

# Données de validation télébillétiques

*Présentation des données Open data*



L'autorité organisatrice de vos  
transports en ile-de-france

**Données disponibles en  
Open Data sur la  
plateforme du STIF**

**Opendata.stif.info**

# 1

D'où viennent ces  
données ?

Depuis plusieurs années, le **Syndicat des Transports de l'Ile de France (STIF)** a décidé de généraliser le passage des tickets magnétiques sur support télébilléttique avec le **passé Navigo**.

Dans ce cadre, il a conçu un système d'information décisionnel lui permettant une **meilleure connaissance de la mobilité des usagers**.



Ce système permet la remontée **des données de validation** depuis les opérateurs jusqu'au STIF, puis leur uniformisation et leur archivage.

Dès lors, à chaque fois qu'un usager valide son passé Navigo sur le réseau francilien, **une information anonyme** est remontée à un système central permettant la collecte et le calcul de statistiques sur la mobilité.

# Des données anonymes

L'ensemble du système télébillétique a été défini avec la **CNIL (Commission Nationale de l'informatique et des libertés)** qui a rendu un avis positif sur le système.

L'anonymat est garanti de deux façons :

1- Par **l'organisation sécurisée** mise en place. Les données de validation sont toujours anonymes et ne sont jamais associées au nom du voyageur. L'organisation est conçue sur la séparation du circuit commercial de celui du traitement des données de validation. Les accès aux données sont strictement réglementées et compartimentées.

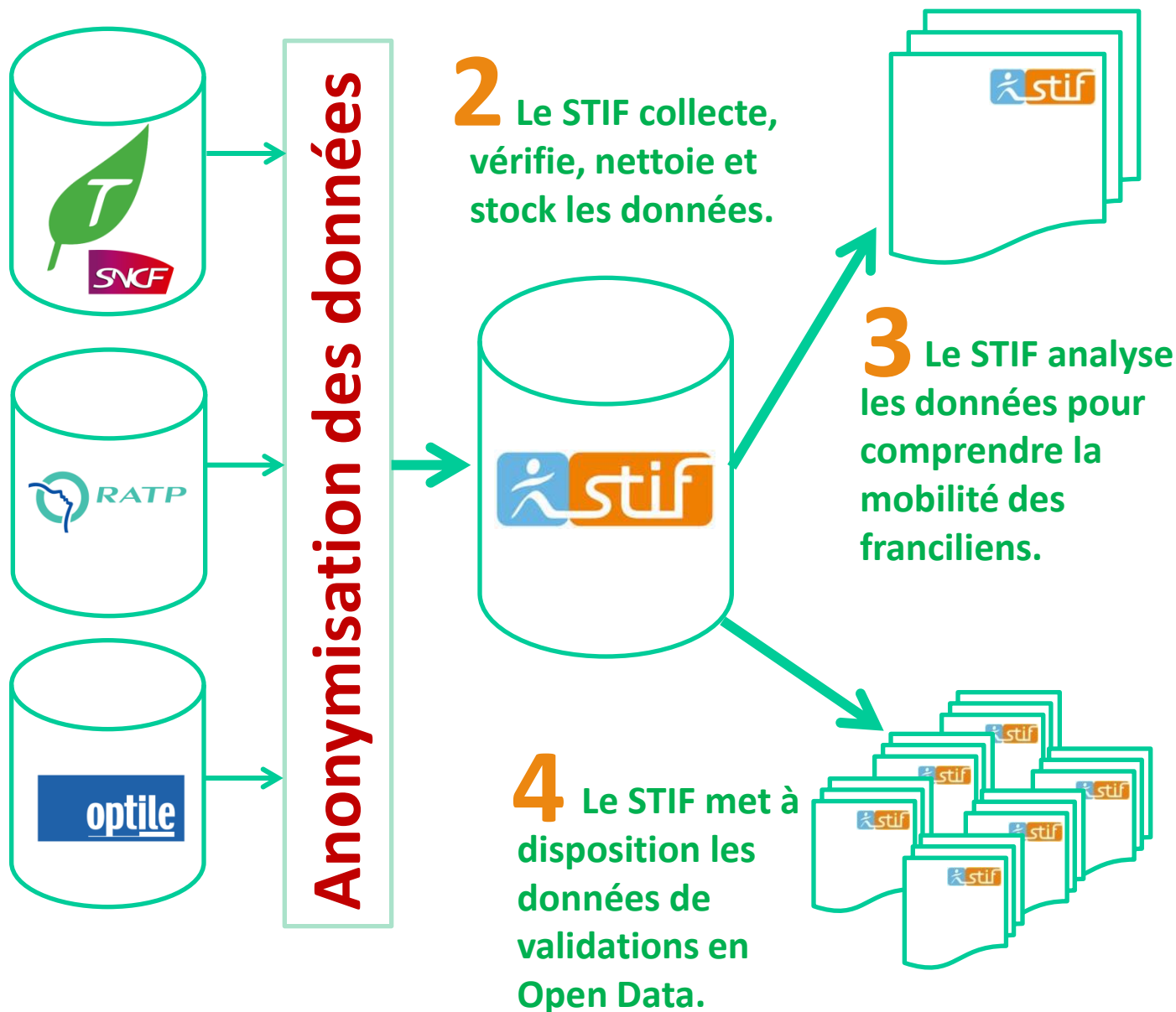
2- Par les **choix technologiques** retenus. Il est impossible d'associer une validation à un passe. Les numéros de passe sont chiffrés de façon irréversible une première fois par les transporteurs et une deuxième fois par le STIF.



