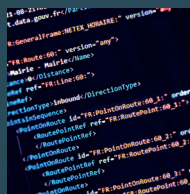
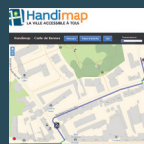


OUVERTURE ET UTILISATION DES DONNÉES DE MOBILITÉ



> **Juin 2023**

État des lieux et bilan des contrôles en 2022/2023
Recommandations et actions pour 2023/2024

INTRODUCTION

BILAN DU CONTRÔLE 2022/2023 DE L'OUVERTURE ET DE L'UTILISATION DES DONNÉES DE MOBILITÉ

Le deuxième rapport de l'Autorité de régulation des transports (ART) concernant l'ouverture et l'utilisation des données sur les déplacements et la circulation s'inscrit dans le cadre des missions de contrôle confiées à l'Autorité en vertu des articles L. 1115-5, qui prévoit également la publication d'un rapport relatif à ces mêmes contrôles, et L. 1264-1 du code des transports. Ce rapport est appelé à alimenter le rapport biennal devant être remis par la France à la Commission européenne sur l'état des lieux de l'ouverture des données et l'évaluation de la conformité aux exigences du règlement délégué (UE) 2017/1926.

Ce rapport présente d'abord un état des lieux en comparant la situation en janvier 2023 au précédent constat (réalisé fin 2021) dressé par l'Autorité et complète les statistiques publiées sur le point d'accès national (transport.data.gouv.fr). Il dresse ensuite les premiers résultats de ses contrôles visant à une publication plus importante et de meilleure qualité des données de mobilité : rappels de la réglementation aux producteurs n'ayant pas publié leurs données sur le point d'accès national, analyse des formats de publication et de la qualité des données elles-mêmes. Un focus a été réalisé sur la publication et l'utilisation des données de perturbations en temps réel en Île-de-France. Enfin, l'utilisation des données est étudiée suivant le support de présentation (site Internet ou application mobile) et suivant la transparence et la neutralité des critères de classement des itinéraires proposés aux voyageurs.

AVERTISSEMENT – DONNÉES UTILISÉES

Ce deuxième rapport repose en partie sur les données publiées sur le point d'accès national (transport.data.gouv.fr) :

- l'état des lieux est effectué à partir d'une extraction faite au 17 janvier 2023 ;
- les contrôles ayant donné lieu à des rappels à la réglementation se fondent sur les données publiées en septembre 2022 ;
- les analyses des formats et de la qualité des données sont fondées sur la sélection aléatoire de jeux de données publiés en décembre 2022 et janvier 2023.

Les analyses de la publication et de l'utilisation des données de perturbations en Île-de-France et, plus généralement, de l'utilisation des données de mobilité sont fondées sur des données publiquement accessibles collectées en janvier et février 2023 sur des sites Internet ou applications mobiles de calcul d'itinéraires.

Les résultats présentés doivent donc être lus à l'aune des périodes et des modalités de collecte (sélection aléatoire, collecte sur Internet) des données analysées ; ils présentent seulement un état de la situation pour les périodes mentionnées.

Un glossaire est disponible en fin de rapport, où les termes suivis d'une astérisque (*) sont repris et explicités.

CHRONOLOGIE



SOMMAIRE

01	État des lieux de l'utilisation du dispositif d'ouverture des données	4
02	Contrôle de l'ouverture des données	6
03	Qualité des données statiques et dynamiques publiées sur le point d'accès national	7
04	Diffusion de l'information en temps réel – Perturbations	10
05	Les supports des calculateurs d'itinéraires : applications mobiles et sites Internet	12
06	Transparence et neutralité des critères de classement des solutions de déplacement	13
	Glossaire	15

SYNTHÈSE

Le contrôle de la publication et de l'utilisation des jeux de données publiés sur le point d'accès national mené par l'Autorité de régulation des transports révèle une publication avancée des jeux de données de transports collectifs et un intérêt certain des utilisateurs pour les données en temps réel. Il n'en demeure pas moins qu'un nombre important de jeux de données statiques et dynamiques ne respectent pas les spécifications de base et sont par conséquent difficilement exploitables. Quelques jeux de données exploitables cumulent aussi les erreurs et complexifient leur traitement afin de proposer une information de qualité aux voyageurs. En outre, la publication des données en temps réel dans un format peu exploitable explique une information sur les perturbations relativement incomplète de la part des calculateurs d'itinéraires et leur difficulté à proposer des itinéraires alternatifs. Par ailleurs, l'Autorité a étendu les contrôles de l'utilisation des données aux applications mobiles, en plus des sites Internet, et constate que les calculateurs d'itinéraires manquent de transparence quant à leurs critères de classement, tant sur les sites Internet qu'au niveau de leurs applications.

Sur la base de ces constats, l'Autorité prévoit, pour sa campagne de contrôle 2023-2024 :

- (1) **de poursuivre ses actions contribuant à une publication plus large des données**, notamment en temps réel, en incluant les données relatives aux perturbations, véhicules et engins partagés, covoiturage et pistes cyclables. À cet effet, elle envisage d'ouvrir des procédures en manquement à l'encontre des producteurs qui ne se sont pas engagés dans un processus de publication de leurs données ;
- (2) **de contribuer à une meilleure qualité et interopérabilité des données**, en contrôlant le respect des exigences techniques des formats de publication des données statiques et dynamiques ; et
- (3) **de s'assurer de la fiabilité des calculateurs d'itinéraires**, en contrôlant les critères de classement non visibles et/ou non transparents pour les itinéraires en transports collectifs, et en étudiant les paramètres de classement des propositions pour les autres modes de transport.

En complément des objectifs de la loi « Climat et résilience » concernant l'utilisation des données publiées par les calculateurs d'itinéraires, l'Autorité prévoit aussi des actions pour accroître la publication des données relatives aux impacts environnementaux des transports et leur réutilisation, pour mieux informer les voyageurs sur cet aspect de leurs déplacements.

01. État des lieux

- Les Autorités organisatrices de la mobilité (AOM) participent activement à la publication majoritaire des jeux de données statiques relatifs aux transports collectifs. Ces données sont principalement publiées sous un format non prévu par la réglementation, ce qui n'entrave pas **une réutilisation des données en temps réel de plus en plus importante**, tant en nombre d'utilisateurs officiels que de requêtes. La loi « Climat et Résilience » prévoit que les données publiées sur le PAN permettent de proposer des trajets en transports moins carbonés.

02. Contrôles

- Les contrôles de l'Autorité ont permis une publication plus large des données de transports collectifs sur le PAN, bien que les AOM rencontrent de réelles difficultés techniques et administratives quant à la publication des données dynamiques.

03. Qualité des données

- Près de 20 % des jeux de données statiques sont partiellement inexploitables car les spécifications techniques des formats ne sont pas pleinement respectées. En revanche, la qualité des jeux de données exploitables demeure bonne, en particulier les jeux au format NeTEx, hormis quelques jeux cumulant plusieurs erreurs pour les jeux de données en format GTFS.
- 61 % des jeux de données dynamiques sont exploitables et proposent de l'information en temps réel mise à jour en moins de 2 minutes. L'absence de correspondance de certaines données dynamiques avec les données statiques sur lesquelles elles reposent est préjudiciable à leur réutilisation.

04. Diffusion des données en temps réel

- Les données des travaux perturbant les itinéraires des bus ne sont pas publiées dans un format réutilisable en Île-de-France. Elles sont donc peu reprises par les calculateurs d'itinéraires, au détriment des voyageurs. À l'inverse, les données relatives aux grèves font l'objet d'une meilleure information des voyageurs par les calculateurs d'itinéraires. Quelle que soit l'information traitée, **les calculateurs d'itinéraires ne proposent pas toujours un itinéraire modifié, faute de publication de données interopérables**, ce qui limite leur développement et leur intérêt, notamment lors de grandes manifestations.

05. Réutilisation des données

- Les applications mobiles présentent une offre servicielle plus étendue et jouissent d'une audience plus large que les sites Internet des calculateurs d'itinéraires. Les solutions de déplacement suggérées sont sensiblement similaires à périmètre constant, indiquant une cohérence entre les informations délivrées par les sites Internet et les applications mobiles.

06. Transparence et neutralité du classement

- Les critères de classement des itinéraires en transports collectifs présentent les mêmes défauts de transparence sur les applications mobiles que ceux identifiés sur les sites Internet. Les critères de classement des itinéraires en transport à la demande (sur les applications mobiles) diffèrent de ceux suggérés en transports collectifs.
- S'ils ne sont pas toujours transparents, les critères de classement des solutions de déplacement semblent être relativement objectifs pour les transports collectifs, mais pas nécessairement pour les autres modes, notamment pour les véhicules de transport avec chauffeur (VTC). Ils reposent aussi sur des hypothèses légèrement différentes selon les calculateurs d'itinéraires en fonction du mode de transport.

