 Description du modèle de données d’échanges

Impulse

Réf : IMPULSE-Data-Exchange-Specifications\_V1.4

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Auteur | Objet |
| 0.1 | 01/04/2022 | Equipe Impulse | Création |
| 0.2 | 02/08/2022 | P. Lagarde | Repris du document |
| 0.3 | 05/08/2022 | P. Lagarde | Compléments après analyse avec  Clara Verhaeghe |
| 0.4 | 17/08/2022 | P. Lagarde | Les paramètres renvoient vers un autre document |
| 0.5 | 21/10/2022 | P. Lagarde | Ajout du lien PowerBi |
| 1 | 16/02/2023 | P. Lagarde | Actualisation avec le contenu PowerBI  Ajout de la table de redressement |
| 1.1 | 15/05/2023 | P. Lagarde | Version sans le détail qui est décrit dans le modèle. |
| 1.2 | 25/08/2023 | P. Lagarde | Mise à jour avec B. Disson + T. Gabillard |
| 1.3 | 07/12/2023 | P. Lagarde | Ajout des attributs sur le SAE pour la récupération des informations sur le matricule des conducteurs |
| 1.4 | 15/07/2024 | P. Lagarde | Ajout des types de support selon les décisions du validation\_support |

Table des matières

[1. Objet 2](#_Toc136531500)

[2. Format technique de l’échange des données 3](#_Toc136531501)

[2.a. Format GTFS 3](#_Toc136531502)

[2.b. Format hors GTFS 3](#_Toc136531503)

[2.b.i. Colonne de type Number, integer 3](#_Toc136531504)

[2.b.ii. Colonne de type Date et Date/Heure 3](#_Toc136531505)

[2.b.iii. Colonne de type texte 3](#_Toc136531506)

[2.c. Modalités de transmission des données 3](#_Toc136531507)

[2.a. Arborescence des fichiers de données 5](#_Toc136531508)

[3. Format sémantique des échanges des données 7](#_Toc136531509)

[3.a. Données de paramétrage 7](#_Toc136531510)

[3.a.i. Paramètres généraux 7](#_Toc136531511)

[3.a.ii. Paramètres de ponctualité-régularité 8](#_Toc136531512)

[3.a.i. Paramètres de redressement 8](#_Toc136531513)

[3.b. Données GTFS 8](#_Toc136531514)

[3.c. Référentiels 9](#_Toc136531515)

[3.c.i. Référentiel des lignes 9](#_Toc136531516)

[3.c.ii. Typologies 9](#_Toc136531517)

[3.d. Référentiel des véhicules 9](#_Toc136531518)

[3.e. Référentiel des produits 10](#_Toc136531519)

[3.f. Données de validations billettiques 10](#_Toc136531520)

[3.g. Données d’exploitation SAE 10](#_Toc136531521)

[4. Evènements 10](#_Toc136531522)

# Objet

La plateforme Impulse fonctionne grâce à l’utilisation de plusieurs sources de données qui constituent le cœur du dispositif. La qualité de ces données fournies assure la qualité des résultats affichées sur la plateforme. Il est donc essentiel que le respect des spécifications et plus encore que la complétude des données soient garantis par la filiale.

Dans le cadre de l’utilisation pour des besoins hors du scope d’Impulse, ce formalisme est utilisable pour des besoins hors du cadre de Impulse notamment pour les informations du SAE.

Les paragraphes suivants présentent chaque datasource avec sa description, son arborescence, sa fréquence de mise à jour, etc. Plusieurs types de données sont utilisées : fichiers de paramètres de la plateforme, référentiels, données de validations, données d’exploitation SAE, données du GTFS, faits marquants. D’autres sources de données sont acquises directement via des appels sur des API et ne sont donc pas à fournir par la filiale/partenaire.

Les principaux dispositifs de fourniture de données sont les suivants :

* L’offre de transport décrit grâce au standard de fait GTFS, aujourd’hui imposé par la plateforme PAN du Ministère ;
* Les compléments d’information sur l’offre nécessaire à la bonne configuration de la plateforme Impulse ;
* La description des véhicules permettant de réaliser l’offre ;
* L’offre réalisée en termes d’horaires réalisées, généralement issues des données du SAE de la filiale ;
* Les données de validation issues des informations des valideurs et donc des billettiques.

D’autres sources de données sont envisagées comme les données de comptage ou les données de production de CO2 par exemple.

Nous faisons ici l’hypothèse que la donnée est donc disponible sous forme numérique et qu’elle peut **être transmise de manière régulière vers la plateforme d’intégration Impulse**.

# Format technique de l’échange des données

## Format GTFS

Les données GTFS doivent être transmises suivant la norme diffusée par Google : <https://developers.google.com/transit/gtfs/reference>.

## Format hors GTFS

Les données hors données GTFS sont attendues au format CSV avec un encodage UTF8 et un séparateur de lignes Windows (CR LF).