# Processus de remonté des données de validation



**1** Les validations sont anonymisées puis envoyées par les opérateurs au STIF.



**11 millions** de validations  
remontées par jour ouvrable

( 5 millions un samedi  
3 millions un dimanche )

→ près de **2,7 milliards** de validations par an.

# Précautions d'usage des données de validation



Les données de validations ne comptabilisent pas :

- les tickets magnétiques (par exemple, les ticket T+, Forfait Mobilis, Forfait Paris Visite, etc.)
- les usagers qui ne valident pas à la montée dans un bus ou tram, ou en entrant dans une gare (par exemple si le quai est accessible sans passer par une barrière de contrôle, ou si les barrières de contrôle sont indisponibles en raison de travaux, etc. )
- les fraudeurs

... et donnent une vision incomplète du trafic sur le réseau de transport.



# 2

Quelles  
informations sont  
disponibles ?

**Le STIF** met à disposition les  
données de validation de  
l'ensemble du réseau francilien.

# Le réseau de transport vu par les données de validation

**725** arrêts ferrés (RER, métro, Transilien)



**Plus de 1300** lignes administratives\* (bus, tram)



\* Plus de précisions sur les lignes administratives dans la FAQ.

# Distinction entre un arrêt et une ligne

Dans les données de validation, des différences existent entre une validation sur le **réseau ferré (Transilien, RER, métro)** et une validation sur le **réseau de surface (bus, tram)**.

Dans le cas du **réseau ferré**, les validations sont rattachées à **une station** (et non à une ligne). En effet, la validation se fait à l'entrée de la station, indépendamment de la ligne précise empruntée. De plus, une fois entré dans le réseau métro ou RER, l'utilisateur n'a en général pas besoin de revalider lorsqu'il change de ligne.



Pour le **réseau de surface (bus, tram)**, le valideur est mobile puisqu'il se situe dans le bus/tram lui-même. La validation peut donc être facilement **associée à une ligne**.

**ATTENTION : le T4 est ici considéré comme un mode ferré**

# Qu'est-ce qu'une ligne ?

Dans les données de validation, la notion de ligne regroupe un ensemble d'arrêts prédéfinis par l'offre de transport. Elle est identifiée par un nom de ligne et un identifiant de ligne.