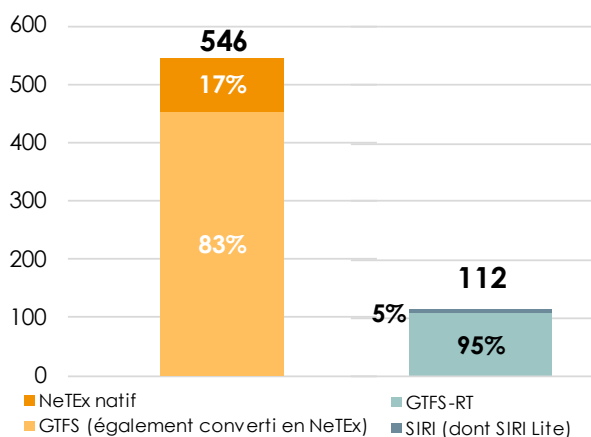
	Niveau 2023 (au 17/01/2023)	Évolution par rapport à 2021 (référence : 19/10/2021)
▪ Nombre de jeux publiés sur le point d'accès national	739	+ 7,3 %
▪ Nombre de producteurs de données de transport collectif (TC)	351	NC
▪ Pourcentage de jeux de transport collectif	89 %	- 3,2 pts de %
▪ Part des jeux de données TC publiés dans un format réglementaire	84,9 %	+ 0,2 pts de %
▪ Part des jeux de données dynamiques TC	17 %	+ 5,2 pts de %
▪ Nombre d'utilisateurs officiels	46	+ 15 %

Les autorités organisatrices de la mobilité (AOM) participent activement à la publication des jeux de données statiques relatifs au transport collectif

Le nombre de jeux de données de transport collectif¹ publiés sur le point d'accès national (PAN*) s'est accru de 2,2 % entre début 2023 et fin 2021 et atteint 658, dont 83 % d'entre eux sont des données statiques. Il en résulte que 320 AOM* sur 335 (hors nouvelles AOM créées en 2022), soit 90 % d'entre elles, sont couvertes par des données statiques publiées sur le PAN. Le taux de couverture des AOM pour les données dynamiques est plus faible et s'élève à 22,4 %. Aujourd'hui, 47 millions de Français vivent dans une AOM (hors AOM régionales) dont les données statiques de transport collectif sont ouvertes².

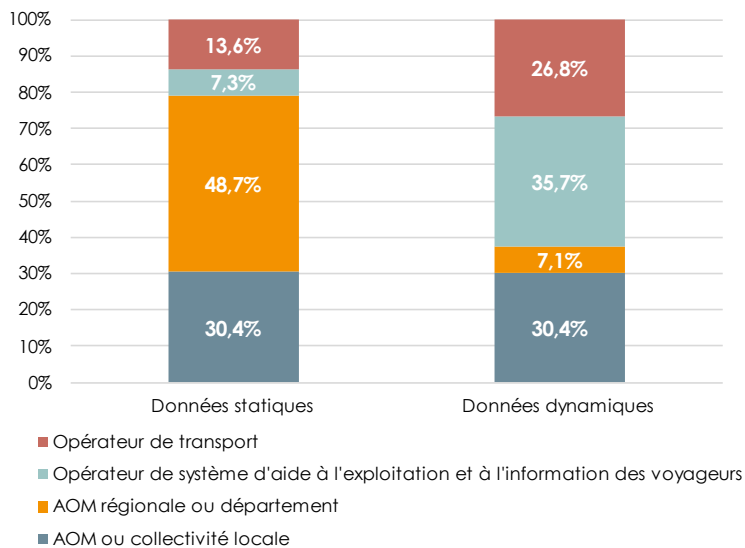
En conséquence, les AOM et les AOMR* demeurent les principales contributrices des données statiques parmi les 351 producteurs de données* de transport collectif recensés. Les régions fournissent 48 % des jeux de données statiques publiés, dont certains d'entre eux sont des jeux « agrégés » de plusieurs réseaux de transport urbains et interurbains d'AOM et de collectivités locales. La mise à disposition des données dynamiques sur le PAN nécessitant une exploitation plus complexe des systèmes d'information, ce sont les opérateurs de système d'aide à l'exploitation et à l'information des voyageurs (SAEIV) qui sont souvent chargés de publier ces données.

Figure 2 – Nombre et répartition des 658 jeux de données de transport collectif publiés sur le PAN selon le format



Source : ART, d'après analyse des jeux de données publiés sur le PAN.

Figure 1 – Répartition des jeux de données de transport collectif publiés sur le PAN selon les catégories de données et de producteurs de données



Source : ART, d'après analyse des jeux de données publiés sur le PAN.

Les données de transport collectif sont principalement publiées sous un format non prévu par la réglementation

Les producteurs publient majoritairement leurs données selon un format non requis par la réglementation (GTFS*), bien que cette tendance soit en diminution. À titre d'exemple, la publication en format SIRI* ne représente que 5 % des jeux de données en temps réel, contre 1 % fin 2021.

Des acteurs tels que Ile-de-France Mobilité (IdFM)* ou le SYTRAL* (Lyon), mais surtout les régions, participent à la diffusion des données en format NeTEx*, en plus de publier leurs données en GTFS, représentant désormais 17 % des jeux statiques. Par ailleurs, le PAN a mis en place une traduction automatique des jeux de données publiés en GTFS en format NeTEx et met ainsi à disposition des utilisateurs 455 jeux de données traduits.

¹Dans ce rapport, on entend par transport collectif tout transport comprenant les services suivants : Transport aérien, transport ferroviaire, autocars longue distance, transport maritime y compris les ferries, métros, trams, bus, trolleybus.

²Statistiques publiées par le PAN. D'autres statistiques sont disponibles à cette page : <https://transport.data.gouv.fr/stats>.

ÉTAT DES LIEUX DE L'UTILISATION DU DISPOSITIF D'OUVERTURE DES DONNÉES

Une utilisation des données en temps réel de plus en plus importante, tant en nombre d'utilisateurs que de requêtes

Alors que l'offre de jeux de données sur le PAN s'accroît, leur réutilisation s'accroît également. D'une part, le nombre d'utilisateurs* officiels augmente, passant de 40 fin 2021 à 46 en 2023. D'autre part, les requêtes sur les jeux de données dynamiques, tant sur les transports en commun au format GTFS-RT* que sur le vélo-partage au format GBFS*, se multiplient et frôlent les 60 millions au premier trimestre 2023. La multiplication par 3 du nombre de requêtes sur les données en temps réel des transports en commun entre le premier trimestre 2022 et celui de 2023 est à mettre en regard avec l'augmentation de 46 % des jeux de données GTFS-RT publiés sur cette même période, attestant d'une intensification de l'usage de ces données et de l'intérêt que leur portent les utilisateurs. Cet intérêt avait été mentionné par ces derniers dans la consultation publique conduite par l'Autorité³.

La loi « Climat et Résilience » vise à ce que les données du PAN permettent de proposer des trajets en transports décarbonés

L'article L. 1115-8-1 du code des transports, créé par l'article 122 de la loi « Climat et résilience »⁴, prévoit que « les services numériques d'assistance au déplacement sont tenus d'informer de façon complète les utilisateurs des impacts environnementaux de leurs déplacements ». Plus particulièrement, l'article D. 1115-19 du code des transports, pris en application de l'article L. 1115-8-1, précise que ces services doivent « au plus tard le 1er décembre 2022, veiller à intégrer l'ensemble des données sur les services de transport réguliers, et à la demande, mises à disposition sur le point d'accès national [...] ainsi qu' au plus tard le 1er décembre 2023, veiller à intégrer l'ensemble des données sur les services de partage de véhicules, de cycles, de cyclomobiles légers, d'engins de déplacement personnels, ou sur les déplacements à pied, mises à disposition sur le point d'accès national [...] ». Il indique aussi que « pour les autres services numériques, lorsqu'ils ne proposent pas de services de transport mais proposent a minima l'utilisation du véhicule individuel, [ils doivent], au plus tard le 1er décembre 2022, veiller à intégrer l'ensemble des données relatives au réseau cyclable, aux aires de covoiturage et au stationnement, mises à disposition sur le point d'accès national mentionné à l'article D. 1115-1 ».

La disponibilité des données mentionnées dans les articles du code des transports issus de l'article 122 de la loi « Climat et résilience » est donc nécessaire pour que les services numériques d'assistance aux déplacements puissent réutiliser ces données et informer correctement les voyageurs quant aux alternatives dont ils disposent.

