d'optimiser les trajets domicile-travail des employés des entreprises avec lesquelles elle collabore), mais aussi de fournisseurs d'aide à l'optimisation et à la gestion des données et de l'information voyageurs (Apitux et Mobi report par exemple) ;
- la catégorie « *services d'informations sur les déplacements* » désigne des fournisseurs de services numériques proposant des itinéraires variés et alternatifs pour un trajet donné, à destination d'un usager final. Cette catégorie intègre notamment des entités ne proposant qu'un service d'informations d'une part, et des fournisseurs de services plus intégrés d'autre part, proposant notamment la vente de titres de transports ou la possibilité de réserver des services additionnels (exemple : hôtellerie, tourisme). Une typologie des services d'informations sur les déplacements est proposée dans la partie suivante ;
- la catégorie « *autres* » inclut par exemple des fournisseurs de services à destination des personnes à mobilité réduite ou mal-voyantes (GoSense et Handisco développent des cannes blanches connectées et intelligentes), ou des fournisseurs d'applications à destination des personnes circulant en voiture (Les Voitures et Macaron proposent respectivement des services de localisation de bornes ou stations-services et de places de parking).

Figure 27 – Classification des utilisateurs de données de mobilité auto-déclarés sur le PAN



Source : ART, à partir des utilisateurs auto-déclarés du PAN. Classification à la discrétion des auteurs.

Ces catégories, qui constituent des marchés distincts, ne sont cependant pas totalement disjointes et certains acteurs se positionnent à l'interface de plusieurs d'entre elles. Kisiio Digital et Here technologies développent par exemple des calculateurs d'itinéraires en « marque blanche »<sup>64</sup> (l'outil « Navitia » pour Kisiio Digital par exemple) mobilisables par des acteurs souhaitant proposer un service d'information sur les déplacements. De même, si le service « Le Localisateur » propose pour l'essentiel des services à destination des personnes mal-voyantes (aide à l'accès à Internet via une interface ultra-simplifiée), il inclut également un calculateur d'itinéraire et s'inscrit dès lors en partie dans la catégorie des services d'informations sur les déplacements. Le service Andyamo se rapproche aussi de cette catégorie en délivrant un service de cartographie permettant aux personnes à mobilité réduite d'identifier les itinéraires qui leur sont accessibles dans le périmètre de la ville de Grenoble.

Pour ce premier état des lieux, l'Autorité a fait le choix d'étudier une des réutilisations des données de mobilité les plus consultées par les voyageurs<sup>65</sup> : les services d'informations sur les déplacements et plus particulièrement ceux fournissant un calcul d'itinéraire. D'autres réutilisations de données, comme la comparaison de prix, seront étudiés dans les prochains rapports.

**Constituant un marché au dynamisme croissant et résultant d'initiatives publiques ou privées, les fournisseurs de services d'informations sur les déplacements sont d'importants utilisateurs des données de mobilité mises à disposition sur le PAN.** L'étude des seuls acteurs déclarés rend compte du dynamisme de ce marché en pleine croissance. Parmi eux, plus de la moitié ont été créés au cours de la dernière décennie.

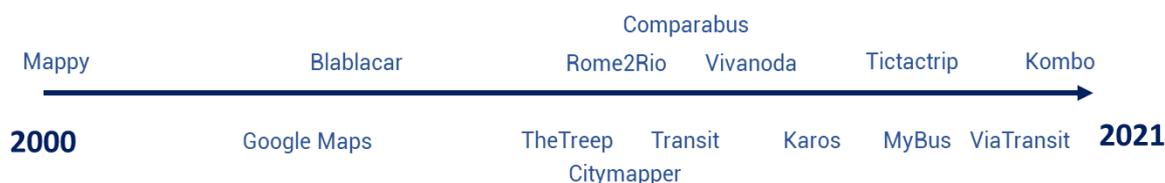
Mappy apparaît comme précurseur (Figure 28). Créée au début des années 2000, l'interface se spécialise dans un premier temps dans l'aide au déplacement *via* des services de cartographie sur Internet, puis diversifie son activité en proposant le calcul d'itinéraire multimodaux. Son principal concurrent, Google Maps, est lancé en 2005 et propose notamment le calcul d'itinéraires en véhicule individuel en temps réel, pour ensuite intégrer une gamme de modes de transport plus large.

<sup>64</sup> Un service (ici calculateur d'itinéraires) est dit fourni en « marque blanche » lorsqu'il est conçu et vendu par une entreprise à une autre, sans que ces dernières n'affichent le nom du fournisseur sur le service final proposé. Il s'agit ainsi d'un service fourni clé en main, que l'exploitant final n'a plus qu'à personnaliser, sans y apporter de nouvelles fonctionnalités.

<sup>65</sup> <https://www.automobile-club.org/espace-presse/communiqués/les-europeens-disent-oui-a-une-application-de-mobilite-unique-regroupant-tous-les-modes-de-transport>

Si ces deux applications jouissent aujourd'hui d'une audience importante<sup>66</sup>, elles évoluent dans un écosystème mouvant où de nouveaux acteurs font leur apparition. Ces derniers proposent des services similaires ou qui s'en différencient horizontalement (offres similaires présentant quelques variations comme la présentation des options ou le paramétrage, mais adressées à une cible de marché équivalente) ou verticalement (offres de services de même type mais adressées à d'autres cibles de marché : offre premium, offre professionnelle).

Figure 28 – La dynamique de création de services d'informations sur les déplacements et la circulation



Source : ART, à partir des utilisateurs auto-déclarés du PAN. Chronologie indicative.

**Différents modes de gouvernance - privés ou publics - co-existent pour le portage des services d'informations sur les déplacements.** S'ils résultent majoritairement pour les services recensés d'initiatives privées (qu'il s'agisse d'opérateurs de transport ou d'autres structures), on peut également identifier quelques portages par des acteurs publics (notamment des collectivités locales et régions), qui en assurent le développement ou font appel à un prestataire extérieur. La quasi-totalité des régions et métropoles ont fait le choix de développer des services d'information multimodale, on constate cependant leur absence parmi les utilisateurs du PAN sans qu'il puisse être déterminé si elles ont fait le choix à la date de publication de ce rapport de ne pas utiliser les données présentes sur le PAN ou d'utiliser ces données sans s'être identifiées sur le PAN.

Figure 29 – Couvertures géographiques de trois services d'informations sur les déplacements



Sources : Site internet de Citymapper, Modalis et Google Maps.

<sup>66</sup> Environ 11 millions d'utilisateurs mensuels, 340 millions de visites par an et 4 milliards d'itinéraires calculées en 2018 pour Mappy (sources Mappy). 36 millions de visiteurs uniques mensuels pour Google Maps ([https://www.challenges.fr/high-tech/calculs-d-itineraires-en-15-ans-google-maps-et-waze-sont-devenus-incontournables\\_699369](https://www.challenges.fr/high-tech/calculs-d-itineraires-en-15-ans-google-maps-et-waze-sont-devenus-incontournables_699369)).

**Le marché des fournisseurs de services d'informations sur les déplacements apparaît également relativement segmenté à plusieurs égards**, par les services proposés, leurs fonctionnalités ou les zones géographiques couvertes. Comme l'illustre la Figure 29, certains services se limitent à une couverture locale (à l'échelle d'une ville et de sa périphérie), d'autres couvrent un périmètre plus large (régional par exemple). Enfin, des acteurs tels Google Maps, Mappy et Rome2Rio déploient leur activité au niveau national, et international.

La couverture géographique est plus ou moins étendue selon les services d'informations sur les déplacements considérés et reflète avant tout les segments de marché visés (Figure 30). Les services dont la couverture est nationale répondent davantage, sans s'y restreindre toutefois, à une demande de planification de trajets longue distance. A contrario, des couvertures géographiques se limitant à l'échelle locale ciblent de fait la planification de trajets courts, voire intra-urbains.

Ainsi, la segmentation observée en matière de couverture géographique va de pair avec une segmentation par type de services de mobilité intégrés dans le service d'informations. La Figure 30 illustre ce constat. Elle permet d'appréhender l'offre de transport proposée par les services d'informations sur les déplacements en la mettant en perspective avec le type de déplacement (longue ou courte distance).

**Figure 30 – Offre de services de transport intégrés et types de déplacement**

		Avion	Ferry	Train	Autocar	TER	Bus	Méto / tramway	Covoiturage	Voiture individuelle	Taxi	Transport à la demande (scooter, vélo, trottinette)	Vélo individuel	Marche
LONGUE DISTANCE	Comparabus	X		X	X	X			X					
	Kombo	X		X	X	X			X					
	Vivanoda	X	X	X	X	X			X					
	theTreep	X		X	X	X								
	Tictactrip			X	X	X			X					
MIXTE	Rome2Rio	X	X	X	X	X	X	X		X	X			X
	GoogleMaps	X		X	X	X	X	X		X	X	X	X	X
	Mappy			X	X	X	X	X	X	X	X	X	X	
COURTE DISTANCE	MyBus					X	X	X				X		X
	Citymapper						X	X		X	X	X	X	X
	Transit						X	X			X	X	X	X
	Moovibreizh						X	X					X	
	Cap Cotentin						X			X			X	X

Note : cette classification n'intègre pas les services monomodaux (voiture uniquement par exemple).

Source : ART, à partir des sites internet des services d'informations sur les déplacements et la circulation.

Les services d'informations sur les déplacements proposant des solutions de déplacement sur de courtes distances se limitent généralement à l'intégration de services de mobilité urbains et des modes actifs (aussi appelés « modes doux », tels que le vélo ou la marche à pied), à l'image de Citymapper et Transit.

A contrario, les services qui proposent une planification d'itinéraires à plus longue distance, considèrent plutôt des modes de transport tels que le train, l'avion et l'autocar. C'est le cas de Comparabus et theTreep qui ont une couverture nationale et proposent des solutions de déplacement en autocar, avion et train essentiellement.

Certains services, enfin, opèrent à différents échelons (national et local), intégrant ainsi une offre de transport plus dense et variée, qui combine à la fois des modes de transport davantage adaptés aux déplacements sur courtes distances et les modes utilisés sur des distances plus longues.

**Outre la diversité des périmètres couverts par les services d'informations sur les déplacements, ces derniers se différencient par les fonctionnalités qu'ils proposent et la diversité des services mis à disposition de l'utilisateur final.** Si l'activité de recherche d'itinéraires constitue le socle commun aux services identifiés, certains proposent également la vente de titres de transport ou d'autres services connexes, intégrant ainsi l'ensemble du processus de planification sur une plateforme unique : information, réservation, achat, validation (différenciation verticale). Ces services se distinguent aussi par la variété des modes de transport qu'ils intègrent et la latitude qu'ils laissent à l'utilisateur dans le paramétrage de leur planification d'itinéraires (différenciation horizontale).

La compréhension de l'écosystème des services d'informations sur les déplacements fournit une grille de lecture sur laquelle capitaliser pour la suite de l'étude. En particulier, les typologies et segmentations de marché identifiées offrent des clés d'analyse essentielles à l'élaboration d'une méthodologie de contrôle de ces mêmes services, notamment en ce qui concerne la comparabilité des calculateurs et des périmètres qu'ils couvrent.