Les séparateurs de champs sont des points virgules « ; ».

Champ optionnel : un champ optionnel non renseigné devra malgré tout apparaitre dans le fichier (on aura donc dans le fichier csv : ;;

La première ligne du fichier reprend la liste des champs.

### Colonne de type Number, integer

* Les séparateurs décimaux des points « . ».
* Pas de séparateur des milles, cent…
* Les nombres négatifs sont exprimés avec «  - »
* L’absence de valeur par un champ vide (différent de 0)

### Colonne de type Date et Date/Heure

* Les dates et date/heure s’expriment au format AAAA-MM-jj
* Les heures s’expriment au format AAAA-MM-JJ hh:mm:ss
* Les heures supérieures à minuit sont gérées (cas de certains SAE comme 25:12 pour dire 1h12 du matin) mais elles doivent être exprimées avec J+1 et heures standard.
* L’absence de valeur par un champ vide (différent de 0)

### Colonne de type texte

Toutes les données textuelles sont encodées en UTF8 et les textes « ; » doivent être saisies par \ ;

## Modalités de transmission des données

Chaque source de données doit être déposée dans un dossier spécifique (dit ADLS) décrit via une arborescence fixée. Le nom des fichiers est composé d’une partie fixe et de la date d’export. Par exemple, parametres\_20220404.csv signifie le fichier parametres.csv exporté par la filiale le 4 avril 2022. Cette règle ne s’applique pas au GTFS.

## Arborescence des fichiers de données

 L’arborescence des fichiers attendu est décrite dans le schéma suivant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| table\_file\_name | file\_frequency | file\_path | file\_extension | file\_encoding |
| agency | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| calendar | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| calendar\_dates | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| evenements\_yyyyMMdd | Quotidien | input/divers/ | Comma-separated values ".csv" | UTF-8 |
| matrice\_ponctualite\_regularite\_yyyyMMdd | A chaque changement | input/parametres/ | Comma-separated values ".csv" | UTF-8 |
| parametres\_yyyyMMdd | A chaque changement | input/parametres/ | Comma-separated values ".csv" | UTF-8 |
| passages\_sae\_{XXXX\_}yyyyMMdd | Quotidien | input/sae/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_arrets\_commerciaux\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_arrets\_physiques\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_lignes\_arrets\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_lignes\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_parcours\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_produits\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_typologies\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| referentiel\_vehicules\_yyyyMMdd | A chaque changement | input/referentiels/ | Comma-separated values ".csv" | UTF-8 |
| routes | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| shapes | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| stop\_times | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| stops | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| trips | A définir avec la filiale | input/gtfs/AAAAMMJJ/{additional\_folder\_if\_needed}/ | Text ".txt" | UTF-8 |
| validations\_{XXXX\_}yyyyMMdd | Quotidien | input/billettique/ | Comma-separated values ".csv" | UTF-8 |

# Format sémantique des échanges des données

Dans tout le modèle de données, deux champs ayant le même nom ont la même définition. Leur présence dans plusieurs sources de données permet la réalisation des jointures.

Les tableaux des attributs sont disponibles dans le tableau de bord Modèle d’échanges IMPULSE disponible au lien pour les équipes Keolis (<https://app.powerbi.com/groups/me/apps/846e906e-659a-4f3a-b6e8-d39ada8cc78b/reports/ddc1a6b3-320c-4615-9678-bdf8bebe634f/ReportSection4b74ec2920ac2510365f>)

Le tableau powerBi fait référence au regard de ce document (actualisation moins fréquente).

## Données de paramétrage

### Paramètres généraux

Table : parametres\_[Date].csv.

La table des paramètres permet à la filiale de définir un certain nombre de comportements en fonction de ces souhaites. La liste des paramètres fait l’objet d’un document spécifique.

A screenshot of a computer

Description automatically generated

### Paramètres de ponctualité-régularité

Table : matrice\_ponctualite\_regularite\_[Date].csv.

Notes : La table de ponctualité-régularité indique le mode de calcul de ponctualité ou de régularité de chaque ligne. Aujourd’hui, les champs 2, 3 et 4 ne sont pas utilisés. Une ligne est soit en ponctualité, soit en régularité.

A screenshot of a computer

Description automatically generated

### Paramètres de redressement

Table : redressement\_\_validations[Date].csv.

Notes : Cette table permet de décrire les paramètres de redressement des données billetiques. Elle permet une information détaillée du redressement par ligne, par période, par type de jour et par tranche horaire. Le redressement est un coefficient.