## Construction de l'identifiant de la ligne

L'identifiant de la ligne est constitué du :

- Code STIF du transporteur de la ligne
- Code STIF du réseau administratif
- Code STIF de la ligne

Possibilité de faire le lien avec d'autres données grâce au référentiel du STIF : `ID_GroupOfLines`

Exemples :

- Tramway 1 : 100;112;011
- Bus RATP 26 : 100;100;026
- Car 3B STAVO : 044;044;003

# Qu'est-ce qu'un arrêt ?

Dans les données de validation, un arrêt est le lieu où le voyageur valide son titre de transport sur le réseau ferré (i.e. dans une station de métro ou une gare). Il est identifié par un nom d'arrêt et un identifiant de l'arrêt.



## Construction de l'identifiant de l'arrêt

L'identifiant de la ligne est constitué du :

- Code STIF du transporteur
- Code STIF du réseau
- Code STIF de l'arrêt

Exemples :

- Station de métro Alesia: 100;110;10
- Gare RER Arpajon 800;803;26

Possibilité de faire le lien avec d'autres données grâce au référentiel STIF : ID\_REFA\_LDA

Mises à jour régulières des  
données sur le portail.



Actualisation  
**tous les 6 mois**  
des jeux de données.

# Données libérées en Open Data

**4 fichiers .csv disponibles**  
sur le portail Open Data du STIF  
**Opendata.stif.info**

Validations sur le réseau ferré : Nombre de validations par jour

NB\_FER

Validations sur le réseau ferré : Profils horaires par jour type

PROFIL\_FER

Validations sur le réseau de surface: Nombre de validations par jour

NB\_SURFACE

Validations sur le réseau de surface : Profils horaires par jour type

PROFIL\_SURFACE

*Le nom des fichiers est suivi de la période sur laquelle porte l'extraction.  
Par exemple **NB\_FER\_2015S1** contient les données du **premier semestre de 2015**.*



# Données libérées en Open Data

Validations sur le réseau ferré : Nombre de validations par jour

NB\_FER

VARIABLES	FORMAT	DEFINITION
JOUR	Date (01/01/2015)	Jour d'exploitation (de 04:00 à 03:59 le lendemain)
COD_STIF_TRNS	Numérique	Code Stif du transporteur
COD_STIF_RES	Numérique	Code Stif du réseau
COD_STIF_ARRET	Numérique	Code Stif de l'arrêt/station
LIBELLE_ARRET	Caractère	Libellé de l'arrêt/station
ID_REFA_LDA	Caractère	Identifiant arrêt référentiel STIF
CATEGORIE_TITRE	Caractère	Titre de transport
NB_VALD	Numérique	Nombre de validations (en entrée sur le réseau)

01/01/2015;100;110;1;PORTE MAILLOT; 71379; AMETHYSTE;117

...

01/01/2015;100;110;12;ECOLE VETERINAIRE DE MAISONS-ALFORT;70522;NAVIGO; 703

# Données libérées en Open Data

Validations sur le réseau ferré : Profils horaires par jour type

PROFIL\_FER

VARIABLES	FORMAT	DEFINITION
COD_STIF_TRNS	Numérique	Code Stif du transporteur
COD_STIF_RES	Numérique	Code Stif du réseau
COD_STIF_ARRET	Numérique	Code Stif de l'arrêt/station
LIBELLE_ARRET	Caractère	Libellé de l'arrêt/station
ID_REFA_LDA	Caractère	Identifiant arrêt référentiel STIF
CAT_JOUR	Caractère	Catégorie du jour (modalités CAT_JOUR dans la FAQ)
TRNC_HORR_60	Caractère	Tranche horaire d'une heure
POURC_VALIDATIONS	Numérique	<b>Pour une station <math>i</math></b> : Rapport entre le nombre de validations à une station $i$ , sur une tranche horaire $TH$ , et le nombre de validations sur la journée entière à cette station $i$ .

100;110;93;BOTZARIS;71906; SAVS;15H-16H;7.00

...