### **3.2 La nature des exigences en matière de réutilisation de données et la diversité des acteurs concernés nécessitent d'examiner le fonctionnement des services fournis par ces derniers**

**Le règlement délégué 2017/1926 précise les exigences auxquelles les utilisateurs de données issues du PAN doivent se conformer.** Outre l'encadrement de la mise à disposition sur le PAN des données de mobilité, le règlement délégué 2017/1926 dresse le cadre de leur réutilisation par des utilisateurs, à des fins multiples telles que la fourniture de services d'informations sur les déplacements, de services de transport à la demande ou de cartes numériques. Le règlement délégué 2017/1926 porte une attention particulière aux fournisseurs de services d'informations sur les déplacements et la circulation (statiques et/ou dynamiques) à destination des utilisateurs et utilisateurs finaux<sup>67</sup>. Ces services d'informations peuvent couvrir un mode unique de transport ou être multimodaux, et suggérer différents itinéraires pour répondre à une demande de planification d'un trajet par un utilisateur final.

L'ensemble de ces utilisateurs, qui incluent des entités publiques ou privées, doivent se conformer plus particulièrement aux exigences spécifiées aux articles 7 et 8 du règlement délégué 2017-1926. **Ils sont ainsi dans l'obligation de réutiliser les données de « manière neutre, sans discrimination ni biais »** (article 8.2) et de renseigner la source des données, ainsi que leur date et heure de mise à jour dans le cas de données statiques renseignées (article 8.3).

Les fournisseurs de services d'informations sur les déplacements doivent se conformer à des exigences additionnelles justifiées par le fait qu'ils peuvent « proposer aux utilisateurs finaux plusieurs options de voyage avec différents opérateurs de transport »<sup>68</sup>. Dans ce cas,

<sup>67</sup> C'est-à-dire les utilisateurs des données du PAN et les utilisateurs d'informations produites à partir des données du PAN. Cf. glossaire pour les définitions détaillées des termes « utilisateurs » et « utilisateurs finaux ».

<sup>68</sup> Considérant 20 du règlement délégué 2017/1926.

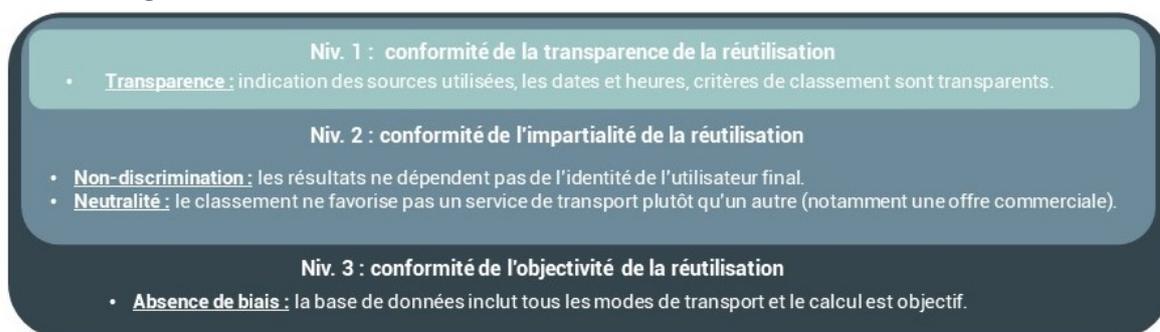
**les critères utilisés pour le classement d'itinéraires doivent – outre le fait d'être transparents et appliqués sans discrimination - ne se fonder « sur aucun facteur directement ou indirectement lié à l'identité de l'utilisateur ou, le cas échéant, à une considération commerciale liée à la réutilisation des données »** (article 8.2). En outre, « *la première présentation d'itinéraires de voyage n[e doit pas] indui[re] l'utilisateur final en erreur* » (article 8.2).

Dans le cas de liaison entre les services d'informations sur les déplacements, afin de fournir notamment des résultats de recherche d'itinéraire plus complets<sup>69</sup>, d'autres exigences s'appliquent (article 7). Ces liaisons peuvent consister en une connexion entre des systèmes d'informations sur les déplacements locaux, régionaux et nationaux. Les résultats de recherche d'itinéraires suggérés – provenant d'interfaces de programmation (API) – sont fondés sur :

- les points de départ et d'arrivée des demandeurs, ainsi que l'heure et la date de départ ou d'arrivée, ou les deux ;
- les options de voyage disponibles ainsi que l'heure et la date de départ ou d'arrivée, ou les deux, y compris les éventuelles correspondances possibles ;
- le point de transfert entre les services d'informations sur les déplacements ;
- si possible, en cas de perturbations, les autres options de voyage disponibles ainsi que l'heure et la date de départ ou d'arrivée, ou les deux, y compris les éventuelles correspondances, le cas échéant.

Ces différentes exigences peuvent être déclinées et hiérarchisées en différents niveaux et critères de contrôle, successifs (Figure 31). Cette démarche permet ainsi de prioriser les manquements aux obligations réglementaires selon une logique d'imbrication en ce que l'évaluation de certains critères constitue une étape nécessaire à l'évaluation des critères suivants. Par exemple, l'évaluation de la transparence des critères de classement des itinéraires est essentielle à la vérification de leur impartialité et objectivité. Le premier niveau se focalise ainsi sur le critère de transparence, comme la transparence des sources de données utilisées ou celle des critères de classement des options de voyage. L'évaluation de la conformité vis-à-vis de ce critère permet d'appuyer le contrôle de deuxième niveau visant à vérifier la non-discrimination et la neutralité de la réutilisation des données<sup>70</sup>. Enfin, le contrôle de niveau 3 vérifie l'absence de biais constituant un critère plus large et englobant ainsi les deux premiers niveaux.

Figure 31 – Schéma d'imbrication des niveaux de contrôle des utilisateurs de données



Source : ART.

<sup>69</sup> Considérant 19 du règlement délégué 2017/1926.

<sup>70</sup> L'imbrication des critères est à envisager comme suit : la vérification d'un critère doit faciliter le contrôle du suivant. Ainsi, il est possible qu'une réutilisation des données puisse être non transparente mais neutre par exemple ; le contrôle de la neutralité est plus pertinent à effectuer une fois le respect de la transparence confirmée.

**Afin de mener à bien les missions confiées à l'Autorité, une étude de l'écosystème des utilisateurs est nécessaire pour délimiter le périmètre des services entrant dans son champ de compétences.** Au titre de l'article 9 du règlement délégué 2017/1926, complété par l'article L. 1115-5 du code des transports, l'Autorité est chargée de s'assurer que les utilisateurs des données publiées sur le PAN se conforment aux obligations qui leur incombent. Ces missions de contrôle de conformité applicables aux utilisateurs nécessitent :

- d'une part, d'appréhender dans son ensemble l'écosystème des utilisateurs afin de mieux en connaître la diversité, en particulier concernant les fournisseurs des services d'informations sur les déplacements singulièrement ciblés par les exigences du règlement délégué 2017/1926. L'intérêt porté à ces acteurs se justifie en particulier par la taille et la dynamique actuelle de ce marché, ainsi que par l'audience qu'il atteint ;
- d'autre part, d'analyser leur manière de réutiliser les données, en particulier au sein des calculateurs d'itinéraires sur lesquels les services d'information reposent généralement en fournissant des solutions de déplacement et aux options de voyages. Ces calculateurs sont souvent basés sur des algorithmes au fonctionnement et aux paramètres opaques (« boîtes noires »), dont la neutralité et l'absence de caractère discriminatoire ne peuvent pas être facilement vérifiées par les usagers. L'encadrement spécifique de ces exigences par le règlement délégué 2017/1926 et la LOM renforce la nécessité d'en comprendre le fonctionnement et les risques associés.

### **3.3 L'étude des calculateurs révèle des fonctionnements hétérogènes, reposant sur des paramètres et des fonctionnalités variés et partiellement opaques**

Afin d'assurer le développement de services d'informations sur les déplacements qui soient de qualité et offrent des informations fiables, le règlement délégué 2017/1926 prévoit que les données de mobilité du PAN soient réutilisées de « *manière neutre, sans discrimination ni biais* ». En particulier, les critères utilisés pour le classement des options de voyages (i.e. des itinéraires) doivent être transparents et ne se fonder sur aucun facteur directement ou indirectement lié à l'identité de l'utilisateur ou à une considération commerciale liée à la réutilisation des données. Compte tenu de ces dispositions, l'étude des calculateurs d'itinéraires, inclus dans les services d'informations sur les déplacements, est orientée de façon à apprécier leur fonctionnement sur la base des résultats proposés aux usagers finaux et sur la manière dont ces résultats sont présentés (classement des voyages et options affichées notamment).

**La compréhension du fonctionnement des calculateurs d'itinéraires exige une analyse préalable de leurs spécificités et périmètres respectifs afin d'évaluer leur comparabilité ou différenciation dans les résultats affichés à des requêtes potentiellement similaires.** L'étude ciblée des calculateurs d'itinéraires de cinq services d'informations sur les déplacements, apporte un éclairage sur leur fonctionnement tout en mettant en lumière leurs similitudes et leurs différences :

- Citymapper, créé en 2011, est présent dans huit villes en France ;
- Vianavigo (aujourd'hui Île-de-France Mobilités) lancé en 2012, est le service d'information de l'autorité organisatrice Île-de-France Mobilités ;
- Modalis, développé en 2018, est disponible en Nouvelle-Aquitaine ;
- Google Maps créé en 2005, couvre l'ensemble du territoire métropolitain et l'international ;
- Rome2Rio, lancé en 2011, est présent sur l'ensemble du territoire métropolitain et à l'international.

La méthodologie employée afin d'étudier le fonctionnement des services d'informations sur les déplacements, et en particulier des calculateurs d'itinéraires sous-jacents, se décline en deux phases.

Une première phase consiste en la définition des périmètres d'analyse pertinents. En effet, l'analyse des fonctionnalités de chaque calculateur, à défaut d'en auditer complètement les algorithmes sous-jacents, peut s'appuyer utilement sur leur comparaison afin d'en dégager les différences et les similitudes interprétables, et de nourrir *in fine* la compréhension de leur fonctionnement. La comparaison de deux calculateurs aux périmètres différents n'apporterait que peu de connaissances sur chacun, faute de pouvoir conclure sur la source des différences ou des similitudes observées : celles-ci pouvant dès lors être imputables à une différence de fonctionnement mais aussi de périmètre.

Une seconde phase consiste en l'analyse des solutions de déplacements affichées aux usagers finaux par les calculateurs étudiés, pour des trajets similaires, à des fins de comparaison.