A screenshot of a computer

Description automatically generated

## Données GTFS

Dossier : AAAAMMJJ avec à l’intérieur les fichiers suivants : agency.txt ; calendar.txt ; calendar\_dates.txt ; routes.txt ; stop\_times.txt ; stops.txt ; trips.txt, shapes. txt

Fréquence de dépôt : A chaque mise à jour de l’offre

Le format sémantique est décrit selon la spécification Google <https://developers.google.com/transit/gtfs/reference>

Il est vivement **conseillé de fournir un fichier tous les jours permettant de collecter la dernière offre disponible pour le client.**

## Référentiels

### Référentiel des lignes

Le référentiel des lignes est construit à partir de la table routes.txt du GTFS. Ce référentiel additionnel permet de compléter les informations du GTFS avec la typologie de chaque ligne (principale, secondaire, etc.) et l’affichage par défaut de la ligne sur la page du rejeu.

Table : referentiel\_lignes\_[Date].csv.

A screenshot of a computer

Description automatically generated

### Typologies

Le référentiel des typologies permet d’associer chaque ligne à une typologie et d’obtenir des informations par groupe de lignes sur les différentes pages de la plateforme

Table : referentiel\_typologies\_[Date].csv.

A screenshot of a computer

Description automatically generated

## Référentiel des véhicules

Le référentiel des véhicules décrit chaque véhicule du réseau avec son type, son mode et sa capacité.

Table : referentiel\_vehicules\_[Date].csv.

A screenshot of a computer

Description automatically generated

## Référentiel des produits

Le référentiel des produits permet d’associer chaque carte et chaque validation à un type de titre de transport et sert à regrouper les validations suivants l’offre tarifaire.

Table : referentiel\_produits\_[Date].csv.

A screenshot of a computer

Description automatically generated

## Données de validations billettiques

La table des validations est constituée de tous les champs opérationnels tels que l’arrêt ou la course. S’il n’est pas possible d’avoir ces informations, les champs peuvent être laissés vides. Le champ code\_vehicule doit être rempli obligatoirementTable : validations\_aaaammjj.csv.

A screenshot of a computer

Description automatically generated

Pour précision, la colonne validation\_support a pour correspondance :

* ticket physique : ex. Billet Sans Contact réutilisable ou non, magnétique, etc.
* carte physique : ex. carte calypso ou équivalent
* support dematerialisé : ex. E-ticket, Flash ticket, NFC par téléphone, QR Code, etc.
* carte bancaire : ex. Open Paiement, Direct Paiement
* support sms

## Données d’exploitation SAE

La table des passages SAE contient toutes les informations disponibles sur la circulation réelle des véhicules sur le réseau. Il s’agit de l’ensemble des passages **commerciaux** qui ont **effectivement été réalisés** que le véhicule ait ou non marqué l’arrêt. Si un passage théorique (graphiqué) n’a pas été réalisé pour une raison ou une autre (exemples : courses non réalisé, course supprimée, arrêt non desservi pour cause de déviation, course non complément effectuée,…) les données associées à ce passage ne seront pas exportées. A l’inverse, une course contient un arrêt temporaire et que celui-ci a été désservi, les données associées seront exportées.

Table : passages\_sae\_[Date].csv

A screenshot of a computer

Description automatically generated

Le fichier peut être complété par les champs suivants pour des usages tiers :

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| passages\_sae | service\_driver | 13 | True | Code du service agent |  | Données non utilisées par Impulse | Optional | String | A2401 |
| passages\_sae | service\_vehicle | 14 | True | Code du service voiture |  | Données non utilisées par Impulse | Optional | String | RM507 |
| passages\_sae | matrice\_driver | 15 | True | Code du matricule de l’agent |  | Données non utilisées par Impulse | Optional | String | 12354 |

**Règles de gestion :**

1°) Heure d’arrivée et de départ – cas général

L’heure d’arrivée correspond à l’heure d’arrivée dans la zone d’arrêt

L’heure de départ correspond à l’heure ou le véhicule quitte la zone d’arrêt

2°) heure d’arrivée et de départ dans le cas d’un terminus départ : les 2 champs correspondant auront la même valeur et correspondront à l’heure de départ réalisé de la course.

3°) heure d’arrivée et de départ dans le cas d’un terminus arrivé : les 2 champs correspondant auront la même valeur et correspondront à l’heure  d’arrivée du véhicule à son terminus départ

4°) trip\_id : identifiant de la course tel que disponible dans le GTFS.

4°) actual\_trip\_id : est l’identifiant SAE de la course. Il doit toujours être renseigné à la différence du trip\_id. Il peut correspondre au trip\_id.

5°) given\_path\_id : identifiant du parcours (parcours type « hastus » ou équivalent) et il doit être fourni lorsque le trip\_id n’est pas connu.

6°) actual\_distance : Distance entre l’arrêt actuel (celui associé à la valeur) et le premier arrêt de la course. C’est une valeur croissante sur une course qui correspond à la distance parcourue par le véhicule depuis le premier arrêt.

# Évènements

Table : evenements\_[Date].csv.

Les faits marquants sont affichés dans la météo du réseau.

A screenshot of a computer

Description automatically generated