Les actions et priorités de contrôle de l'Autorité pour la campagne 2023-2024

L'Autorité constate le nombre important de données publiées concernant les transports collectifs et l'intérêt marqué pour la réutilisation des données* en temps réel des transports collectifs et des véhicules, vélos et engins personnels en libre-service. Elle note également l'importance de la mise à disposition de ces données, tout comme de celles relatives aux facteurs environnementaux liés aux transports pour l'application de l'article 122 de la loi « Climat et Résilience ».

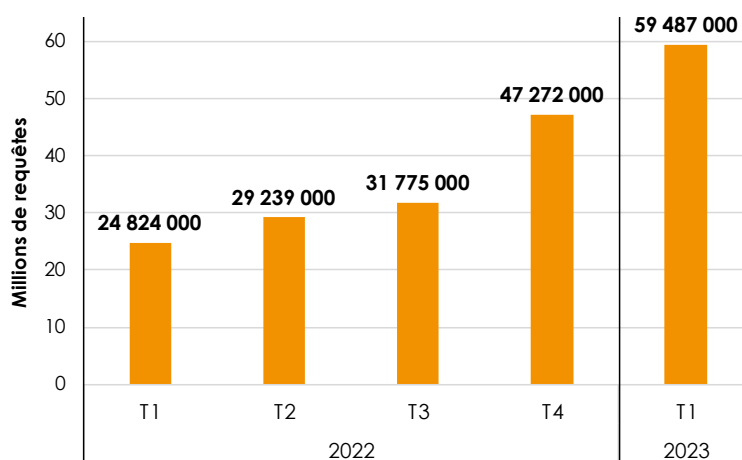
L'Autorité engage donc les producteurs de données, conformément à la réglementation, à poursuivre leurs efforts de publication et envisage de contribuer à une publication plus large des données lors de sa campagne de contrôles 2023/2024, en :

- poursuivant les actions préventives de rappel à la réglementation pour les données statiques et en temps réel de transports collectifs ;
- étendant ces actions préventives auprès des fournisseurs de services de partage de véhicules, cycles et engins de déplacement personnel*, mais aussi des gestionnaires d'infrastructure* (notamment en ce qui concerne les données liées aux transports collectifs, au réseau cyclable, aux aires de covoiturage et au stationnement).

Conformément à la réglementation, l'Autorité engage aussi les producteurs de données à publier les facteurs environnementaux des transports, et les utilisateurs à proposer aux usagers une information sur les impacts environnementaux de leurs déplacements. Elle invite ainsi les fournisseurs de calculateurs d'itinéraires* à proposer, en complément des critères tels que le prix ou la durée de trajet, un classement des itinéraires sur la base de critères environnementaux. Ainsi, l'Autorité prévoit, lors de la campagne de contrôle 2023/2024 :

- des actions préventives de rappel à la réglementation relative à la publication des facteurs environnementaux ;
- une évaluation de la prise en compte de ces facteurs environnementaux par les utilisateurs.

Figure 3 – Nombre de requêtes sur les jeux de données dynamiques au format GTFS-RT et GBFS par trimestre



Source : ART d'après l'analyse des données transmises par le PAN.

³La synthèse de la consultation publique et des auditions concernant l'ouverture et la réutilisation des données de mobilité sur le point d'accès national est disponible ici : <https://www.autorite-transport.fr/wp-content/uploads/2022/06/synthese-consultation-publique-donnees-de-mobilite.pdf>.

⁴Loi n° 2021-1104 du 22 août 2021 portant lutte contre le dérèglement climatique et renforcement de la résilience face à ses effets, dite loi « Climat et résilience ».

02

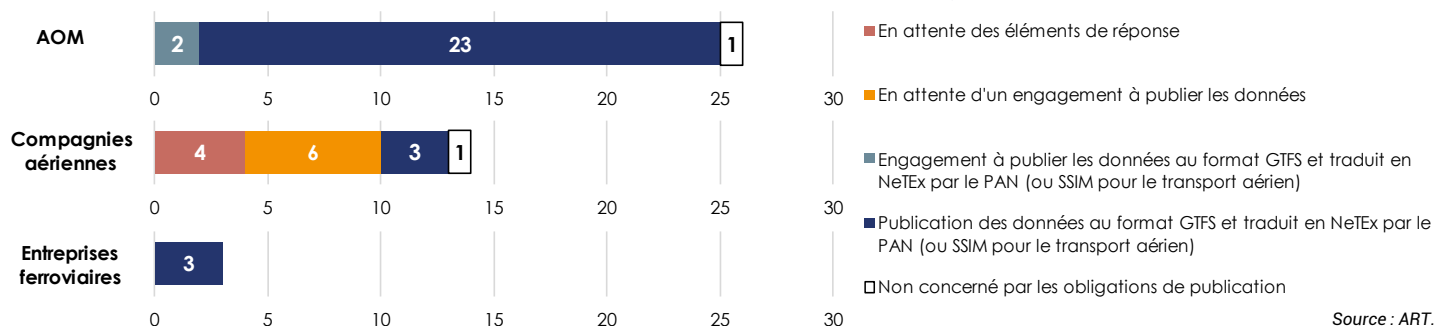
CONTRÔLE DE L'OUVERTURE DES DONNÉES

Les contrôles de l'Autorité ont permis une publication plus large des données de transport collectif sur le PAN

Au titre de ses missions de contrôle du respect des obligations relatives au règlement délégué (UE) 2017/1926, l'Autorité a identifié des AOM et opérateurs de transports possédant des jeux de données qui ne seraient pas publiés sur le PAN ou mis à jour. Ainsi, l'Autorité a contacté 43 producteurs de données manquant à leur obligation de publication des jeux de données statiques, 13 producteurs n'ayant pas mis à jour leur jeu de données statiques et 41 AOM n'ayant pas publié leurs données dynamiques.

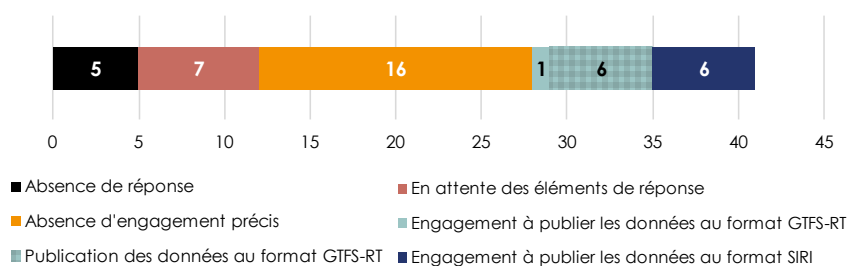
Ces actions préventives de rappel à la réglementation ont permis une mise à jour des données mais ont eu des résultats différenciés concernant la publication des données sur le PAN en fonction du type de producteurs. Si les AOM et les entreprises ferroviaires se sont mises en conformité rapidement, les compagnies aériennes sont peu informées de leurs obligations et n'ont pas toujours fourni à l'Autorité les réponses attendues (délai et format de publication notamment).

Figure 4 – Répartition du retour des producteurs de données manquant à leur obligation de publication des données statiques de transport collectif selon le type de producteur



Les réponses des AOM quant à la publication des données en temps réel témoignent de réelles difficultés techniques et administratives

Figure 5 – Répartition du retour des producteurs de données manquant à leur obligation de publication des données dynamiques de transport collectif



La mise en conformité des AOM au regard de la publication des données dynamiques au format SIRI* se heurte à des obstacles liés à la publication tardive du profil national et à des difficultés techniques et administratives. Les AOM contactées indiquent souvent l'absence de données au format SIRI, tandis que d'autres mettent en avant les difficultés à fournir le flux de données liées à de nombreuses requêtes, la plupart d'entre elles ignorant qu'un serveur peut être mis à leur disposition par le PAN pour supporter la charge de requête.

En tout état de cause, les AOM prévoient en majorité d'inscrire dans leur délégations de service public (DSP*) avec leur opérateur de SAEIV* la publication du jeu de données SIRI. En attendant le renouvellement de la DSP*, les producteurs proposent de publier un jeu de données GTFS-RT* à destination des utilisateurs. Cette solution palliative permettra de répondre aux besoins exprimés par les utilisateurs.

Les actions et priorités de contrôle de l'Autorité pour la campagne 2023-2024

L'Autorité constate que quelques producteurs n'ont pas fourni de réponses à la demande de l'Autorité et/ou ne se sont pas mis en conformité en publiant leurs données statiques. Elle relève aussi la difficulté des AOM à fournir les données dynamiques.

L'Autorité engage donc les producteurs de données, conformément à la réglementation, à poursuivre leurs efforts de publication et à fournir les réponses circonstanciées nécessaires aux contrôles de l'Autorité. Elle prévoit de contribuer à une publication plus large des données lors de la campagne 2023/2024 en :

- poursuivant ses actions préventives de rappel à la réglementation à destination des producteurs de données n'ayant pas encore publié leurs données ;
- ouvrant des procédures en manquement à l'encontre des producteurs qui ne se sont pas engagés dans un processus de publication de leurs données statiques ;
- étudiant avec les acteurs concernés les possibilités de publication des données en temps réel, y compris sous forme de solution palliative temporaire.