La définition des périmètres d'analyse s'appuie sur les typologies et les segmentations de marché identifiées et recensées dans la partie précédente. Une attention particulière est portée sur les segments de marché ciblés par chaque calculateur étudié, à savoir le type d'itinéraires proposé (longue ou courte distance) ainsi que le nombre et la nature des modes de transport intégrés. La Figure 32 restitue les informations relatives à chaque calculateur et permet d'identifier les calculateurs pouvant être jugés comparables sur certains de leurs segments.

Figure 32 – Type d'itinéraires et modes de transport proposés par calculateur

Type de déplacements	Calculateurs	Type d'offre	Mode transport				
			Longue distance	En commun urbains	TAD (hors doux)	TAD (doux)	Indivi - duel
Longs ↑ ↓ Courts	Rome2Rio	Courte (peu) et longue distance	X	X	X		X
	Google Maps	Courte et longue distance	X	X	X	X	X
	Citymapper	Courte distance		X	X	X	X
	Vianavigo	Courte distance		X	X	X	X
	Modalis	Courte distance		X	X	X	X

Source : ART, classification à partir des sites internet des services d'information multimodale.

Deux catégories de calculateurs se dégagent :

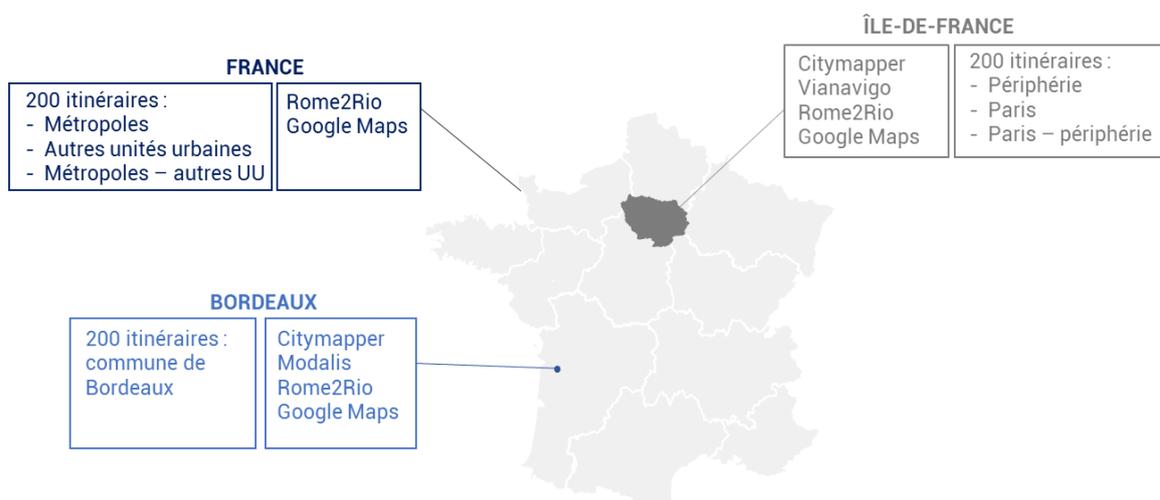
- Rome2Rio et Google Maps se distinguent par leur couverture géographique nationale (internationale si l'on ne se restreint pas au périmètre de la France métropolitaine). Dès lors, ces services intègrent des modes de transport aptes à la longue distance tels que l'avion, le train et l'autocar ;
- Citymapper, Modalis et Vianavigo couvrent quant à eux un périmètre moins étendu : régional ou à l'échelle d'une agglomération.

Ainsi, sur des trajets longues distance, seuls Rome2Rio et Google Maps sont étudiés. En revanche, ces derniers ne se restreignant pas à la planification d'itinéraires longue distance, ils peuvent aussi entrer dans une comparaison avec la deuxième catégorie de calculateurs.

Par ailleurs, la couverture géographique constitue un élément essentiel à la définition des périmètres d'analyse. Trois zones ou périmètres géographiques différents ont été retenus : l'Île-de-France, Bordeaux et la France métropolitaine. Enfin, la délimitation des périmètres d'analyse tient également compte des modes de transport intégrés par les calculateurs. En particulier, les transports en commun constituant un dénominateur commun à l'ensemble des calculateurs étudiés, ceux-ci sont ciblés dans le périmètre d'analyse.

Afin d'analyser le fonctionnement des services d'informations sur les déplacements, les résultats de 200 recherches d'itinéraires, se distinguant par la zone géographique couverte et le service fourni, ont été étudiés (Figure 33). Dans le but de diversifier les requêtes, des trajets variés ont été considérés. En Île-de-France par exemple, des itinéraires Paris-périphérie, périphérie-périphérie et Paris-Paris ont été sélectionnés. Chaque itinéraire est défini par une adresse de départ et une adresse d'arrivée. L'étude s'appuie sur les résultats de recherches d'itinéraires en transport en commun proposés sur les sites internet des services considérés.

Figure 33 – Périmètre et nombre d'itinéraires étudiés



Source : ART.

**Les calculateurs d'itinéraires sont régis par des instructions et des règles - respectivement renseignées par l'utilisateur final et paramétrées par le concepteur du calculateur – qui déterminent les suggestions d'options de voyage finales. Couplé à la présentation de l'interface usager, le choix de ces instructions et règles semble notamment refléter le positionnement initial de chaque service parmi les différents segments de marché. Ces instructions et ces règles sont conçues à des fins d'aide à la décision des voyageurs via le calcul, la prévision, le tri et le classement de solutions de déplacement.**

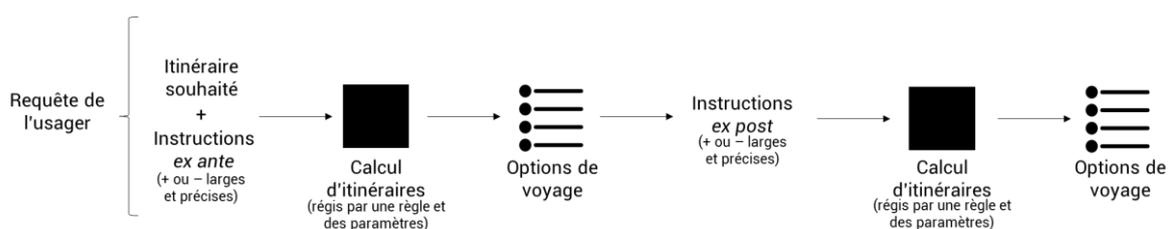
Les *instructions* renvoient à l'ensemble des informations qu'il est possible, pour l'utilisateur final, de renseigner sur l'interface lors d'une recherche d'itinéraires, afin notamment d'orienter la nature et la composition des options de voyage qui lui sont proposées. Outre l'intitulé du trajet renseigné en premier lieu (spécification des points de départ et d'arrivée), des fonctionnalités complémentaires plus ou moins variées peuvent être proposées. Certains services permettent de préciser une heure de départ, la date et l'heure de voyage souhaitée ou de renseigner certaines de ses préférences (minimisation de la durée du trajet, du nombre de correspondances, filtres de modes de transport...). Ces instructions peuvent être spécifiées en amont de la recherche d'itinéraire ou bien sélectionnées (via des filtres) après l'affichage d'une

première série d'options de voyage. L'exemple de Vianavigo présenté en annexe (Annexe 3 – Des instructions ex ante et/ou ex post : l'exemple de Vianavigo) illustre ce point.

Les *règles* sont définies par le concepteur du calculateur d'itinéraire. Il s'agit d'un ensemble de paramétrages orientant la manière dont les options de voyages sont triées/sélectionnées puis classées. En d'autres termes, à chaque calculateur sa règle et les paramètres la caractérisant. A sa discrétion, le fournisseur du calculateur peut en effet choisir de plafonner le nombre de correspondances composant un voyage, sa durée ou le nombre de modes de transport différents qui le composent afin, notamment, de ne sélectionner que les options de voyage satisfaisantes à son sens. Outre la dimension de tri et de sélection, ces règles peuvent donc influencer le classement des options suggérées de telle sorte, par exemple, qu'elles soient présentées dans l'ordre croissant en termes de durée.

Comme le décrit la Figure 34, la combinaison des instructions et règles aboutit *in fine* à la suggestion d'une liste d'options de voyage, en réponse à une requête initiale de l'utilisateur.

Figure 34 – Instructions et règles dans le processus de calcul d'itinéraires



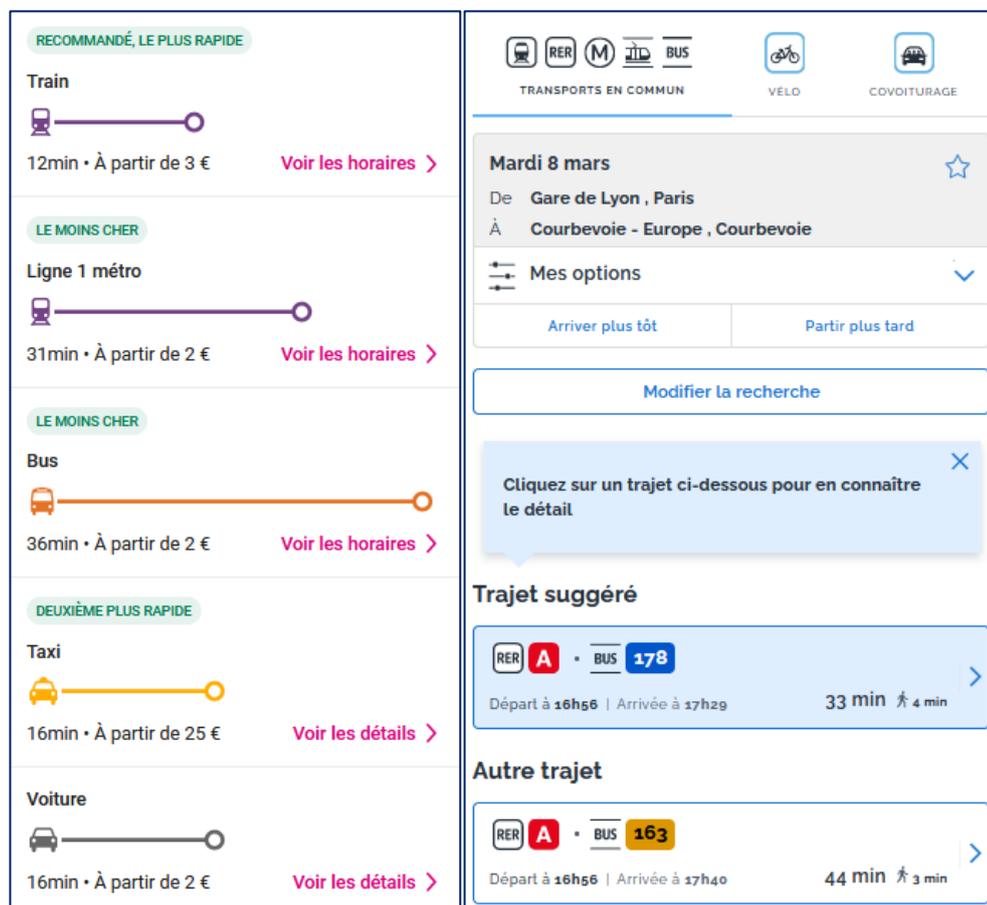
Source : ART.

