

# I-Choix du Schema

Tout d'abord, nous avons choisi un schema qui n'est ni en construction, ni uniquement publié mais qui a été adopté afin de garantir un minimum de cohérence et de qualité dans les données choisies.

Ensuite, pour être à peu près sûre qu'on n'aura pas beaucoup de problème de nettoyage : nous avons regardé dans la partie « données » du schéma proposé, les jeux de données qui ont adopté ce schéma. Le schéma choisi est « Prénoms des nouveaux-nés » (accessible via le lien ci-après <https://schema.data.gouv.fr/scdl/prenoms/> vérifié le 04-12-2023).

[Accueil](#) [Schémas](#) [Contribuer](#) [Initiatives similaires](#) [Statistiques](#) [A propos](#)

[Saisir ou valider mes données](#) [Schéma](#) [Données](#)

[Informations](#) [Documentation](#) [Changements](#)

Version du schéma :

## Prénoms des nouveaux-nés

Spécification de la liste annuelle des prénoms des nouveaux-nés déclarés à l'état-civil

### Contexte

Le modèle de données relatif aux prénoms des nouveaux-nés déclarés à l'état-civil spécifie la structure et le contenu d'un jeu de données simple apprécié du public. Il consiste en une liste de prénoms déclarés à l'état-civil de la commune de naissance, avec l'occurrence de chacun pour une donnée. Les prénoms listés correspondent au premier prénom donné dans chaque acte de naissance.

Nous avons repéré quelques noms de villes qui ont adopté le schéma « Prénoms des nouveaux-nés ». On a repéré sur Open data la ville de Torcy et d'Antibes par exemple.


<https://www.data.gouv.fr/fr/datasets/?schema=scdl/prenoms>

2022


Recherche

3 résultats

Trier par : Pertinence

**Prénoms des nouveaux-nés à Antibes**  
Par [Ville d'Antibes](#)  
Prénoms des nouveaux-nés déclarés à l'état-civil à Antibes.  
Qualité des métadonnées  — Mis à jour aujourd'hui

0 réutilisations  
0 favoris

**Prénoms des enfants nés à Torcy**  
Par [Ville de Torcy](#)  
Liste de prénoms déclarés à l'état-civil de Torcy, pour les enfants nés à Torcy.  
Qualité des métadonnées  — Mis à jour aujourd'hui

0 réutilisations  
0 favoris

Nous avons ensuite vérifié les informations sur les jeux de données et on a choisi Antibes car la période couverte était plus large (2011-2022) que celle de Torcy (2018-2019)

## Ville de Torcy

The screenshot shows the 'Informations' tab of a dataset on data.gouv.fr. The browser address bar shows the URL: <https://www.data.gouv.fr/fr/datasets/prenoms-des-enfants-nés-a-torcy/#/information>. The page has tabs for 'Fichiers (1)', 'Réutilisations (0)', 'Discussions (0)', 'Ressources communautaires (0)', and 'Informations'. Under 'Informations', there are sections for 'Mots-clés' (prenoms, scdl, scdl-prenoms), 'Licence' (Licence Ouverte / Open Licence version 2.0), 'ID' (62e1655d42efa92720b3b58f), 'TEMPORALITÉ' (Création: 27 juillet 2022, Dernière mise à jour: 4 décembre 2023, Fréquence: Annuelle, Couverture temporelle: 2018 à 2019).

INFORMATIONS		
<b>Mots-clés</b>		
prenoms scdl scdl-prenoms		
<b>Licence</b>	<b>ID</b>	
Licence Ouverte / Open Licence version 2.0	62e1655d42efa92720b3b58f	
<b>TEMPORALITÉ</b>		
<b>Création</b>	<b>Fréquence</b>	<b>Couverture temporelle</b>
27 juillet 2022	Annuelle	2018 à 2019
<b>Dernière mise à jour</b>		
4 décembre 2023		

## Ville d'Antibes

The screenshot shows the 'Informations' tab of a dataset on data.gouv.fr. The browser address bar shows the URL: <https://www.data.gouv.fr/fr/datasets/prenoms-des-nouveaux-nés-a-antibes/#/information>. The page has tabs for 'Fichiers (1)', 'Réutilisations (0)', 'Discussions (0)', 'Ressources communautaires (0)', and 'Informations'. Under 'Informations', there are sections for 'Licence' (Licence Ouverte / Open Licence version 2.0), 'ID' (5e7c8525048a0a8a64d46188), 'TEMPORALITÉ' (Création: 26 mars 2020, Dernière mise à jour: 4 décembre 2023, Fréquence: Annuelle, Couverture temporelle: 2011 à 2023), and 'COUVERTURE SPATIALE' (Couverture territoriale: Antibes, Granularité de la couverture territoriale: Commune française).

INFORMATIONS		
<b>Licence</b>	<b>ID</b>	
Licence Ouverte / Open Licence version 2.0	5e7c8525048a0a8a64d46188	
<b>TEMPORALITÉ</b>		
<b>Création</b>	<b>Fréquence</b>	<b>Couverture temporelle</b>
26 mars 2020	Annuelle	2011 à 2023
<b>Dernière mise à jour</b>		
4 décembre 2023		
<b>COUVERTURE SPATIALE</b>		
<b>Couverture territoriale</b>	<b>Granularité de la couverture territoriale</b>	
Antibes	Commune française	

Et enfin on s'est rendu sur le site de la ville pour retrouver le jeu de données. Prénoms des nouveaux nés de la ville d'Antibes accessible via le lien ci-après : <https://trouver.datasud.fr/dataset/prenoms-des-nouveaux-nés-de-la-ville-antibes-juan-les-pins/resource/c49b3327-31d1-4d34-b90b-2d1585a20327>

## II-Analyse du jeu de données

Nous avons procédé au Sprint Qualité du jeu de données puis nous l'avons vérifié sur <https://publier.etalab.studio/fr/>. Nous avons nommé le fichier corrigé « antibes\_prenoms\_naissance\_corrige.csv ».