03

QUALITÉ DES DONNÉES STATIQUES ET DYNAMIQUES PUBLIÉES SUR LE POINT D'ACCÈS NATIONAL

Près de 20 % des jeux de données statiques sont partiellement inexploitable car les spécifications techniques des formats ne sont pas pleinement respectées

Un peu plus de 80 % des jeux de données statiques relatifs aux transports en commun respectent les spécifications techniques, qui reposent notamment sur leur accessibilité, leur lisibilité, leur mise à jour et le respect de la structuration des données au sein de ces jeux. Les 20 % des jeux restants demeurent difficilement exploitables voire inutilisables en raison de problèmes de téléchargement, de métadonnées* inexactes, de données expirées et d'une structure des fichiers constituant les jeux de données non conformes.

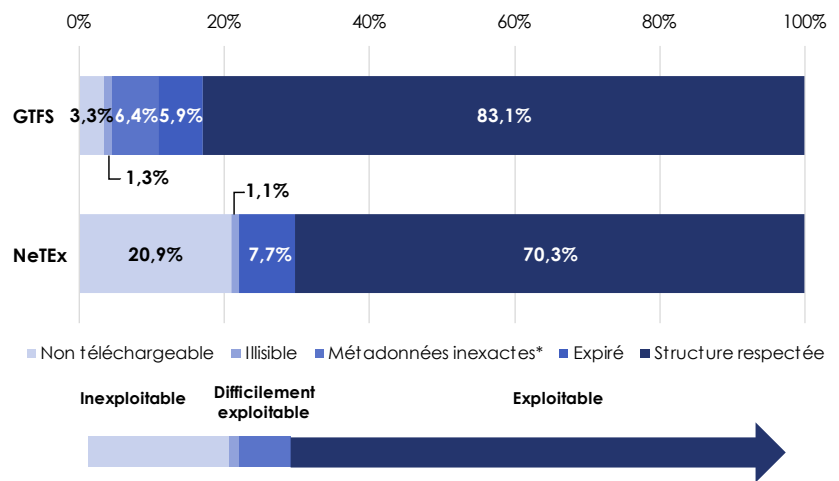
Il subsiste une différence de qualité entre les jeux de données GTFS et NeTEx⁵. Seulement 70,3 % des jeux de données NeTEx respectent l'ensemble des critères de validation de base et ce constat résulte d'un problème de téléchargement pour 20 % des jeux de données contrôlés (absence d'URL, lien URL expiré, etc.).

La qualité des jeux de données exploitables est bonne, à l'exception de quelques jeux cumulant plusieurs erreurs

La majorité des jeux de données exploitables délivrent une information de bonne qualité. Ainsi 88 % d'entre eux ne possèdent aucune erreur liée au manque de données ou au respect de leur format.

A contrario, 26 jeux de données GTFS possèdent un taux d'erreurs relatives à l'absence de données ou à une syntaxe incorrecte supérieur à 5 %, dont un jeu pour lequel 37 % des données ne respectent pas les règles de syntaxe. En outre, certains jeux cumulent les erreurs d'absence de données obligatoires et de syntaxe incorrecte, les rendant ainsi difficilement utilisables.

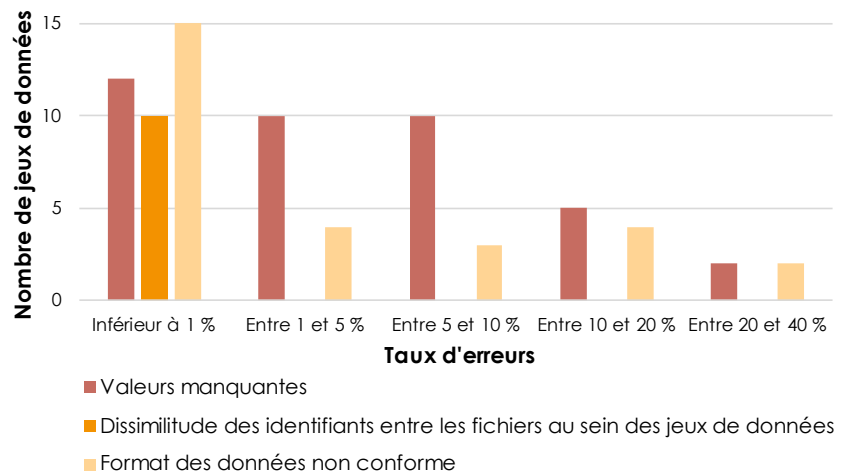
Figure 6 – Répartition des jeux de données statiques selon les erreurs de base détectées et le format de données



Note : le PAN ne met pas à disposition les métadonnées relatives au calendrier d'exploitation pour les jeux de données NeTEx produits par les producteurs eux-mêmes. La statistique relative à la validation de la structure pour les jeux de données au format GTFS se réfère uniquement aux fichiers et champs obligatoires.

Source : ART, d'après analyse des jeux de données publiés sur le PAN.

Figure 7 – Répartition des jeux de données au format GTFS selon les taux et types d'erreurs



Source : ART, d'après analyse des jeux de données publiés sur le PAN.

⁵ Dans cette partie, on entend par jeu de données au format NeTEx un jeu élaboré par le producteur de données lui-même, et non un jeu de données traduit par le PAN à partir d'un jeu de données au format GTFS.

QUALITÉ DES DONNÉES STATIQUES ET DYNAMIQUES

61 % des jeux de données en GTFS-RT respectent les spécifications de bases et sont donc exploitables

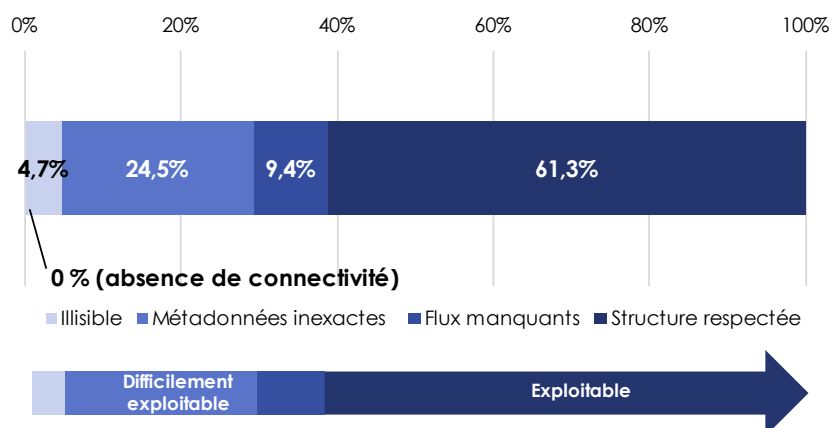
Un peu plus de 61 % des jeux de données dynamiques au format GTFS-RT relatifs aux transports collectifs respectent les spécifications de base, qui reposent sur leur accessibilité, leur lisibilité, leurs métadonnées exactes et le respect de la structuration des données au sein de ces jeux. Les jeux restants demeurent difficilement exploitables. En outre, 25 jeux de données possèdent des métadonnées indiquant une mise à jour du jeu dans un délai inférieur à 30 secondes, tandis que l'horodatage des flux montre que ces derniers ne se mettent pas à jour.

L'information en temps réel est mise à jour en moins de 2 minutes, mais l'absence de correspondance avec les données statiques des flux de données GTFS-RT est préjudiciable à leur réutilisation

Les jeux de données exploitables proposent une information actualisée en moins de deux minutes pour 84 % d'entre eux et 59 % des jeux contrôlés rafraîchissent leurs informations en moins de 30 secondes. Les jeux dont le temps de rafraîchissement dépasse le seuil de deux minutes contiennent, pour la majorité d'entre eux, un seul flux d'information relatif à la mise à jour des trajets. Or, le PAN recommande un temps de rafraîchissement inférieur à deux minutes (les recommandations de la communauté à l'origine des spécifications techniques GTFS-RT sont de 1 min 30).