Ces instructions et règles, mises à disposition de l'utilisateur ou paramétrées par le concepteur, sont plus ou moins variées selon les services d'informations sur les déplacements. Tout comme la présentation de l'*interface usager*, elles reflètent notamment le positionnement initial du service sur le marché et sont ainsi choisies de façon à satisfaire et à répondre à leur segment de marché (multimodal ou monomodal ; local, régional ou national).

Les interfaces usagers, c'est-à-dire les interfaces que ces derniers visualisent et remplissent pour chercher un itinéraire (sur un site internet ou une application), fonctionnent généralement en modules. Autrement dit, elles affichent les résultats de recherche d'itinéraires associés à différents modes de transport dans différents onglets : c'est le cas de Vianavigo, Modalis, Google Maps et Citymapper.

La Figure 35 illustre ce point et montre que Rome2Rio est le seul service parmi ceux analysés dont l'interface ne dispose pas de module, affichant ainsi l'ensemble des itinéraires qu'il propose sur un même onglet sans mettre en avant un mode de transport spécifique. Ces interfaces semblent refléter – du moins partiellement – le positionnement initial sur le marché du fournisseur de services d'informations sur les déplacements. En ce sens, Vianavigo ou Modalis – deux services créés par des AOM – mettent en avant les résultats d'itinéraires en transports en commun, l'utilisateur pouvant naviguer dans les autres onglets pour accéder aux itinéraires à vélo ou en voiture. A contrario, Google Maps suggère en premier lieu des résultats d'itinéraires en voiture.

Figure 35 – Interfaces usagers de Rome2Rio (à gauche) et Vianavigo (à droite) pour l'itinéraire Gare de Lyon – Courbevoie



Sources : Sites internet de Rome2Rio et Vianavigo.

Parce qu'elles reflètent les fonctionnalités mises à disposition de l'utilisateur par l'interface – et sont donc clairement affichées – les options d'instruction sont plus aisément identifiables que les règles suivies par les algorithmes. Leur influence sur les suggestions finales d'options de voyage est par ailleurs plus facilement neutralisable lors du processus de comparaison des calculateurs. Comme à l'instar des périmètres, il est en effet possible de raisonner à instruction égale en n'en spécifiant aucune qui soit de nature à orienter ou filtrer les résultats, à l'exception de celles qui sont indispensables pour effectuer la recherche d'itinéraire (en premier lieu l'intitulé du trajet souhaité). La restitution des résultats qui suit se restreint donc à la qualification des règles inhérentes à chaque service étudié, soit à périmètre constant<sup>71</sup> et instructions neutres et identiques.

**L'identification des paramètres de tri et de classement des résultats de recherche d'itinéraires suggérés par les calculateurs d'itinéraires étudiés met en évidence l'existence de paramètres visibles et invisibles** – parmi lesquels certains peuvent être « déduits » – qui caractérisent les règles du calculateur.

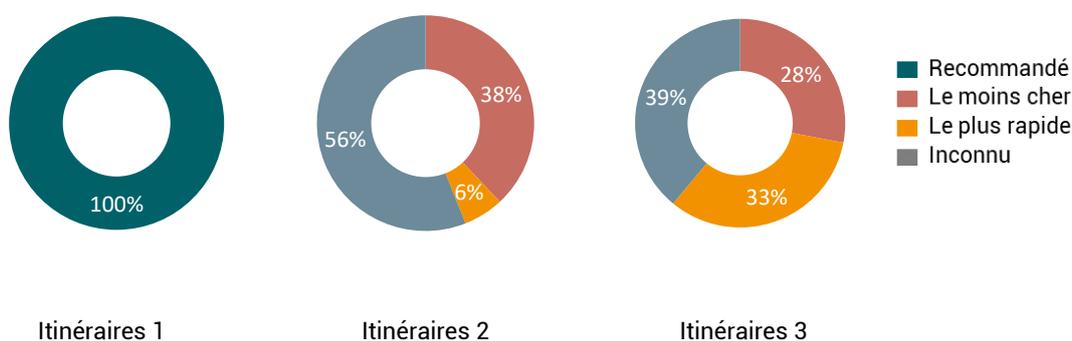
Les *paramètres dits « visibles »* renvoient à des critères de classement clairement affichés par les services d'informations sur les déplacements sur l'interface usager et donc identifiables

<sup>71</sup> L'étude du fonctionnement des calculateurs d'itinéraires considère des trajets et périmètres géographiques identiques afin de garantir une certaine comparabilité entre eux. Voir méthodologie.

par l'utilisateur final. Il s'agit en particulier de « tags » associés à chaque itinéraire suggéré, spécifiant s'il s'agit du « plus court en termes de distance à parcourir », du « plus rapide » ou du « plus confortable » par exemple. L'utilisateur peut ainsi aisément comprendre comment sont classés les itinéraires qui lui sont proposés. La Figure 36 montre comment se caractérisent les itinéraires proposés par Rome2Rio puis Citymapper, selon leur ordre d'apparition.

Figure 36 – Paramètres visibles de classement des itinéraires

Rome2Rio



Citymapper



Source : ART, à partir des résultats d'itinéraires en Île-de-France.

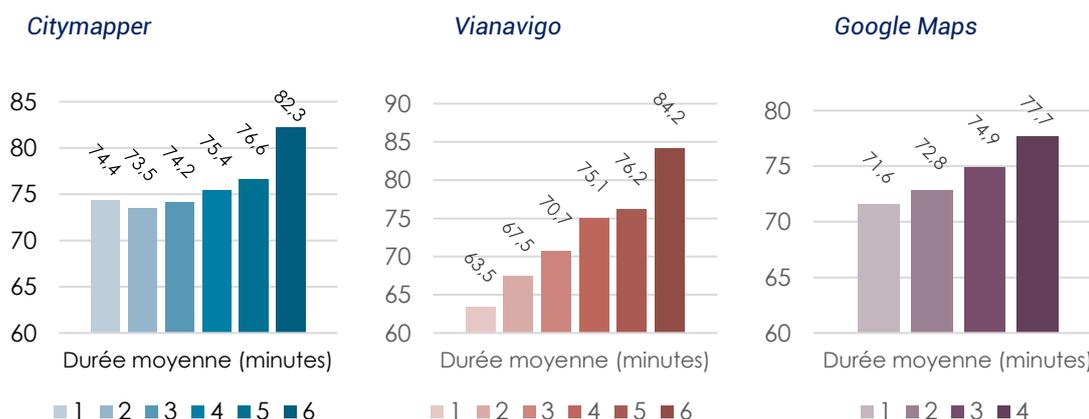
Les premiers itinéraires suggérés par Rome2Rio, pour l'ensemble des 200 trajets étudiés en Île-de-France, sont « tagués » par le paramètre « recommandé ». Les deuxièmes itinéraires présentés correspondent quant à eux, dans 38 % des cas, aux solutions les moins chères. 6 % sont les plus rapides et 56 % ne se caractérisent par aucun « tag » visible. Au vu de ces résultats, la règle sous-jacente au fonctionnement du calculateur Rome2Rio pourrait consister à privilégier les itinéraires les moins chers puis les plus rapides. **Le terme « recommandé » est quant à lui plus difficilement interprétable.**

**Les paramètres dits « invisibles » renvoient à des critères de classement qui ne sont pas renseignés sur l'interface usager et ne sont donc pas visibles par ce dernier.** Ils consistent, et ce de façon similaire aux paramètres « visibles », en des méthodologies de sélection et ordonnancement des itinéraires proposés en résultat des calculateurs. Ces paramètres sont

dans la majorité des cas<sup>72</sup> inaccessibles en pratique pour l'utilisateur en étant masqués de l'interface usager, voire complètement inaccessibles en ne figurant pas dans le code source accessible de l'application. Le classement affiché d'un itinéraire peut donc à la fois être défini par des paramètres visibles mais aussi invisibles, par exemple dans le cas d'un classement de deux options d'itinéraires ayant des paramètres visibles (de prix, durée, distance...) identiques et dont on ne parvient pas à comprendre l'ordre relatif d'apparition.

Certains de ces paramètres « invisibles » peuvent cependant être déduits par le biais d'une analyse statistique simple. En effet, une manière d'appréhender la façon dont sont classés les itinéraires est l'étude des caractéristiques moyennes des itinéraires proposés selon leur ordre d'apparition. Il est ainsi possible d'établir si le premier itinéraire suggéré est, en moyenne, plus court et direct que les itinéraires suivants. Ceci permet d'extraire des tendances de classements des options de voyage et d'établir les paramètres déterminants dans ce processus de classement. Un exemple illustre cette idée dans la Figure 37 indiquant la durée moyenne des itinéraires suggérés par Citymapper, Vianavigo et Google Maps selon leur ordre de classement (c'est-à-dire d'apparition pour l'utilisateur).

Figure 37 – Durée moyenne des itinéraires par calculateur, selon leur ordre d'apparition



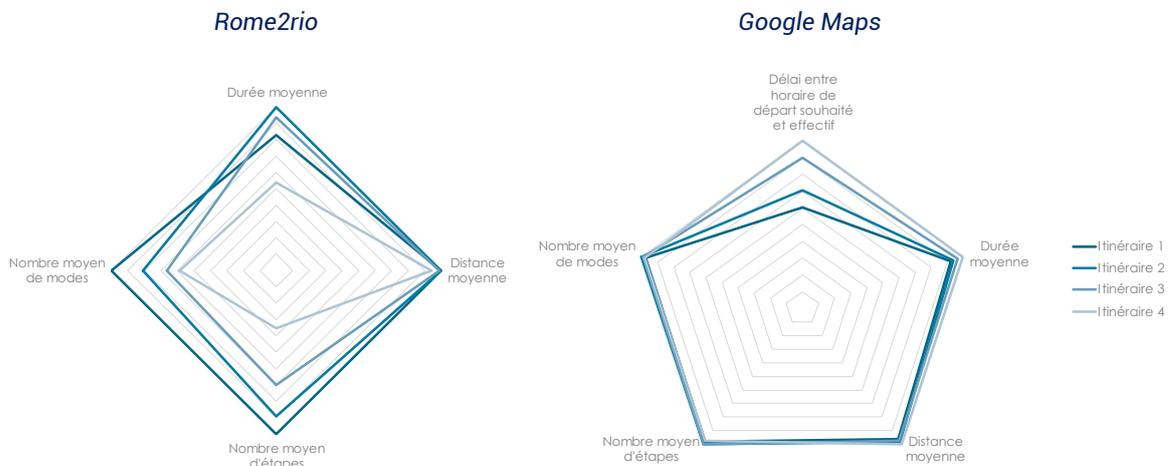
Source : ART, à partir des résultats d'itinéraires en Île-de-France.

**La durée semble être un critère discriminant puisque les itinéraires suggérés en premier sont, en moyenne, plus rapides que les suivants.** Ceci est particulièrement vrai pour Vianavigo et Google Maps. Si cette observation confirme un résultat largement identifié dans la littérature sur les calculateurs d'itinéraires<sup>73</sup>, d'autres critères influencent le classement des options de voyage (Figure 38). Des itinéraires composés d'un nombre de correspondances réduit peuvent par exemple être privilégiés.