800;852;720;RAMBOUILLET; 60665;DIJFP;12H-13H;5.61

# Données libérées en Open Data

Validations sur le réseau de surface: Nombre de validations par jour

NB\_SURFACE

VARIABLES	FORMAT	DEFINITION
JOUR	Date (01/01/2015)	Jour d'exploitation (de 04:00 à 03:59 le lendemain)
COD_STIF_TRNS	Numérique	Code Stif du transporteur
COD_STIF_RES	Numérique	Code Stif du réseau
COD_STIF_LIGNE	Numérique	Code Stif de la ligne
LIBELLE_LIGNE	Caractère	Libellé de la ligne
ID_GroupOfLines	Caractère	Identifiant STIF de référence de la ligne administrative
CATEGORIE_TITRE	Caractère	Titre de transport
NB_VALD	Numérique	Nombre de validations (en entrée sur le réseau)

01/01/2015;11;11 ;320;LES MUREAUX (GARE SNCF) - LES MUREAUX (BIZET);A00123;AMETHYSTE;40

...

01/01/2015;4;4;10 ; VIROFLAY (Chaville Velizy RER C) - VELIZY-VILLACOUBLAY (Velizy 2);A00040;FGT;6

# Données libérées en Open Data

Validations sur le réseau de surface : Profils horaires par jour type

PROFIL\_SURFACE

VARIABLES	FORMAT	DEFINITION
COD_STIF_TRNS	Numérique	Code Stif du transporteur
COD_STIF_RES	Numérique	Code Stif du réseau
COD_STIF_LIGNE	Numérique	Code Stif de la ligne
LIBELLE_LIGNE	Caractère	Libellé de la ligne
ID_GroupOfLines	Caractère	Identifiant ligne référentiel STIF
CAT_JOUR	Caractère	Catégorie du jour (modalités CAT_JOUR dans la FAQ)
TRNC_HORR_60	Caractère	Tranche horaire d'une heure
POURC_VALIDATIONS	Numérique	<b>Pour une ligne <math>j</math></b> : Rapport entre le nombre de validations sur une ligne $j$ , sur une tranche horaire $TH$ , et le nombre de validations sur la journée entière sur cette ligne $j$ .

2;2;9;THIAIS (E. LEVASSEUR) - CHOISY-LE-ROI (GARE RER);A00005;JOHV;0H-1H;0.53

...

11;11;2;LES MUREAUX (SNCF) - PONTOISE;A00098;JOHV;16H-17H;2.18

# Les titres de transport

Les données de validations sont classées par catégorie de **titre de transport** :

- « **IMAGINE R** » : regroupe les forfaits annuels **Imagine R Scolaire** et **Imagine R Etudiant**, forfaits réservés aux élèves, apprentis et étudiants qui permet de voyager à volonté toute l'année et dans toute l'Ile-de-France.
- « **NAVIGO** » : regroupe les forfaits **Navigo Annuel**, **Navigo Mois** et **Navigo Semaine**.
- « **AMETHYSTE** » : comptabilise les **forfaits Améthyste** : forfait réservé aux personnes âgées ou handicapées sous conditions de ressources ou de statut, et résidant en Île-de-France. Ce forfait annuel permet à son bénéficiaire de se déplacer de manière illimitée sur tous les modes de transport à l'intérieur des zones de validité.
- « **TST** » : regroupe les forfaits hebdomadaires et mensuels à tarif réduit, accordé aux bénéficiaires de la **Réduction Solidarité Transport**, pour voyager à l'intérieur des zones choisies sur tous les modes de transport en Île-de-France.
- « **FGT** » : comptabilise les **Forfaits Navigo Gratuité Transport**, forfait qui permet à certain bénéficiaire d'aides sociales de voyager gratuitement dans toute l'Île-de-France.
- « **AUTRE TITRE** » : comptabilise les forfaits spéciaux.
- « **NON DEFINI** » : comptabilise les validations dont le type de titre n'est pas défini (anomalies).