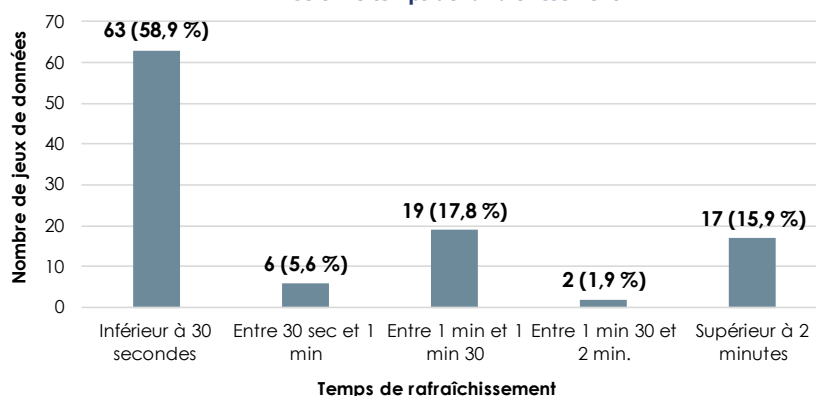
La qualité de l'information délivrée est altérée par des erreurs de concordance avec le jeu de données statiques sur lequel ils reposent. Une minorité de flux d'informations sont ainsi déconnectés des données issues des fichiers GTFS. D'une part, pour 5 % des jeux de données GTFS-RT, la plupart des informations ne correspondent pas aux données du jeu GTFS; d'autre part, 15 % des jeux de données créent des identifiants qui n'existent pas dans les données statiques.

Figure 8 – Répartition des jeux de données au format GTFS-RT selon les taux d'erreurs et le type d'erreur



Source : ART d'après analyse des jeux de données publiés sur le PAN.

Figure 9 – Répartition des jeux de données au format GTFS-RT selon le temps de rafraîchissement*



Note : Le rafraîchissement du flux correspond à la mise à jour de l'horodatage de l'entité correspondant aux métadonnées.

Source : ART, d'après analyse des jeux de données publiés sur le PAN. Analyse réalisée le 17/05/2023.

Les actions et priorités de contrôle de l'Autorité pour la campagne 2023-2024

L'Autorité constate ainsi un respect insuffisant des spécifications techniques des formats des données publiées tant statiques que dynamiques.

L'Autorité engage donc les producteurs de données de transports collectifs à respecter les exigences techniques des formats de publication statiques et dynamiques et à s'assurer de la qualité des données publiées notamment en utilisant plus systématiquement les validateurs de données existants. L'Autorité va contribuer à une publication de données exploitables et de meilleure qualité lors de la campagne 2023/2024 en :

- lançant des actions préventives de rappel à la réglementation concernant les exigences techniques des formats de publication pour les jeux de données statiques non téléchargeables, illisibles, expirés ou dont la structuration ne respecte pas les exigences de base, ainsi qu'aux jeux statiques de mauvaise qualité (données incohérentes) ;
- vérifiant la cohérence des informations pour les jeux de données dynamiques de mauvaise qualité, notamment pour ceux qui proposent des informations ne correspondant pas à leur jeu de données statiques ;
- lançant des actions préventives de rappel à la réglementation concernant les exigences techniques des formats de publication pour les jeux de données dynamiques illisibles, sans flux d'informations dont les métadonnées sur la mise à jour des flux sont inexactes.

L'Autorité prévoit d'étendre le contrôle des spécifications techniques aux jeux de données des services de transport à la demande*, aux services de partage de véhicules, cycles et engins de déplacement personnel.

04

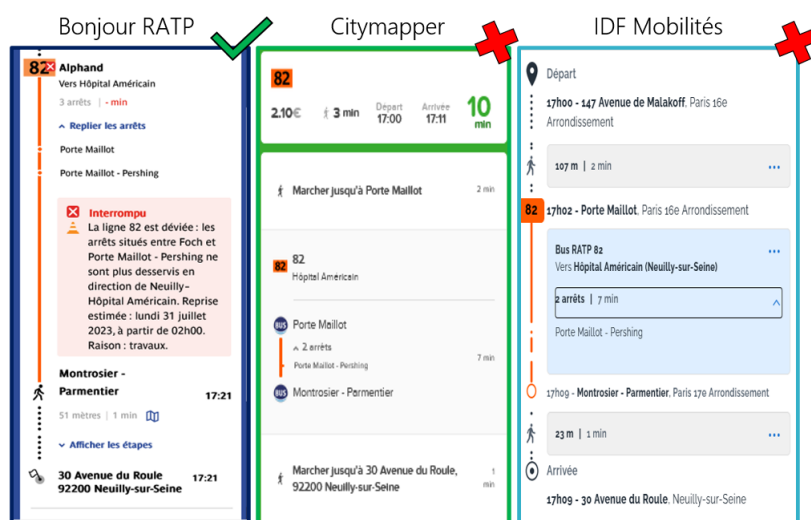
DIFFUSION DE L'INFORMATION EN TEMPS RÉEL - PERTURBATIONS

Les données des travaux perturbant les itinéraires de bus n'étant pas publiées dans un format réutilisable en Île-de-France, elles sont peu reprises par les calculateurs d'itinéraires, au détriment des voyageurs

La RATP* fournit, sur son site Internet, des informations sur les travaux affectant son réseau de bus mais ne diffuse pas ces informations dans un format lisible en machine et interopérable. L'analyse des trajets en bus suggérés par Citymapper, Bonjour RATP et IDF Mobilités (IDFM) témoigne d'un usage de l'information relative aux travaux du réseau de bus RATP publiée par la RATP différent selon les calculateurs d'itinéraires. Si Bonjour RATP relaie une information complète (l'ensemble des déviations sont renseignées), cette information est affichée de manière variable pour les autres calculateurs d'itinéraires. En effet, seuls 27,5 % des résultats de calculs d'itinéraires affichent une information sur les travaux (contrôle opéré sur un ensemble de six itinéraires affectés par des travaux et testés à plusieurs reprises parmi cinq calculateurs d'itinéraires).

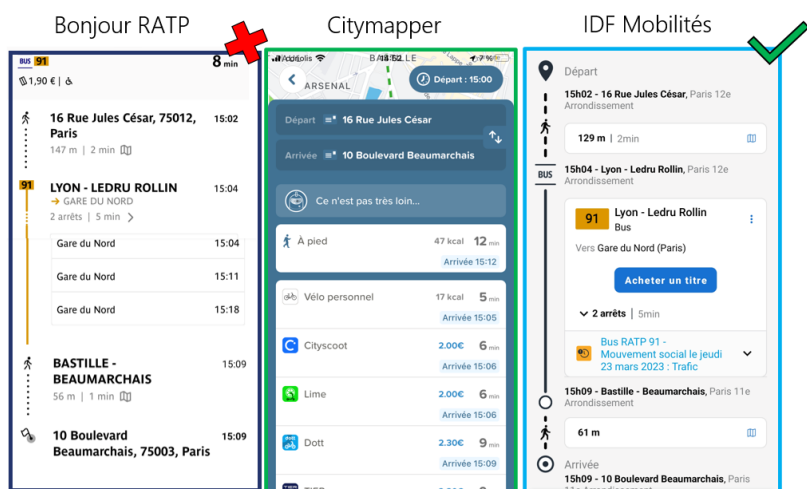
En l'absence de mise à disposition des données de perturbations sous un format lisible et interopérable, notamment au travers d'une API*, les calculateurs d'itinéraires testés rencontrent des difficultés à traiter l'information sur les travaux perturbant les services RATP de bus, ce qui limite sa diffusion aux voyageurs.

Figure 10 – Affichage de l'information relative aux travaux affectant le réseau de bus RATP



Source : ART à partir des captures d'écran des calculateurs d'itinéraires de Bonjour RATP, Citymapper, Google Maps, IDF Mobilités et Moovit.

Figure 11 – Affichage de l'information relative aux mouvements sociaux affectant le réseau francilien



Source : ART à partir des captures d'écran des calculateurs d'itinéraires de Bonjour RATP, Citymapper, Google Maps, IDF Mobilités et Moovit.

Les données relatives aux grèves font l'objet d'une meilleure information des voyageurs par les calculateurs d'itinéraires

IDFM* publie, au travers de plusieurs API, les informations textuelles, reprenant les messages d'annonce, relatives aux grèves, dans un format non conforme mais réutilisable. Ainsi, et contrairement aux informations sur les travaux du réseau de bus RATP, les informations relatives aux grèves sont largement réutilisées. Sur le panel d'itinéraires étudiés, seul Bonjour RATP ne relaie pas ces informations.

Malgré une plus grande réutilisation de ces données, elles restent toutefois dans un format non conforme (sous un format JSON* ne respectant pas les spécifications techniques du format SIRI prévu par la réglementation). Il en résulte que 53 % seulement des résultats d'itinéraires intègrent une alerte relative aux perturbations liées aux mouvements sociaux (contrôle sur un ensemble de cinq itinéraires affectés par des mouvements sociaux et testés à plusieurs reprises sur cinq calculateurs d'itinéraires).