<sup>72</sup> Dans le cas de certaines applications, la publication ou non-protection du code source de l'application peut permettre d'identifier – pour un panel cependant très restreint d'utilisateurs finaux - certains des paramètres invisibles sur l'interface de l'application.

<sup>73</sup> Iglesias (2016) Calcul d'itinéraire multicritère en transport multimodal (Thèse de doctorat) ; Scano (2016) Calcul d'itinéraires multiples et de trajets synchronisés dans des réseaux de transport multimodaux (Thèse de doctorat).

Figure 38 – Approche multicritère des itinéraires par ordinateur, selon leur ordre d'apparition



*Note de lecture : chaque axe se lit de manière indépendante et renvoie à un critère de classement potentiel des itinéraires suggérés. Le premier itinéraire proposé à l'utilisateur apparaît ici en bleu et est caractérisé, dans le cas de Google Maps, par un délai entre l'horaire de départ souhaité et effectif plus faible que les itinéraires 2, 3 et 4. Les différents itinéraires proposés par l'interface ne diffèrent en revanche pas en matière de nombre de correspondances. Dans le cas de Rome2Rio, les itinéraires suggérés diffèrent largement en matière de durée moyenne, nombre moyen de modes de transports qu'ils incluent et nombre moyen de correspondances.*

*Note méthodologique : les données de chaque axe ont été normalisées sur une échelle de 0 à 20 par rapport à chaque critère pour l'ensemble des itinéraires. Ainsi, pour chaque critère, plus la valeur est élevée plus le point du graphe tend vers l'extérieur.*

*Source : ART, à partir des résultats de recherche d'itinéraires en Ile-de-France.*

*In fine*, une unique règle ne s'impose pas pour un ordinateur d'itinéraires donné et de multiples critères sont considérés dans le processus de classement des résultats de recherche d'itinéraires. Leur pondération diffère selon les préférences et les objectifs suivis par les fournisseurs de services d'informations sur les déplacements rendant ainsi la comparaison difficile et la mise en évidence d'éventuelles non-conformités complexe.

### 3.4 L'analyse du marché et du fonctionnement des calculateurs, au travers du prisme de la transparence et la non-discrimination, permettra d'orienter l'action de l'Autorité de régulation des transports dans l'exercice de ses missions

Plusieurs enseignements peuvent être tirés des analyses exposées précédemment :

- Si l'opacité partielle des algorithmes des calculateurs d'itinéraires limite les possibilités de pleine analyse de leur fonctionnement interne, une observation et une analyse des interfaces usagers, instructions, règles visibles et règles invisibles (dont déduites) des calculateurs permet cependant d'en faire une analyse ciblée sur le respect de la réglementation en termes de transparence et de non-discrimination ;
- Les options d'instructions proposées par les calculateurs, qui influent sur le tri (la sélection) et le classement des options de voyage proposées, apparaissent très diversifiées au sein de l'ensemble des applications. Les options d'instructions multiplient et diversifient les possibilités de choix des usagers et présentent donc par construction peu de risque de non-transparence ou discrimination. **Ces options peuvent en revanche induire, en cas de mauvaise visibilité ou interprétation, des biais d'usages.** Ce pourrait

être le cas par exemple lorsqu'une pré-sélection de mode, horaire, ou préférence de voyages serait effectuée sur l'interface usager ;

- La diversité des typologies d'*interfaces usagers*, compte tenu du fait qu'elle permet des présentations différenciées et potentiellement orientées des résultats d'itinéraires (par exemple sur des onglets différents), peut également avoir un impact sur la bonne lisibilité des résultats et comporter des risques de biais des usagers finaux ;
- Le caractère « public » des *instructions* et *interfaces usagers* permet cependant facilement une évaluation, une comparaison et un arbitrage direct par les usagers de la qualité de ces calculateurs pour répondre à leur besoin, leur permettant un usage et un choix éclairé du service d'information multimodale le plus adapté à leur besoin ;
- L'analyse des *règles visibles* et *règles déduites* qui influent sur l'affichage et le classement des options de voyage proposées apparaît en revanche plus complexe, mais également plus susceptible de générer des biais de transparence et discrimination. Concernant les *règles visibles* on peut en particulier noter la qualification d'itinéraires « recommandés », dont la bonne interprétation par l'ensemble des usagers peut être questionnée. Concernant les *règles déduites*, on peut remarquer que **les classements suggérés par les calculateurs (i.e. sans instruction supplémentaire des usagers) peuvent se différencier nettement entre applications suivant des critères divers (durée, priorisation de mode, correspondances) qui ne sont pas systématiquement rendus transparents auprès des usagers, et peuvent donc nuire également à leur bonne interprétation** - sans biais - des résultats affichés ;
- L'analyse des *règles déduites* ne constitue qu'une partie - qu'il est possible d'appréhender et rendre ainsi « visible » via l'analyse statistique - des *règles invisibles* pour les usagers qui concentrent par nature une part importante des risques de non-transparence des calculateurs. La présence de *règles invisibles* peut notamment être confirmée par une comparaison des résultats et des classements d'itinéraires sur différents calculateurs à des requêtes similaires. **Les premières analyses de l'Autorité confirment ainsi que des différences de résultats peu ou pas interprétables par des usagers finaux peuvent être observées entre calculateurs, et pourraient témoigner de la présence de biais de transparence voire de cas de discrimination.**

Les travaux présentés dans ce premier rapport éclairent ainsi les manquements susceptibles d'être identifiés. La mise en évidence de paramètres « visibles » parfois peu explicites et de paramètres « invisibles » par nature masqués aux usagers finaux illustrent a minima des manquements à l'exigence de transparence dans les tris et le classement des options de voyage. Dès lors, il semble que **certains fournisseurs de services d'informations sur les déplacements et la circulation ne répondent pas pleinement aux exigences de la réglementation**. La distinction entre les paramètres « visibles » et « invisibles » invite notamment à aborder la notion de transparence sous deux angles : la visibilité et le caractère explicite (Figure 39). La combinaison de ces deux dimensions semble essentielle à la qualification d'un critère de transparence. En d'autres termes, un critère visible n'est pas nécessairement un critère transparent.

Figure 39 – Matrice de qualification d'un critère de transparent ou non



*Note de lecture : un critère à la fois compréhensible et visible peut être qualifié de transparent (en haut à droite). Toutefois, tout critère dérogeant à l'une ou l'autre – ou les deux – des conditions que constituent la visibilité et le caractère explicite ne peut être considéré comme étant transparent. Par exemple, la qualification de certains itinéraires de « recommandés », tout en constituant un critère visible, demeure complexe à interpréter. D'autres critères peuvent être compréhensibles et pour autant non-visibles sur l'interface des résultats (l'analyse du code source d'un calculateur affiche ainsi par exemple un tag de classement « rapide » non affiché sur l'interface usager et permettant d'expliquer des classements relatifs d'itinéraires).*

Source : ART.

Si ces travaux rendent compte de la complexité associée à l'évaluation de la manière dont les utilisateurs de données de mobilité se conforment aux obligations qui leur incombent, ils permettent d'orienter l'action de l'Autorité dans sa démarche de compréhension des services d'informations sur les déplacements et la circulation. Les analyses menées permettent ainsi d'observer de possibles biais « visibles » de transparence ou de manque de neutralité quant aux offres proposées par les services d'information sur les déplacements et la circulation. Ces analyses s'inscrivent bien dans l'intérêt de l'utilisateur final qui ne doit pas être induit en erreur par les solutions proposées et doit pouvoir parfaitement comprendre ce qui lui est présenté. Ces travaux permettent en outre de construire de premières méthodologies et des outils (via le parangonnage de calculateurs notamment) permettant une évaluation de la présence de biais « invisibles » quant aux offres proposées par les services d'information sur les déplacements et la circulation, qui nécessiteront des analyses complémentaires ciblées de l'Autorité et pourront appeler à des clarifications ou des évolutions de l'usage des données de mobilité dans le respect des exigences réglementaires relatives à l'ouverture des données d'intérêt collectif.

## GLOSSAIRE

Libellé	Définition
Autorité chargée des transports	Autorité publique chargée de la gestion de la circulation ou de la planification, du contrôle ou de la gestion d'un réseau de transport ou de modes de transport donnés, ou des deux, relevant de sa compétence territoriale (article 2 du règlement délégué 2017/1926). Les entités concernées par cette définition sont les autorités organisatrices de la mobilité au sens du code des transports, l'Etat, les régions, les départements, les communes, les établissements publics de coopération intercommunale, les autorités désignées à l'article L. 1811-2 du même code (pour les collectivités territoriales de Guyane et de Martinique), les syndicats mixtes, Ile-de-France Mobilités et la métropole de Lyon (L.1115-1 du code des transports).
Carte numérique	Carte géographique permettant de révéler des informations sur des éléments localisés, par exemple les arrêts, la géométrie routière et des horaires de passage associés à ces arrêts. Une carte numérique peut être « matricielle » (c'est-à-dire statique et constituant un simple équivalent d'image numérique d'une carte papier), ou « vectorielle », c'est-à-dire incluant des données sous-jacentes potentiellement en temps réel et pouvant permettre également de la navigation (cas de l'usage pour des appareils GPS).
Calculateur d'itinéraires	Outil informatique reposant sur des algorithmes de calcul et permettant de déterminer un ensemble d'itinéraires pertinents en réponse à une requête de recherche d'itinéraires.
Données de mobilité	Données sur les déplacements et la circulation définies à l'annexe du règlement délégué 2017/1926.
Fournisseur de service de transport à la demande	Fournisseur public ou privé de service de transport à la demande à des utilisateurs et des utilisateurs finaux, y compris les informations correspondantes sur les déplacements et la circulation (article 2 du règlement n° 2017/1926).
Fournisseur de services de partage de véhicules, de cycles et d'engins de déplacement personnel	Fournisseur mettant à disposition un service de partage de véhicules (voitures, scooters) ; cycles, engins de déplacement personnel (tels que les différents modèles de trottinettes et patinettes électriques, gyropodes ou monoroues) sans stations prédéfinies (« free-floating »).
Gestionnaire d'infrastructure	Organisme public ou privé ou entreprise chargé(e) notamment de l'établissement et de l'entretien de l'infrastructure de transport ou d'une partie de celle-ci (article 2 du règlement délégué 2017/1926).
Informations sur les déplacements et la circulation	Informations créées à partir de la réutilisation des données de mobilité (par exemple, les itinéraires suggérés à partir d'une requête).
Itinéraire	Résultat fourni par un service d'information sur les déplacements en réponse à une requête de recherche consistant en une description détaillée et horodatée d'un parcours d'un lieu de départ à un horaire donné vers un lieu d'arrivée à un horaire donné.
Liaison de services	Connexion de systèmes d'informations sur les déplacements locaux, régionaux et nationaux qui sont reliés par des interfaces techniques afin de fournir des résultats de recherche d'itinéraire ou d'autres résultats provenant d'interfaces de programmation (API), fondés sur