# 3

**Quels liens avec  
les autres données  
du STIF ?**

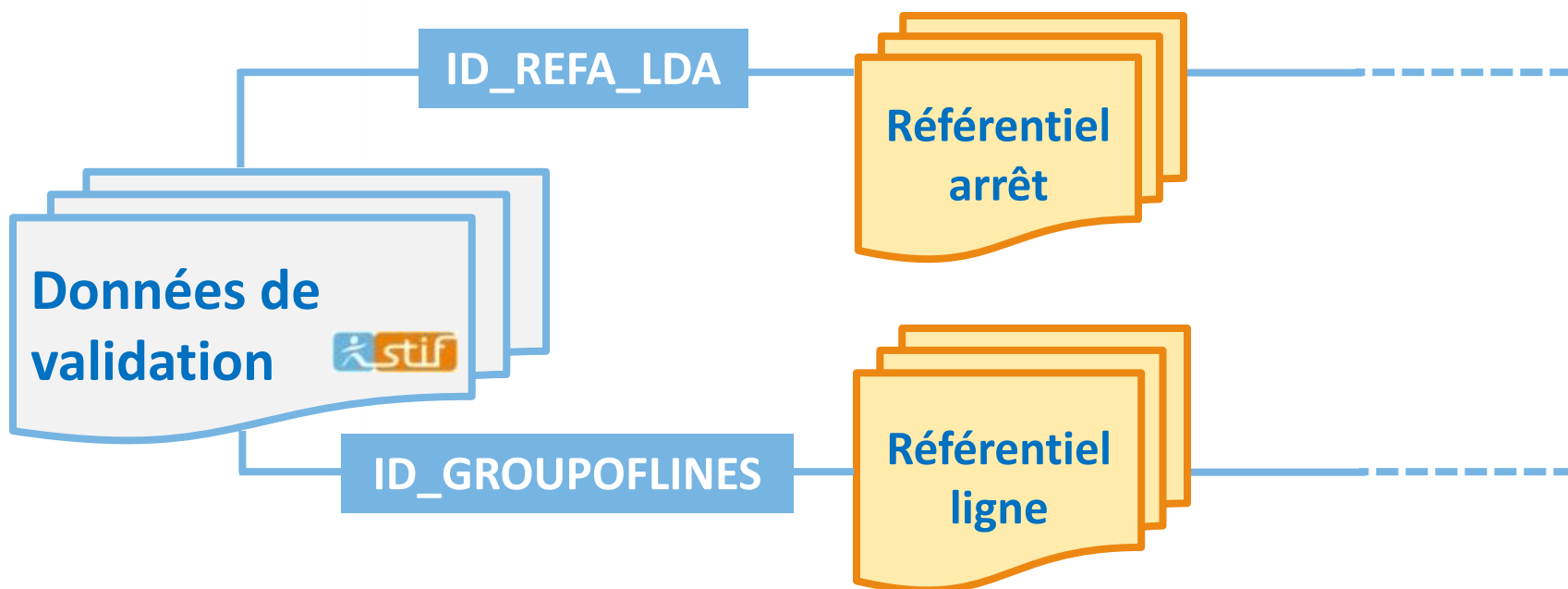
Le **STIF** déploie son référentiel  
de lignes et arrêts.

Aujourd'hui, plusieurs codifications  
coexistent encore pour désigner  
une ligne ou un arrêt. Seule la  
codification issue des référentiels  
est pérenne.