DIFFUSION DE L'INFORMATION EN TEMPS RÉEL

Le traitement différencié des informations de perturbations ne montre pas de situation plus favorable pour les calculateurs d'itinéraires de la RATP et d'IdFM

Le différentiel d'informations relatives aux perturbations relayées par les calculateurs d'itinéraires étudiés ne semble pas être le fait d'un accès privilégié des applications Bonjour RATP et Île-de-France Mobilités aux données de perturbations du réseau RATP. Ainsi, l'absence d'informations relatives aux grèves publiées par Bonjour RATP implique que les fermetures de stations ne sont pas indiquées sur l'application.

Faute de publication de données de perturbation interopérables, les calculateurs d'itinéraires ne proposent pas toujours un itinéraire alternatif, ce qui limite leur intérêt pour les voyageurs

Les calculateurs d'itinéraires ne proposent généralement pas de solutions de déplacement alternatives en présence de perturbations avérées et affichent seulement un message de perturbation quand ils en disposent. Seul Citymapper suggère des alternatives de déplacements à pied en période de grève tandis que Moovit est l'unique calculateur d'itinéraires proposant des parcours tenant compte des déviations liées à des travaux.

Toutefois, la mise à disposition des données de perturbations dans un format conforme est en voie d'amélioration. De nouveaux services de données temps réel SIRI sont en cours de développement, notamment un service « Situation Exchange », qui pourraient permettre aux calculateurs d'itinéraires d'offrir aux voyageurs une information relative aux perturbations plus complète et adaptée.

Néanmoins, le rythme d'ouverture de ces services et de leur mise en qualité entrave la capacité des calculateurs d'itinéraires à proposer des itinéraires alternatifs à partir des données non interopérables déjà publiées. Cette situation pose question quant à la possibilité de proposer rapidement une information voyageur fiable en situation perturbée, notamment en cas d'afflux de voyageurs occasionnels lors de grandes manifestations (JO 2024). L'ouverture limitée et tardive de données interopérables sur les perturbations est ainsi de nature à limiter le développement des calculateurs d'itinéraires, dont la capacité à orienter les voyageurs en situation perturbée constitue un facteur d'audience important.

Cette intégration est, en tous les cas, complexe pour les calculateurs d'itinéraires, en raison de l'absence de données interopérables. De ce fait, ils intègrent manuellement, souvent partiellement et de manière différenciée, les informations disponibles et fournissent en conséquence une information incomplète qui peut induire les voyageurs en erreur.



Source : ART, à partir des captures d'écran des calculateurs d'itinéraires de Bonjour RATP, Citymapper, Google Maps, IDF Mobilités et Moovit.

Figure 13 – Récapitulatif du constat des contrôles de mars 2023 sur la présence des informations et leur impact sur l'itinéraire

Résumé du constat mars 2023	Infos grèves		Infos travaux	
Calculateurs d'itinéraires	Présence	Itinéraire modifié	Présence	Itinéraire modifié
Bonjour RATP	≈	X	O	X
Citymapper	O	≈	≈	X
Google Maps	O	X	≈	X
IDF Mobilités	O	X	≈	X
Moovit	O	X	O	≈

O : Oui X : Non ≈ : Partiel

Source : ART, à partir des captures d'écran des calculateurs d'itinéraires de Bonjour RATP, Citymapper, Google Maps, IDF Mobilités et Moovit.

Les actions et priorités de contrôle de l'Autorité pour la campagne 2023-2024

L'Autorité regrette tant une publication insuffisante et non interopérable des données relatives aux perturbations, que la réutilisation incomplète des données déjà publiées.

L'Autorité engage donc les producteurs de données à publier au plus tôt les informations relatives aux travaux et autres perturbations dans un format interopérable. Elle prévoit de contribuer à une publication et à la fiabilité des calculateurs d'itinéraires lors de la campagne 2023-2024 en :

- lançant des actions préventives de rappel à la réglementation concernant la publication de données de perturbations interopérables ;
- étudiant avec les utilisateurs concernés les possibilités d'intégration des données déjà publiées afin de proposer des itinéraires alternatifs lors de perturbations et pas uniquement un message d'avertissement.

05

LES SUPPORTS DES CALCULATEURS D'ITINÉRAIRES : APPLICATIONS MOBILES ET SITES INTERNET

Les applications mobiles présentent une offre servicielle plus étendue et jouissent d'une audience plus large que les sites Internet des calculateurs d'itinéraires

Les applications mobiles constituent le principal support utilisé pour la recherche d'itinéraires (baromètre du numérique de l'ARCEP⁶), justifiant leur prise en compte par l'Autorité au même titre que les sites Internet

Le périmètre couvert par les applications mobiles est plus large que celui des sites Internet, les premières ayant généralement une offre servicielle plus étendue (modes et opérateurs intégrés, services proposés) (comparaison intra-calculateur).

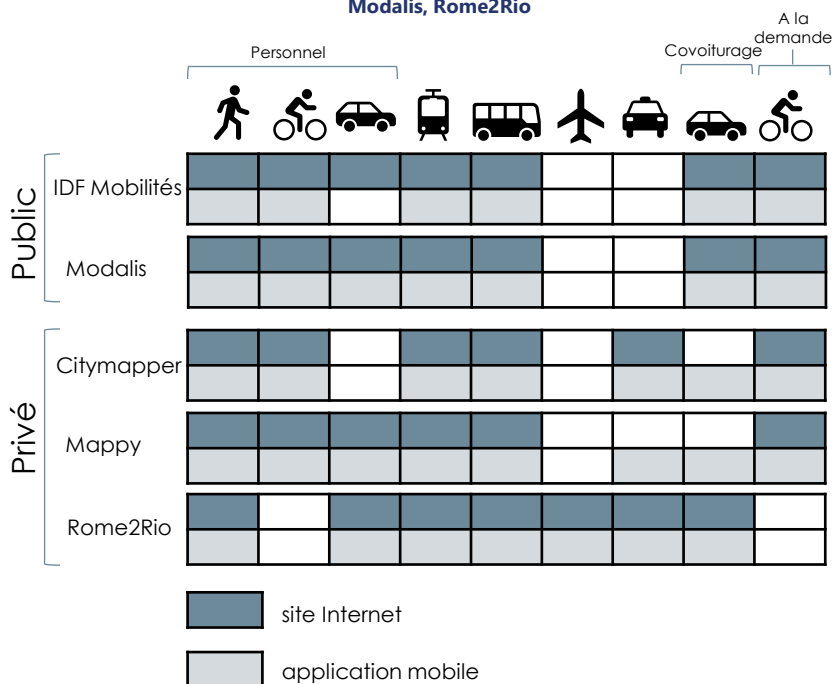
Les applications mobiles des calculateurs d'itinéraires résultant d'initiatives privées intègrent plus de services de transport à la demande que les celles d'initiatives publiques (comparaison inter-calculateur).

Les solutions de déplacement suggérées par les sites Internet et applications mobiles des calculateurs d'itinéraires sont sensiblement similaires

Les suggestions d'itinéraires en transports collectifs des applications mobiles et sites Internet des calculateurs d'itinéraires présentent des caractéristiques moyennes identiques pour tous les paramètres testés, en termes de nombre de solutions de déplacements, de durée moyenne des déplacements, de classement des itinéraires et de critères de classement.

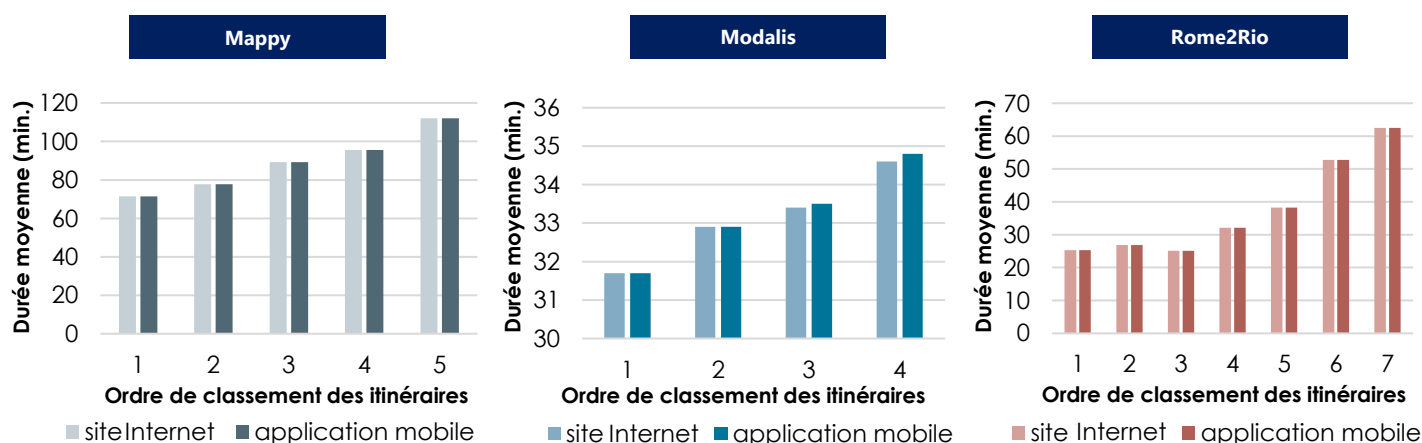
Ainsi, le temps moyen des différents itinéraires proposés par les sites Internet et applications mobiles de Mappy, Modalis et Rome2Rio, montre que les critères de classement et les algorithmes de calcul sont identiques pour ce paramètre.