	des informations statiques et/ou dynamiques sur les déplacements et la circulation (article 2 du règlement délégué 2017/1926).
Métadonnée	Description structurée du contenu des données qui aide à les retrouver et à les utiliser (article 2 du règlement délégué 2017/1926).
Opérateur de système d'aide à l'exploitation et à l'information des voyageurs	Opérateur en charge d'un système de localisation en temps réel des véhicules de transport en commun.
Norme	Au sens de la réglementation européenne, standard validé par le Comité européen de normalisation (CEN) dans le cas des standards de publication des données de mobilité.
Opérateur de transport	Entité publique ou privée responsable de l'entretien et de la gestion du service de transport (article 2 du règlement délégué 2017/1926)
Point d'accès	Interface numérique qui donne accès au moins aux données statiques sur les déplacements ou aux données historiques concernant la circulation, avec les métadonnées correspondantes, en vue de leur réutilisation par les utilisateurs, ou aux sources et métadonnées de ces données.
Producteurs de données	Ensemble regroupant les autorités chargées des transports, les opérateurs de transport, les gestionnaires d'infrastructure, les fournisseurs de services de transport à la demande, au sens du règlement délégué 2017/1926, et les fournisseurs de services de partage de véhicules, cycles et engins de déplacement personnel (article L. 1115-1 du code des transports).
Réseau transeuropéen de transport global	Infrastructure de transport qui fait partie du réseau global tel que défini dans le règlement (UE) 1315/2013 (article 2 du règlement délégué 2017/1926).
Requête de recherche d'itinéraire	Ensemble des informations renseignées par un usager final (trajet souhaité, horaire de départ, mode de transport souhaité, etc.) sur l'interface d'un service d'information sur les déplacements, afin d'obtenir des suggestions d'itinéraires pour un trajet donné.
Réutilisation des données	Exploitation des données du PAN par un utilisateur afin de créer une information sur les déplacements et la circulation à destination d'autres utilisateurs, comme les usagers des services de transport. Elle se distingue de l'utilisation dont l'information créée est destinée à l'utilisateur lui-même.
Service de transport	Service de transport public ou privé ou service disponible pour une utilisation collective ou privée par le public et couvrant différents modes de transport (article 2 du règlement délégué 2017/1926).
Standard	Publication de lignes directrices qui fournissent des spécifications techniques pour les produits, services et processus utilisés dans un champ donné (par exemple le standard GTFS pour les données de mobilité). Certains standards « de fait » peuvent être transformés en standards « de droit » en devenant des normes (par exemple le format de fichier PDF, devenu norme ISO en 2008).
Service d'informations sur les déplacements et la circulation	Service de transport intelligent (STI), comprenant des cartes numériques, qui fournit aux utilisateurs et aux autres utilisateurs finaux des informations sur les déplacements et la circulation (des suggestions d'itinéraires) pour au moins un mode de transport (article 2 du règlement délégué 2017/1926).
Systèmes de Transport Intelligents (STI)	Systèmes dans lesquels des technologies de l'information et de la communication sont appliquées, dans le domaine du transport routier, y compris les infrastructures, les véhicules et les usagers, et dans la

	<p>gestion de la circulation et la gestion de la mobilité, ainsi que pour les interfaces avec d'autres modes de transport; ce sont des applications avancées qui visent à fournir des services innovants liés aux différents modes de transport et à la gestion de la circulation et qui permettent à différents utilisateurs d'être mieux informés et de faire un usage plus sûr, plus coordonné et plus « intelligent » des réseaux de transport (Directive 2010/40/UE). Les STI associent les télécommunications, l'électronique et les technologies de l'information à l'ingénierie des transports afin de planifier, concevoir, exploiter, entretenir et gérer des systèmes de transport. Par exemple, ces systèmes sont capables de générer des données grâce à des technologies de géolocalisation des véhicules, des capteurs utilisés dans la gestion de la fréquentation des transports en commun, etc., et de communiquer entre eux. A partir des données dont ils disposent ou en intégrant des données externes, les STI génèrent des flux d'informations permettant une analyse de la mobilité et la prévision du comportement des usagers.</p>
Transport à la demande	<p>Service de transport de voyageurs qui se caractérise par la flexibilité des itinéraires et des horaires, tels que le covoiturage, le partage de voitures, les vélos en libre-service, le partage de trajets, les taxis, les services de trajet à la demande. Ces services requièrent normalement l'interaction préalable du fournisseur de transport à la demande et des utilisateurs finaux (article 2 du règlement délégué 2017/1926).</p>
Utilisateur de données	<p>Entité publique ou privée qui utilise le « point d'accès national », telle que les autorités chargées des transports, les opérateurs de transport, les fournisseurs de services d'informations sur les déplacements, les producteurs de cartes numériques, les fournisseurs de services de transport à la demande et les gestionnaires d'infrastructure.</p>
Utilisateur final	<p>Une personne physique ou morale, qui a accès aux informations sur les déplacements.</p>
Valideur de données	<p>Programme informatique visant à vérifier la conformité d'un jeu de données aux exigences définies en paramètres. Par exemple, un valideur couramment utilisé est le correcteur orthographique qui vérifie la conformité d'un mot écrit à partir du dictionnaire qu'il a en mémoire.</p>



## ANNEXES

### **L'utilisation des formats de données** 62

Exemple de lecture d'un jeu de données  
GTFS et NeTEx 62

Structuration d'un jeu de données GTFS 66

### **Précisions sur l'état des lieux de l'ouverture des données** 68

Couverture géographique des données statiques  
des transports en commun selon le standard utilisé 68

Les jeux de données GTFS  
incluent fréquemment des fichiers « facultatifs » 68

### **Des instructions *ex ante* et/ou *ex post* : l'exemple de Vianavigo** 70

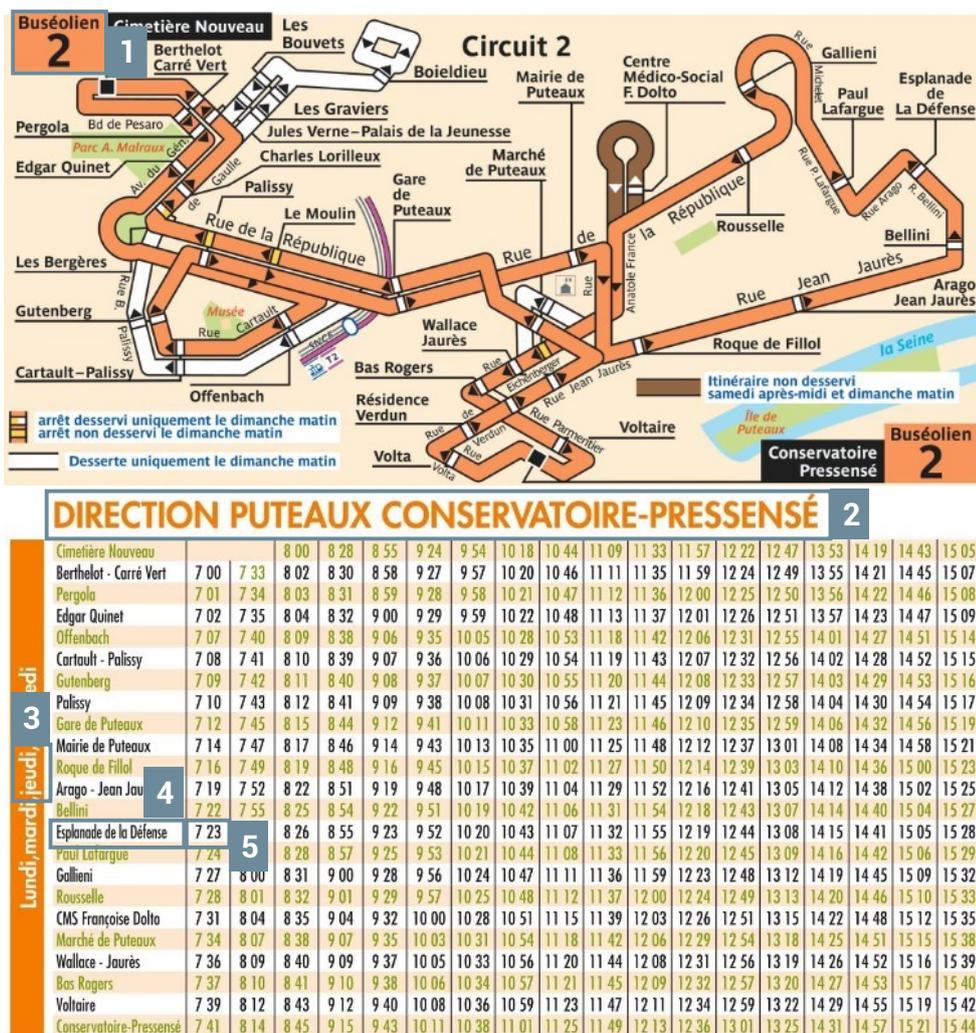
## Annexe 1 - L'utilisation des formats de données

### 1 - Exemple de lecture d'un jeu de données GTFS et NeTeX

Afin de mieux appréhender le fonctionnement des standards présentés, l'exemple suivant permet de comparer la lecture d'une fiche horaires papier et de fichiers GTFS et NeTeX. L'exercice consiste en la recherche d'une information précise, à savoir : à quelle heure passe le premier bus (ligne 2) en direction de Puteaux Conservatoire-Pressensé, le jeudi 13 janvier 2022 à l'arrêt Esplanade de La Défense ?

### Lecture de la fiche horaire papier du service de transport Buséolien de Puteaux

Figure 40 – Capture d'image de la fiche horaire de la ligne 2 « Buséolien » de Puteaux



Source : ART, à partir des fiches horaires publiés sur le site de la ville de Puteaux.