# Liens avec les autres jeux de données

Pour **établir des liens** entre les données de validation et les autres données du STIF, il faut utiliser les identifiants issus des **référentiels du STIF** :

- **ID\_GROUPOFLINES** pour les identifiants des lignes
- **ID\_REFA\_LDA** pour les identifiants des arrêts



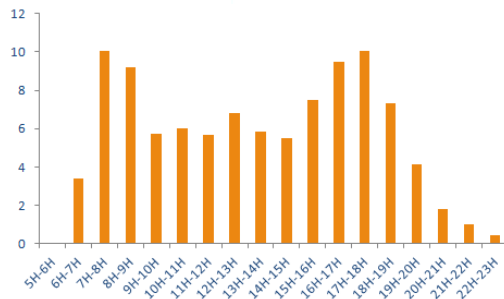


# FAQ

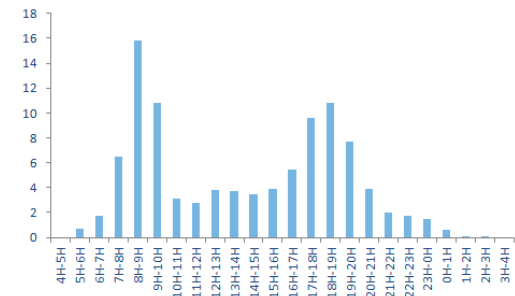
## ❏ Comment est calculé le profil horaire ?

La variation du volume de validations selon l'heure est exprimée en rapportant le volume des validations d'une tranche horaire, au volume des validations de la journée entière.

**Pour une station  $i$  :** Rapport entre le nombre de validations à une station  $i$ , sur une tranche horaire  $TH$ , et le nombre de validations sur la journée entière à cette station  $i$ .



*Profil horaire (JOHV) de la ligne 22 - Daniel Meyer  
SAVIGNY-SUR-ORGE (gare RER) - GRIGNY (tuileries)*



*Profil horaire (JOHV)  
station Porte Maillot - RATP*

**Pour une ligne  $j$  :** Rapport entre le nombre de validations sur une ligne  $j$ , sur une tranche horaire  $TH$ , et le nombre de validations sur la journée entière sur cette ligne  $j$ .

# FAQ

## ❑ Que signifie « Moins de 5 » dans les colonnes NB\_VALD ?

Par respect des exigences sur la protection des données personnelles, la mention « Moins de 5 » est indiquée pour les volumes inférieurs à 5 non nuls. La mention « 0 » est indiquée pour les volumes nuls.

## ❑ Est-ce que les validations en sortie et en correspondance sont incluses dans les données ?

Non. Dans les données mises à disposition par le STIF, il n'y a que les validations en entrée sur le réseau. Les validations en sortie et correspondance sont exclues. Seules les validation des entrants depuis la voirie sur les différents réseaux sont comptabilisées, c'est-à-dire toutes les validations des personnes entrants dans un bus et des personnes entrants dans le réseau ferré.

## ❑ Comment sont calculés les *catégories du jour* ?

Les calculs sont réalisés sur l'ensemble des données du semestre selon leur catégorie :

**JOHV** : Jour Ouvré Hors Vacances Scolaires

**SAHV** : Samedi Hors Vacances Scolaires.

**JOVS** : Jour Ouvré en période de Vacances Scolaires.

**SAVS** : Samedi en période de Vacances Scolaires.

**DIJFP** : Dimanche et Jour Férié et les ponts .

# FAQ

## ❑ Est-il possible de suivre une personne avec ces données ?

Non. Il est impossible de retrouver une personne au travers de ces données. En effet, les données mises à disposition par le STIF sont des agrégations de validations qui ne permettent pas de cibler une personne en particulier. De plus, toutes les données sont anonymes et les clés d'anonymisation sont renouvelées tous les 3 mois.

## ❑ Qu'est ce qu'une « ligne administrative » ?

Il existe deux types de lignes :

- La **ligne commerciale** est une ligne de transport identifiée par le voyageur sur le terrain par son nom ou son code commercial. Dans les données de référence publiées par le STIF, cette ligne possède une identification unique et pérenne (ID\_Line) permettant de faire le lien avec les autres données.
- La **ligne administrative** est la vision contractuelle de la ligne. Une ligne administrative peut comprendre 1 ou plusieurs lignes commerciales. Dans les données de référence publiées par le STIF, la ligne administrative possède une identification unique et pérenne (ID\_GroupOfLines) permettant de faire le lien avec les autres données.

# Opendata.stif.info