Figure 14 – Périmètres couverts par Mappy, Citymapper, IDF Mobilités, Modalis, Rome2Rio



Source : ART, à partir de l'étude des interfaces web et applications mobiles des calculateurs d'itinéraires.

Figure 15 – Durée moyenne (en minutes) des itinéraires par calculateur selon leur ordre d'apparition (site Internet et application mobile)



Note de lecture : les premiers itinéraires suggérés par le site Internet et l'application mobile de Mappy ont une durée moyenne identique.

Source : ART, à partir des résultats d'itinéraires collectés grâce à une API du PERen*.

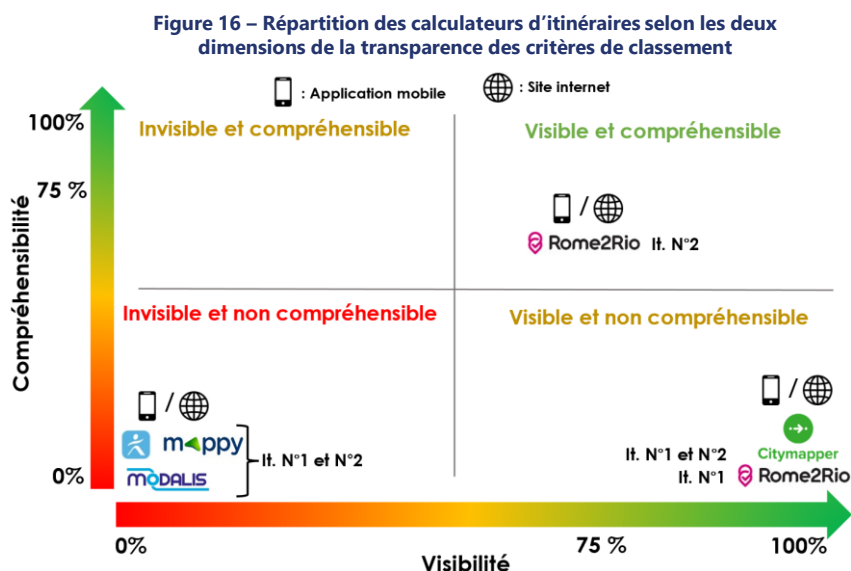
⁶ Baromètre du numérique – édition 2022, ARCEP : https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-numerique-edition-2022-Rapport.pdf

TRANSPARENCE ET NEUTRALITÉ DES CRITÈRES DE CLASSEMENT DES SOLUTIONS DE DÉPLACEMENT

Les critères de classement des itinéraires en transports en commun présentent les mêmes défauts de transparence sur les applications mobiles que ceux identifiés sur les sites Internet

Deux dimensions permettent de qualifier les critères de classement des itinéraires suggérés par les calculateurs d'itinéraires de transparents ou non : leur visibilité et leur caractère compréhensible. Des critères de classement « visibles », « invisibles » et « déductibles » avaient été identifiés lors du précédent rapport. Ce rapport avait en outre montré que les critères de classement ne répondaient pas toujours à la notion de transparence, soit qu'ils soient visibles mais pas objectivés (« recommandés », « optimal »), soit qu'ils soient invisibles.

Les calculateurs manquent de transparence quant au classement des itinéraires et ce constat est similaire à périmètre constant quel que soit le support numérique (site internet ou applications mobiles). Les résultats illustrent des manquements à l'exigence de transparence dans les tris et le classement des options de voyage, que ce soit du fait de critères non visibles ou non directement compréhensibles (comme « recommandé »). Cet aspect peut nuire à la bonne interprétation – sans biais – des résultats affichés par les voyageurs.



Note de lecture : Pour Rome2Rio, 100 % des itinéraires proposés en premier sont visibles mais non compréhensibles tandis que 75 % des itinéraires proposés en deuxième sont visibles et compréhensibles.

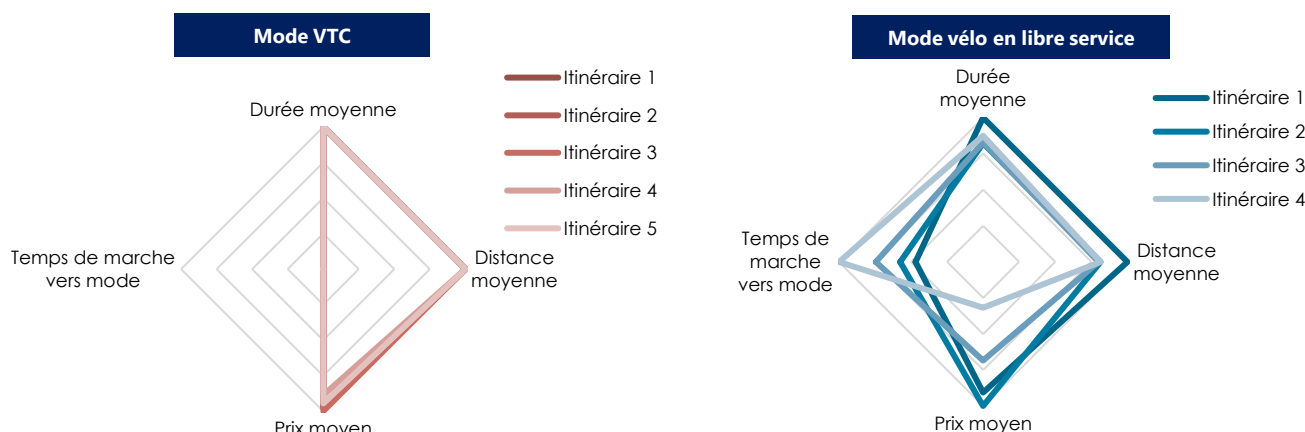
Source : ART, à partir des résultats d'itinéraires suggérés sur application mobile et collectés grâce à une API du PEReN.

Les critères de classement des itinéraires en transport à la demande (sur les applications mobiles) diffèrent de ceux suggérés en transport en commun

À défaut d'être classés selon des critères visibles pour l'utilisateur, l'étude des caractéristiques des solutions de déplacement en transports à la demande révèle des classements différenciés selon les modes. Néanmoins, cette étude n'a permis ni de comprendre la méthodologie utilisée pour classer ces solutions, ni d'établir tous les paramètres

déterminant ce processus de classement. Le temps de marche vers le mode et le prix moyen semblent être ainsi déterminants dans le classement des itinéraires suggérés en vélo. Les paramètres étudiés de classement des itinéraires suggérés en VTC n'ont pas permis de conclure dans ce cas.

Figure 17 – Approche multicritère des itinéraires suggérés selon leur ordre d'apparition pour Citymapper



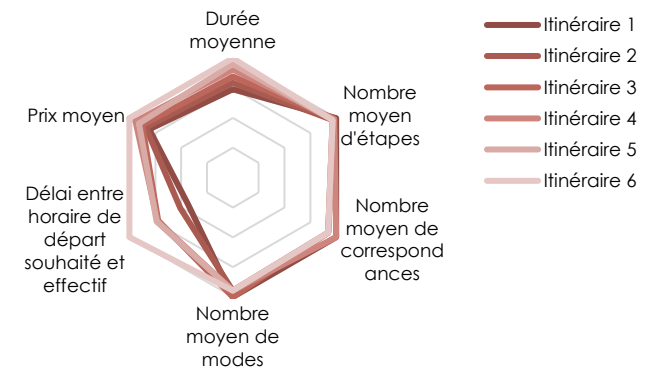
Note de lecture : pour le mode vélo en libre service, le premier itinéraire suggéré minimise le temps de marche vers la station par rapport aux itinéraires suivants. A l'inverse, la durée et la distance sont en moyenne plus élevées. Pour le mode VTC, aucun critère ne semble discriminer le classement des itinéraires.

Source : ART, à partir des résultats d'itinéraires suggérés sur application mobile et collectés grâce à une API du PEReN.