Pour répondre à cette question, la fiche horaires indique (Figure 40) :

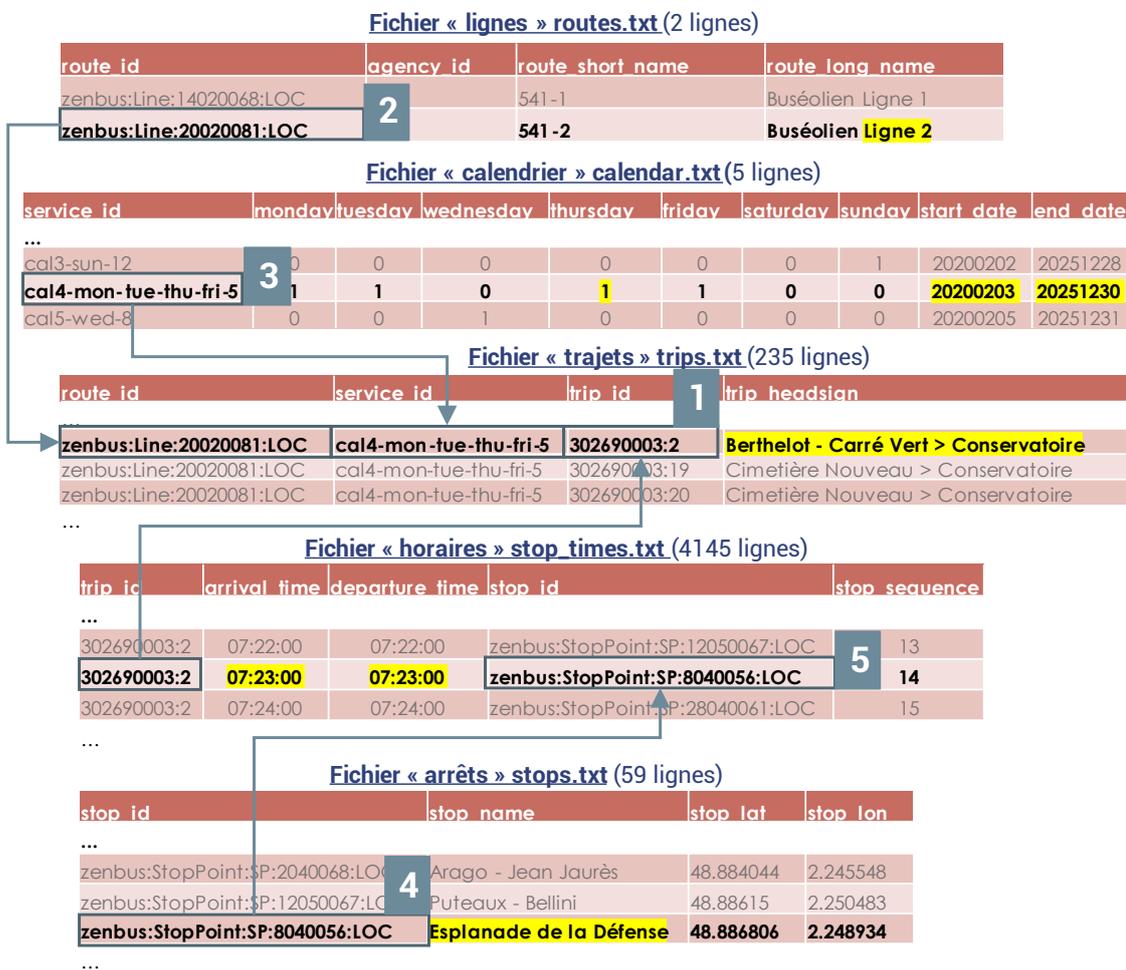
- (1) La ligne 2 du service « Buséolien »
- (2) en direction « Puteaux Conservatoire-Pressensé ».
- (3) passant le jeudi
- (4) par l'arrêt « Esplanade de la Défense »
- (5) à 7 h 23 pour le premier passage.

#### **Lecture du jeu de données GTFS décrivant le service de transport Buséolien de Puteaux.**

Un jeu de données GTFS se compose de plusieurs fichiers, chacun renvoyant à un élément précis du service de transport. Dès lors, répondre à la question posée dans cet exemple nécessite de combiner plusieurs fichiers : en particulier les fichiers *lignes* (« routes.txt »), *arrêts* (« stops.txt »), *calendrier* (« calendar.txt »), *horaires* (« stop\_times.txt ») et *trajets* (« trips.txt ») présentés dans la Figure 41. La combinaison de ces fichiers se réalise grâce à l'utilisation d'identifiants communs (ou clés primaires) présents dans les champs-clés des fichiers. En particulier, le fichier *trajets* permet la jonction entre les fichiers *calendrier*, *lignes* et *horaires* à partir des variables clés « route\_id », « service\_id » et « trip\_id ». Ainsi, l'horaire de passage du bus recherché est indiqué par le jeu de données GTFS de la manière suivante :

- (1) le trajet n° 302690003:2 en direction de « Conservatoire », correspondant à
- (2) la ligne 2 et
- (3) à la circulation valable le jeudi entre le 23/02/2020 et le 30/12/2025,
- (4) dont un de ces arrêts est « Esplanade de la Défense »
- (5) pour lequel il passe à 7h 23.

Figure 41 – Extraits des fichiers du jeu de données GTFS du service Buséolien de Puteaux



Note : Certains fichiers, champs et lignes ont été tronqués pour une question de lisibilité.

Source : ART, à partir du jeu de données « Buséolien » GTFS publié sur le PAN.

### Lecture du jeu de données NeTEx décrivant le service de transport Buséolien de Puteaux

La lecture du jeu de données NeTEx, décrivant le même service de transport, permet de retrouver la même réponse (Figure 42). Néanmoins, la constitution du jeu de données en un fichier « .xml » principal, qui décrit l'offre de transport par ligne (« offre.xml »), et de plusieurs fichiers secondaires, apportant les informations sur le calendrier (« calendrier.xml ») et sur les arrêts (« arrêts.xml »), implique une lecture différente par rapport à un jeu de données GTFS. A l'inverse des tables du jeu de données GTFS distinguant les éléments du service de transport, l'ensemble de ces éléments sont généralement inclus ici dans un fichier unique. Dans le cas du service « Buséolien », le fichier offre est divisé en parties lignes, services, horaires, etc. Chaque partie contient des clés primaires permettant la liaison entre les parties du même fichier mais aussi avec les fichiers secondaires. Dès lors, la lecture du jeu de données NeTEx indique que :

- (1) la ligne 2 en direction de « Conservatoire »,
- (2) dont une circulation le jeudi et

- (3) (2 bis) valable entre 27/01/2020 et le 04/01/2026,
- (4) possède un arrêt
- (5) (3 bis) « Esplanade de la Défense »
- (6) où le premier bus passe à 7 h23.

Figure 42 – Extraits des fichiers du jeu de données NeTEx du service « Buséolien » de Puteaux

### Fichier XML principal

Offre ligne 2.xml (32271 lignes)

Partie « lignes »...

```

<Route id="FR:Route:zenbus_Line_20020081_LOC:" version="any">
  <Name>Cimetière Nouveau - Conservatoire</Name>
  <Distance>0</Distance>
  <LineRef ref="FR:Line:zenbus_Line_20020081_LOC:">
  </LineRef>
  <DirectionType>inbound</DirectionType>

```

**1**

...sous-partie « services »...

```

<ServiceJourney id="FR:ServiceJourney:302690003_20:" version="any">
  <dayTypes>
    <DayTypeRef ref="FR:DayType:cal4-mon-tue-thu-fri-5:">
    </DayTypeRef>
  </dayTypes>

```

**2**

...sous-partie « arrêts »...

```

</PassengerStopAssignment>
  <PassengerStopAssignment
id="FR:PassengerStopAssignment:302690003_10_14:" order="15" version="any">
    <StopPlaceRef
ref="FR:StopPlace:Navitia_zenbus_StopPoint_SP_8040056_LOC_bus:">
    </StopPlaceRef>
  </PassengerStopAssignment>

```

**3**

...sous-partie « horaires ».

```

<TimetabledPassingTime>
  <ArrivalTime>07:23:00</ArrivalTime>
  <DepartureTime>07:23:00</DepartureTime>
</TimetabledPassingTime>

```

**4**

## Fichiers XML secondaires

### Calendriers.xml (75 lignes)

```
<ValidBetween>
  <FromDate>2020-01-27T00:00:00+00:00</FromDate>
  <ToDate>2026-01-04T23:59:59+00:00</ToDate>
</ValidBetween>
<UicOperatingPeriod id="FR:UicOperatingPeriod:cal4-mon-tue-thu-fri-5:"
version="any">
```

2 bis

### Arrets.xml (1698 lignes)

```
<Quay id="FR:Quay:zenbus_StopPoint_SP_8040056_LOC:"
version="any">
  <Name>Esplanade de la Défense</Name>
  <Centroid>
    <Location>
      </Location>
    </Centroid>
  <TransportMode>bus</TransportMode>
</Quay>
```

3 bis

Note : les lignes des fichiers ont été tronquées pour une meilleure lisibilité.

Source : ART, à partir du jeu de données « Buséolien » NeTeX publié sur le PAN.

## 2 - Structuration d'un jeu de données GTFS

Un jeu de données GTFS est composé de plusieurs fichiers en format « .txt » dont chacun d'eux renferme les données, sous la forme d'une table de données, sur un élément précis d'un service de transport en commun : les lignes (« routes.txt »), les horaires (« stops\_times.txt »), etc. Le standard GTFS définit dans quelle mesure ces fichiers ont l'obligation d'être publiés reflétant ainsi le caractère essentiel ou non des données qui y sont stockées. Tout d'abord, un jeu de données GTFS doit contenir 5 fichiers « obligatoires » (figure X) sans lesquels un utilisateur ne pourrait obtenir une information sur les déplacements et la circulation. Ensuite, d'autres fichiers sont définis comme « obligatoires sous conditions », dont les fichiers « calendar.txt » et « calendar\_dates.txt ». Leur présence est conditionnée notamment par la complexité du service de transport décrit. Par exemple, le fichier « calendar.txt » peut suffire pour la description d'un calendrier d'exploitation dont le service est régulier, tandis que l'utilisation de « calendar\_dates.txt » est souhaitable lorsque le calendrier possède de nombreuses exceptions. Enfin, la publication en GTFS autorise l'ajout de fichiers « facultatifs » permettant d'enrichir le jeu de données.