Nous avons essayé de partir d'un fichier xls qu'on a enregistré sous le format CSV pour palier au problème d'affichage des caractères accentués (voir notre rendu sprint qualité). Mais cela n'était pas la bonne approche car ce fichier ne pouvait s'ouvrir qu'avec Microsoft Excel. Nous avons donc désinstaller LibreOffice et l'avons réinstallé avec un JRE 64 bit car c'était un warning du logiciel. Ce qui explique pourquoi il y avait des problèmes d'affichage. Une fois ces mises à jour faites, il n'y avait plus de problème avec l'accentuation. Nous avons changé le format de la colonne « COLL\_INSEE » afin que cela puisse correspondre au format du code postal d'Antibes.

Formatage des cellules

Nombres | Police | Effets de caractère | Alignement | Bordures | Arrière-plan | Protection de cellule

Catégorie	Format	Langue
Tout		Par défaut - Français (France)
Défini par l'utilisateur		
Nombre		
Pourcentage		
Monnaie		
Date		
Heure		
Scientifique		
Fraction		
Valeur logique		
Texte		

Options

Nombre de décimales :    ☐ Nombres négatifs en rouge

Zéros non significatifs :    ☐ Séparateur de milliers

Code de format

#


Défini par l'utilisateur

Ensuite, nous nous sommes rendus sur <https://publier.etalab.studio/fr/upload?schema=scdl%2Fprenoms>. D'après l'analyse faite, il y avait 2 erreurs dans la colonne « ENFANT\_PRENOM » et on les trouvées grâce au rapport d'erreurs proposé :

- 2 syntaxes de prénoms non valides aux lignes 7310, 7355.

## ⚠ Votre fichier contient des erreurs.

Nous avons trouvé 2 erreur(s). Vous pouvez voir la liste des erreurs et corriger le fichier.

Corriger le fichier 

Voir le rapport d'erreur 

Erreurs de contenu (2)

- **Format incorrect** : colonne *ENFANT\_PRENOM*, ligne 7310

Prénom de nouveau(x)-né(s) constaté comme premier prénom dans les actes d'état-civil de l'année correspondante. Un acte de naissance peut désigner un nouveau-né avec plusieurs prénoms et le législateur a prévu que "*tout prénom inscrit dans l'acte de naissance peut être choisi comme prénom usuel*". La plupart du temps le premier prénom est le prénom d'usage initialement choisi par le(s) parent(s). Cette spécification ne retient donc que le premier prénom : si un nouveau-né est appelé 'Armelle Julia Blanche', seul 'Armelle' sera retenu pour constituer ce jeu de données. Un prénom composé comme 'Marie-Jeanne' compte pour un prénom complet. Attention, "*l'alphabet utilisé doit être celui qui sert à l'écriture du français. Les caractères alphabétiques étrangers ne sont donc pas autorisés (par exemple le « ñ »)*". Outre les caractères alphabétiques, un prénom peut posséder un trait d'union, voire deux, comme dans 'Lou-Anne' ou 'Mohamed-El-Amine'. Des prénoms peuvent posséder une apostrophe comme dans Gwenc'Hlan ou N'Deye, voire peut-être deux. Nous considérons aussi qu'un prénom pourrait se terminer, voire débiter, par une apostrophe - cette dernière étant parfois utilisée en français pour marquer la suppression de la finale ou le début d'un mot, comme dans Boul' Mich'.

- **Format incorrect** : colonne *ENFANT\_PRENOM*, ligne 7355

## Correction des syntaxes incorrectes :

Sachant que c'était le même problème de contenus (syntaxe des prénoms) pour les 2 lignes citées plus haut, on a rectifié selon le problème .

- A la ligne 7310 : un prénom composé séparé par un espace et pas par deux tirets comme spécifié dans la documentation du schema. Donc, on a remplacé « Jean-Eric Thierry » par « Jean-Eric-Thierry ».

D7310		$f_x \sum$	=	Jean-Eric Thierry
	A	B	C	D
7308	Antibes	06004	M	Jassim
7309	Antibes	06004	M	Jayden
7310	Antibes	06004	M	Jean-Eric Thierry

7309	Antibes	6004	M	Jayden
7310	Antibes	6004	M	Jean-Eric-Thierry
7311	Antibes	6004	F	Jeanne

- A la ligne 7355 : on a gardé uniquement la première lettre du prénom en majuscule.

7354	Antibes	06004	M	Kiyansh
7355	Antibes	06004	M	KiyoYnigo

7354	Antibes	6004	M	Kiyansh
7355	Antibes	6004	M	Kiyoynigo
7356	Antibes	6004	F	Kléo

## Conclusion:

Après la correction des 2 erreurs : on a eu un jeu de données qui valide le schéma « Prénoms des nouveaux-nés » et qui est donc conforme, prêt à être utilisé.

<https://publier.etalab.studio/fr/upload?schema=scdl%2Fprenoms>

Tableau de bord Courrier - Dieynaba K... Choix GED, Visual Wor... GitHub Gestion de projet | Trello Gestion de projet - Go... seance\_M2

# Vérifier et charger un fichier existant

[Prénoms des nouveaux-nés](#)



Nom du fichier : antibes\_prenoms\_naissance\_corrige.csv  
Taille : 257.797 ko

✔ Bravo ! Votre fichier est parfaitement conforme.

Prévisualiser votre fichier 

Editer le fichier 

Publier sur data.gouv.fr 