S'ils ne sont pas toujours transparents, les critères de classement des solutions de déplacement en transports collectifs semblent être relativement objectifs

Le temps de trajet, le prix et le délai entre les horaires de départ souhaité et effectif semblent être des critères déterminants du classement des itinéraires suggérés en transport en commun. Ainsi, si des paramètres comme le nombre d'étapes, de correspondances et de modes utilisés sont sensiblement équivalents pour tous les trajets proposés, le temps de trajet (notamment au travers du délai entre l'horaire de départ souhaité et l'horaire du départ effectif du moyen de transport) et, de façon plus mineure, le prix, apparaissent comme véritablement discriminants dans les itinéraires proposés.

Figure 18 – Approche multicritères des itinéraires suggérés par IDF Mobilités en transport en commun selon leur ordre d'apparition



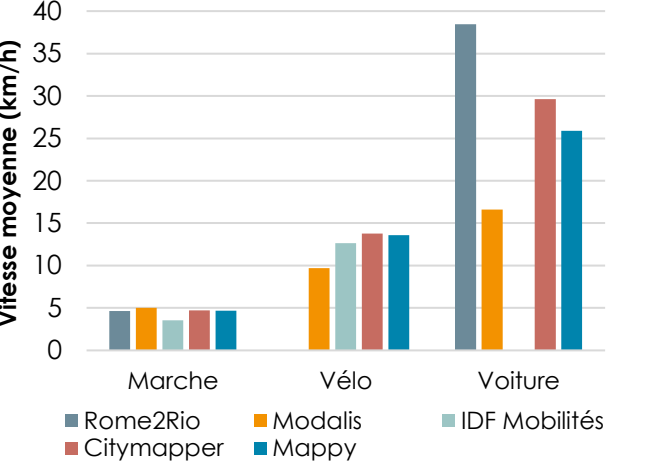
Note de lecture : le premier itinéraire suggéré minimise le délai entre l'horaire de départ souhaité et l'horaire de départ du parcours, la durée moyenne du parcours ainsi que le prix moyen par rapport aux itinéraires suivants. Le nombre moyen d'étapes, de correspondances et de modes ne semblent pas être des critères discriminants dans le classement des itinéraires.

Source : ART, à partir des résultats d'itinéraires suggérés sur application mobile et collectés grâce à une API du PEREn.

Les critères de classement des itinéraires reposent sur des hypothèses légèrement différentes selon les calculateurs d'itinéraires

Le différentiel de vitesses moyennes des modes marche, vélo et voiture observé selon le calculateur suggère des paramètres par défaut (hypothèses sur lesquelles repose le calcul des temps de trajet, notamment) variant d'un calculateur à un autre. Ce différentiel rend les comparaisons entre calculateurs – pourtant essentielles à la détection d'absence de biais dans le classement des itinéraires qu'ils proposent – plus complexes (approche inter-calculateur).

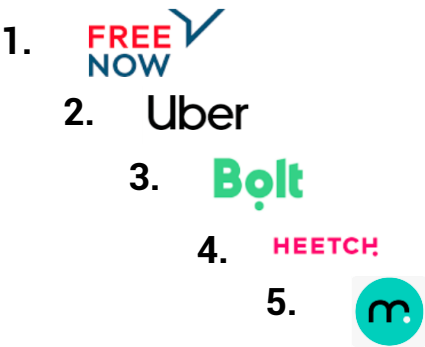
Figure 19 – Vitesse moyenne selon le mode (marche, vélo, voiture) et le calculateur d'itinéraires



Note méthodologique : moyenne des vitesses de l'ensemble des itinéraires suggérés par mode par calculateurs d'itinéraires. IDF Mobilités et Rome2Rio n'intègrent respectivement pas le mode voiture et vélo.

Source : ART, à partir des résultats d'itinéraires suggérés sur application mobile et collectés grâce à une API du PEREn.

Figure 20 – Ordre de classement des opérateurs de VTC par Citymapper



Source : ART, à partir des résultats d'itinéraires suggérés sur application mobile et collectés grâce à une API du PEREn.

L'étude du classement des itinéraires en VTC interroge quant au caractère neutre et sans biais des critères sur lesquels repose ce classement

Le classement systématiquement identique des opérateurs de VTC par certains calculateurs (exemples de Citymapper et Mappy) nécessite de poursuivre l'analyse conduite afin de déterminer si ce classement repose sur des critères objectifs. Parmi les hypothèses étudiées figurent le temps d'attente avant la prise en charge ou l'ordre d'inscription des opérateurs VTC à l'application, qui pourraient être des critères de classement des itinéraires.

Les actions et priorités de contrôle de l'Autorité pour la campagne 2023-2024

L'Autorité constate le manque de transparence des calculateurs d'itinéraires et s'interroge sur la neutralité de classement des propositions de trajets pour les modes à la demande.

L'Autorité engage donc les utilisateurs de données de mobilité à rendre visibles et transparents les critères de classement des calculateurs d'itinéraires. Elle prévoit de contribuer à la fiabilité des calculateurs d'itinéraires lors de la campagne 2023/2024 en :

- lançant des actions préventives de rappel à la réglementation à destination des calculateurs d'itinéraires dont les critères de classements des itinéraires proposés en transports collectifs ne sont pas transparents ;
- approfondissant l'analyse de l'objectivité du classement des itinéraires en investiguant leur caractère neutre et sans biais, notamment sur les applications mobiles.

- AOM : autorité organisatrice de la mobilité. 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, à savoir l'État, 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 (article L.1115-1 du code des transports).
- AOMR : autorité organisatrice de la mobilité régionale.
- API : « application programming interface ». Interface logicielle qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service afin d'échanger des données et des fonctionnalités.
- ARCEP : autorité de régulation des communications électroniques, des postes et de la distribution presse.
- 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
- DSP : délégation de service public. Contrat par lequel une personne morale de droit public confie la gestion d'un service public dont elle a la responsabilité à un délégataire public ou privé.
- 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), de cycles et d'engins de déplacement personnel (tels que les différents modèles de trottinettes et patinettes électriques, gyropodes ou monoroues) sans stations prédéfinies (« free-floating »).
- GBFS : « [general bike share feed specification](#) ». Format d'échange de données pour les véhicules partagés dont le développement a été piloté par MobilityData (association de producteurs de données nord-américaine).
- 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).
- GTFS : « [general transit feed specification](#) ». Format d'échange de données pour transport public développé initialement par Google puis repris par MobilityData (association de producteurs de données nord-américaine).
- GTFS-RT : « [general transit feed specification – realtime](#) ». Format d'échange de données en temps réel pour le transport public développé en complément au format GTFS.
- IDFM : Île-de-France mobilités.
- JSON : javascript object notation. Format d'échange de données hiérarchisé en « objets », décrits par des « attributs », eux-mêmes hiérarchisés au sein de chaque objet.
- Métadonnées : description structurée du contenu des données qui aide à les retrouver et à les utiliser (article 2 du règlement délégué 2017/1926).
- NeTEx : « [network timetable exchange](#) ». Norme européenne (NeTEx CEN/TS 16614) définissant le format d'échange de données pour le transport public. NeTEx est basé sur un langage XML (eXtended Markup Language) et suit le modèle de données Transmodel (<https://www.transmodel-cen.eu>).
- PAN : point d'accès national (<https://transport.data.gouv.fr>).
- PEReN : pôle d'expertise de la régulation numérique.
- 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).
- RATP : régie autonome des transports parisiens.
- 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.
- SAEIV : système d'aide à l'exploitation et à l'information des voyageurs.
- Spécification technique : ensemble de règles relatives à la structure et la syntaxe des données au sein de chaque format d'échange.
- SIRI : « [service interface for real time information](#) ». Norme européenne (SIRI CEN/TS 15531) définissant le format d'échange de données en temps réel pour le transport public. SIRI est basé sur un langage XML (eXtended Markup Language) et suit le modèle de données Transmodel (<https://www.transmodel-cen.eu>).
- SSIM : « [standard schedules information manual](#) ». Format d'échange de données aéronautique de programme de vols, dont le développement a été piloté par l'Association internationale du transport aérien (IATA).
- SYTRAL : syndicat mixte des transports pour le Rhône et l'agglomération lyonnaise.
- Transport à la demande : 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).
- Utilisateur de données : 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.
- VTC : voiture de transport avec chauffeur.

Directeur de la publication : Philippe Richert

Pilotage et coordination : Fabien Couly / Brewenn Métayer

Auteurs et contributeurs : Alban Gougoua, Alexandre Li, Julie Rouault

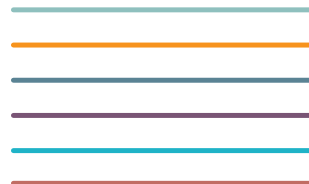
Impression : Imprimerie de la direction de l'information légale et administrative en 150 exemplaires

Dépôt légal : juin 2023

ISSN : 2678 - 6575

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

L'édition du rapport ouverture et utilisation des données de mobilité (juin 2023) est consultable en ligne sur le site Internet de l'Autorité de régulation des transports : <https://www.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