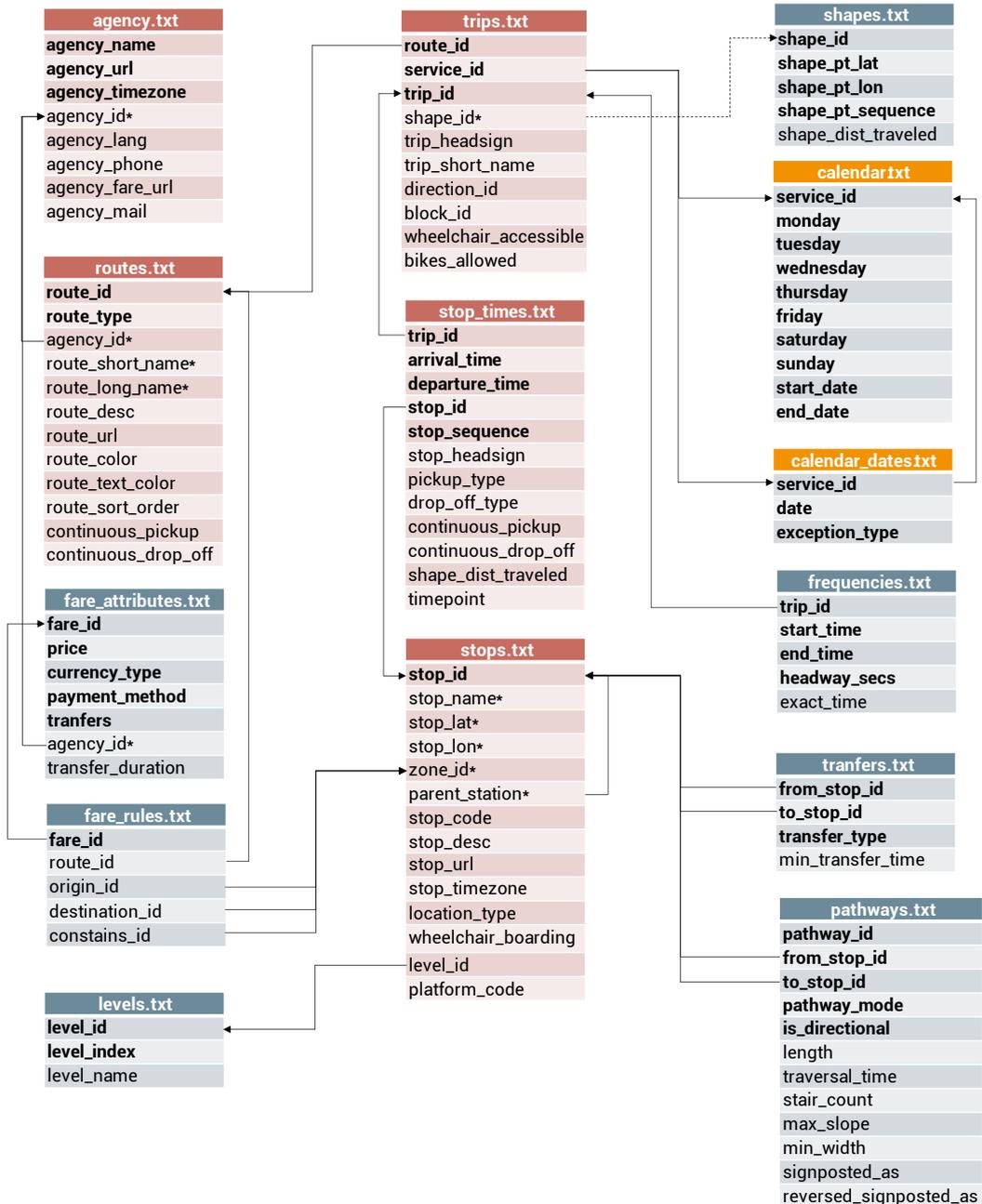
Par ailleurs, chaque fichier est divisé en champs (colonnes dans la table de données) et à l'instar des fichiers, le standard GTFS définit des champs « obligatoires », « obligatoires sous conditions » et « facultatifs ». Parmi eux, les « champs-clés » ont la particularité de relier les fichiers entre eux et de rendre possible la combinaison des données afin de construire l'information sur les déplacements<sup>74</sup>. Chaque champ-clé d'un fichier possède son équivalent dans un autre fichier et par les clés-primaires ou identifiants identiques qu'ils contiennent, les fichiers sont liés entre eux.

<sup>74</sup> Voir notamment l'Annexe 1.

Figure 43 – Extrait du schéma de structuration d'un jeu de données GTFS

**Légende :**

- agency.txt** Fichier obligatoire
- calendar.txt** Fichier obligatoire sous conditions
- levels.txt** Fichier facultatif
- trips.txt** Champ obligatoire
- shape\_id\*** Champ obligatoire sous condition
- route\_desc** Champ facultatif
- Liaison entre clés primaires



Note : L'ensemble des fichiers « obligatoires sous conditions » et « facultatifs » n'apparaissent pas dans ce schéma.

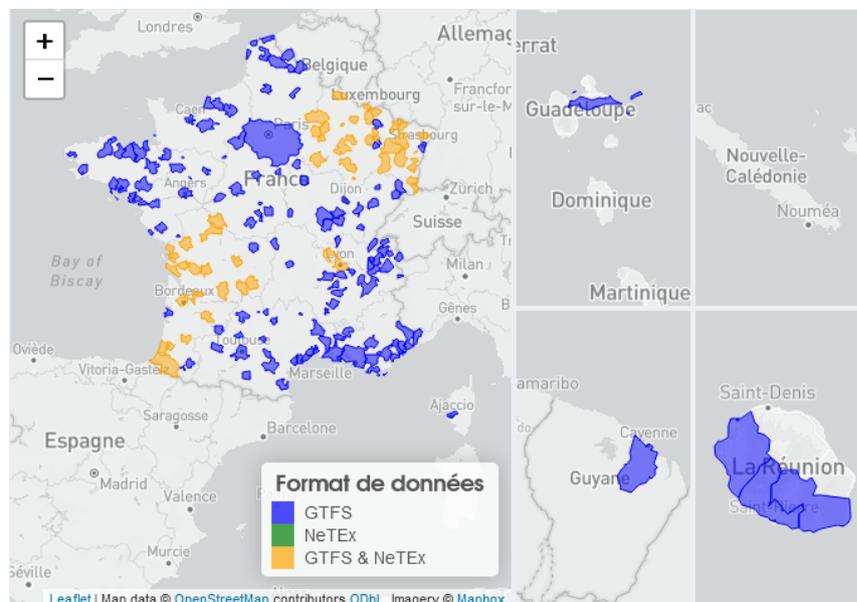
Source : ART, à partir des spécifications techniques GTFS publiées sur <https://developers.google.com/transit/gtfs/reference>.

## Annexe 2 - Précisions sur l'état des lieux de l'ouverture des données

### 1 - Couverture géographique des données statiques des transports en commun selon le standard utilisé

La couverture géographique des jeux de données statiques des transports en commun demeure contrastée selon le format considéré. Bien que les jeux de données GTFS couvrent l'ensemble des AOM locales ayant publié leurs données, la duplication de ces jeux en format NeTEx « natifs » est mise en œuvre dans les régions Grand-Est et Nouvelle-Aquitaine et dans la métropole lyonnaise (Figure 44). Dans les deux premiers territoires, les régions se sont investies et ont collaboré avec une majorité d'AOM locales pour publier les données en format NeTEx.

Figure 44 – Couverture géographique des jeux de données des AOM (hors régions) selon le standard

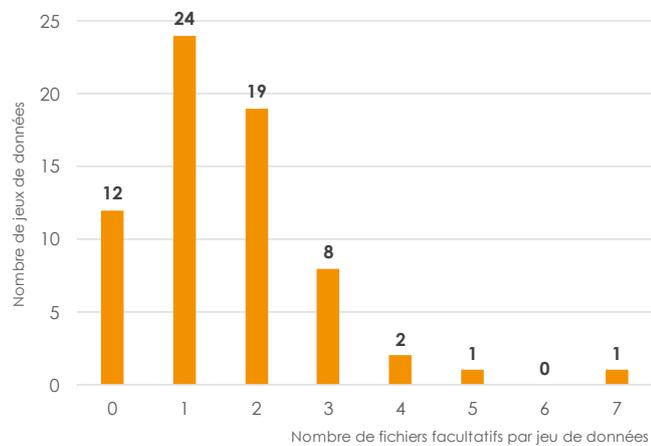


Source : PAN, capture d'écran le 01/03/2022.

### 2 - Les jeux de données GTFS incluent fréquemment des fichiers « facultatifs »

Il ressort du contrôle des jeux de données statiques des transports en commun que les producteurs cherchent à enrichir leurs jeux de données par l'ajout de fichiers « facultatifs », même si cet enrichissement demeure partiel au regard de l'incomplétude de ces fichiers ou du manque de pertinence des données (partie 2.5). Sur 67 jeux de données, seulement 12 d'entre eux ne contiennent que les fichiers nécessaires à leur lecture (soit 18 %) (Figure 45). La plupart des producteurs inclut soit 1 fichier « facultatif » (pour 24 jeux de données) soit 2 (pour 19 jeux de données).

**Figure 45 – Répartition des jeux de données GTFS selon le nombre de fichiers facultatifs présents au sein de chaque jeu**



Source : ART, à partir des jeux de données contrôlés.

### Annexe 3 – Des instructions ex ante et/ou ex post : l'exemple de Vianavigo

La figure 46 illustre la notion d'*instructions*, au travers l'exemple de Vianavigo. Les interfaces pré et post requête sont respectivement présentées à gauche et à droite et rendent compte des instructions pouvant être renseignées par l'utilisateur final.

En particulier, avant d'effectuer une recherche d'itinéraire, l'utilisateur peut :

- Fixer un horaire de départ et d'arrivée ;
- Spécifier s'il souhaite trouver un trajet adapté aux voyageurs en fauteuil roulant ;
- Sélectionner les modes de transport qu'il souhaite inclure dans sa recherche ;
- Renseigner son profil voyageur (niveau en marche à pied et vélo notamment).

Une fois la recherche d'itinéraire effectuée et une première vague d'options de voyage proposée, l'utilisateur peut à nouveau procéder à des réglages en :

- Sélectionnant les modes de transports qu'il souhaite inclure dans sa recherche ;
- Renseignant son profil voyageur (niveau de marche à pied et vélo notamment).

Figure 46 – Interfaces pré et post requête de Vianavigo

Ex ante la recherche d'itinéraire

DÉPART (adresse, arrêt, lieu...) - obligatoire  
ex. : Gare de Lyon

ARRIVÉE (adresse, arrêt, lieu...) - obligatoire  
ex. : Courbevoie

Inverser le point de départ et la destination

- Quand **Instructions ex ante**
- Facilités d'accès
- Modes de transport
- Profil voyageur

Rechercher

Ex post la recherche d'itinéraire

TRANSPORTS EN COMMUN VELO COVOITURAGE

Mardi 1 mars

De Gare de Lyon, Paris

À Courbevoie, Courbevoie

Mes options

Facilités d'accès :

Connaître les facilités présentes sur mon trajet

Trouver des trajets adaptés aux voyageurs en fauteuil roulant

**Instructions ex post**

Je souhaite inclure les modes suivants :

- Train - RER
- Métro
- Tramway
- Bus
- Vélo

J'ai un rythme de marche :

Avec difficultés

Normal

Bon

Je suis un cycliste :

Débutant

Intermédiaire

Expérimenté

Arriver plus tôt Partir plus tard

Modifier la recherche

Source : Site internet de Vianavigo.

**Directeur de la publication** : Bernard Roman

**Pilotage et coordination** : Geneviève Lallemand-Kirche, Fabien Couly

**Auteurs et contributeurs** : Gregory Basnier, Alban Gougoua, Anthony Martin, Brewenn Métayer, Elisabeth Montero, Julie Rouault, Sukhee Lee

**Impression** : Imprimerie de la Direction de l'information légale et administrative en 50 exemplaires

**Dépôt légal** : juin 2022

**ISSN** :

L'édition du [l'état des lieux de l'ouverture et de l'utilisation des données de mobilité](#) est consultable en ligne sur le site [autorite-transport.fr](http://autorite-transport.fr).



11 Place des Cinq Martyrs du Lycée Buffon  
CS 30054 - 75675 Paris Cedex 14  
Tél. +33 (0)1 58 01 01 10

 @ART\_transports

Retrouvez toute l'actualité, les avis et décisions,  
les textes de référence et les publications  
de l'Autorité **sur le site internet**

**autorite-transports.fr**